Étude de la perception de la voix chez le patient sourd postlingual implanté cochléaire unilatéral et le normo-entendant en condition de simulation d'implant. Psychophysique et imagerie

Zoé Massida

▶ **To cite this version:**

Zoé Massida. Étude de la perception de la voix chez le patient sourd postlingual implanté cochléaire unilatéral et le normo-entendant en condition de simulation d'implant. Psychophysique et imagerie. Neurosciences [q-bio.NC]. Université Paul Sabatier - Toulouse III, 2010. Français. NNT : . tel-00803654

HAL Id: tel-00803654
https://theses.hal.science/tel-00803654

Submitted on 22 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par *l'Université de Toulouse 3 Paul Sabatier*
Discipline ou spécialité : *Neurosciences*

Présentée et soutenue par *Zoé Massida*
Le *20 octobre 2010*

Titre : *Etude de la perception de la voix chez le patient sourd post lingual implanté cochléaire unilatéral et le sujet normo-entendant en condition de simulation d'implant. Psychophysique et imagerie*

JURY

Pascal BARONE	*Directeur de thèse*
Olivier DEGUINE	*Co-directeur de thèse*
Nicole BRUNEAU	*Rapporteure*
Daniel PRESSNITZER	*Rapporteur*
Eric TRUY	*Examinateur*
Pascal BELIN	*Examinateur*
Pier-Giorgio ZANONE	*Président du jury*

Ecole doctorale : *CLESCO*
Unité de recherche : *CerCo (UMR 5549)*
Directeur(s) de Thèse : *Pascal BARONE et Olivier DEGUINE*

Remerciements

Tout d'abord, je tiens à remercier Pascal Barone et Olivier Deguine d'avoir accepté de diriger ce travail de thèse. Merci à Pascal Barone pour son travail d'encadrement tout au long de cette étude, merci pour son intérêt, sa confiance et son respect. Merci d'avoir cru en mes facultés et mes ressources, souvent ignorées de moi-même. Merci pour ses compétences scientifiques et son soutien pendant ces 4 années. Merci également à Olivier Deguine pour m'avoir ouvert avec enthousiasme les portes du service ORL de Purpan, merci d'avoir cru en moi, merci pour son expertise clinique et ses conseils scientifiques avisés. Tous deux ont été une aide précieuse.

Je remercie également Michelle Fabre-Thorpe de m'avoir accueillie au sein de son laboratoire (CerCo, UMR 5549) durant cette thèse. Merci de m'avoir permis de découvrir le monde de la recherche publique avec une vraie réflexion sur son insertion politique, économique et sociale. Merci également à elle pour l'énergie qu'elle injecte dans le fonctionnement du laboratoire, avec de vrais soucis d'harmonie et de dynamisme.

Ma gratitude s'adresse également à Pascal Belin, pour sa collaboration fructueuse sur l'ensemble de ce travail, pour son temps, son implication et son expertise scientifique indispensable.

Je remercie les membres du jury pour avoir accepté de donner de leur temps pour lire, commenter et examiner mon travail de thèse. Merci donc à Daniel Pressnitzer et Nicole Bruneau pour avoir accepté d'être rapporteurs de ce travail, ainsi qu'à Eric Truy, Pascal Belin et Pier-Giorgio Zanone pour avoir rejoint le jury en tant qu'examinateurs.

Au cours de mon séjour au sein du laboratoire du CerCo, nombreux ont été ceux qui m'ont apportés leur aide scientifique. A ce titre je tiens à remercier particulièrement Julien Rouger, Olivier Joubert et Kuzma Strelnikov pour leur patience au cours de mon laborieux apprentissage de Matlab. Merci à Jean-Michel Hupé pour son avis et ses conseils statistiques, un point de repère dans cet univers toujours très complexe. Merci à Ruffin VanRullen également pour sa précision dans l'appréhension de la théorie de la détection du signal. Et encore merci Kusma, pour sa pédagogie à m'expliquer l'analyse d'images fonctionnelles en TEP et IRMf. Un grand merci à Catherine Marlot pour son aide efficace et rapide dans mes recherches d'articles introuvables, à Claire Thoueilles pour ses compétences logistiques, ainsi qu'à Carmen Olmos et Joël Ferrer pour leurs compétences informatiques. Je n'oublie pas non plus toute la sollicitude de beaucoup à vouloir intégrer au CerCo malgré eux, les habitants de la Sibérie. Ce ne fut jamais simple, mais toujours apprécié. Pas de remerciement personnel à ce titre, mais je sais que chacun se reconnaîtra.

Mes visites nombreuses au service ORL de Purpan m'ont amenée à côtoyer cet univers centré autour du patient et les rapports qu'il entretient avec le monde de la recherche. Je voudrais remercier tous ceux qui ont participé de près ou de loin à ce travail, par leur accueil ou leur collaboration. Merci donc à Bernard Fraysse de m'avoir accueillie au sein du service ORL, merci à Marie-Laurence Laborde de m'avoir introduite auprès des patients, de m'avoir fait partager son expérience et son travail, sans lequel ce rapport n'aurait pu aboutir. Merci à double titre à Mathieu Marx pour sa présence au sein du service ORL et sa collaboration fructueuse au sein du laboratoire du CerCo. Merci à Gaëtan, Jean-Philippe Bousquet, Marie-France Martin et l'ensemble du personnel du service ORL pour m'avoir fait de la place pour travailler dans ce manque de place.

Ce travail n'aurait pas pu être mené à bien sans la convention CIFRE mise en place avec et grâce à la société Cochlear®. J'en profite pour remercier Chris James pour ses conseils, ses questions et ses réponses. Merci également à Sandrine Cabrol, pour l'intérêt constant qu'elle a porté à ce travail, le soutient qu'elle m'a manifesté et sa chaleur humaine. Je n'oublie pas Agnès Duvignau, qui a permis le bon déroulement administratif de cette collaboration.

Enfin, merci à toutes les personnes qui ont participé aux nombreuses et longues expériences, et sans qui ce travail n'existerait pas. Merci pour leur bénévolat et leur disponibilité qu'ils soient patients ou sujets sains, tous volontaires, merci pour leur contribution, indispensable à l'avancement de la recherche en général.

Table des Matières

CHAPITRE 1 : L'AUDITION, LES SONS ET LA VOIX 13
1 L'audition 15
1.1 La structure du récepteur auditif 15
1.1.1 Oreille externe 16
1.1.2 Oreille moyenne 17
1.1.3 Oreille interne : cochlée 19
1.1.3.1 Anatomie de la cochlée 19
1.1.3.2 Organe de Corti 22
1.1.3.3 Fonctionnement de la cochlée 24
1.2 Mécanismes centraux de l'audition 27
1.2.1 Anatomie des voies auditives 27
1.2.1.1 Relais sous-corticaux 27
1.2.1.2 Le cortex auditif 29
1.2.2 Physiologie/propriétés des voies auditives 32
1.2.2.1 Codage de la fréquence sonore ou tonotopie 32
1.2.2.2 Codage de la localisation sonore 33
1.2.2.3 Codage de l'intensité sonore 33
2 Les sons 35
2.1 Le pitch 37
2.1.1 Définition 37
2.1.2 Caractéristiques physiques du pitch 37
2.1.3 Perception de la hauteur et perception du pitch 38
2.1.4 Corrélats neuronaux de la perception du pitch 41
2.1.5 Le cas particulier de l'oreille absolue 45
2.2 Timbre 50
2.2.1 Définition 51
2.2.2 Caractéristiques physiques du timbre 52
2.2.2.1 Les formants ou la répartition de l'énergie dans le spectre 52
2.2.2.2 L'enveloppe temporelle ou les transitoires 55
2.2.2.3 Autres paramètres 56
2.2.3 La perception du timbre 57
2.2.4 La neuropsychologie du timbre 60
2.3 Pas de pitch sans timbre ? 62
2.4 Durée 63

2.5 Intensités 64
2.5.1 Définition et seuils 64
2.5.2 Echelle des phones ou échelle d'isosonie 65
2.5.3 Facteur de variation de la sonie 66
2.6 La source sonore : nature et localisation 67
2.6.1 Nature de la source 68
2.6.2 Localisation 68
2.6.2.1 La localisation en azimut 68
2.6.2.2 La localisation en élévation 69
2.6.2.3 La localisation en distance 69
3 La voix 71
3.1 Formation de la voix 71
3.1.1 La soufflerie pulmonaire 71
3.1.2 Le larynx 72
3.1.2.1 Anatomie 72
3.1.2.2 Production sonore 74
3.1.3 L'espace supralaryngé 76
3.2 Caractéristiques physiques de la voix 77
3.2.1 Le pitch de la voix 79
3.2.2 Le timbre de la voix 80
3.2.3 Autres caractéristiques 82
3.3 La perception de la voix 83
3.3.1 Dimensions perceptives et leurs corrélats acoustiques 84
3.3.2 Jugement de dissemblance et corrélats acoustiques 85
3.4 Le traitement cérébral de la voix 86
3.4.1 La phonagnosie : un argument en faveur d'un traitement cérébral spécifique pour la voix ……87
3.4.2 Une aire spécifique à la voix, la TVA (Temporal voice area) 87
3.4.3 Autres arguments en faveur de la TVA 89
3.5 La voix, un visage auditif ? 92
3.5.1 Perception de l'information langagière 93
3.5.2 Perception de l'information affective 94
3.5.3 Perception de l'information d'identité 94
CHAPITRE 2 : LA SURDITÉ ET LES IMPLANTS COCHLÉAIRES 97
4 La surdité 99
4.1 Diagnostic de surdité 99
4.1.1 Classification des surdités 99
4.1.1.1 Classification fonctionnelle 99
4.1.1.2 Classification audiométrique 100

4.1.1.3 Autres classifications 101
4.1.2 Diagnostic de la surdité 101
4.1.2.1 Outils diagnostiques 102
4.2 Causes de la surdité 104
4.2.1 Etiologie des surdités de transmission 104
4.2.1.1 Surdités de transmission avec tympan normal 104
4.2.1.2 Surdités de transmission avec tympan anormal 105
4.2.2 Etiologie des surdités de perception 105
4.2.2.1 Les surdités endocochléaires 106
4.2.2.2 Les surdités rétrocochléaires 107
4.3 Prothèses auditives 107
4.3.1 Les prothèses auditives externes 108
4.3.1.1 Les prothèses à conduction aérienne 108
4.3.1.2 Les prothèses à conduction osseuse 109
4.3.2 Les prothèses implantables 110
4.3.2.1 Les prothèses ossiculaires 110
4.3.2.2 Les prothèses à ancrage osseux 110
4.3.2.3 Les prothèses vibratoires d'oreille moyenne (Vibrant) 110
4.3.2.4 Les implants cochléaires 111
4.3.2.5 Les implants du tronc cérébral 111
5 Les implants cochléaires 113
5.1 Historique 113
5.2 Technique 114
5.2.1 Principe et fonctionnement 114
5.2.1.1 Partie externe 115
5.2.1.2 Partie interne 116
5.2.2 Stratégies de codage 119
5.2.2.1 Les stratégies temporelles 119
5.2.2.2 Les stratégies spectrales 120
5.2.2.3 Les stratégies hybrides 120
5.2.2.4 D'autres stratégies nécessaires ? 121
5.3 L'implantation cochléaire 122
5.3.1 Critères d'implantation 122
5.3.1.1 Critères audiométriques 123
5.3.1.2 Critères oto-rhino-laryngologiques 123
5.3.1.3 Critères psychologiques et environnementaux 125
5.3.2 L'étape chirurgicale 125
5.3.3 Suivi et rééducation 127

5.4 Conséquences de l'implantation cochléaire sur les performances auditives des patients ……128
5.4.1 Le langage 128
Facteurs pronostiques 129
5.4.2 Les informations paralinguistiques 131
5.4.3 Les sons de l'environnement 132
5.4.4 La musique 132
6 Réorganisation 135
6.1 Effet de la déprivation auditive sur l'organisation neurofonctionnelle 135
6.2 Effet de l'implantation cochléaire sur l'organisation neurofonctionnelle 137
6.2.1 Réactivation des voies auditives centrales 137
6.2.2 Réorganisation des voix auditives 138
CHAPITRE 3 : ÉTUDE EXPÉRIMENTALE DE LA PERCEPTION DE LA VOIX APRES IMPLANTATION COCHLÉAIRE 141
7 Problématique 143
8 Méthodes 145
8.1 Traitement des stimuli comportementaux 145
8.1.1 Génération de stimuli 145
8.1.2 Simulation acoustique de l'implant cochléaire ou vocoder 145
8.2 Analyses 146
8.2.1 Rappels sur la théorie de détection du signal 146
8.2.2 Statistiques 149
9 Tests comportementaux de la reconnaissance de la voix humane chez le sujet sain 151
9.1 Test de détection de la voix humaine 151
9.1.1 Voix humaine versus sons environnementaux, stimulation binaurale 152
9.1.1.1 Matériel et méthodes 152
9.1.1.2 Résultats 155
9.1.1.3 Résumé et discussion préliminaire 161
9.1.2 Voix humaine versus sons de l'environnement, stimulation monaurale 162
9.1.2.1 Résultats 163
9.1.2.2 Résumé et discussion préliminaire 167
9.1.3 Voix humaine versus voix animale 167
9.1.3.1 Matériel et méthodes 167
9.1.3.2 Résultats 170
9.1.3.3 Résumé et discussion préliminaire 174
9.2 Tests comportementaux de la reconnaissance de l'identité de la voix 175
9.2.1 Test de catégorisation du genre 175
9.2.1.1 Matériel et méthodes 175
9.2.1.2 Résultats 179

9.2.1.3 Résumé et discussion préliminaire 187
9.2.2 Test de discrimination entre deux voix (du continuum) 189
9.2.2.1 Matériel et méthodes 189
9.2.2.2 Résultats 191
9.2.2.3 Résumé et discussion préliminaire 194
9.2.3 Test de la reconnaissance de l'âge 195
9.2.3.1 Matériel et méthodes 195
9.2.3.2 Resultats 198
9.2.3.3 Résumé et discussion préliminaire 201
9.3 Discussion 201
9.3.1 Synthèse des résultats 201
9.3.2 Latéralité du traitement de la voix 203
9.3.3 Le statut paticulier du langage 203
9.3.4 Différents indices acoustiques pour différentes tâches 204
9.3.5 Conclusion 205
10 Tests comportementaux de la reconnaissance de la voix chez le patient implanté cochléaire 207
10.1 Test de détection de la voix humaine 209
10.1.1 Voix humaine versus sons environnementaux 209
10.1.1.1 Matériel et méthodes 209
10.1.1.2 Résultats du groupe transversal 210
10.1.1.3 Résultats du groupe en suivi 217
10.1.1.4 Rôle de la prothèse auditive controlatérale sur les performances avec l'implant seul 228
10.1.2 Résumé et discussion préliminaire 231
10.2 Tests comportementaux de la reconnaissance de l'identité de la voix 233
10.2.1 Test de catégorisation du genre 233
10.2.1.1 Matériel et méthodes 233
10.2.1.2 Résultats 236
10.2.1.3 Rôle de la prothèse controlatérale sur les performences avec l'implant seul 270
10.2.2 Test de discrimination entre deux voix 273
10.2.2.1 Résultats 274
10.3 Discussion 276
10.3.1 La détection de la voix 276
10.3.2 La catégorisation du genre 277
10.3.3 La discrimination entre deux voix 279
10.3.4 Dissociation de récupération entre voix et langage 280
10.3.5 Intérêt du suivi 281

10.3.6 Effet de l'audition résiduelle sur la perception des informations paralinguistiques 282
10.3.7 Implications technologique et clinique 282
11 Étude d'imagerie fonctionnelle chez le sujet sain en condition de simulation d'implant 285
11.1 Méthodes 285
11.1.1 Sujets 285
11.1.2 Stimuli 285
11.1.3 Protocole expérimental 286
11.1.3.1 Mesures comportementales 286
11.1.3.2 Voice localizer 287
11.1.3.3 Activation de l'aire de la voix avec le vocoder 287
11.1.3.4 Acquisition des images 288
11.1.3.5 Prétraitement des images 288
11.1.3.6 Analyses IRMf 288
11.2 Résultats 289
11.2.1 Résultats comportementaux 289
11.2.1.1 Tâche voix versus non-voix 289
11.2.1.2 Tâche mot versus non-mot 291
11.2.2 Voice localizer 292
11.2.3 Effet de la simulation d'implant par le vocoder 295
11.2.3.1 Réponse aux stimuli originaux 295
11.2.3.2 Contraste Voix-Non Voix et vocoder 297
11.2.3.3 Réponse aux stimuli dégradés par le vocoder 299
11.2.3.4 Corrélation avec les données comportementales 301
11.3 Discussion 302
12 Discussion générale et conclusion 309
12.1 Synthèse des résultats 309
12.2 Discussion et perspectives 310
12.3 Conclusion 313
13 Références 315
14 Annexes 337
15 Publications 347

CHAPITRE 1 : L'AUDITION, LES SONS ET LA VOIX

1 L'AUDITION

Le système auditif est apparu très tôt dans l'évolution. En effet, « la sortie des eaux » des vertébrés s'est accompagné de l'apparition d'un système auditif très rudimentaire, avec notamment l'apparition de l'oreille interne qui était une structure interne du crâne et à qui les ondes sonores étaient le plus souvent transmises par des vibrations de la peau, des os et des muscles (Tanzarella & Mamecier, 2006). Puis est apparu, chez les batraciens l'oreille moyenne et enfin l'oreille externe très développée chez les mammifères. Chez les vertébrés, s'il existe des espèces animales aveugles, aucune espèce sourde n'a encore été découverte. L'audition a joué un rôle très important dans notre adaptation à l'environnement. En effet, elle aurait permis de développer notamment une meilleure perception de l'espace et une meilleure localisation des différentes sources sonores, mais est aussi le pré-requis essentiel pour le développement de la communication intra et inter-espèces.

Aujourd'hui chez l'homme, l'audition est un sens très privilégié. Par exemple certaines personnes peuvent avoir un odorat inexistant ou extrêmement diminué, on parle alors d'anosognosie, sans que cela soit un handicap important ou que cela les privent de leur autonomie. Certaines personnes anosognosique le sont sans même s'en apercevoir. En ce qui concerne l'audition, il paraît évident qu'on ne peut perdre l'ouïe sans s'en apercevoir ou sans que son environnement s'en aperçoive. L'audition est un pilier de la communication humaine, des relations sociales et joue également un rôle important dans la perception de l'espace. Dans ce chapitre nous essaierons de décrire et d'expliquer les mécanismes qui sous-tendent le bon fonctionnement de l'appareil auditif, puis de l'onde sonore jusqu'à la sensation auditive.

1.1 LA STRUCTURE DU RECEPTEUR AUDITIF

Le récepteur auditif, appelé communément oreille, est composé de plusieurs structures que l'on a tendance à distinguer en 3 parties : l'oreille externe, l'oreille moyenne et l'oreille interne. Cette distinction se fait sur la base du mode de transmission de l'information acoustique, qui arrive en onde acoustique et va être transformée en information nerveuse pouvant être interprétée par le cerveau.

Nous verrons dans cette partie que les rôles essentiels de ce récepteur auditif, outre celui de transmetteur, sont des rôles d'amplificateur, de transformateur de l'onde acoustique, éventuellement de protecteur si le son est trop « fort » et enfin de récepteur et décomposeur en ce qui concerne l'oreille interne. Nous verrons aussi que ces récepteurs auditifs sont capables d'expression.

1.1.1 OREILLE EXTERNE

L'oreille externe reçoit une information acoustique brute sous forme d'onde sonore et son rôle est d'assurer la bonne transmission de cette information jusqu'au tympan.

Les structures qui composent l'oreille externe (voir Figure 1.1) sont le pavillon et le conduit auditif.

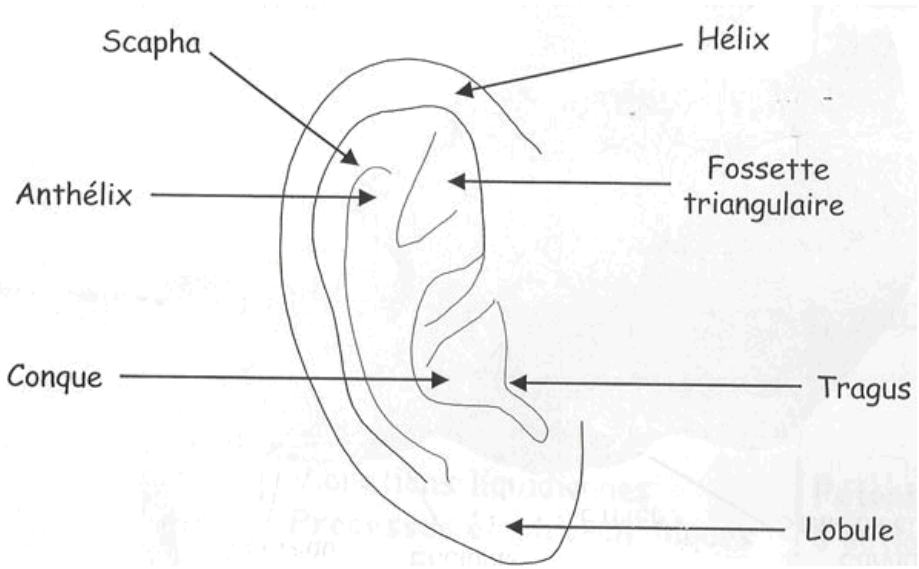


Figure 1.1 : Schéma anatomique du pavillon de l'oreille externe. Le conduit auditif jusqu'au tympan n'est pas représenté mais appartient tout de même à l'oreille externe.

Le pavillon, mesurant 60 à 65 mm pour la hauteur et 25 à 36 mm pour la largeur, est constitué par un cartilage recouvert de peau. Il présente de nombreux replis qui captent les sons et les dirigent, on dit qu'il a un rôle de réflecteur. Le conduit auditif est un petit tuyau de 2 à 3 cm de long et de 0,5 à 1 cm de diamètre, creusé dans l'os temporal. La forme du conduit cylindrique permet un phénomène de résonance, ou d'amplification. Mais on peut d'ors et déjà parler de filtre en ce qui concerne l'oreille externe, car s'il y a bien amplification des fréquences, cette amplification s'applique tout particulièrement aux fréquences les plus significatives pour l'homme, notamment pour le langage. L'amplification est la plus importante, soit 10 à 15 dB, pour les fréquences comprises entre 1 500 et 7 000 Hz, particulièrement lorsque l'onde est émise latéralement dans l'axe du conduit auditif (voir Figure 1.2).

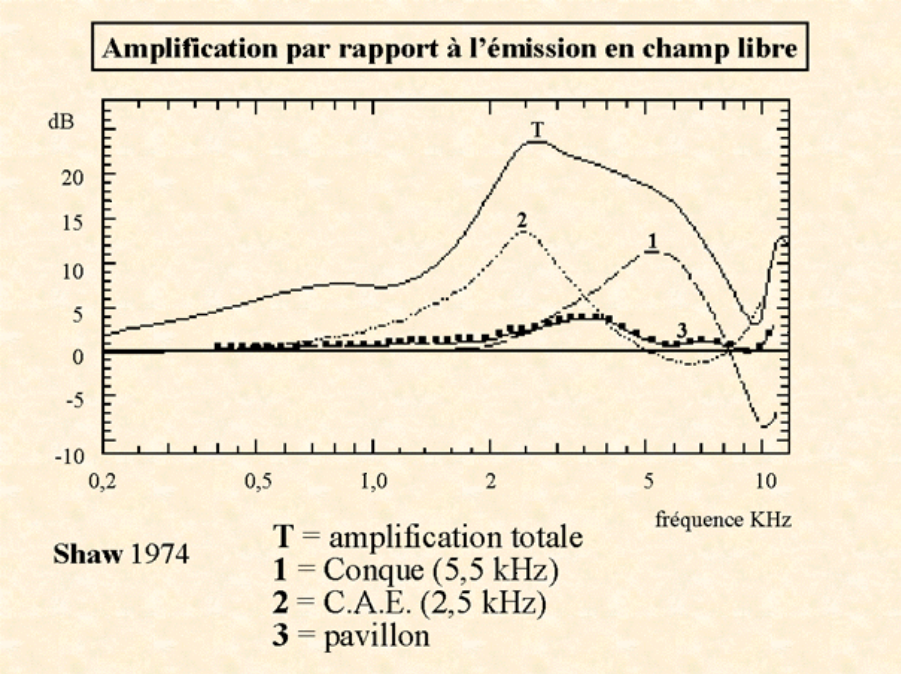


Figure 1.2 : Fonction d'amplification par différentes parties de l'oreille externe : la conque (1), le conduit auditif externe (2) et le pavillon (3). La courbe (T) représente l'amplification totale des fréquences. Graphique adapté d'après Shaw *et al* (1974) (Shaw, 1974).

1.1.2 OREILLE MOYENNE

L'oreille moyenne est la première étape de transformation de l'énergie sonore : l'information reçue par le tympan va être transformée en information mécanique et acheminée jusqu'à l'oreille interne.

L'oreille moyenne (voir Figure 1.3) est une petite cavité remplie d'air située dans l'os temporal qui se constitue du tympan, de 3 osselets (le marteau, l'enclume et l'étrier) et de 2 muscles striés. Elle communique avec l'oreille interne via les deux ouvertures membraneuses que sont la fenêtre ronde et la fenêtre ovale. Elle communique également avec le pharynx par la trompe d'Eustache. Généralement fermée par une valve, cette trompe peut s'ouvrir lors du bâillement ou de la déglutition, afin d'équilibrer la pression de l'air de part et d'autre du tympan, soit entre l'oreille moyenne et le milieu ambiant.

Le tympan est une fine membrane qui vibre en fonction des ondes sonores qu'il reçoit. Le marteau est, d'un côté articulé sur le tympan, afin que toutes les vibrations de ce dernier lui soient transmises et de l'autre fixé à l'enclume par un lien rigide. Le dernier maillon de la chaîne des osselets, l'étrier, est relié à l'enclume par une rotule très mobile. La partie plate de l'étrier repose sur la fenêtre ovale et s'articule à la façon d'un piston pour transmettre les vibrations aux fluides de l'oreille interne.

Avec 55mg à eux 3, les osselets sont les plus petits os du corps humain et ont deux fonctions principales. Tout d'abord, ils assurent un rôle d'amplification de l'onde sonore pour qu'elle soit transmise et audible dans le milieu aqueux de la cochlée où l'impédance est plus élevée.

L'amplification est rendue possible d'une part grâce au principe de levier selon lequel les osselets fonctionnent mais également grâce au fait que la pression appliquée à la surface tympanique relativement étendue soit reportée à la surface beaucoup plus petite de la fenêtre ronde. Ces deux processus permettent de donner une pression environ 200 fois plus élevée sur la fenêtre ovale que sur le tympan.

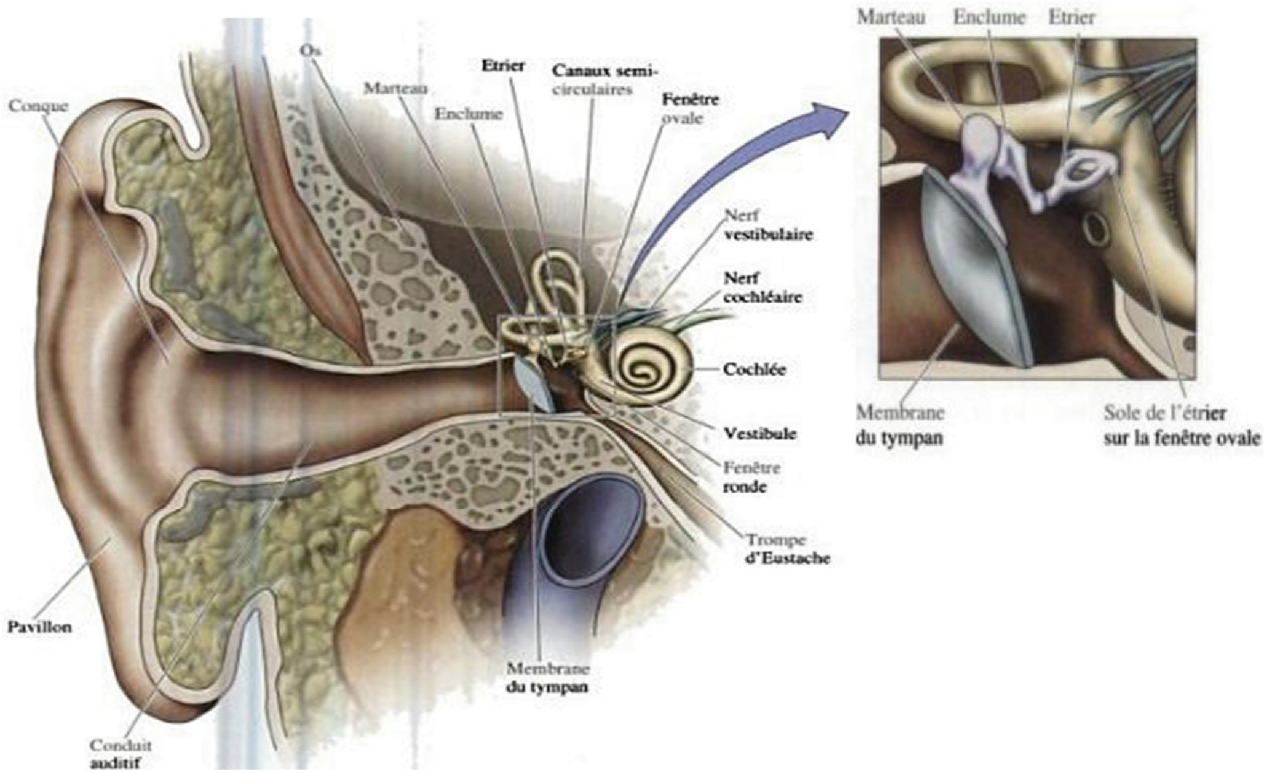


Figure 1.3 : Schéma anatomique de l'oreille (à gauche) et vue grossie (à droite) sur l'oreille moyenne comprenant le tympan, les osselets (marteau, enclume et étrier), la fenêtre ovale et la fenêtre ronde

Mais également, en cas d'amplitude sonore intense, l'oreille moyenne tient aussi un rôle d'atténuation pour préserver ce mécanisme des traumatismes. Le réflexe d'atténuation fait intervenir deux petits muscles : le muscle tenseur du tympan, fixé au marteau d'un côté et à l'os de la cavité de l'autre, et le muscle stapedius, fixé entre l'étrier et l'os. En présence d'un stimulus sonore fort, l'amplitude des mouvements des osselets est très grande et pourrait endommager le récepteur auditif. L'amplitude de ces mouvements est sous le contrôle de ces muscles qui, une fois contractés, rendent la chaine des osselets moins mobiles et protège l'oreille interne des sons trop violents en absorbant une partie de l'énergie sonore. Ce réflexe d'atténuation se produit lorsqu'un son violent et soudain apparait, mais il n'est malheureusement effectif que 50 à 100 ms après le début du son. Il ne peut donc pas empêcher les dommages causés pendant ce laps de temps. Il semble qu'il soit également

présent lorsque nous parlons, de sorte que nous ne puissions pas entendre notre propre voix aussi forte qu'elle est émise.

1.1.3 OREILLE INTERNE : COCHLEE

L'oreille interne est considérée comme une partie de l'oreille mais tous ses constituants ne sont pas en relation avec le système auditif. En réalité seule la cochlée est impliquée dans l'audition et retiendra notre attention (voir Figure 1.4). C'est grâce à celle-ci que l'information sonore mécanique transmise par l'oreille moyenne va être transformée en influx nerveux. Son mécanisme peut être comparé à celui du prisme pour la lumière ; en effet la cochlée va décomposer les sons complexes en les différentes fréquences qui la composent.

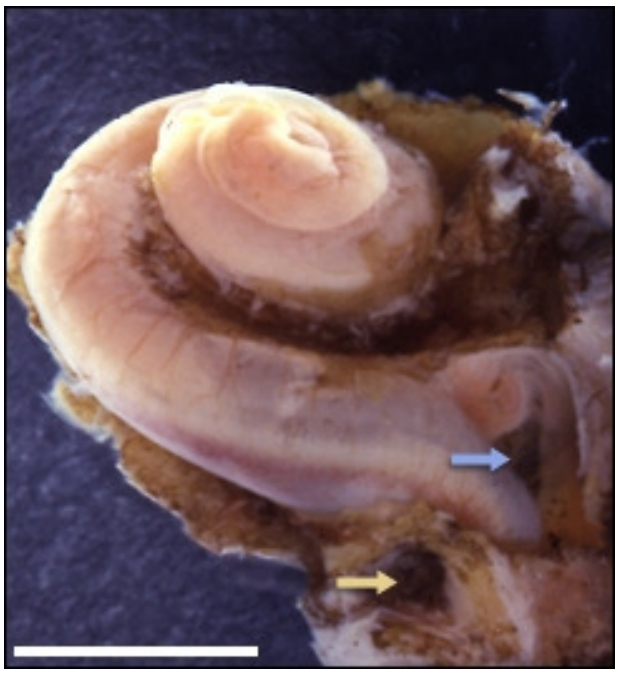


Figure 1.4 : Cochlée de fœtus humain de 5 mois vue en microscope électronique. On peut voir les 2,5 tours de spire du labyrinthe membraneux (35 mm de longueur). Les fenêtres ovale (flèche bleue) et ronde (flèche jaune) sont mentionnées. Notons que dès ce stade précoce, le développement morphologique de la cochlée est achevé. D'après Pujol *et al.* (1999)

1.1.3.1 ANATOMIE DE LA COCHLEE

La cochlée (ou limaçon) se présente sous forme d'un tunnel, le tube cochléaire, enroulé sur deux tour et demi autour d'une structure conique occupée par les fibres auditives, le modiolus (voir Figure 1.5). La cochlée mesure environ 32 mm de long pour un diamètre de 2 mm.

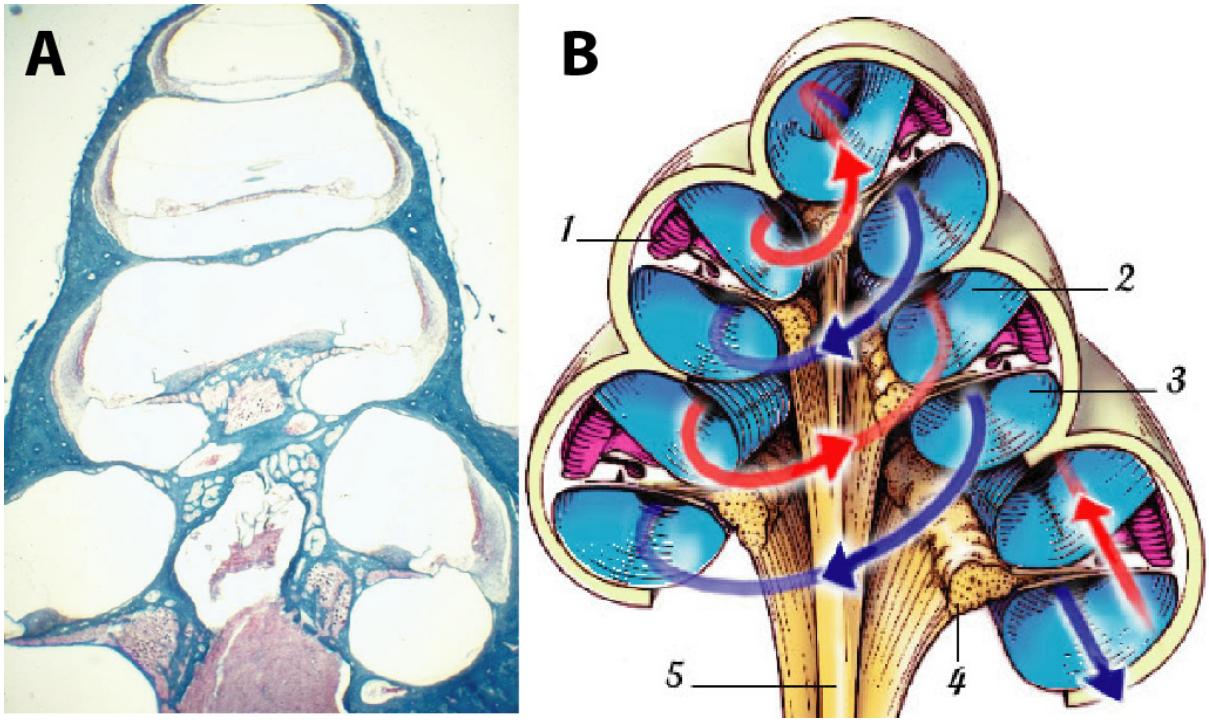


Figure 1.5 : Coupe axiale de la cochlée. (A) Vue au microscope électronique. (B) Schéma. On peut voir l'enroulement du canal cochléaire (1) contenant l'endolymphe, et celui des rampes vestibulaire (2) tympanique (3) contenant la périlymphe. La flèche rouge vient de la fenêtre ovale et la bleue aboutit à la fenêtre ronde. Au centre, (modiolus) le ganglion spiral (4) et les fibres du nerf cochléaire (5) apparaissent en jaune. D'après Pujol *et al.* (1999).

A la base de la cochlée se trouvent les deux ouvertures membraneuses, la fenêtre ovale qui est liée à la partie plate de l'étrier, et la fenêtre ronde. Une coupe transversale de la cochlée nous permet de voir qu'elle est constituée de 3 compartiments : la rampe vestibulaire, le canal cochléaire et la rampe tympanique (voir Figure 1.6). La rampe tympanique et la rampe vestibulaire contiennent toutes deux la périlymphe, qui est riche en ions sodium. Le canal cochléaire contient l'endolymphe, qui est riche en ions potassium. Cette différence de composition des liquides entre les différentes rampes provoque une différence de potentiel d'environ 80 mV. Ce potentiel est dit potentiel endocochléaire, et il représente un facteur important pour la transduction auditive.

A l'apex de la cochlée, le canal cochléaire se referme, tandis qu'un orifice dans la membrane basilaire permet de relier les rampes vestibulaires et tympaniques. Cet orifice s'appelle l'hélicotrème. La périlymphe des rampes tympaniques et vestibulaires sont donc en continuité.

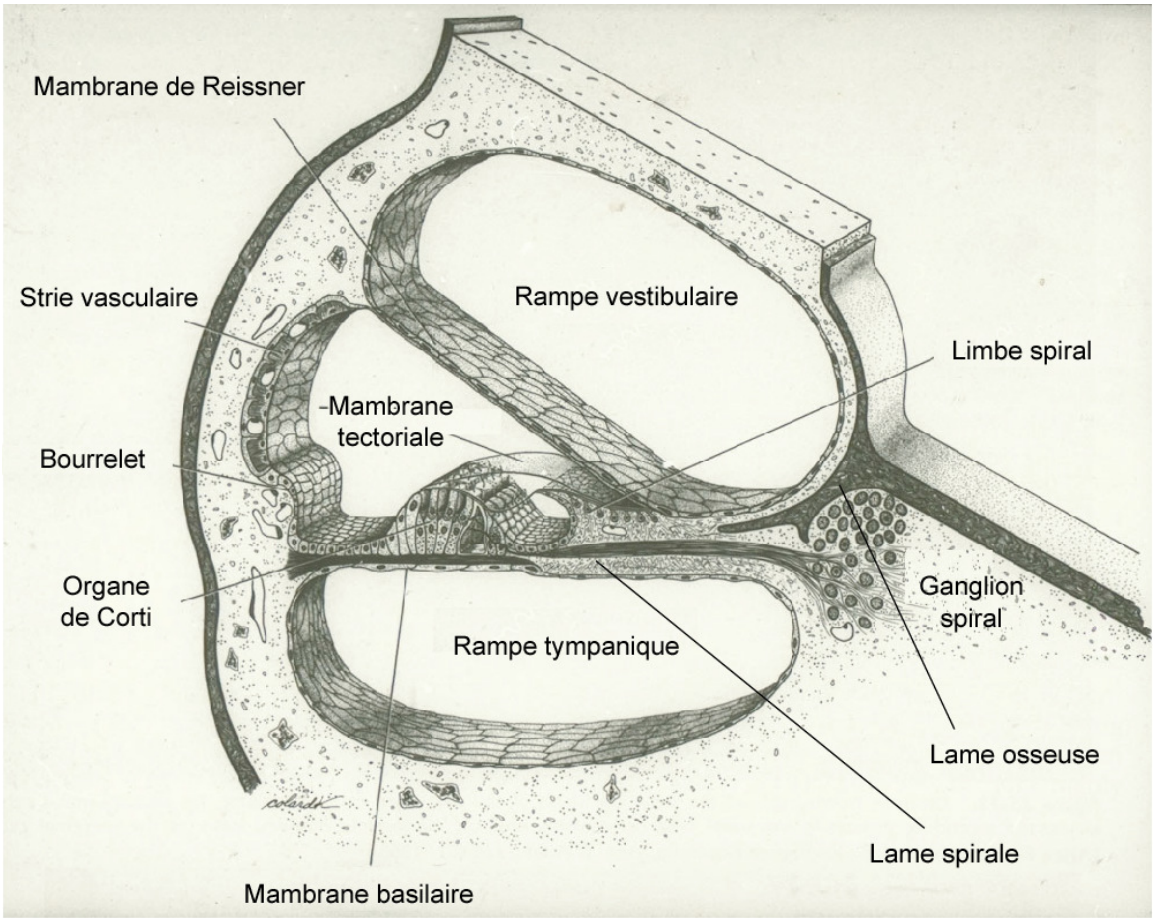


Figure 1.6 : Schéma représentant la coupe d'un tour de spire de la cochlée.

La rampe tympanique et la rampe vestibulaire sont séparées par la lame osseuse spirale à proximité du modiolus et par le canal cochléaire à proximité de la périphérie. Le canal cochléaire est séparé de la rampe vestibulaire par la membrane de Reissner, tandis que la membrane basilaire le sépare de la rampe tympanique. La paroi externe du canal cochléaire comporte la strie vasculaire, qui participe à l'élaboration de l'endolymphe, et le bourrelet, dont on pense qu'il participe à la résorption de l'endolymphe. La paroi inférieure du canal cochléaire est constituée par l'extrémité de la lame spirale, sur laquelle repose le limbe spiral, et par la membrane basilaire, sur laquelle repose l'organe de Corti. Le limbe spiral est un amas fibro-conjonctif recouvert par un épithélium dont les cellules supportent l'extrémité interne de la membrane tectoriale qui recouvre l'organe de Corti. La membrane basilaire est constituée de fibres rayonnant à partir de la lame spirale et noyées dans une substance amorphe. Sur sa face supérieure repose l'organe de Corti, où siègent les neurones récepteurs auditifs. Dans l'épaisseur de la membrane basilaire, sous l'organe de Corti, chemine un volumineux capillaire sanguin. Bien que la cochlée se rétrécisse progressivement de la base à l'apex, la membrane basilaire s'élargit en allant vers l'apex.

1.1.3.2 ORGANE DE CORTI

L'organe de Corti est composé de plusieurs types de cellules dites cellules de soutien, et par des cellules sensorielles accessoires.

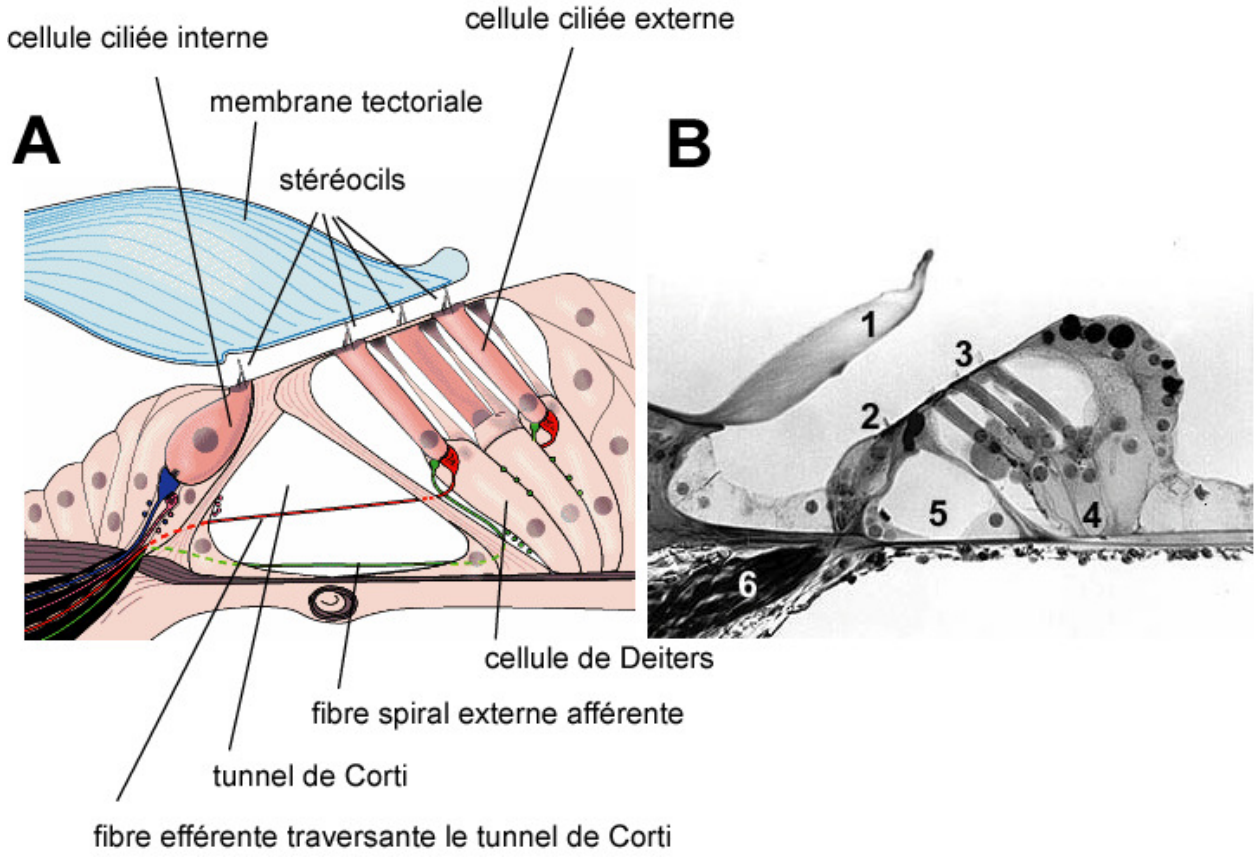


Figure 1.7 : Schéma (A) et vue au microscope (B) de l'organe de corti. On peut distinguer les cellules ciliées internes (2) et externes (3), le tunnel de Corti (5), les cellules de Deiters (4). On peut également voir la membrane tectoriale (1), les fibres nerveuses (6) et les stéréocils aux extrémités des cellules ciliées.

Les cellules sensorielles sont les cellules ciliées qui sont situées de part et d'autre du tunnel de Corti (voir Figure 1.7). Les parois de ce dernier sont constituées de deux rangées de cellules rigides, "les Piliers de Corti", s'appuyant l'une à l'autre par leur sommet. Entre les piliers de Corti, se trouve de la cortilymphe, dont la composition ionique est voisine de celle de la périlymphe. Les cellules ciliées sont disposées en rangées parallèles de la base au sommet de la cochlée : une rangée de cellules ciliées internes (CCI), situées entre le modiolus et le tunnel de Corti, et trois rangées de cellules ciliées externes (CCE) situées au-delà du tunnel de Corti (voir Figure 1.8). Les 4 rangées de cellules sont régulièrement espacées, tous les 10 µm, le long de la membrane basilaire. Les cellules ciliées ne se renouvellent pas. L'organe de Corti est donc très fragile.

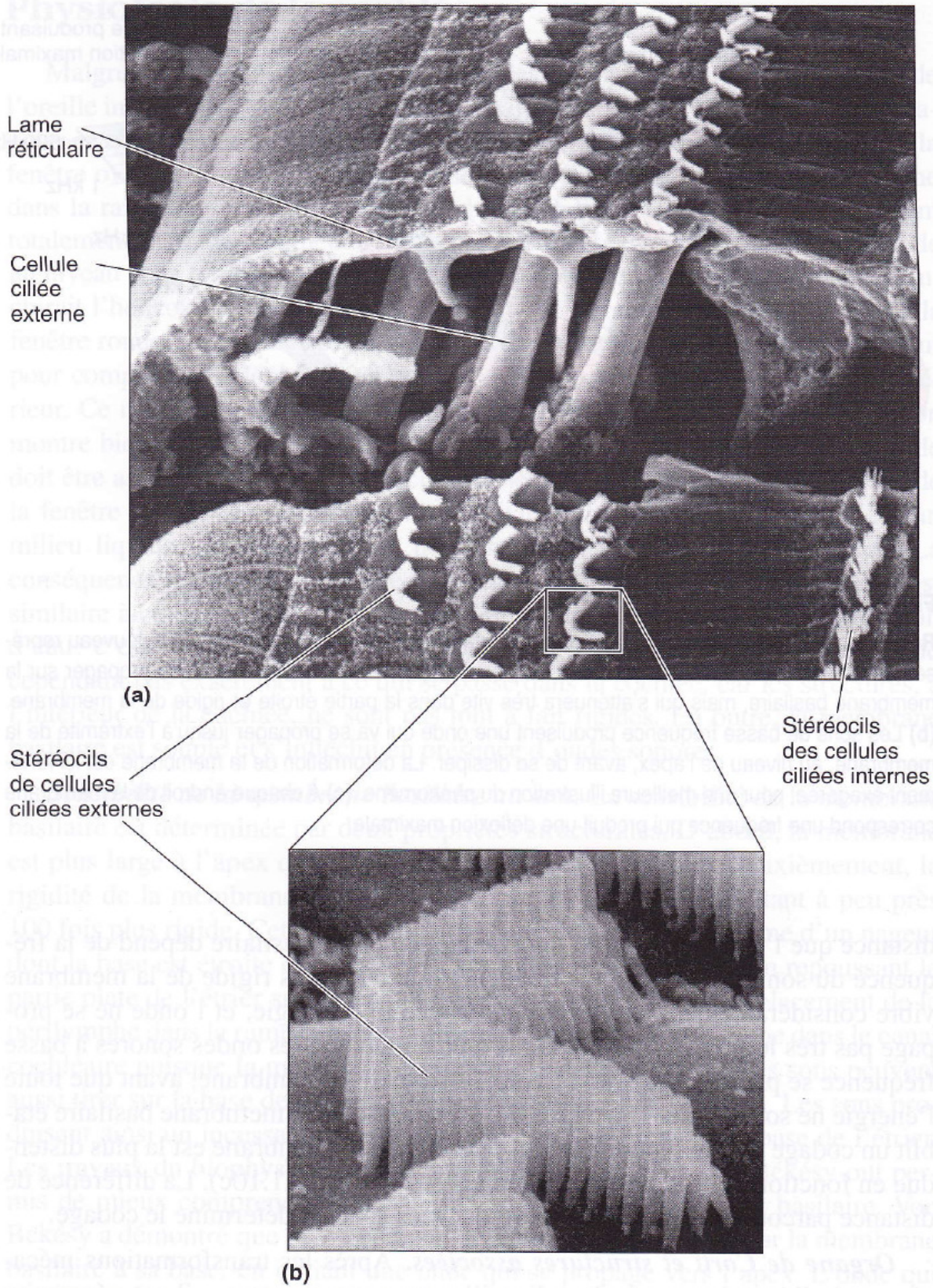


Figure 1.8 : Cellules ciliées observées au microscope à balayage. (a) Les cellules ciliées et leurs stéréocils. (b) Microphotographie à haute résolution des stéréocils d'une cellule ciliée externe. Les stéréocils ont une longueur d'environ 5 pm. Extrait de *Neurosciences : A la découverte du cerveau* (Bear, Connors, & Paradisio, 2002).

Les cellules ciliées externes (CCE), aussi appelées cellules de Corti, existent au nombre de 12 à 15 000. A leur base elles reposent fermement sur des cellules de soutien, les cellules de Deiters, qui forment un siège sans les entourer. Les CCE sont disposées en 3 rangés et sont séparées les unes des autres par les espaces de Nuel. Leur taille varie le long de la cochlée, avec une taille croissante de la base vers l'apex. La partie qui se raccorde à la membrane réticulaire porte environ 80 stéréocils (disposés en V ou W), eux-mêmes disposés en 3 ou 4 rangées de taille croissante (voir Figure 1.8, (b)) et immobiles. Ces cils viennent s'insérer dans la membrane tectoriale, où ils laissent des empreintes en creux parfaitement visibles. Les stéréocils sont liés entre eux par des liens dont l'intégrité fonctionnelle est primordiale. Au niveau de la base de ces cellules, deux types de fibres nerveuses font synapse. Les premières, voies afférentes, sont les terminaisons dendritiques des cellules nerveuses sensorielles situées dans le ganglion spiral, ou ganglion de Corti. Une proportion d'environ 5 000 neurones bipolaires reçoit les informations des 15 000 CCE. Les deuxièmes types de fibres nerveuses sont les terminaisons de fibres olivo-cochléaires effectrices, évaluées à 500 pour l'ensemble de l'organe de Corti.

Les 3 500 cellules ciliées internes (CCI), reposent également sur des cellules de soutien, qui entourent complètement leurs parois latérales. Au niveau de la lame réticulaire, une centaine de stéréocils sont alignés, en forme de U, en 3 ou 4 rangées de taille croissante. Ils plongent en totalité dans l'endolymphe. Pour les cellules ciliées internes, la liaison avec la membrane tectoriale est réalisée indirectement, de manière assez lâche par un système de filaments. Les CCI seront activées secondairement à l'amplification active par les CCE. A leur base, c'est une proportion très importante (environ 90 %) des neurones du ganglion spiral qui font synapse. Les synapses avec les cellules du ganglion spiral sont plus nombreuses en ce qui concerne les CCI. Le nombre des neurones connectés, ainsi que la position protégée des cellules auditives internes explique la meilleure résistance de la partie interne de l'organe de Corti aux agressions sonores et au vieillissement.

1.1.3.3 FONCTIONNEMENT DE LA COCHLEE

Lorsque les osselets entrent en mouvement, ils transmettent leurs vibrations à la fenêtre ovale. En raison de l'incompressibilité de la périlymphe, cette vibration se transmet alors à la périlymphe dans la rampe tympanique, puis à la rampe vestibulaire et enfin à la fenêtre ronde. Mais, la membrane basilaire étant souple, elle va également s'infléchir en présence des ondes sonores. La partie de la membrane qui va entrer en vibration avec les ondes sonores dépend d'une part de la fréquence du son et d'autre part des propriétés de la membrane basilaire. Comme cela a déjà été dit, la membrane basilaire s'élargie de la base à l'apex, de même que sa rigidité diminue. Ces paramètres impliquent donc que ce sera la base de la membrane basilaire qui va vibrer en réponse aux hautes fréquences, tandis que les ondes sonores à basses fréquences se propageront jusqu'à l'apex pour

faire vibrer la membrane (voir Figure 1.9). Le mouvement ondulatoire de la membrane basilaire est maximal dans la région où sa fréquence de résonance est identique à celle de l'onde sonore. A ce point précis, la membrane basilaire ondule avec une si grande facilité que toute l'énergie de l'onde sonore est dissipée. La réponse nerveuse de la membrane basilaire se fera en fonction des différents sites qui auront vibré en réponse aux ondes sonores. Chaque fréquence, ou distance sur la membrane basilaire, sera codée indépendamment. Cette répartition continue des fréquences le long de la cochlée est appelée tonotopie. Cette représentation tonotopique des sons sera conservée à tous les étages du traitement sous-cortical et cortical du son.

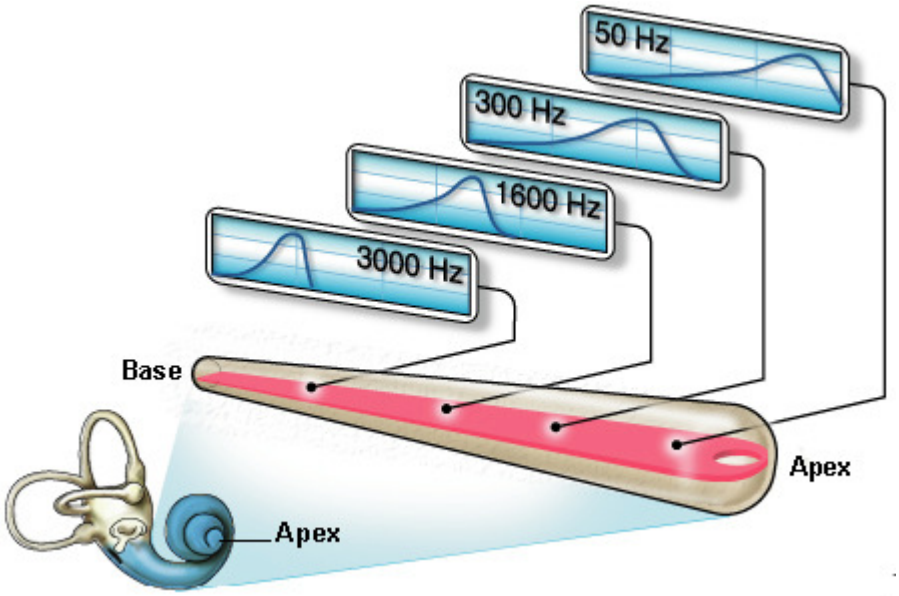


Figure 1.9 : Répartition de l'amplitude des vibrations de la membrane basilaire en fonction de la fréquence de stimulation. D'après Pujol *et al* (1999)

Une fois que la membrane basilaire est entrée en vibration, ses déformations vont se communiquer aux cils de l'organe de Corti. La déformation de la membrane basilaire vers le haut va entraîner une inclinaison des cils vers l'extérieur de la cochlée, tandis qu'une déformation vers le bas va entraîner une inclinaison vers l'intérieur (voir Figure 1.10). Les stéréocils sont reliés par des filaments, de sorte que tous les cils d'une cellule se déplacent ensemble. Lorsque la membrane basilaire se déplace, et que les cils sont inclinés, cela va entraîner l'ouverture des canaux potassiques, situés aux extrémités des cils. Hyperpolarisation et dépolarisation des cellules ciliées vont alors se succéder en fonction du sens de déformation des cils, autant que durera la vibration de l'onde sonore. Ces potentiels d'action vont ensuite être transmis aux fibres nerveuses cochléaires correspondantes, stimulant ainsi le nerf auditif.

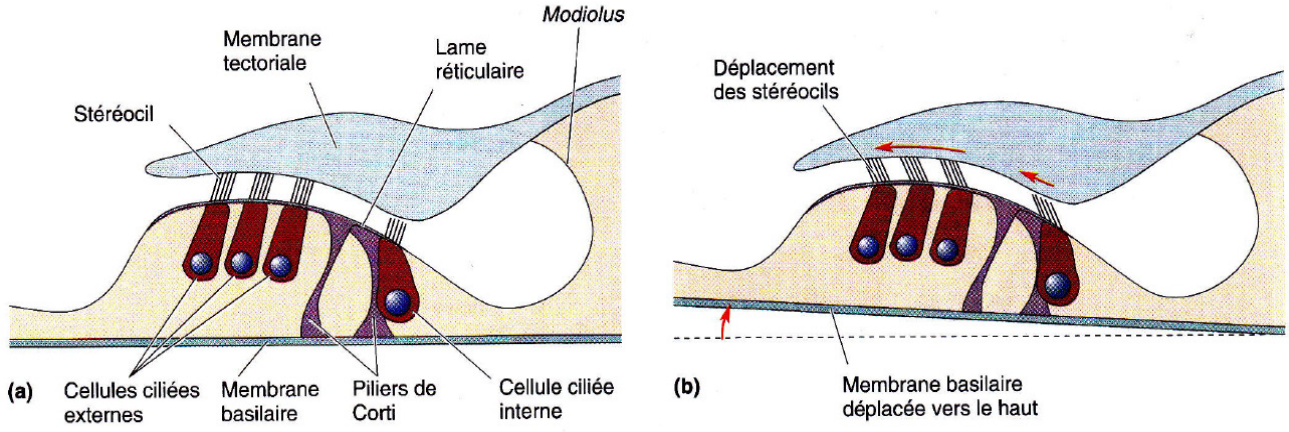


Figure 1.10 : Mouvements des stéréocils produits par le déplacement de la membrane basilaire. **(a)** Au repos, les cellules ciliées sont maintenues entre la lame réticulaire et la membrane basilaire, et l'extrémité des stéréocils des cellules ciliées externes est attachée à la membrane tectoriale. **(b)** Quand un son provoque une déformation vers le haut de la membrane basilaire, la lame réticulaire se déplace vers le haut et vers la partie interne de la cochlée, entrainant un déplacement des stéréocils vers l'extérieur. Extrait de *Neurosciences : A la découverte du cerveau* (Bear et al., 2002).

Les différences fonctionnelles entre les cellules ciliées internes et les cellules cillées externes ne sont pas encore totalement établies. Toutefois, on sait que la majeure partie de l'information acoustique transmise au nerf cochléaire le sera par les CCI, ce qui semble paradoxal étant donné que les CCE sont 3 fois plus nombreuses. Jusqu'à présent on considère que les CCI vont transmettre au nerf cochléaire le message sonore en informations nerveuses. Elles vont alors coder fréquences et intensité. On dit que ce sont les véritables récepteurs sensoriels du système auditif. Mais on sait aujourd'hui que les CCE jouent un rôle majeur dans les processus cochléaires mécaniques actifs. Bien que les CCE ne soient à l'origine que de 5 à 10% des influx auditifs afférents, leur élimination induit une perte auditive et une diminution de la sélectivité fréquentielle des réponses des CCI. Les CCE semblent jouer un rôle d'amplificateur. Elles permettent également une meilleure discrimination des fréquences car elles amplifient le déplacement de la membrane basilaire dans une zone restreinte où les CCI sont activées.

La preuve de ces processus actifs des CCE a été apportée par la mise en évidence d'émission oto-acoustique (EOA) : enregistrement de sons dans le conduit auditif externe en dehors de toute stimulation cochléaire. En effet des mouvements liquidiens se propagent en réponses à ces contractions des cils des CCE, et sont à l'origine des EOA. Quand on émet un son très bref dans le conduit auditif externe, on peut enregistrer à l'aide d'un microphone situé dans ce conduit, des sons émis par la cochlée. Cet enregistrement des oto-émissions provoquées constitue un moyen simple et rapide de vérifier l'intégrité des mécanismes actifs. Il s'git d'une sorte d'écho. Dans des cas pathologiques, suite à des expositions à des sons trop intenses ou encore à l'absorption de substances ototoxique, ces productions acoustiques, si elles sont suffisamment fortes peuvent être à l'origine d'acouphènes.

1.2 MECANISMES CENTRAUX DE L'AUDITION

L'espèce humaine étant dotée de deux oreilles, situées de part et d'autre de la tête, nous verrons qu'elles sont en interaction l'une avec l'autre, et nous verrons aussi les bénéfices apportés par cette dualité.

1.2.1 ANATOMIE DES VOIES AUDITIVES

Le traitement de l'information auditive à partir de la cochlée est assez hiérarchisé. En partant de la cochlée, les voies auditives sont assez dispersées après les noyaux cochléaires, et toutes ces voies ne sont pas toujours très bien décrites. Nous avons choisi de décrire la voie auditive primaire, qui nous intéresse le plus, en partant de la cochlée et en allant jusqu'au cortex auditif (voir Figure 1.11).

1.2.1.1 RELAIS SOUS-CORTICAUX

LE GANGLION SPIRAL

Le ganglion spiral, ou ganglion de Corti se situe au niveau de la cochlée, et reçoit directement les informations provenant des cellules ciliées, puisqu'il contient les corps cellulaires des cellules qui font synapse avec les cellules ciliées. Les quelques 130 000 axones qui en partent forment le nerf cochléaire, principale composante du nerf VIII, et projettent de façon uniquement ipsi-latérale vers un premier relai, les noyaux cochléaires.

LES NOYAUX COCHLEAIRES

Les noyaux cochléaires se situent à un trajet d'environ 1 cm (chez l'homme) du ganglion spiral, dans le bulbe à l'entrée du tronc cérébral. Les noyaux cochléaires sont subdivisés en plusieurs zones qui répondent spécifiquement à certaines fréquences. Les basses fréquences sont traitées dans les parties ventrales, rostrales et latérales, tandis que les hautes fréquences sont traitées dans les parties dorsales, caudales et médiales (Galifret, 1991). On distingue généralement la partie antéroventrale et la partie postéroventrale. Les noyaux cochléaires vont projeter de façon bilatérale vers le complexe olivaire supérieur. Autrement dit une partie de l'information sera envoyé vers le complexe olivaire supérieur ispsilatéral et l'autre partie sera envoyé vers le complexe olivaire supérieur controlatéral.

LE COMPLEXE OLIVAIRE SUPERIEUR

Le complexe olivaire supérieur (COS) se situe dans le canal médullaire. Il est composé de neuf noyaux dont trois principaux constituent ce deuxième relais du système ascendant. Chaque COS reçoit des informations des deux oreilles, dites binaurales. A ce niveau s'effectue donc la première confrontation des messages peu modifiés depuis le ganglion de corti, en provenance des

deux oreilles. C'est dans ce COS qu'ont lieu les premiers traitements de localisation de la source sonore. Pour cela, la confrontation des afférences bilatérales va se faire sur des parties de l'information sonore différentes. Ces confrontations concernent les variations d'intensité binaurale, les différences d'intensité entre les deux oreilles et le délai interaural. De plus, des études ont montré que le traitement de ces différentes informations était supporté par différents noyaux COS. L'olive supérieure latérale serait sensible aux différences interaurales d'intensité entre les sons (Park, Grothe, Pollak, Schuller, & Koch, 1996), qui permettrait essentiellement de mesurer la distance de la source sonore, tandis que l'olive supérieure médiane serait plus sensible aux différences interaurales de temps d'arrivée des sons (Yin & Chan, 1990) ce qui va permettre de déterminer la direction de la source sonore.

LES NOYAUX DU LEMNISQUE LATERAL

Les axones des neurones du complexe olivaires supérieur empruntent le lemnisque latéral mais ne s'y arrêtent pas. Toutefois, les noyaux du lemnisque latéral reçoivent des informations monaurales en provenance des noyaux cochléaires controlatéraux Les neurones des ces noyaux sont sensibles à des indices tels que l'amplitude ou la durée du son. Comme celles qui proviennent des noyaux olivaires supérieurs, ces voies convergent vers le mésencéphale.

LE COLLICULUS INFERIEUR

Le colliculus fait partie du mésencéphale. On considère actuellement que toutes les voies auditives ascendantes convergent vers le colliculus inférieur. Il reçoit donc les projections directes bilatérales en provenance du complexe olivaire supérieur, des noyaux du lemnisque latéral, des noyaux cochléaires ou encore d'autres voies ascendantes que nous n'avons pas décrites. Grace à cette diversité d'afférences, un nouveau traitement va avoir lieu, à savoir une représentation topographique de l'espace auditif. La sensibilité aux disparités interaurales d'intensité ou de phase y est plus fine que dans le complexe olivaire supérieur. On dit qu'il est le centre de traitement de l'information spatiale. En outre, le colliculus inférieur poursuit l'analyse de la structure du signal sonore amorcée lors des étapes précédentes. Certains neurones sont également sensibles au sens de déplacement d'une source sonore par rapport au plan médian. On pense que le colliculus inférieur intervient dans l'orientation de la tête et des yeux en direction du son. Enfin le colliculus inférieur va projeter vers plusieurs structures, dont notamment le corps genouillé médian.

LE CORPS GENOUILLE MEDIAN

Le corps genouillé médian (CGM) fait partie du thalamus. Trois parties le composent : les parties dorsales et médiales entourent la partie ventrale. Cette dernière constitue le relai thalamo-cortical principal, c'est le relais auditif spécifique. Elle possède une organisation laminaire en couches concentriques, parallèles à la surface. Les couches superficielles correspondent aux basses fréquences, et les couches profondes aux hautes fréquences. Certains neurones présentent une sensibilité spécifique à des sons complexes particuliers et un nombre important sont inhibés par la stimulation acoustique. Comme dans le colliculus inférieur, de nombreuses cellules reçoivent des

informations des deux oreilles et sont sensibles aux disparités d'intensité ou de phase. En résumé, les neurones du corps genouillé médian reçoivent des afférences convergentes séparées du point de vue spectral et temporel, ce qui implique une intégration de ces deux aspects au niveau du CGM. La partie ventrale du CGM va projeter vers le cortex auditif primaire

Les parties médiale et dorsale du noyau répondent moins bien à la stimulation acoustique ; ces deux aires sont considérées comme des aires de convergence multisensorielle permettant l'établissement de liaison entre des informations hétérogènes. Ce phénomène de convergence multisensorielle est à la base des mécanismes d'intégration qui sous tendent les conduites perceptives. La partie dorsale va projeter vers le cortex auditif secondaire.

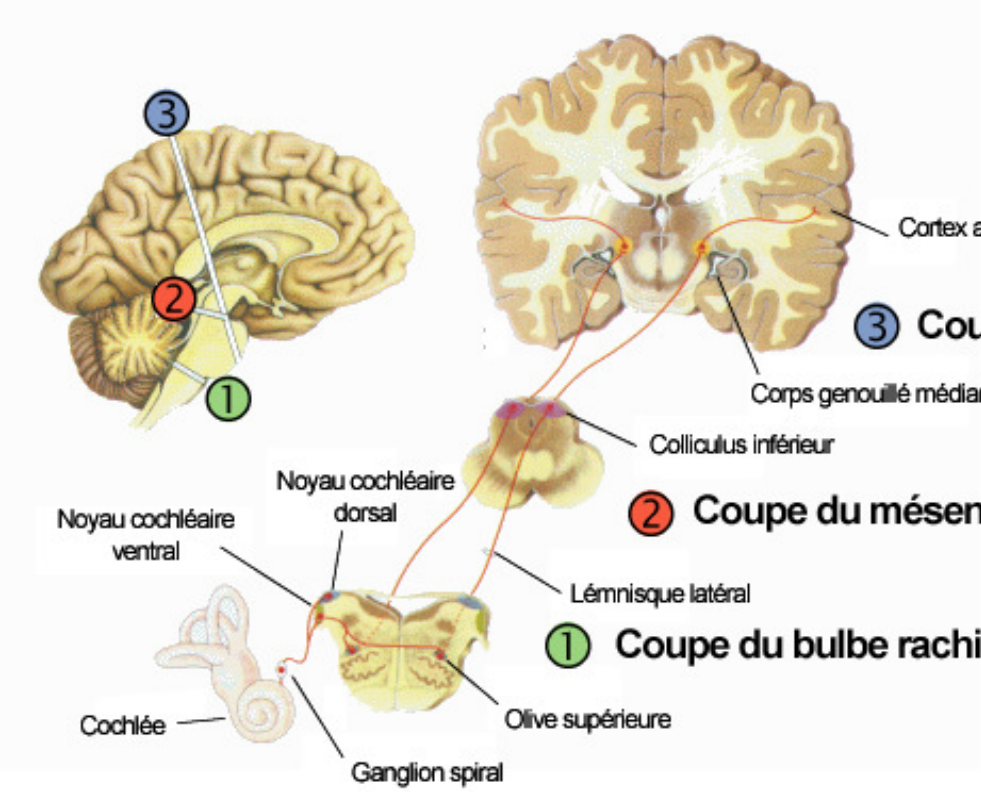


Figure 1.11 : Schéma d'ensemble des voies nerveuses auditives principales. Adapté d'après Bear *et al.* (2002).

1.2.1.2 LE CORTEX AUDITIF

Chez l'homme le cortex auditif est situé dans la partie postérieure supérieure du lobe temporal, profondément enfoui dans la scissure de Sylvius. On distingue classiquement une aire primaire (le cortex auditif primaire) et des ceintures d'aires périphériques (les cortex auditifs secondaires et associatif). Selon l'ancienne classification de Brodmann, ces aires impliquent les aires de Brodmann (BA) 41, 42 et 22 (voir Figure 1.11). La subdivision fonctionnelle comprend classiquement deux portions : le noyau (*core*) et la ceinture (*belt*). Le noyau correspond au gyrus de

Heschl (BA 41) alors que la ceinture (BA 42 et 22) correspond plutôt au cortex associatif. On distingue même la *belt* (BA 42) et la *parabelt* (BA 22).

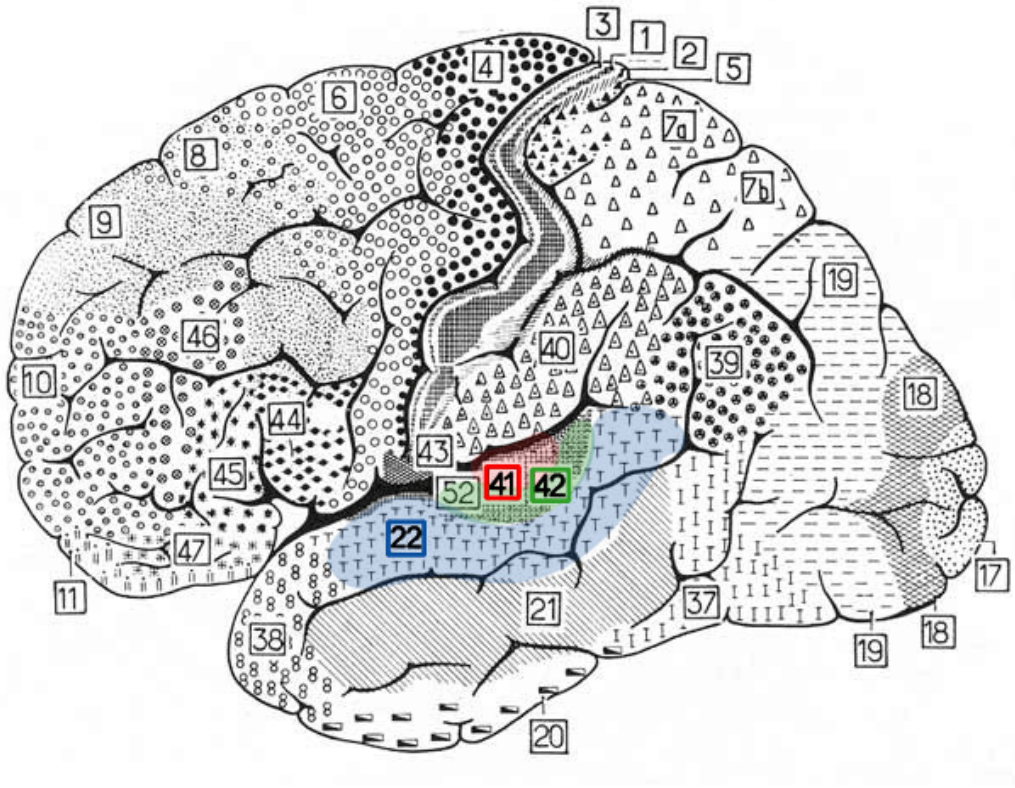


Figure 1.12 : Schéma des aires cytoarchitectoniques définies par Brodmann. On peut visualiser les aires de Brodmann 41 cortex auditif primaire), 42 et 22 (les cortex auditifs secondaire et associatif), impliquées dans l'analyse auditive.

LE CORTEX AUDITIF PRIMAIRE

Le cortex auditif primaire (AI) se situe le long de la face ventrale du lobe temporal supérieur, et plus précisément dans le gyrus transverse, ou gyrus de Heschl, et correspond à l'aire 41 de Brodmann (voir Figure 1.12 et Figure 1.13). Son aire de surface est d'environ 3,4 cm², mais cette surface moyenne est à relativiser car il existe des variations cytoarchitectoniques importantes entre les individus ; le cortex auditif primaire peut représenter au moins 30% et jusqu'à 80% du gyrus de Heschl selon les individus (Morosan et al., 2001). Le cortex auditif primaire est capable de répondre à des fréquences allant de 20 à 20 000 Hz. Mais on note que la majorité des neurones auditifs ont des courbes de sélectivité fréquentielle étroites lorsqu'ils sont situés dans l'aire AI. La représentation tonotopique sur cette aire y est donc la plus fine. De plus, on peut se rendre compte qu'A1 comporte des bandes isofréquencielles disposées selon un plan médiolatéral. En d'autres termes, il y a dans A1 des rangées transversales de neurones qui répondent tous à la même fréquence caractéristique. On constate que les bandes correspondant à ces fréquences sont de plus en plus longues, et que leur écart s'accroît à mesure que croît la fréquence, d'où une surreprésentation des fréquences élevées (tour basal de la cochlée). Les cellules appartenant à une bande donnée sont excitées par les deux oreilles (cellules EE), tandis que celle de la bande adjacente sont excitées par une oreille et inhibée par

l'autre (cellules EI). Les bandes EE et EI s'alternent tout le long de V1. Les colonnes EE ou EI sont dans chaque tranche correspondant à une fréquence, rassemblées en plusieurs ensembles, de telle manière que l'on peut tracer sur le cortex, quasi perpendiculairement aux lignes isofréquencielles, les limites de bande EE ou EI. Les interactions de type EE ou EI conditionnent la sensibilité d'environ 50 % des neurones de l'aire AI à la localisation du son dans l'espace. On peut donc considérer que, sur l'aire AI, la direction rostrocaudale concerne la discrimination des fréquences, et la direction orthogonale concerne la spatialisation.

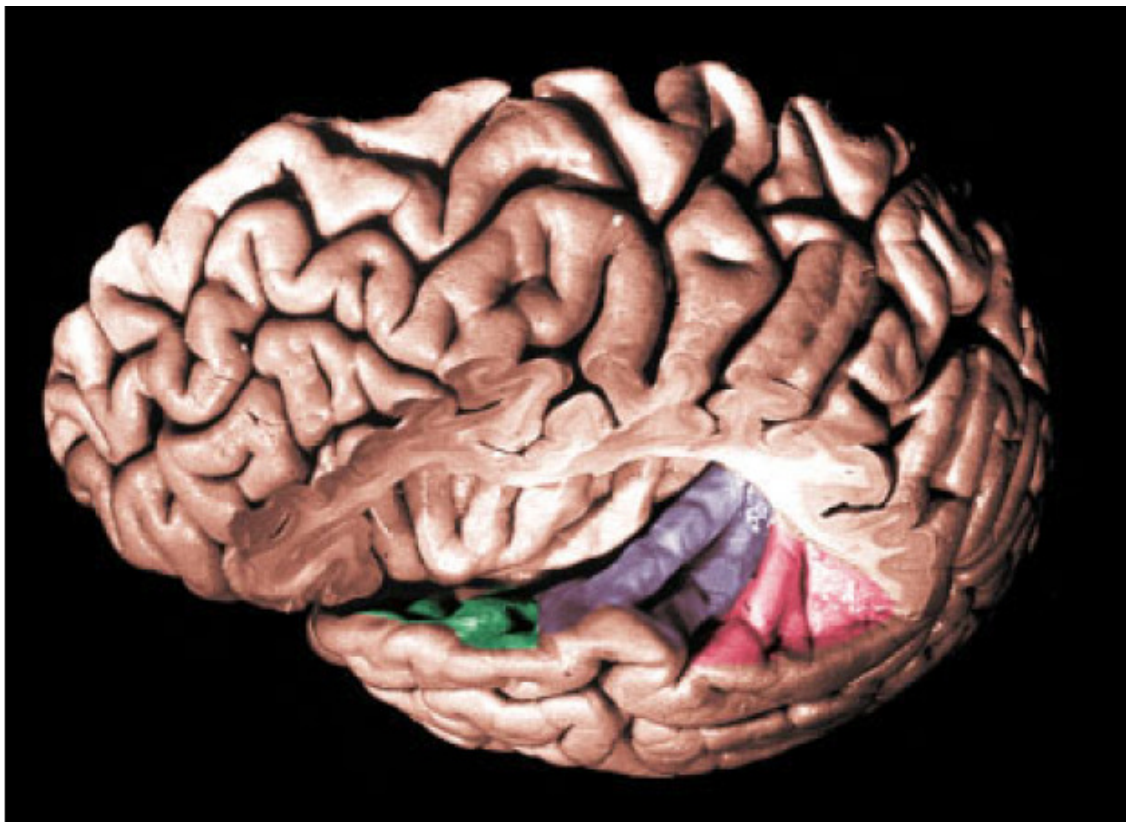


Figure 1.13 : Vue du gyrus temporal gauche dans la scissure de Sylvius : *planum polare* (vert), gyrus temporal transverse antérieur et postérieur (mauve) et *planum temporale* (magenta). Adapté d'après (Duvernoy, 1999)

LE CORTEX AUDITIF SECONDAIRE

Le cortex auditif secondaire entoure le cortex auditif primaire en recouvrant le bord supérieur de la surface latérale du lobe temporal et correspond principalement à l'aire de Brodmann 42 et à une petite partie de l'aire 22 de Brodmann (voir Figure 1.12, p30). Le cortex auditif secondaire comprend le *planum polare* (PP) en avant de A1, et le *planum temporale* (PT) en arrière de A1 (voir Figure 1.13). Le PT est impliqué dans le traitement du langage, mais on sait depuis plus récemment qu'il joue également un rôle dans le traitement des sons et de la musique, de même que le PP.

LE CORTEX AUDITIF ASSOCIATIF

Le cortex auditif associatif comprend l'ensemble des aires associatives, et correspond approximativement à l'aire 22 de Brodmann. Ces aires sont fortement impliquées dans le traitement des informations auditives. Elles reçoivent également des projections d'aires visuelles et

somesthésiques. Une aire du cortex auditif associatif particulièrement importante est l'aire de Wernicke, fortement impliquée dans la compréhension du langage. Toutefois l'aire de Wernicke, bien qu'essentiellement située dans l'aire 22 de Brodmann, déborde sur l'aire 39 située dans le lobe pariétal.

1.2.2 PHYSIOLOGIE/PROPRIETES DES VOIES AUDITIVES

Le cortex auditif est le support de différents traitements du son, qui concernent la fréquence, l'intensité ou encore la localisation du son. Nous allons tenter de décrire les étapes de ces différentes analyses.

1.2.2.1 CODAGE DE LA FREQUENCE SONORE OU TONOTOPIE

Comme cela a déjà été dit le codage de la fréquence sonore existe dès les premiers stades du traitement perceptif, à savoir dès la cochlée. Cette tonotopie est conservé à tous les étages du traitement auditif, devenant de plus en plus fine et précise jusqu'au cortex auditif primaire. La perception de la fréquence est la principale dimension autour de laquelle est organisé le cortex auditif primaire.

Chez l'animal, l'organisation tonotopique a pu être étudiée en détail. Le modèle couramment accepté distingue (Schonwiesner, von Cramon, & Rubsamen, 2002) une région centrale, le *core* (cœur auditif), un ensemble de régions directement adjacentes au cœur, la *belt* (ceinture auditive) et un ensemble de régions adjacentes et latérales à la ceinture auditive, la *parabelt* (paraceinture auditive). Le cœur est caractérisé par une couche IV granulaire dense et comprend plusieurs cartes tonotopiques avec des bandes isofréquencielles orientées médio-latéralement (Hackett, Stepniewska, & Kaas, 1998; Kaas & Hackett, 1998, 2000; Morel, Garraghty, & Kaas, 1993; Morel & Kaas, 1992). La *belt* entoure le cœur et est composée de plus de sept aires corticales cyto-architectoniquement distinctes, dont certaines semblent être également organisées de façon tonotopique (Kaas & Hackett, 1998, 2000; Pandya & Sanides, 1973). Mais ces régions, qui permettent un second niveau de traitement auditif et répondent à des sons plus complexes que pour le *core*, dépendent en grande partie des afférences en provenance du *core* (Rauschecker, 1997). Enfin, la *parabelt*, située latéralement à la *belt*, sur la surface dorsale et dorso-latérale du gyrus temporal supérieur (Pandya & Sanides, 1973; Rauschecker, 1997; Rauschecker, Tian, & Hauser, 1995).

Toutefois, si l'existence d'une organisation tonotopique du cortex auditif a été parfaitement démontrée chez l'animal, paradoxalement, cette spécificité chez l'homme, reste moins bien décrite. Mais les études utilisant les potentiels évoqués ou encore les méthodes TEP ou IRMf ont permis de confirmer le principe d'une représentation tonotopique (Lauter, Herscovitch, Formby, & Raichle, 1985; Lockwood et al., 1999; Pantev et al., 1995; Strainer et al., 1997; Talavage, Ledden, Benson, Rosen, & Melcher, 2000; Wessinger, Buonocore, Kussmaul, & Mangun, 1997). Elles ont en effet permis de déterminer chez l'homme des régions postéro-médiales répondant plus aux stimuli de

basses fréquences et des régions plus antéro-médiales sensibles aux stimuli de hautes fréquences. Malgré l'absence de reproductibilité essentiellement due à une grande variabilité anatomique et fonctionnelle du gyrus de Heschl, toutes ces études plaident en faveur d'un gradient médio-latéral proportionnel à l'augmentation de la fréquence. Il y aurait donc bien chez l'homme une tonotopie du cortex auditif primaire à l'image de celle qui existe depuis la cochlée jusqu'au colliculus inférieur.

1.2.2.2 CODAGE DE LA LOCALISATION SONORE

Au niveau cérébral, le décodage de la source sonore est fondé sur l'analyse de la différence de la phase et de l'amplitude du son qui parvient séparément au niveau des deux oreilles. Nous avons déjà vu l'implication des neurones EI et EE du cortex auditif primaire, juxtaposés et ordonnés en fonction de la plus ou moins grande latéralité de la source sonore qui les concerne. Il semble donc que le cortex auditif primaire soit impliqué de manière bilatérale dans cette localisation (Griffiths, Green, Rees, & Rees, 2000; Griffiths et al., 1996). Zatorre *et al.* ont même montré que le cortex auditif de l'hémisphère mineur apparaissait comme essentiel dans cette tâche (Zatorre & Penhune, 2001). Alors que ce cortex auditif primaire est dédié à l'analyse structurelle de l'onde sonore, certaines autres zones semblent impliquées dans la comparaison et le traitement de la phase et de l'amplitude du son [18, 19]. L'IRMf a encore une fois permis de déterminer l'interaction de ces différentes structures avec les régions auditives postérieures et des régions pariétales dans la localisation spatiale (Bushara et al., 1999; Griffiths et al., 1998; Hall, 2003; J. D. Warren, Zielinski, Green, Rauschecker, & Griffiths, 2002; Weeks et al., 1999; Zatorre, Bouffard, Ahad, & Belin, 2002).

1.2.2.3 CODAGE DE L'INTENSITE SONORE

Quelques études ont recherché une possible organisation du codage de l'intensité sonore, ou ampliotopie, dans les régions auditives. Une telle organisation correspondrait à des cartes sensorielles neuronales distribuées spatialement en fonction de l'intensité sonore, avec des neurones répondant sélectivement à une intensité donnée. Par potentiels évoqués magnétiques, Pantev *et al.* (Pantev, Hoke, Lehnertz, & Lutkenhoner, 1989) ont suggéré l'existence d'une telle représentation ampliotopique dans le cortex auditif. Ces résultats semblent corroborés par des études ultérieures en IRMf, qui ont révélé une activation corticale en réseau bidimensionnel lors de l'augmentation de l'intensité sonore (Bilecen, Seifritz, Scheffler, Henning, & Schulte, 2002). Ces activations ont été observées à l'intérieur du gyrus temporal transverse (gyrus de Heschl), avec une réponse pour des niveaux sonores croissants s'étendant du bord ventral au bord dorsal et de la partie latérale à la partie médiale. Ce maillage latéro-médial serait donc une carte topographique des intensités sonores, analogue à l'organisation tonotopique du cortex auditif.

2 LES SONS

Le son est une sensation auditive perçue via les oreilles. Il s'agit donc d'un concept cognitif qui nécessite que le récepteur ait capté et groupé un ensemble (ou non) de fréquences sonores. Depuis les travaux de Bregman (Bregman, 1990), il est communément admis que l'analyse d'un stimulus sonore par le système perceptif commence par la ségrégation de l'information sensorielle en plusieurs flux (analyse de la scène auditive).

Les fréquences audibles pour l'oreille humaine sont généralement comprises entre 20 et 20 000 Hz (soit une période comprise entre 0,05 et 50 ms), bien que cette gamme tende à diminuer avec l'âge.

Milieu	Vitesse du **son** (en $m.s^{-1}$)
Caoutchouc souple	70
Oxygène	316
Air sec à 0°C	331
Azote	334
Air à 20°C	340
Vapeur d'eau	402
Hélium	970
Hydrogène à 0°C	1270
Eau	1400
Eau douce à 8°C	1435
Eau de mer à 15°C	1500
Eau salée	1500 à 1800
Béton	3000
Bois	4000
Acier, verre	5000

Tableau 2.1 : de propagation du son dans différents milieux. Extrait de (Tanzarella & Mamecier, 2006)

Avant d'être capté et identifié comme tel, le(s) son(s) a(ont) pour origine une (ou des) variation(s) de la pression du milieu ce qui entraîne une (ou des) vibration(s) des molécules, généralement de l'air, qui peut être transmise à travers toute matière que cette variation peut déformer pour se propager. Cette vibration des molécules s'accompagne d'une modification alternative de la pression. C'est cette modification de la pression qui, une fois arrivée à l'oreille, va donner une sensation sonore. On sait que la vitesse de propagation du son dans l'air est de 331,4 m/s, mais dans des milieux plus solides cette vitesse augmente (voir Tableau 2.1). Par exemple elle est de 1400 m/s dans l'eau et de 5000 m/s dans l'acier (Tanzarella & Mamecier, 2006). Cette variation du milieu peut être provoquée par différents phénomènes, un choc, un frottement, le passage de l'air, etc. Une fois entrée en mouvement les sources sonores vont vibrer à des fréquences qui leurs sont propres. Notons que la ou les fréquence(s) émise(s) par la source sonore va (vont) dépendre de la

manière dont elle est excitée. Par exemple, une bouteille en verre n'émettra pas le même son selon que l'on souffle dedans ou que l'on tape dessus avec un objet métallique. La multitude de sources sonores et de manières de les exciter expliquent ainsi les infinies possibilités de sons existant. C'est également pour cette raison que l'oreille humaine, ou plus exactement le cerveau humain, doit être expert dans la reconnaissance de source sonore. Il doit être capable de reconnaître le son du verre de la bouteille en verre, que l'on souffle dedans ou que l'on tape dessus avec un objet métallique. Ou bien encore, il doit être capable de reconnaître le son provenant de la voix d'un individu particulier au milieu d'une foule de gens qui parlent. Il doit regrouper en un seul, les évènements acoustiques appartenant à la même source sonore, il doit créer un objet sonore et le reconnaître dans le temps. On parle alors d'analyse de la scène auditive (Bregman, 1990). Une fois que le cerveau humain a regroupé les différentes fréquences pour les attribuer à l'objet source, on peut alors parler de son.

Lorsqu'un son est composé d'une seule et unique fréquence on parle de son pur, autrement on parle de son complexe. Parmi les sons complexes, on distingue les sons périodiques et les sons non périodiques. On parle de son périodique complexe lorsque le son contient au moins deux fréquences qui se recouvrent dans le temps (Ciocca, 2008). Les sons périodiques complexes, sont aussi les sons les plus couramment produits par une source physique. Il semble donc logique que l'oreille humaine ait évolué de façon à être le plus apte à traiter ce type de son. Les sons non périodiques sont ceux qu'on appelle couramment des bruits. Un son non périodique peut aussi être considéré comme un son dont la période aurait une durée infinie.

Du point de vue acoustique, il existe deux sortes de bruit :

- le bruit aléatoire = dans la nature il s'agit du vent ou de la pluie, dans la musique, du balai de la batterie ou des maracas, et dans le langage les consonnes constrictives où le passage de l'air se rétrécit mais n'est pas interrompu (i.e. [s] [ʃ] et [f]).
- le bruit impulsionnel = par exemple un objet qui se brise, ou bien en musique la batterie, et enfin dans le langage les consonnes occlusives sourdes c'est à dire que le passage de l'air est fermé et que le son résulte de son ouverture subite (i.e. [p], [t] et [k]).

En 2008, Ciocca (Ciocca, 2008) accordait aux sons complexes 5 attributs : (1) le pitch c'est-à-dire la fréquence fondamentale en Hz à laquelle varient les particules du milieu de transmission du son, (2) le timbre c'est-à-dire les différentes fréquences qui composent un son et dont la nature sera propre à une source sonore, (3) l'intensité ou l'amplitude en décibel (dB), (4) l'origine c'est-à-dire la localisation et nature de la source et (5) la durée. Le timbre, l'intensité, l'origine et la durée sont des attributs que possèdent tous les sons. En revanche, on ne peut parler de pitch que pour les sons périodiques complexes comme la voix ou encore les instruments de musique. Dans ce chapitre nous nous attarderons plus longuement sur les aspects du pitch et du timbre car, comme nous le verrons par la suite, il semble qu'ils soient deux aspects particulièrement importants de la voix. En revanche nous passerons rapidement sur les autres attributs du son.

2.1 LE PITCH

2.1.1 DEFINITION

Si la fréquence est une définition physique, en Hertz, le pitch, lui, est une notion perceptive de l'attribut d'un son. Selon l'Association Française de Normalisation (AFNOR), le pitch, ou hauteur tonale, est « le caractère de la sensation auditive lié à la fréquence d'un son périodique, qui fait que le son est aigu ou grave selon que cette fréquence est plus ou moins élevée » (AFNOR, 1977). En résumé, le pitch est la sensation de hauteur perçue, qui va permettre à l'auditeur de situer un son sur l'échelle grave/aigu. Toutefois la relation entre la fréquence fondamentale d'un son et la sensation auditive perçue n'est pas linéaire, et comme nous le verrons par la suite, la seule notion de grave/aigu n'est pas satisfaisante en termes de définition. La définition de l'American National Standards Institute (www.ainsi.org) est un peu plus large et propose que le pitch soit « cet attribut de la sensation auditive selon lequel les sons peuvent être ordonnée sur une échelle musicale ». Cette dernière définition, ne restreint donc pas le pitch ni à la dimension grave/aigu, ni aux sons complexes périodiques, mais laisse vaste le champ de ce à quoi correspond cette sensation auditive.

2.1.2 CARACTERISTIQUES PHYSIQUES DU PITCH

En acoustique, ce qui défini le pitch est, la fréquence fondamentale, c'est à dire la fréquence de base (la plus basse) émise par une source, dont tous les harmoniques d'un son complexe périodique seront des multiples. Les harmoniques peuvent être des multiples entieres lors d'un son harmonique (voir Figure 2.1), ou partielles lors d'un son inharmonique, on parle alors de partiels plutôt que d'harmoniques.

C'est donc sur la base de cette fréquence fondamentale, que l'on peut déterminer la hauteur du son. Lorsque la fréquence est basse il s'agira d'un son grave, tandis que pour une fréquence élevée on entendra un son aigu.

La différence conceptuelle entre la fréquence et le pitch réside dans le fait que le pitch peut être perçu par le cerveau humain, même en l'absence d'énergie à cette fréquence fondamentale. Lorsque la fondamentale est absente du son (c'est à dire qu'elle a une énergie nulle ou quasi-nulle), on parle de fréquence fondamentale manquante. Une bonne illustration à ceci est l'exemple du téléphone. En effet, la bande passante du téléphone, c'est-à-dire la gamme de fréquence que peut restituer le système, est comprise entre 300 Hz et 3400 Hz. La faible bande passante du téléphone auquel s'ajoute son rapport signal/bruit (de l'ordre de 40 dB) limitent la qualité du signal analogique transmis. Toutefois, nous avons tous l'expérience du téléphone, et sommes en mesure d'affirmer que nous reconnaissons en majorité notre locuteur, éventuellement son humeur, ou pouvons savoir s'il est malade, les différentes intonations de sa voix, etc. Or la fréquence fondamentale moyenne de la

voix masculine tombe généralement autour de 120Hz, tandis que celle de la voix féminine tomberait entre 210Hz (Baken & Olikoff, 2000; M. P. Gelfer & Mikos, 2005; Guimaraes & Abberton, 2005; Iseli, Shue, & Alwan, 2007; Pegaroo Krook, 1988; Russell, Penny, & Pemberton, 1995). Nous arrivons donc à reconstruire en son absence la fréquence fondamentale produite par l'interlocuteur, et ce grâce aux indices présents des autres fréquences qui sont des multiples de cette fréquence fondamentale. Ceci explique pourquoi nous pouvons « entendre » le pitch au téléphone, alors que les basses fréquences sont filtrées.

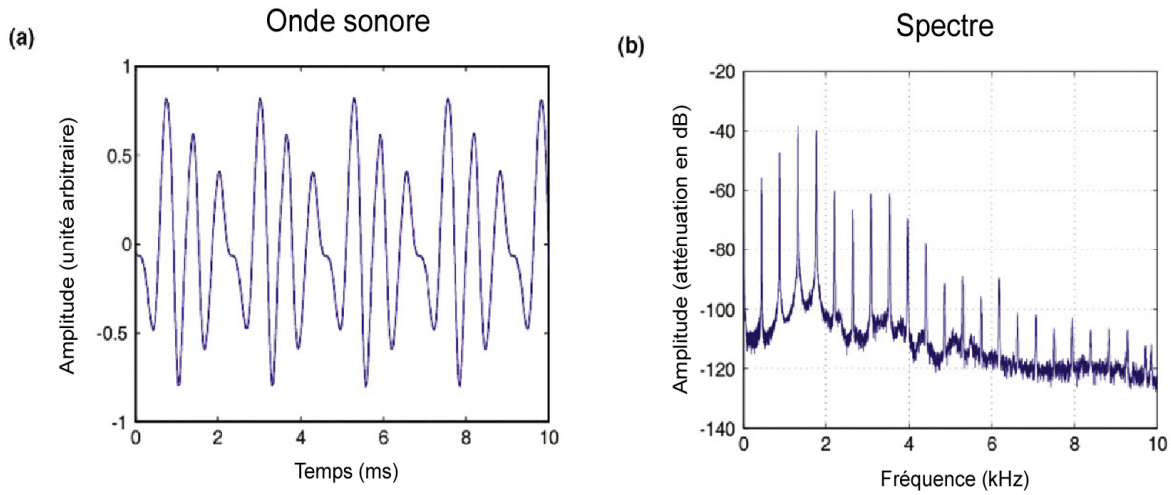


Figure 2.1 : Illustration d'une note de musique jouée par un hautbois à 440 Hz. (a) On peut voir que l'onde sonore est périodique, avec une période de 2,27 ms. (b) On peut voir sur le spectre que les harmoniques sont des multiples de la fréquence fondamentale (440 Hz). Cette dernière est physiquement présente mais les second, troisième et quatrième harmoniques ont une plus grande amplitude.

2.1.3 PERCEPTION DE LA HAUTEUR ET PERCEPTION DU PITCH

La perception de la hauteur a essentiellement été étudiée et définie dans le cas des sons purs. On sait que certains facteurs influencent la hauteur tonale perçue, et notamment le niveau d'intensité. Il est connu depuis le XIXe siècle que, la hauteur tonale perçue d'un son pur de fréquence donnée est susceptible de varier lorsqu'on modifie son niveau d'intensité. La «règle de Stevens» (S. S. Stevens, 1935), veut qu'une élévation de niveau élève (rende plus aigu) la hauteur tonale perçue des sons de fréquence élevée, laisse à peu près inchangée la hauteur tonale perçue des sons de fréquence moyenne et diminue la hauteur tonale perçue des sons de basse fréquence. Les résultats de Walliser (Walliser, 1969) confirment la validité de cette règle (voir Figure 2.2).

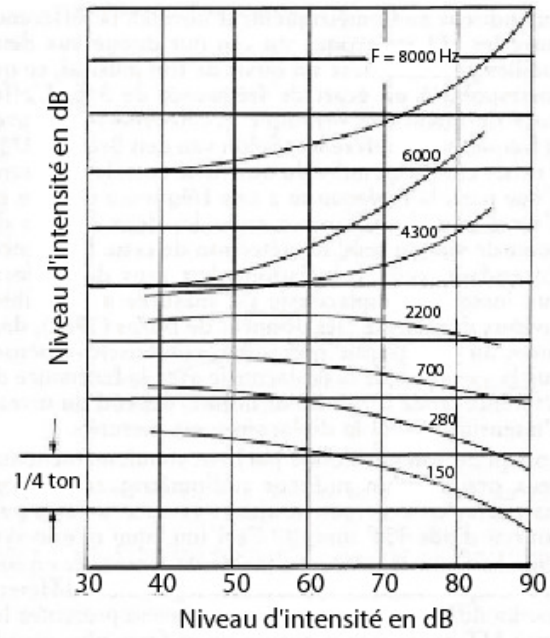


Figure 2.2 : Effet du niveau d'intensité d'un son pur sur sa hauteur tonale perçue. Adapté d'après Walliser *et al* 1969.

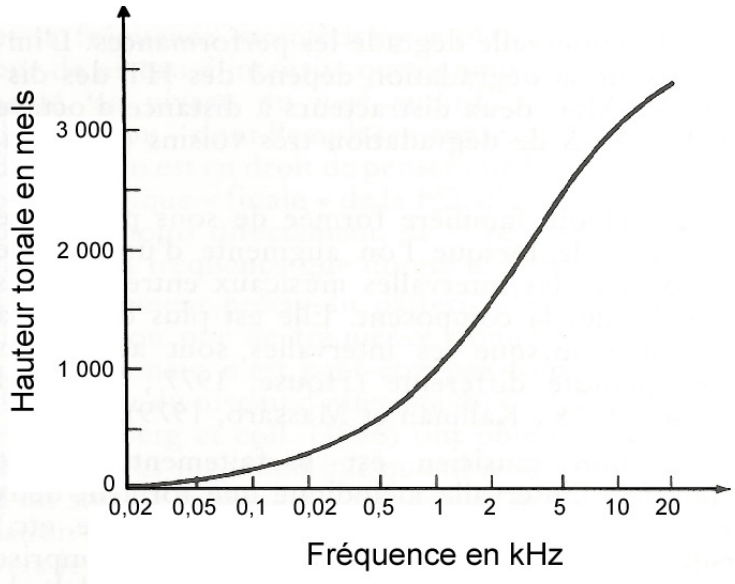


Figure 2.3 : Hauteur tonale d'un son pur en fonction de sa fréquence (S. S. Stevens & Volkmann, 1940). Un mel est sensé correspondre à une différence constante entre deux hauteurs tonales. L'échelle de la fréquence en abscisse est logarithmique. A noter que la fonction représentée n'est pas une droite, ce qui signifie selon Stevens et Volkmann, que la hauteur d'un son pur ne varie pas comme le logarithme de sa fréquence.

Une question essentielle de la psychoacoustique est de comprendre la relation entre phénomène physique et perception. La question de la sensation produite par le spectre fréquentielle a donc été largement étudiée. S'il est relativement facile de mesurer physiquement la fréquence fondamentale d'un son, il a été plus difficile de mesurer la sensation perceptive, la hauteur perçue ou tonie. Deux échelles de mesures semblent avoir particulièrement éclairé ce point ; la tonie et la chroma. Stevens (S. S. Stevens & Volkmann, 1940) a été le premier à établir une échelle de tonie, l'échelle de Mel. Cette échelle a été construite par mesurage direct, en demandant à des sujets d'ajuster un son de façon à ce qu'il soit, par exemple, deux fois moins aigu qu'un son de référence, ou deux fois plus aigu. L'échelle de Mel avait donc pour but de mettre en évidence la position des perceptions à l'intérieur du continuum grave/aigu.

Pour Stevens et ses collaborateurs, la fonction résultante est non logarithmique. C'est essentiellement pour cette raison que cette échelle de Mel a été fortement critiquée et essentiellement par les musiciens. En effet, Attneave et Olson en 1971, ont démontré que la règle selon laquelle la perception de la hauteur suivait une fonction logarithmique, était vraie même pour les auditeurs qui ne pratiquaient pas la musique (Attneave & Olson, 1971). Une autre échelle, développée un peu plus tard définie ce qu'on appelle la chroma. Il s'agirait d'une qualité selon laquelle deux sons purs, dont le rapport de fréquence est deux, sont similaires ou identiques. En musique, on parle d'octave, c'est-à-dire que deux sons séparées d'une octave (on double la fréquence) ont la même chroma. Les notes séparées d'une octave portent le même nom et sont toujours jugées très similaires, même par les non musiciens. Certains vont même jusqu'à les confondre (Lockhead & Byrd, 1981). Actuellement on considère que ces échelles de mesures ne sont pas exclusives et peuvent très bien se compléter. La perception de la hauteur relève à la fois des chroma et de la dimension. Par exemple, tous les La d'une gamme ont le même chroma, mais varient entre eux sur la dimension grave/aigu.

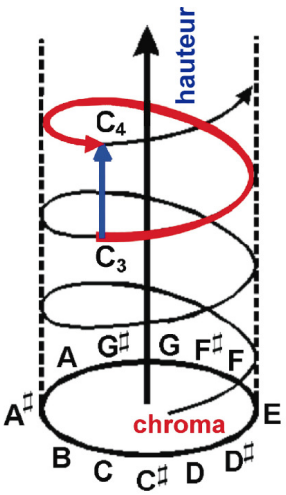


Figure 2.4 : Echelle des pitch en forme d'hélice. Sur le rayon est décrit l'échelle musicale des chroma, et sur la hauteur de l'échelle est définie l'échelle de la hauteur du son (dimension grave/aigu). Adapté d'après Warren *et al.* (2003).

En ce qui concerne la perception du pitch à proprement parler, elle correspond à la sensation de hauteur tonale produite par un son complexe périodique, on parlera dans ce cas de hauteur fondamentale ou de pitch. La perception du pitch d'un point de vue psychophysique n'est pas encore totalement claire. Plusieurs modèle explicatifs de l'extraction du pitch ont été proposés, tels que le modèle de Terhardt (Terhardt, 1974, 1979; Terhardt, Stoll, & Seewann, 1982) ou celui de More-van Noorden (B. C. J. Moore, Glasberg, & Peters, 1982; van Noorden, 1982). Ces modèles s'opposent essentiellement sur la question de l'innée ou de l'acquis de la perception du pitch, mais ils se retrouvent sur un point essentiel, à savoir que la perception du pitch se ferait à partir d'une analyse des composantes spectrales. Ceci semble être l'hypothèse la plus plausible actuellement. Des auteurs (B. C. Moore, Glasberg, & Peters, 1986), ont montré que la perception du pitch d'un son complexe à 12 es était sensible aux variations des harmoniques qui le composent. C'est à dire que si les harmoniques sont « déviantes », le pitch est perçu avec une variation allant dans le même sens que celle du ou des harmoniques déviantes. Les auteurs concluent donc que la perception du pitch est une opération de groupement des harmoniques. Une étude plus récente de Ciocca, V. 1999 a montré que lors de l'écoute d'un son complexe à 12 harmoniques, présenté avec toutes ses harmoniques ou avec seulement les harmoniques impaires, la déviance de la 3ème et 5ème harmoniques entraînait une modification de la perception du pitch (Ciocca, 1999). Ceci confirme pour les auteurs que la perception du pitch relève donc bien d'un groupement des composants d'un son sur la base de leur proximité des harmoniques de la fréquence fondamentale F0. On considère donc qu'un son complexe périodique peut être apparié à un son pur dont la fréquence est égale à celle de la fondamentale du son complexe périodique.

Quelques études ont pu mettre en évidence une sensation perceptive de pitch dans le cas de l'écoute de bruits (Formby, 1985; R. M. Warren & Wrightson, 1981), mais par convention on considère que, puisque le pitch est la sensation physique correspondante à la fréquence fondamentale (manquante ou non) par groupement des harmoniques qui composent le son, cette caractéristique ne peut être attribuée qu'aux sons complexes périodiques.

2.1.4 CORRELATS NEURONAUX DE LA PERCEPTION DU PITCH

Nous avons vu que la cochlée opérait déjà une ségrégation des fréquences en termes de perception, et que cette tonotopie perdurait à tous les relais du traitement auditif, jusqu'au cortex auditif. Le fait qu'une composante spectrale, ici la fréquence fondamentale, puisse être perçue isolément, avec sa propre hauteur, est très probablement lié à l'organisation tonotopique du système auditif. Toutefois, dans cette partie nous ne nous intéresserons pas au codage fréquentiel, mais à la perception du pitch à proprement parler. Autrement dit, nous essaierons de décrire les aires susceptibles d'être le substrat neuronal à la perception du pitch, c'est-à-dire à la sensation de tonie produite par un son. Nous parlerons donc de corrélats entre perception et activité neuronale.

Des processus corticaux de haut niveau étant probablement impliqués dans la perception du pitch, différentes études se sont intéressées au substrat neuronal susceptible de supporter la perception du pitch. Quelques études mettent en avant le rôle du *planum temporale* dans le traitement du pitch. Par exemple des auteurs ont étudié l'asymétrie anatomique du *planum temporale*, qui est naturellement en faveur de l'hémisphère gauche (Schlaug, Jancke, Huang, & Steinmetz, 1995). Ils ont montré que les sujets musiciens qui ont une perception particulièrement précise du pitch – les sujets ayant l'oreille absolue – montraient une asymétrie particulièrement marquée en comparaison de sujets non musiciens. Toutefois il s'agissait d'une étude anatomique, corrélée avec une aptitude essentiellement musicale, ce qui ne présage pas forcément de la perception du pitch en tant que telle. Une autre étude plus récente (Hyde, Peretz, & Zatorre, 2008) a montré que seul le *planum temporale* s'activait en réponse des changements de pitch relativement petits (i.e., 200 cents, où 100 cents = ½ ton).

D'avantage d'études, mettent en avant un autre candidat pour soutenir le traitement du pitch ; le gyrus de Heschl. En 1988, Zatorre a montré que les performances d'extraction de pitch pour des sons où la fréquence fondamentale était absente était déficitaire chez des patients ayant subi une lobectomie unilatéral du lobe temporal, notamment à droite et incluant le gyrus de Heschl (Zatorre, 1988). Ce résultat a été confirmé plus tard dans une autre étude, avec une tâche de discrimination de la direction de changement de pitch (direction montante ou descendante). Les patients ayant une ablation du gyrus de Heschl droit, bien que capables de discriminer les pitch avait un déficit dans la tâche de reconnaissance de direction de changement de pitch (Johnsrude, Penhune, & Zatorre, 2000). Des études anatomiques ont montré tout d'abord une relation entre l'anatomie du gyrus de Heschl et les performances de perception du pitch. Par exemple, Schneider *et al.* se sont basé sur l'idée que la détection du pitch pouvait se faire soit à partir de la fréquence fondamentale soit à partir du spectre harmonique (par exemple dans le cas de la fondamentale manquante). Ces auteurs ont montré que cette aptitude à traiter le pitch était corrélée au volume du gyrus de Heschl latéral, plutôt celui de droite lorsque le traitement se fait sur la base du spectre, et plutôt celui de gauche lorsque la perception du pitch se fait sur la base de la fondamentale (Schneider et al., 2005). D'autres auteurs, Wong *et al.*, ont étudié la relation entre le traitement du pitch et la taille du gyrus de Heschl dans une étude IRM anatomique. Les sujets apprenaient une langue artificielle sur des variations de pitch. Pour cela ils étaient entraînés à associer des pseudo-mots à des images. Les auteurs ont montré une corrélation positive et significative entre les performances à apprendre cette langue et le volume des matières grise et blanche du gyrus de Heschl droit. La corrélation était négative et significative, entre ce volume et le temps mis par les sujets pour atteindre leurs meilleures performances. Cela signifiait que les sujets ayant un gyrus de Heschl droit plus important en volume, étaient meilleurs en terme de performances et de durée d'apprentissage d'une langue basée sur le pitch (P. C. Wong et al., 2008). Enfin des études d'imagerie fonctionnelle ont aussi montré un rôle prédominant du gyrus de Heschl dans la perception du pitch. Dans une étude IRMf, Patterson *et al.* ont montré que seul le gyrus latéral de Heschl bilatéral était préférentiellement activé par la régularité temporelle, c'est-à-dire le pitch, de stimuli acoustique (Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002). Dans une autre

étude, Penagos *et al.* ont comparé le signal BOLD produit par des sons complexes harmoniques, selon que leur pitch était saillant ou non. Leurs résultats ont montré qu'une aire du cortex auditif secondaire, recouvrant l'extrémité antérolatérale du gyrus de Heschl, s'activait plus faiblement lorsque la saillance du pitch du stimulus sonore était faible et plus fortement lorsque la saillance était importante (Penagos, Melcher, & Oxenham, 2004). Cette étude a donc confirmé la localisation de l'aire de traitement du pitch identifié par Patterson *et al.* (2002) en utilisant des stimuli différents. Dans une étude de 2006, Bendor et Wang ont défini par approximation à partir de plusieurs études mettant en évidence le rôle du gyrus latéral de Heschl dans le traitement du pitch, la localisation potentielle de l'aire responsable de la perception du pitch (Bendor & Wang, 2006). Cette approximation est présentée sur la Figure 2.5.

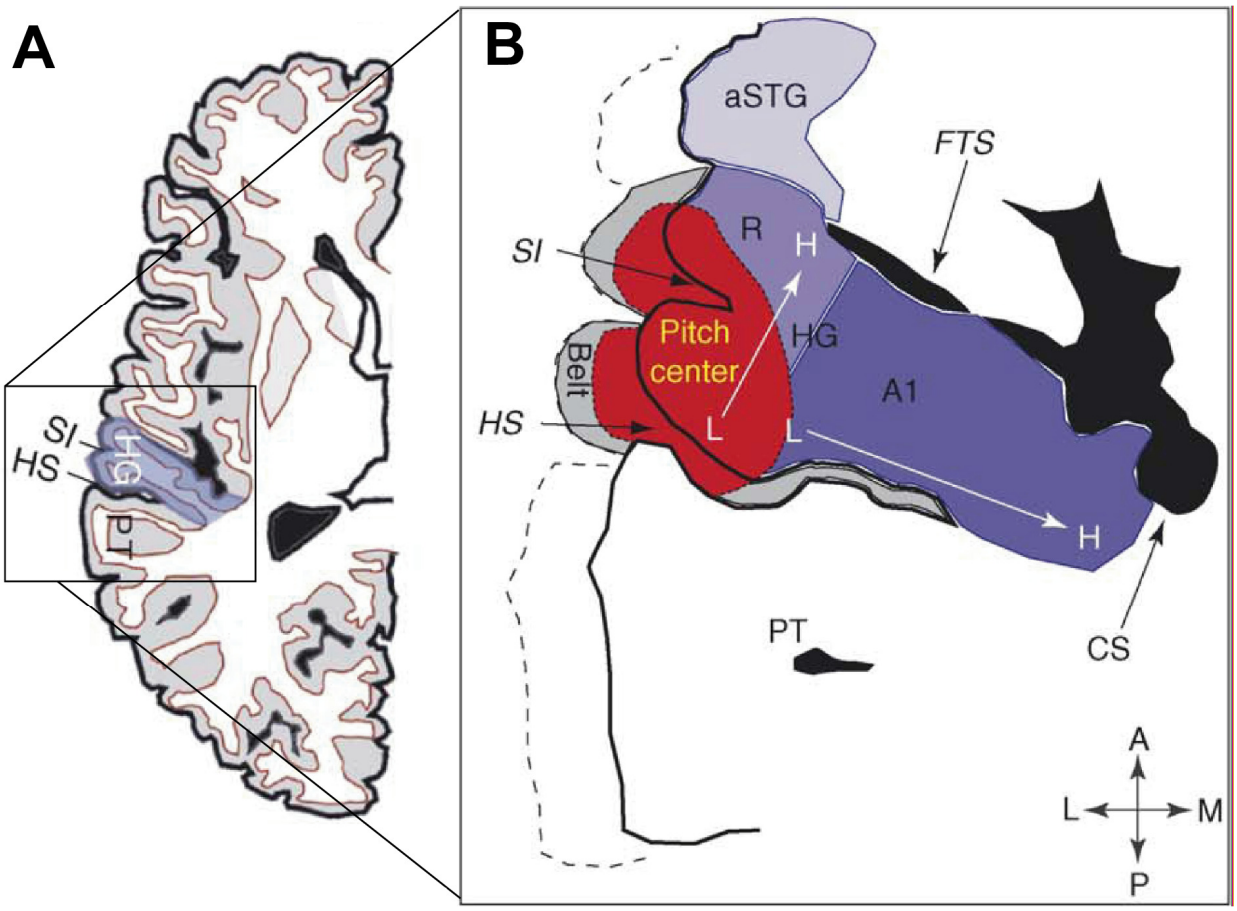


Figure 2.5 : **(A)** Schéma d'une section horizontale du lobe temporal humain. **(B)** Vue grossie du gyrus de Heschl. Le cortex auditif est présumé occuper la portion médiale du gyrus de Heschl. La localisation des aires voisines (R, Pitch center, lateral belt) est une approximation basée sur les études Schneider *et al.* (2005) Formisano *et al.* (2003), et Patterson *et al.* (2002) (Formisano et al., 2003; Patterson et al., 2002; Schneider et al., 2005). Abréviations : **AI**, cortex auditif primaire ; **aSTG**, gyrus temporal supérieur antérieur ; **Belt**, ceinture auditive ; **CS**, sillon cingulaire ; **FTS**, premier sillon transverse ; **H**, haute fréquence ; **HG**, gyrus de Heschl ; **HS**, sillon de Heschl ; **L**, basse fréquence ; **Pitch center**, centre de traitement du pitch ; **PT**, *planum temporale*; **R**, cortex auditif rostral ; **SI**, sillon intermédiaire. Adapté d'après Bendor et Wang (2006).

Une étude plus récente, a montré des résultats en accord avec les études précédentes (Puschmann, Uppenkamp, Kollmeier, & Thiel). En plus de montrer une activation bilatérale du gyrus latéral de Heschl pour tous les stimuli évoquant un pitch en comparaison du bruit, les auteurs ont montré un probable rôle du *planum temporale* et du *planum polare* dans le traitement de la mélodie (contour de pitch) dans la localisation. De façon assez anecdotique, une étude IRMf s'est intéressé en particulier à la notion de pitch telle qu'on l'entend le plus souvent dans le contexte musical, à savoir une définition en fonction de la dimension grave/aigu (hauteur) et de la chroma (J. D. Warren, Uppenkamp, Patterson, & Griffiths, 2003). Les auteurs ont présentés à des auditeurs normo-entendants des sons complexes harmoniques en faisant varier indépendamment le chroma ou la hauteur du son. Les résultats ont montré que des activations dans le gyrus de Heschl latéral et dans une partie du *planum temporale* et du *planum polare* étaient associé spécifiquement à l'écoute passive de sons complexes harmoniques par rapport à du bruit. De manière plus précise encore, il s'est avéré que des aires particulières du *planum temporale* s'activaient spécifiquement lorsque les stimuli sonores présentaient des changements en hauteur et non en chroma. A l'inverse, des régions particulières du *planum polare* s'activaient spécifiquement lorsque les stimuli sonores présentaient des changements de chroma et pas de hauteur. Mais, étant donné que peu d'études de neuroimagerie se sont intéressées au pitch dans ses dimensions de hauteur et de chroma, il n'y a pas d'autre étude qui ait confirmé ces résultats.

La question de la latéralisation du traitement du pitch dans l'hémisphère droit n'est pas encore tout à fait consensuelle dans la littérature. En effet, parmi les études que nous avons déjà citées, certaines montraient une latéralisation à droite des structures associées à la perception du pitch (Hyde et al., 2008; Johnsrude et al., 2000; Schneider et al., 2005; P. C. Wong et al., 2008; Zatorre, 1988) alors que d'autres non (Hall & Plack, 2009; Patterson et al., 2002; Penagos et al., 2004; Puschmann et al., 2010). De plus, quelques rares études suggéraient une préférence de l'hémisphère gauche pour le traitement du pitch (Mazziotta, Phelps, Carson, & Kuhl, 1982; Phelps & Mazziotta, 1985; Platel et al., 1997). Dans une revue de la littérature, Zatorre *et al.* recensaient les nombreuses études en faveur d'une latéralisation à droite (Zatorre & Gandour, 2008). L'ensemble des résultats de ces études, sont à mettre en parallèle avec le modèle de la spécialisation hémisphérique développé par Zatorre *et al.* (Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002). Les auteurs ont en effet mis en évidence une spécialisation de l'hémisphère gauche pour le traitement des changements temporels rapides, qui seront alors plus impliqués pour le traitement du langage, et une spécialisation pour l'analyse spectrale de l'hémisphère droit, qui serait alors plus impliqué dans le traitement des sons musicaux. Des auteurs ont par ailleurs montré que le cortex auditif gauche était sensible à des fenêtres temporelles plus courte, de l'ordre de 25 à 50 ms, tandis que le cortex auditif droit serait sensible à des fenêtres temporelles plus grande, de l'ordre de 20 à 300 ms (Boemio, Fromm, Braun, & Poeppel, 2005; Rupp, Gutschalk, Uppenkamp, & Scherg, 2004). Il semble probable que cette spécialisation hémisphérique puisse expliquer les résultats des nombreuses études portants sur la spécialisation hémisphérique pour le traitement du pitch, car la

plus part des études utilises des stimuli musicaux, c'est-à-dire des variations de pitch. Cela ne retire rien au fait qu'une réponse à la perception du pitch à pu être mise en évidence dans d'autres études. Schneider *et al.* suggéraient qu'un traitement du pitch puisse être bilatéral, mais qu'il ne se ferait pas sur la base des mêmes informations (Schneider et al., 2005). L'extraction du pitch par l'hémisphère gauche se ferait à partir de la fondamentale (périodicité, traitement temporel), tandis que l'extraction du pitch par l'hémisphère droit se ferait sur la base du spectre harmonique (analyse spectrale, fréquentielle). Cette hypothèse serait donc en accord avec l'ensemble de la littérature, à savoir qu'un traitement du pitch puisse avoir lieu bilatéralement mais que l'hémisphère droit soit spécialisé dans le traitement fréquentiel, et donc des stimuli musicaux, y compris les variations de pitch.

En résumé, l'ensemble des études de neuroimagerie sont plutôt en accord pour dire qu'une région corticale proche de A1, située dans le cortex auditif secondaire, probablement le gyrus de Heschl latéral, répondrait préférentiellement aux sons périodiques et pourrait être spécialisée dans le traitement du pitch. Il y a également plus d'argument en faveur d'une latéralisation à droite pour l'analyse des contours de pitch (stimuli musicaux). Mais, nous devons relativiser ces données car toutes ne sont pas concordantes. Hall *et al.*, ont montré que la présentation de différents types de sons périodiques à des auditeurs humains, activait différentes régions du cortex auditif (Hall & Plack, 2009). Pour certains auteurs cela illustre bien le fait que les études utilisent pour la plupart des paradigmes très différents, et cela laisse entendre que la perception du pitch est plutôt constituée d'un réseau, qui s'activerait différemment selon la nature du stimulus présenté et/ou la nature de la tâche demandée au sujet (Walker, Bizley, King, & Schnupp, 2010). En effet, la nature du traitement peut considérablement varier, selon que l'on réalise une tâche de reconnaissance de mélodie, on parle alors de contour de pitch, ou que l'on réalise une tâche passive de perception du pitch. Si dans tous les cas le traitement et la perception du pitch sont engagés, d'autres fonctions cognitives le seront également. De plus, comme on a pu le voir, les auteurs divergent parfois dans leur définition du pitch. Pour certain, il s'agit de la dimension grave/aigu, pour d'autre de chroma, pour d'autre encore il s'agit de la perception de la fondamentale d'un son complexe harmonique, etc. Ces différentes définitions peuvent également expliquer les différents résultats entre les études. Enfin une des plus grandes difficultés des études de neuroimagerie est de savoir si l'activation mise en évidence avec une tâche particulière et des stimuli particuliers, correspond réellement à la perception consciente du stimulus ou à l'encodage des propriétés du stimulus qui permettent cette perception.

2.1.5 LE CAS PARTICULIER DE L'OREILLE ABSOLUE

DEFINITION

La plupart des gens sont capables de reconnaître une mélodie, sans pour autant connaître ou reconnaître la fréquence exacte des notes qui la composent. Le cerveau humain est alors capable de reconnaître la relation, ou l'écart, entre les différentes notes. On parle d'oreille relative. Par

opposition, on parle d'oreille absolue, pour définir la capacité à reconnaître et nommer instantanément et précisément la fréquence d'un son tonal, ou périodique, en l'absence de toute référence. La Figure 2.6 illustre les résultats d'une étude de Zatorre en 2003. On peut voir que lors de la dénomination d'une note l'écart par rapport à la note écoutée en demi-tons a une variance très faible pour les possesseurs de l'oreille absolue, par rapport à des non-possesseurs. Autrement dit, les possesseurs de l'oreille absolue sont toujours plus exacts.

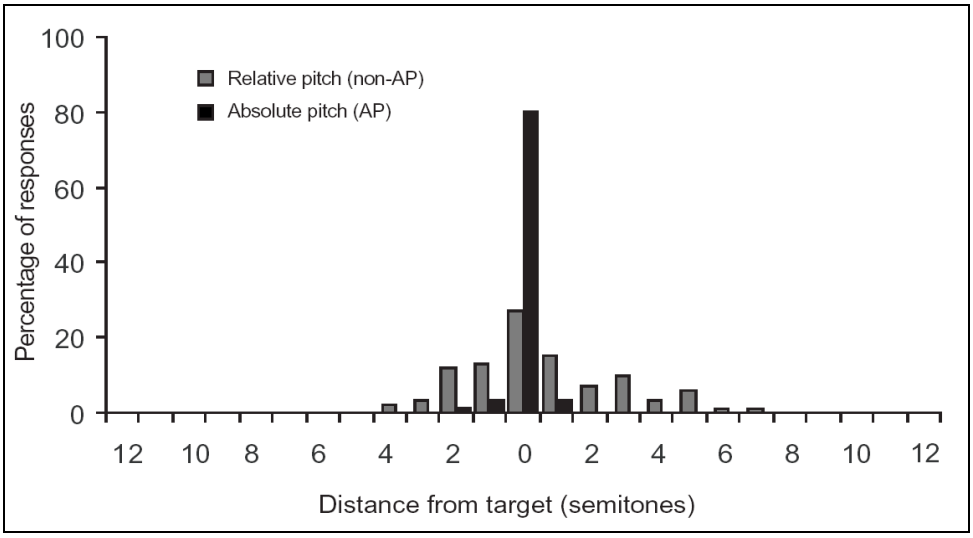


Figure 2.6 : Distribution des réponses de dénomination du pitch de sons présentés aléatoirement chez des musiciens non possesseurs et des possesseurs du pitch absolu. En abscisse, la distance par rapport à la cible indique que les possesseurs du pitch absolu sont plus exacts que les musiciens non possesseurs de pitch absolu. Figure tirée de (Zatorre, 2003).

Cette habileté est particulièrement intéressante chez les musiciens qui peuvent alors transcrire ou mémoriser une partition de musique puisqu'ils sont capables d'identifier directement chaque note. Par exemple, on peut penser que Mozart avait l'oreille absolue, puisque à l'âge de 14 ans, le 11 avril 1770, il a été capable de retranscrire en entier le célèbre *Miserere* d'Allegri, après l'avoir entendu une seule fois à la chapelle Sixtine (Lechevalier, 2003). Beaucoup de musiciens s'entraînent dans le but de développer leur oreille jusqu'à avoir l'oreille absolue, car en musique comme cela a été dit, cela présente un atout majeur. Mais la plus grosse majorité des études ont montré que l'entrainement de sujets adultes à acquérir l'oreille absolue est resté vain jusqu'à présent (Takeuchi & Hulse, 1993; Ward, 1999). Cette capacité auditive est rare et sa prévalence en Amérique du Nord et en Europe est estimée de 1/5000 à 1/10 000 selon les études (Profita & Bidder, 1988). Toutefois aucune n'a encore pu montrer que les possesseurs de l'oreille absolue possédaient en sus des performances particulières dans d'autres taches auditives ou musicales (Burns & Campbell, 1994; Fujisaki & Kashino, 2002; Miyazaki, 2004; Miyazaki & Rakowski, 2002). Beaucoup d'auteurs se sont interrogés sur l'origine de ce phénomène, notamment sur la question de l'inné et de l'acquis. Les deux théories trouvent leurs défenseurs et leurs arguments, mais aujourd'hui l'ensemble des auteurs sur le sujet, s'accordent à dire que l'oreille absolue, dépend à la fois des gènes et de

l'environnement. En faveur d'une origine génétique, on peut notamment parler du fait que les musiciens possédant l'oreille absolue ont plus facilement dans leur famille un ou plusieurs autres individus possédant cette caractéristique (Baharloo, Johnston, Service, Gitschier, & Freimer, 1998; Baharloo, Service, Risch, Gitschier, & Freimer, 2000). Actuellement, aucun gène ni même chromosome n'a (n'ont) été identifiés comme support de la transmission génétique de ce phénomène, et il semblerait plus probable selon Zatorre (Zatorre, 2003), que l'étiologie de l'oreille absolue ne soit pas liée à un seul déterminant génétique, et qu'elle soit donc multiple et complexe. De plus, comme nous l'avons déjà cité, l'entrainement n'a pas d'influence sur le développement de l'oreille absolue. Toutefois ces auteurs ont étudié ce phénomène chez des adultes. D'autres études ont montré l'existence d'une période critique pour l'apparition de ce phénomène, qui est similaire à celle du langage. Cette période critique (voir Figure 2.7) selon les études est aux alentours de 7 ans (Levitin & Menon, 2003; Levitin & Rogers, 2005) ou bien autour de 4-5 ans (Baharloo et al., 1998; Deutsch, Henthorn, Marvin, & Xu, 2006; Deutsch, Le, Shen, & Henthorn, 2009). Pour ces auteurs, l'existence d'une période critique, dans l'apparition de l'acquisition de l'oreille absolue signifie qu'il y à dans ce phénomène des facteurs d'apprentissages, sans exclure pour autant des facteurs génétiques, puisqu'ils rappellent la présence plus importantes de possesseurs de l'oreille absolue au sein d'une même famille. De Deutsch *et al.* ont montré qu'un environnement linguistique tonal, comme le chinois, augmentait la probabilité de trouver des individus ayant l'oreille absolue (Deutsch, Henthorn, & Dolson, 2004; Deutsch et al., 2006).

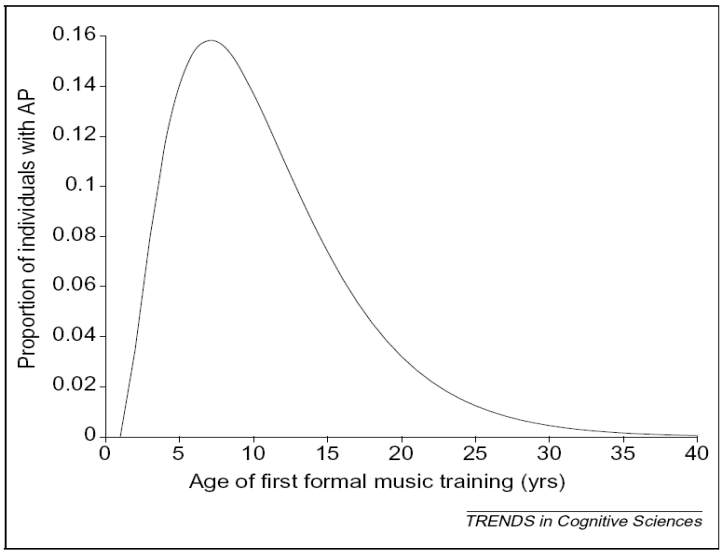


Figure 2.7 : Données d'une étude rétrospective sur l'âge de l'acquisition du langage, représentées par une fonction Gamma. Cela suggère l'existence d'une période critique de l'acquisition de l'oreille absolue. Le mode de cette distribution est autour de 7 ans. Extrait de (Levitin & Rogers, 2005).

SUPPORT NEURONAL

Au-delà de la considération génétique ou innée de l'oreille absolue, plusieurs différences corticales anatomiques et fonctionnelles ont été mises en évidence entre les possesseurs et les non possesseurs de l'oreille absolue. En effet, des études en potentiel évoqués (EEG) ont montré que les non-possesseurs de l'oreille absolue, lorsqu'ils étaient dans une tâche de transposition semblaient plutôt activer la mémoire de travail (Shahin, Bosnyak, Trainor, & Roberts, 2003), alors que les possesseurs d'oreille absolue semblaient activer plutôt la mémoire à long terme (Hirose et al., 2002). Des auteurs font l'hypothèse que ces deniers n'utiliseraient pas la mémoire de travail pour garder une représentation mentale active du pitch, mais qu'ils sont capables d'encoder la fréquence en labellisation verbale (Zatorre, 2003; Zatorre & Beckett, 1989). D'un point de vue anatomique, d'autres auteurs ont montré que les régions du cortex supérieur droit et du *planum temporale* sont plus petites chez les possesseurs de l'oreille absolue que chez les non-possesseurs (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2005; Keenan, Thangaraj, Halpern, & Schlaug, 2001), bien que, comme cela à déjà été dit précédemment, on n'ait pas relevé d'autres performances particulières chez les possesseurs de l'oreille absolue. D'autres auteurs encore parlent d'une asymétrie plus importante chez ces derniers, en faveurs de l'hémisphère gauche, par rapport à des sujets contrôles (Goldberger, 2001; Schlaug et al., 1995). Sur le plan fonctionnel, plusieurs études se sont intéressées au phénomène. Une étude a notamment montré que des sujets ayant l'oreille absolue dans une tâche de dénomination de note montraient une activation focale dans le cortex frontal droit, tandis que les sujets contrôles activaient cette région uniquement lors d'une tâche de dénomination d'intervalles (Zatorre, 1988). Or, il a été montré que cette région était fortement impliquée dans le maintient d'une information de pitch en mémoire de travail, les auteurs de cette étude concluent donc que les possesseurs de l'oreille absolue, plutôt que de maintenir constante une trace sensorielle, utilisent des représentations catégorielles des notes dans chaque tâches (voir Figure 2.8). A l'inverse, une autre aire du lobe frontal, le cortex frontal dorsolatéral postérieur, s'active préférentiellement chez les sujets ayant l'oreille absolue que chez les sujets contrôles, lors d'une tâche passive d'écoute de notes (Zatorre, 1988). Cette aire, connue pour être impliquée dans l'établissement et le maintien d'association conditionnée en mémoire (Zatorre, Perry, Beckett, Westbury, & Evans, 1998), est pour les auteurs la candidate idéale pour être le substrat neuronal particulier aux possesseurs de l'oreille absolue, et ferait le lien entre le pitch et son étiquetage en facilitant cette association (Levitin & Rogers, 2005; Zatorre, 2003).

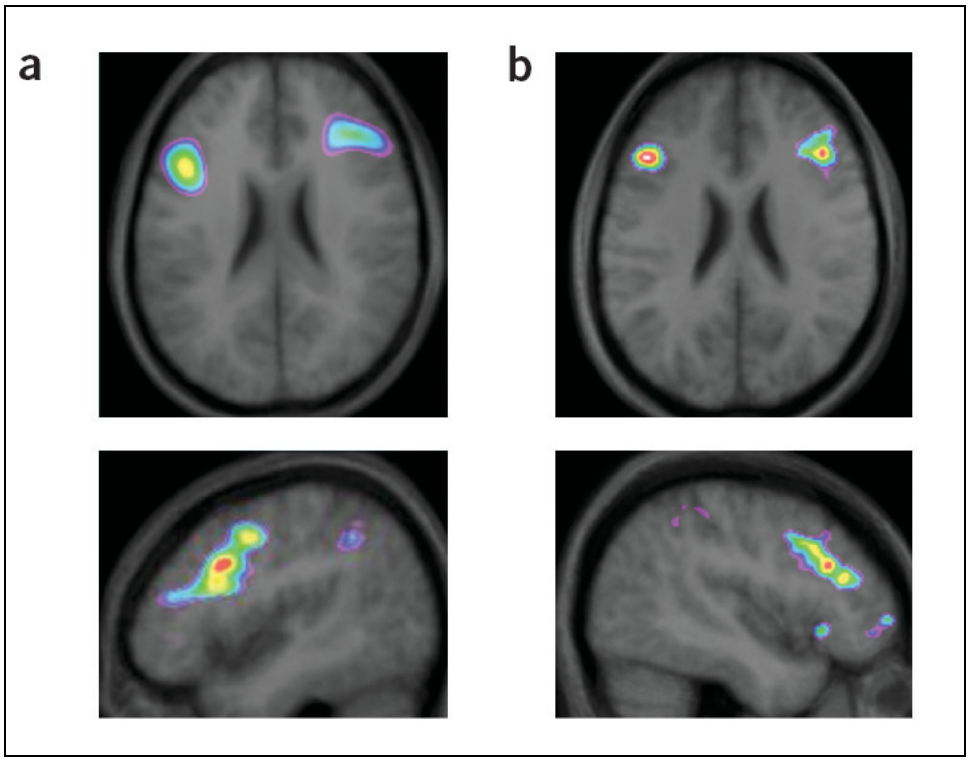


Figure 2.8 : Images issues de TEP et d'IRMf. (a) Vue horizontale (haut) et sagittale de l'hémisphère gauche (bas) montrant les aires cérébrales dont le flux sanguins à augmenté dans le cortex frontal dorsolatéral chez des possesseurs de l'oreille absolue, lors d'u d'une tache d'écoute de notes en TEP. (b) Vues comparables (hémisphère droit en bas) dans une étude d'IRMf chez des non-musiciens qui avaient pour tâche d'identifier des accords à l'aide de touches choisies arbitrairement. Figure tirée de (Zatorre, 2003).

Enfin, une étude, plus récente en EEG a montré que des sujets ayant l'oreille absolue montraient une plus grande activation des hémisphères droit et gauche lors d'une tâche de dénomination de notes, par rapport à des sujets contrôles (Wu, Kirk, Hamm, & Lim, 2008). Les auteurs concluent que les possesseurs de l'oreille absolue sont capables d'activer un plus grand réseau cortical que les sujets contrôles.

En résumé, les possesseurs de l'oreille absolue seraient capables d'associer très facilement une note (ou une fréquence) avec son nom, par un encodage particulier, et seraient également capables d'activer plus facilement et plus largement les réseaux sous-jacents.

ORIGINE ET MODELE

Le débat sur la question de l'innée et de l'acquis de l'oreille absolue étant arrivé dans l'ensemble à un consensus selon lequel, ces deux aspects participeraient dans l'émergence de ce phénomène, sans qu'il soit encore clairement défini quelle part ils tiennent précisément. Ross *et al.* en 2005, ont proposé un modèle tentant de rendre compte de l'origine de l'oreille absolue (Ross, Gore, & Marks, 2005). Pour cela ils ont essayé d'adopter une vision plus large du phénomène. Par exemple ils remarquaient en 2005 que certains possesseurs de l'oreille absolue avaient la capacité de nommer les notes uniquement en musique, tandis que d'autre étaient capable d'étendre cette habileté à d'autres signaux acoustiques (Ross et al., 2005). De plus, ils critiquaient le fait que, dans la plupart

des études, le critère pour définir l'oreille soit la dénomination de note, et induise donc un biais tel que seuls les musiciens occidentaux pouvaient faire un tel test. Autrement dit, il est difficile de s'intéresser à l'existence de l'oreille absolue chez les musiciens orientaux ni même chez les non musiciens. Cela fait de l'oreille absolue un phénomène exclusivement musical. Or si l'on considère un part génétique à ce phénomène, on peut envisager de telles capacités chez des non musiciens, bien que la question de la mise en évidence ne soit pas simple. Enfin, pour Ross et ses collègues, la tâche de dénomination de notes est tellement complexe, qu'elle pourrait s'expliquer par de nombreuses stratégies différentes (Baharloo et al., 1998; Miyazaki, 1988; Profita & Bidder, 1988; Ward, 1999), qui elles-mêmes pourraient rendre compte des différents profils.

Pour Ross *et al.*, il y aurait deux types d'oreille absolue qui se dégagent, selon les profils et en fonction de la littérature. Le premier serait les « vraies » oreilles absolues, qui peut exister non seulement chez des musiciens ayant eu une expérience musicale précoce, mais aussi chez des musiciens n'en ayant pas eu, chez des non musiciens ou même chez des enfants qui ne connaissent pas encore le nom des notes. Pour les auteurs la part génétique dans ce type d'oreille absolue serait prépondérante, par rapport à l'environnement et ne serait pas directement lié à l'expérience musicale, et l'habileté pourrait s'étendre au-delà du domaine musical. Ils nomment ce type d'oreille absolue les APE, pour *« ability to perceptually encode »*. Le phénomène serait alors d'ordre plus perceptuel, et ils invoquent la capacité particulière des oreilles absolues APE à encoder la fréquence des stimuli auditifs avec un rôle important du traitement préattentif. Ross *et al.*, suggèrent pour cela l'existence d'un réseau additionnel chez les APE, qui permettrait cet encodage spécifique (Ross et al., 2005). Le deuxième type d'oreille absolue, serait en revanche lui, beaucoup plus lié et corrélé à l'intensité et la précocité de l'apprentissage musical, et donc au type d'instrument pratiqué (Miyazaki, 1988, 1989; Takeuchi & Hulse, 1991), mais aussi à de fortes capacités mnésiques. Les auteurs les appellent les HTM, pour *« heightened tonal memory »*. Dans ce cas, les sujets mémoriseraient de façon très précise des « exemples » auditifs de référence et compareraient ensuite les notes entendues. Ils auraient la capacité d'extraire les informations sans réseau neuronal additionnel.

Ce model proposé par Ross *et al.* en 2005, est intéressant car il permet de rendre compte des différents profils d'oreille absolue que l'ont peut rencontrer et donne une nouvelle dimension, non exclusivement musicale, à ce phénomène, en lien avec la part de la génétique et de l'expérience.

2.2 TIMBRE

Le timbre reste aujourd'hui difficile à caractériser, tant du point de vue perceptif que du point de vue physique. La partie stable du son semble tenir une place particulière, pour percevoir et définir, d'un point de vue physique, le timbre. Nous allons tenter dans cette partie de définir le timbre, et nous pourrons voir que, contrairement au pitch, les caractéristiques qui servent à le définir

sont complexes et qu'il est assez difficile de définir clairement la relation entre les paramètres physique et la sensation perceptive du timbre.

2.2.1 DEFINITION

Le timbre, à l'instar du pitch, est également une notion assez conceptuelle et surtout bien plus complexe, si bien qu'il est assez difficile à définir. Actuellement on le définit par soustraction. Lorsque deux sons possèdent la même fréquence et la même intensité, mais qu'ils paraissent différents, on dit alors que leur timbre est différent. Par exemple si deux instruments de musique différents, un violon et piano, jouent exactement la même note, par exemple 440 Hz (i.e. un La3), à intensité égale, par exemple 70 Décibels, on s'apercevra aisément qu'il ne s'agit pas des mêmes instruments. En effet, leur timbre diffère. Cette différence de timbre est aussi vraie pour deux instruments identiques, par exemple pour deux violons qui jouent un La3 à 440 Hz à 70 dB SPL. Mais cela nécessite souvent une oreille plus attentive et peut-être aussi plus expérimentée. Le timbre est donc l'empreinte digitale du son, il est unique. De plus le timbre d'un instrument est aussi fonction de la manière dont le son est excité, par exemple, la façon de jouer du musicien, l'outil utilisé (archet, doigt, etc.), ou encore des conditions atmosphériques (humidité, température, etc.) qui, en modifiant la qualité des matériaux de résonance, peuvent jouer un rôle dans la qualité du timbre. Sans le timbre acoustique, des événements sonores ayant les mêmes propriétés de durée, de fréquence et d'amplitude, resteraient indistincts sur le plan auditif. Dans le langage, le timbre est également ce qui permet de distinguer les voyelles, à intensité et fréquence fondamentale égales. De plus chaque personne possède son propre timbre de voix, ce qui permet à chacun d'entre nous d'être reconnu et de reconnaître les personnes de notre entourage uniquement sur le plan acoustique. Une étude de Coleman en 1971, a montré que même lorsque la fréquence fondamentale émise par les cordes vocales était la même chez 10 hommes et 10 femmes, à l'aide d'un électrolarynx, les sujets à qui il était demandé d'identifier le genre de la voix qu'ils entendaient étaient capables d'identifier les genres (Coleman, 1971). Ceci signifie que le timbre contient de l'information acoustique associée au genre du locuteur. Une partie du timbre de la voix est difficilement modifiable, elle contient des constantes sur l'identité du locuteur. Eventuellement lors d'un rhume, d'une chirurgie ou d'une atteinte ORL, le timbre de la voix peut changer de façon temporaire ou permanente. Une autre partie du timbre est modifiée et est le support, entre autre, du langage, de l'état affectif, etc.

Finalement cette définition du timbre, qui ne serait ni le pitch, ni la durée, ni l'intensité laisse l'auditeur à sa sensation perceptive pour savoir ce qu'est le timbre du point de vue perceptif.

2.2.2 CARACTERISTIQUES PHYSIQUES DU TIMBRE

Les critères physiques du timbre ne sont pas encore complètement définis. La véritable complexité à définir les caractéristiques physiques du timbre, réside dans le fait qu'on ne sait pas encore très exactement quels paramètres physiques d'un son sont importants dans la perception du timbre, qui elle-même est très difficile à verbaliser. La manière la plus simple pour connaître les éléments physiques du son qui participent à la sensation de timbre, est de faire varier des paramètres et de voir si la sensation de timbre varie. Aujourd'hui l'ensemble des auteurs s'accorde à dire que le timbre est un attribut multidimensionnel. La question du nombre de dimensions du timbre et des paramètres acoustiques de ces dimensions reste un sujet de débat. Plusieurs caractéristiques physiques du son participeraient donc de façon prédominante à définir son timbre. Par exemple l'enveloppe temporelle et essentiellement les transitoires, c'est à dire l'attaque du son ou son extinction. Il s'agit de la mise en place ou de la diminution progressive des fréquences du son. Les formants, c'est à dire les partiels du son qui contiennent de l'énergie, jouent également un rôle important. D'autres auteurs avancent aussi le rôle des modulations, des distorsions, des microvariations, etc. pour définir le timbre. Nous allons tentez de mieux décrire les contributions de ces paramètres sur la sensation perceptive de timbre.

2.2.2.1 LES FORMANTS OU LA REPARTITION DE L'ENERGIE DANS LE SPECTRE

Pour bien comprendre ce que sont les formants et en quoi ils sont spécifiques du timbre, nous devons d'abord comprendre comment ils se forment.

Les sont purs se décrivent par l'onde la plus élémentaire qui est l'onde sinusoïdale (voir Figure 2.9).

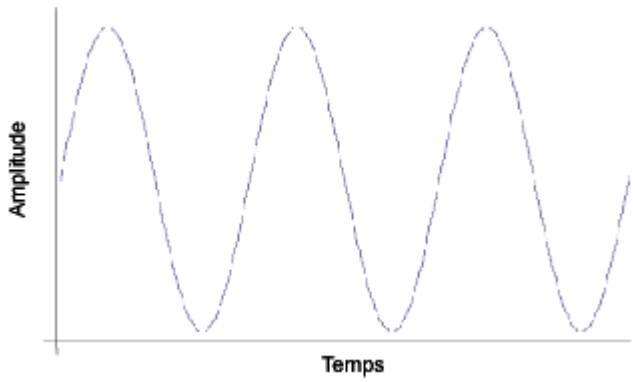


Figure 2.9 : Exemple d'onde sinusoïdale.

Les sons qui nous entourent sont plus fréquemment de nature complexe, c'est à dire composés de plusieurs sons purs de fréquences différentes. Or la loi de Fourier nous dit qu'un signal complexe périodique, c'est à dire qui se reproduit identique à lui même à intervalles de temps réguliers, peut se décomposer en un certain nombre de sinusoïdes dont les valeurs sont des multiples entiers de la plus petite d'entre elles (voir Figure 2.10). Cette définition permet donc la décomposition spectrale de tous les sons complexes périodiques.

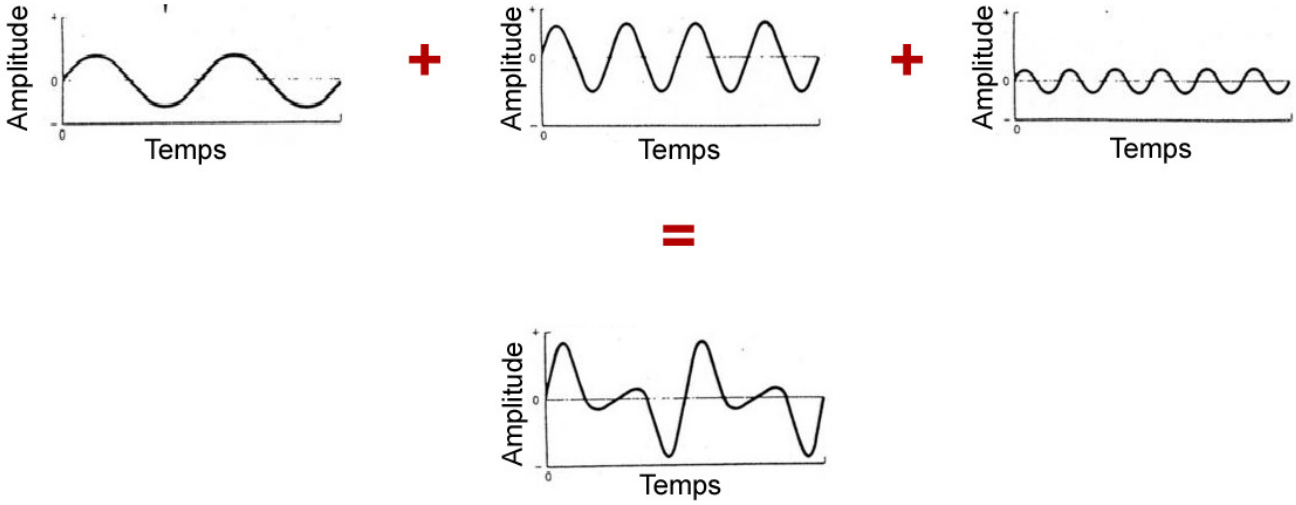


Figure 2.10 : Composition s'un son complexe périodique. La somme des 3 fréquences (3 graphiques du haut) donne un signal complexe périodique.

Autrement dit, à l'intérieur de la masse sonore que peut être un son complexe périodique et que nous percevons comme unique, il y a un mélange de très nombreuses petites sinusoïdes. C'est la nature de ce mélange qui détermine la forme de l'onde résultante et qui participe à l'identité timbrale (voir Figure 2.10).

D'un point de vue acoustique, la présence de toutes ces sinusoïdes s'explique très bien. Lorsqu'une vibration est crée, par exemple une corde qui vibre, elle vibre à une fréquence fondamentale bien précise, par exemple 440 Hz. Cette vibration peut être audible, ou non. Par exemple la corde excitée peut être audible sans rien autour, tandis que le diapason vibre à une intensité très faible, si bien qu'il est difficile de l'entendre directement. Cette vibration va être capable, en fonction de son intensité et de sa proximité, de faire entrer un corps en résonance (i.e. phénomène par lequel un système physique en vibration peut atteindre une très grande amplitude lorsque la vibration excitatrice se rapproche d'une fréquence naturelle de ce système), par exemple la caisse de résonance d'une guitare. En entrant en résonance, le corps va vibrer, à des fréquences qui lui sont propres, et qui seront toujours, soit des multiples entiers, soit des multiples partiels de la fréquence fondamentale qui va faire entrer ce corps en vibration. En général, chaque harmonique ou partiel correspond à la vibration de différentes parties du corps. Ainsi si un corps de résonance est capable de vibrer à toutes les harmoniques de la fréquence fondamentale, alors on les entendra toutes. Mais bien souvent, les harmoniques ou partiels ne sont pas tous présents avec la même

intensité (voir Figure 2.11), et ceci dépend de la nature du corps de résonance. Dans le cas de la voix, la résonance est fonction de l'appareil supralaryngé : la fréquence fondamentale à laquelle elles vont vibrer va être capable d'exciter toutes les matériaux suffisamment proches et capables d'entrer en résonance soit à la fréquence fondamentale soit à un multiple de cette fréquence. Généralement, dans le cas de la voix, vont entrer en résonance différentes parties de la cavité orale et de la cavité nasale. Ces fréquences de résonances sont fonction de la nature, et donc de la taille et de la forme, de l'appareil supralaryngé, y compris les positions de la langue et des lèvres qui peuvent changer la longueur de ce conduit vocal.

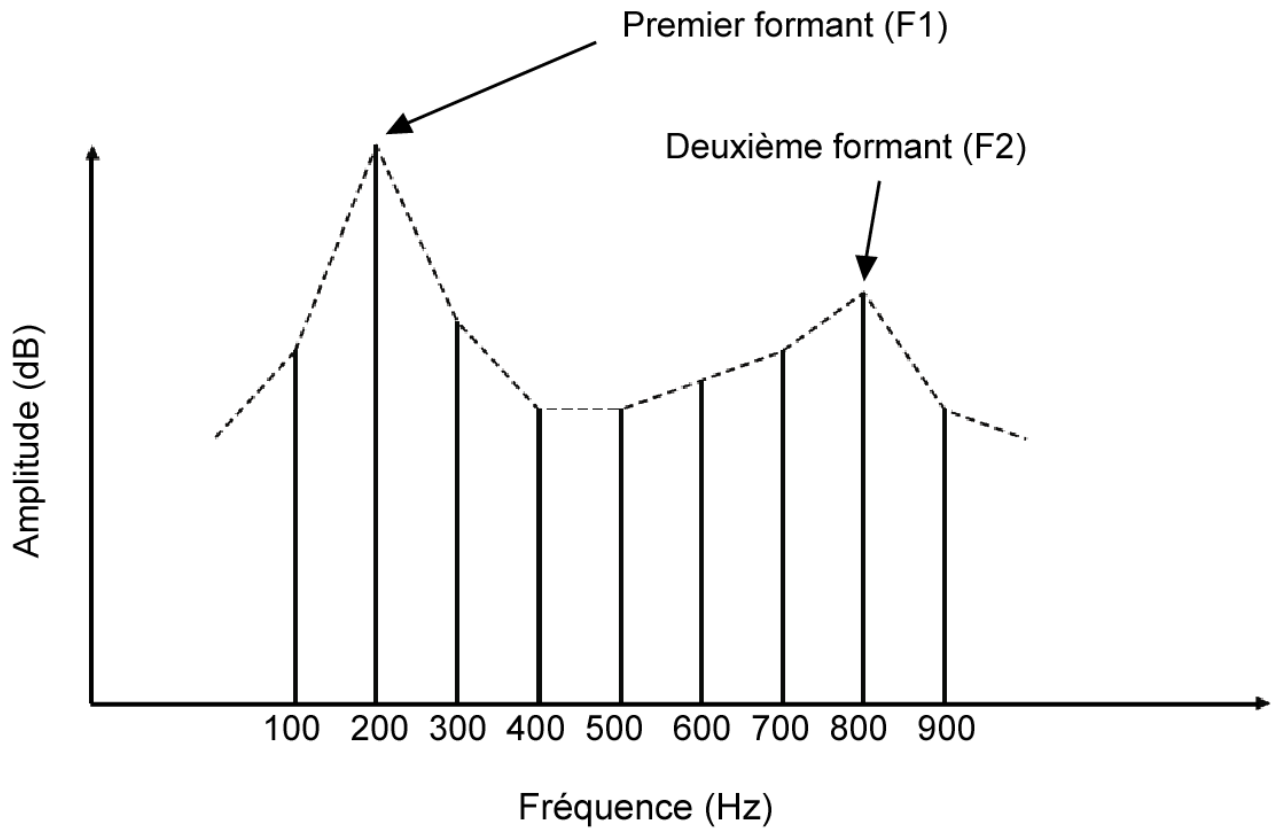


Figure 2.11 : Illustration d'un exemple de son complexe harmonique. On peut voir que la fréquence fondamentale est à 100Hz et que toutes les harmoniques sont des multiples de cette fréquence. Le premier formant, c'est-à-dire le premier pic d'énergie dans le spectre, correspond à la deuxième harmonique, tandis que le deuxième formant correspond à la 8ème harmonique.

Les fréquences, ou harmoniques qui vont vibrer à une amplitude plus importante, vont être plus représentées dans le son. D'un point de vue physique, on dit que ce sont les fréquences qui ont le plus d'amplitude, qui ont le plus d'énergie. Ces zones ou régions du spectre d'un signal vocal correspondant à un maximum d'énergie, sont appelés des formants (voir Figure 2.11). Ils représentent des pics d'énergie dans le spectre, ce sont les fréquences les plus proéminentes. Par convention, on nomme le premier formant F1, le deuxième F2,… L'ordre des formants se fait toujours, à partir de la fréquence fondamentale, des fréquences les plus graves vers les plus aigues.

2.2.2.2 L'ENVELOPPE TEMPORELLE OU LES TRANSITOIRES

L'enveloppe temporelle, ou courbe ADSR, est la visualisation de l'évolution de l'amplitude du son complexe harmonique en fonction du temps (voir Figure 2.12). C'est la représentation temporelle (intensité en fonction du temps) de ce son. On se rend compte que le son a quatre étapes : le transitoire d'attaque (*Attack* en anglais), la retombée (*Decay*), le régime stationnaire (*Sustain*) et le transitoire d'extinction (*Release*). Le nom ADSR provient de l'appellation anglaise. Chacune de ces phases ont des durées très variables, selon les sons.

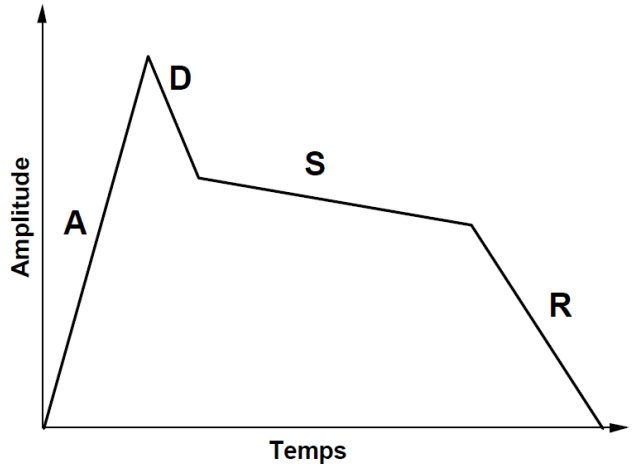


Figure 2.12 : Illustration de l''enveloppe temporelle d'un son sur laquelle nous pouvons différencier 4 phases ; l'attaque (A), la retombée (D), le régime stationnaire (S) et l'extinction (R). Ces paramètres sont souvent libellés A.D.S.R.

Le transitoire est un paramètre spectro-temporel. Il correspond à la partie instable du son, c'est-à-dire la mise en place et la disparition du son. Il y a donc deux transitoires, un correspond à l'attaque du son et le second à son extinction. Le transitoire d'attaque, ou l'attaque, dure quelques dizaines de millisecondes, selon les sons. Il correspond à la période au tout début du son, pendant laquelle la vibration dans l'instrument commence et atteint son régime stationnaire (stable). C'est par exemple entre le moment où le marteau du piano touche la corde et le moment où le son atteint sa plénitude et sa puissance. Durant cette période, les harmoniques naissent d'une manière anarchique : elles ne sont pas calées les unes par rapport aux autres. Cette période contient aussi beaucoup de bruit lié à l'excitation de la source sonore (choc du marteau du piano, frottements excessifs de l'archet, attaque de gorge ou de langue du clarinettiste…). Le transitoire d'extinction, ou l'extinction, dure entre le moment où la source sonore n'est plus excitée et le moment où le son disparaît vraiment, ce qui correspond également à quelques dizaines de millisecondes. Par exemple, entre le moment où un trompettiste arrête de souffler et où l'instrument se "vide" effectivement de son son.

La durée des transitoires et leur forme jouent un rôle considérable dans le timbre. Ils sont également primordiaux pour la reconnaissance de la source sonore surtout lors de l'identification d'instrument. Des auteurs on montré que lorsqu'on coupe ou modifie la durée ou la forme de ces

transitoires lors de l'écoute d'une note jouée par un instrument, l'oreille humaine ne reconnaît plus l'instrument (Castellengo, 1987; Grey & Moorer, 1977). Dans le transitoire il semble que plusieurs paramètres jouent un rôle important. Dans le transitoire d'attaque qui a été beaucoup plus étudié que celui d'extinction, il semble que la durée et l'amplitude aient une place prépondérante dans la perception du timbre. Egalement, mais de façon moindre, l'ordre de mise en place des différentes harmoniques semble avoir son importance (voir Figure 2.13).

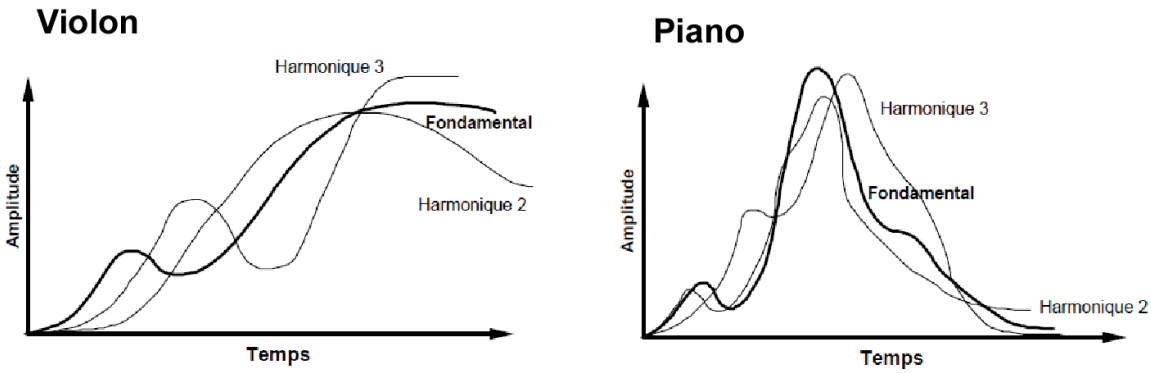


Figure 2.13 : Illustration de deux exemples d'attaque. Le premier son (à gauche) montre la mise en place progressive de la fréquence fondamentale ainsi que des deuxième et troisième harmoniques pour un son de violon et le deuxième son (à droite) montre la mise en place pour les mêmes composantes lorsque le son provient d'un piano.

Les transitoires, très importants pour reconnaître la source sonore, permettrons aussi, dans le cas d'un instrument de musique, de reconnaître l'interprète. Si l'on prend l'exemple du violon, on sait que si l'instrumentiste est débutant, l'attaque sera hésitante, et le régime établi du son long à se mettre en place, alors que dans le cas d'un musicien chevronné, l'attaque sera nette, franche, et le contenu harmonique très vite stabilisé. Les transitoires peuvent donc aussi nous informer sur la façon dont la source sonore est excitée et, si chaque résonateur à son propre timbre, il dépend aussi de la nature de l'excitation. Les transitoires, de part le fait qu'ils dépendent aussi de la façon dont est excité le son, sont aussi dits timbre causal.

Pour conclure, même si des auteurs s'accordent à dire que les transitoires sont probablement très importants dans le timbre (Iverson & Krumhansl, 1993) et particulièrement dans l'identification de la source sonore, les transitoires restent un phénomène peu étudié.

2.2.2.3 AUTRES PARAMETRES

Finalement, on a pu voir que les attributs physiques qui participent à la sensation de timbre sont multiples et multidimensionnels. Nous avons déjà parlé des formants (dimension spectrale), de

l'enveloppe temporelle (dimension temporelle) et plus particulièrement des transitoires (dimension spectro-temporelle), mais il semblerait que ces attributs ne suffisent pas à expliquer totalement la sensation de timbre. En effets, les modulations d'amplitude qui donnent des sensations de battement ou de vibrato et les modulations de fréquence ou les distorsions, semblent participer à la sensation de timbre. Toutefois ces paramètres et leur impact sur la sensation de timbre restent peu étudiés dans la littérature.

Dans l'ensemble, différentes informations, de nature spectrale, temporelle ou spectro-temporelle sont prises en compte pour identifier un timbre. Néanmoins, ces informations ne sont pas nécessairement toutes utilisées simultanément. Par exemple, les informations spectrales semblent suffisantes pour reconnaître un instrument, mais seulement si elles sont disponibles suffisamment longtemps dans le signal acoustique (Hajda, 1999).

2.2.3 LA PERCEPTION DU TIMBRE

Nous avons vu que la sensation de timbre dépend de plusieurs paramètres acoustiques, et nous avons vu aussi qu'il n'est pas possible de relier le percept de timbre à une seule caractéristique de l'onde sonore. Il semble donc évident que le timbre est bien aussi un concept cognitif basé sur l'analyse de plusieurs composants spectro-temporels du son. D'un point de vue perceptif, le timbre, reste difficile à qualifier. Par exemple, un son peut être décrit comme fort ou faible du point de vue de son amplitude perçue, ou encore comme aigu ou grave, du point de vue de sa hauteur tonale perçue (fréquence fondamentale et uniquement pour les sons périodiques). En revanche le timbre d'un son reste difficile à décrire et les quelques descriptions qui en sont faites sont difficiles à corréler à des paramètres physiques. En 1930, Helmholtz a essayé d'associer certaines variables physiques avec des perceptions acoustiques (Helmholtz, 1930). Il concluait que la qualité du son dépend des fréquences qui composent ce même son. Par exemple, selon lui, les sons purs sont perçus comme doux, plaisant, et « lisse », tandis que les sons complexes dont les partiels au-delà de la 6ème harmonique sont prédominants, c'est-à-dire ayant les plus fortes amplitudes, sont perçus comme « cinglant », brut et pénétrant. En 1988, Gelder à développé un outil descriptif de 18 items, chacun mesuré sur une échelle en 8 points, pour décrire des voix (M.P. Gelfer, 1988). Les items proposés étaient par exemple : « haut ou bas », « fort ou faible », « doux ou rêche », « plaisant ou déplaisant », « mélodieux ou grinçant », « animé ou monotone », etc. Les deux premiers exemples servent donc davantage à décrire respectivement la hauteur et l'intensité. Les autres qualificatifs proposés pour décrire la voix font appel soit à des sensations tactiles (« doux-rêche »), des émotions « plaisant-déplaisant », etc. La description des sons reste donc pauvre en qualificatifs auditifs, et est réduite à sa hauteur et son intensité. Avec cet outil, Andrews et Schmidt en 1997 avaient demandé à des sujets de décrire des voix prononcées par des hommes (Andrews & Schmidt, 1997). Ces derniers devaient lire un texte soit avec leur voix masculine normale, soit en version féminine. De façon générale, les

qualificatifs qui revenaient le plus souvent pour les voix féminines par rapport aux voix masculines étaient « aéré », et « animé » (par opposition à monotone).

Le timbre semble donc être un attribut sonore très difficile à qualifier d'un point de vue perceptif, tant il n'existe pas de vocabulaire spécifique à sa description. Cette difficulté à le qualifier et la multiplicité des paramètres physiques participant à ce phénomène, rend l'étude de sa perception d'autant plus compliquée. Toutefois certaines études, et notamment les études de dissemblance de timbre, ont pu mettre en corrélation des indices acoustiques avec des sensations de timbre. Les méthodes d'analyse multidimensionnelle (MDS) ont été largement appliquées aux jugements de dissemblance sur des paires de timbres. Le postulat de ces études est que la MDS permet de mettre en évidence un petit nombre de dimensions qui distinguent les timbres les uns des autres, ces dimensions étant, de par la méthode même, les plus saillantes dans l'ensemble des stimuli utilisés. Ces études de jugements de dissemblance ont été réalisées avec des sons enregistrés, ou des sons de synthèse généralement conçus pour imiter des instruments réels. La majorité des études ont montré que des espaces à deux ou trois dimensions rendaient bien compte des données. Ces dimensions, qui résultent uniquement des jugements des sujets, sont appelées dimensions perceptives. L'espace perceptif représentant les relations entre les différents timbres est appelé espace de timbre. Plusieurs auteurs ont réussi à définir 3 dimensions perceptives correspondant à 3 paramètres acoustiques. En effet, une série d'expériences avec des sons de synthèse imitant des instruments naturels a été conduite avec les sons d'un synthétiseur. Dans une première étude, des sujets musiciens ont jugé les dissemblances dans un ensemble de 21 sons synthétiques imitant des instruments de l'orchestre ou des hybrides de ces instruments (Krumhansl, 1989). Un modèle d'espace de timbre à trois dimensions rendait bien compte des données. Qualitativement selon Krumhansl, les trois dimensions communes seraient respectivement : la rapidité de l'attaque, la brillance, le flux spectral. Une autre étude a ensuite conduit des analyses acoustiques des sons de cette première expérience, afin de proposer des descripteurs acoustiques qui rendent bien compte des axes de l'espace de timbre (Krimphoff, McAdams, & Winsberg, 1994). La dimension perceptive appelée "rapidité d'attaque" semblait être liée aux aspects temporels des sons et le logarithme du temps d'attaque semblait bien en rendre compte. La brillance était corrélée avec le centre de gravité spectral, c'est-à-dire la fréquence qui comportait autant d'énergie en amplitude en-deçà qu'au-dessus d'elle (voir Figure 2.14).

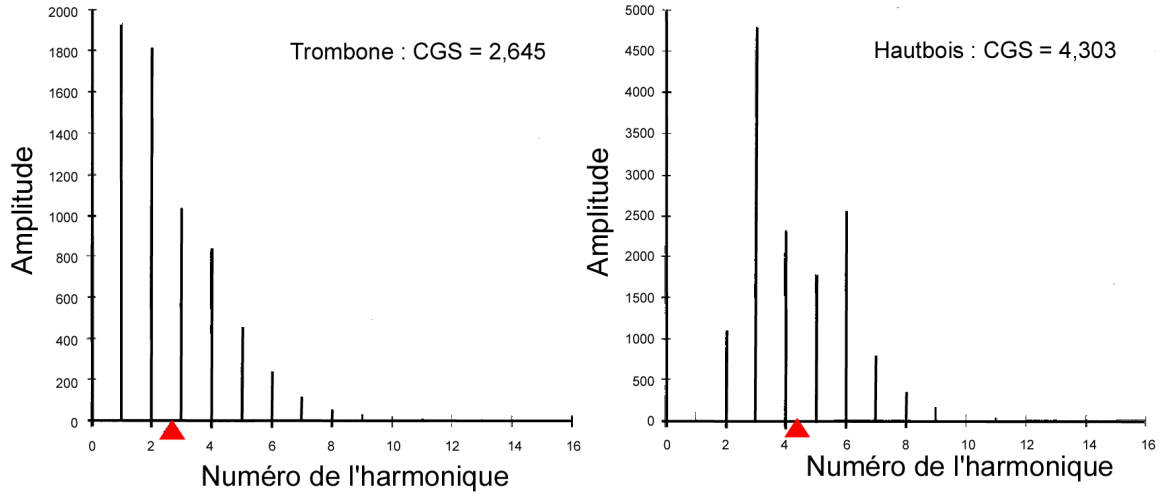


Figure 2.14 : Exemple de deux spectres moyens. Le premier est pour le timbre du trombone (gauche) et le second pour celui d'un hautbois (droite). Leur centre de gravité spectral respectif est indiqué par un triangle rouge. Extrait de Krimphoff *et al.* (1994)

Enfin ce que Krumhansl a appelé le flux spectral correspondait à l'irrégularité du spectre (ou structure fine du spectre), qui reflétait l'écart du spectre réel à un spectre lissé, où les amplitudes d'harmoniques voisines sont peu différentes (voir Figure 2.15). Les corrélations entre les dimensions perceptives et ces trois descripteurs acoustiques étaient élevées (r(19) = 0; 94 ; r(19) = 0; 94 et r(19) = 0; 85 respectivement, p < 0; 01 dans les trois cas (Krimphoff et al., 1994). Ces résultats confirment l'interprétation qualitative de Krumhansl (1989) pour les deux premières dimensions, et proposent une troisième dimension de nature spectrale plutôt que spectro-temporelle.

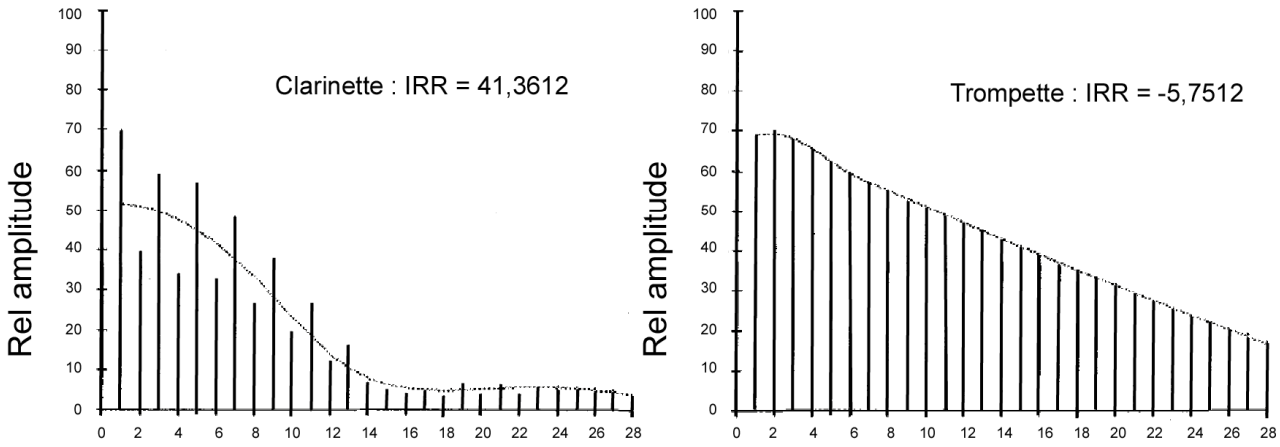


Figure 2.15 : Spectres de raies d'un son de clarinette (gauche) et d'un son de trompette (droite). La ligne continue en gris représente une enveloppe, calculée en prenant la moyenne de trois raies adjacentes. Le paramètre IRR représente le logarithme de la somme des différences entre le spectre de raies et l'enveloppe. Extrait de Krimphoff *et al.* (1994).

Une étude de McAdams *et al.* (1995) a utilisé 18 de ces 21 sons de synthèse de Krumhansl, toujours avec un protocole de jugements de dissemblance (McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995). Les résultats étaient en faveur d'un espace perceptif tridimensionnel, dont les deux premières étaient les mêmes que celle de Krimphoff et la troisième différait un peu. Pour McAdams, la troisième dimension de flux spectral était corrélée avec un paramètre acoustique du même nom qui correspondait à la variation du spectre au cours du temps. Dans cet espace, on a donc une dimension temporelle, une dimension spectrale et une dimension spectro-temporelle, ainsi que d'autre spécificités que nous ne détaillerons pas ici. De façon intéressante, la répartition des sujets dans les différentes classes latentes n'était pas liée au degré d'expertise musicale des auditeurs. Néanmoins, l'expertise musicale réduisait la variabilité des données.

L'ensemble des études avec des jugements de dissemblance rapportées dans les paragraphes précédents montrent qu'un espace de timbre avec deux ou trois dimensions et éventuellement des spécificités permet de bien modéliser les réponses des sujets. Mais ces études montrent également que la perception du timbre s'appuie sur plusieurs indices acoustiques. Il est, à l'heure actuelle, difficile de dire si ces différents espaces de timbre rendent comptent de tous les indices participant à la sensation de timbre. Toutefois, on peut noter qu'il n'y a pas de mots spécifiques à la description du timbre, et que les termes pour le décrire se rapportent souvent à d'autres sens, tels que la vision ou le toucher. S'il est difficile de le décrire, il est donc aussi difficile d'en étudier sa perception. Le fait que la manipulation de divers paramètres acoustiques fasse varier la sensation de timbre chez des sujets (voir paragraphes « Caractéristiques physiques du timbre », p52) nous laisse penser que le timbre est bien perçu. En revanche il n'est que très peu verbalisé.

2.2.4 LA NEUROPSYCHOLOGIE DU TIMBRE

Tant les caractéristiques physiques pour définir le timbre sont nombreuses, tant on peut comprendre que les structures impliquées pour le percevoir peuvent être multiples. De ce fait, peu d'études ont abordé frontalement la question des bases cérébrales de la perception du timbre. En réalité, cette question recouvre un ensemble de sous-questions comme de savoir quelles sont les structures cérébrales mises en jeu dans la perception des différentes dimensions du timbre, ou encore comment le système auditif sépare-t-il et/ou combine-t-il l'information relative aux différentes dimensions du timbre.

Une revue de la littérature de 2003 recense les diverses études traitant du traitement cortical du timbre, et essentiellement de la latéralisation de ce traitement (S. Samson, 2003). Les premières études s'intéressant au traitement cortical du timbre étaient des études de lésions. En 1962, Milner avait comparé des patients épileptiques ayant subi une lobectomie droite à des patients ayant subi une lobectomie gauche dans une tâche de discrimination d'information spectrale dans des sons

harmoniques. Les patients ayant une lobectomie à droite, contrairement aux autres patients, montraient un déficit significatif pour cette tâche, suggérant ainsi que le traitement des informations spectrales impliquées dans le traitement du timbre reposait sur l'intégrité du lobe temporal supérieur droit (Milner, 1962).

Par la suite, Zatorre et ses collaborateurs se sont intéressés à la discrimination des enveloppes spectrales et temporelles du timbre (S. Samson & Zatorre, 1994). Les auteurs ont testés des patients avec une lésion du lobe temporal dans une tâche de discrimination de timbre. Le timbre pouvait varier selon une dimension spectrale, c'est-à-dire ayant un nombre d'harmoniques variable, allant de 3 à 5, ou selon une dimension temporelle avec une durée d'attaque variable, de 1, 50 ou 190 ms. La tâche du sujet était de dire s'il percevait une différence de timbre ou non entre les différentes paires de stimuli présentées. Les résultats ont montré que seuls les patients ayant une lobectomie droite avaient des déficits dans la discrimination de timbres que ce soit sur le plan spectral ou temporel, suggérant ainsi que le lobe temporal droit était essentiel pour le traitement de l'information spectrale et temporelle impliqué dans le timbre. Les auteurs remarquent que ceci remet en question le rôle prédominant de l'hémisphère droit pour le traitement temporel. Cependant, dans sa revue de 2003, Samson, note également que les délais temporels manipulés dans leur étude sont des changements temporel « lent », de l'ordre de la centaine de millisecondes, à l'échelle de ce qu'est capable de traiter l'hémisphère gauche, de l'ordre de la dizaine de milliseconde (S. Samson, 2003). Dans une étude de 2002, Samson *et al.* ont modifier un peu le paradigme de l'étude précédente. Ils ont fait varier la durée de l'attaque, 1, 100 ou 190 ms, et le nombre d'harmoniques, 1, 4 ou 8 harmoniques, créant ainsi 9 timbres différents (S. Samson, Zatorre, & Ramsay, 2002). Ces 9 timbres ont permis de synthétiser 9 sons et 9 mélodies de 8 notes. Les résultats de l'étude ont confirmé que la lésion du lobe temporal droit perturbait la perception des deux dimensions du timbre, temporelle et spectrale. Toutefois, dans cette étude, les auteurs ont aussi montré que la lobectomie du lobe temporal gauche affectait des aspects subtils de la perception du timbre dans la condition mélodique, suggérant que leur représentation du timbre pouvait être moins « stable » que chez des sujets sains. Les auteurs concluent donc que la perception du timbre dépend principalement de l'intégrité du lobe temporal droit, et que le lobe temporal gauche semble aussi contribuer à cette perception.

Des études de neuroimagerie vont aussi dans le sens d'une spécialisation du traitement du timbre par l'hémisphère droit. Dans une étude utilisant la TEP, Platel *et al.* ont montré que l'attention sélective à des changements de timbres de séquences de notes engendrait des activités spécifiques dans des régions frontales de l'hémisphère droit (autour de l'aire 6 de Brodmann), par comparaison à des tâches d'attention sélective à la hauteur ou au rythme (Platel et al., 1997). Egalement, une étude d'IRMf (Halpern, Zatorre, Bouffard, & Johnson, 2004) a permis de caractériser le réseau cérébral mis en jeu lors de jugements de dissemblance sur des paires de timbres présentées auditivement (condition de perception) ou imaginées par le sujet (condition d'imagination). En condition de perception, plusieurs aires du gyrus temporal antéropostéral étaient activées, dont le gyrus de Heschl, le *planum temporale*, et gyrus temporal supérieur, avec une asymétrie en faveur de l'hémisphère droit. Quand le jugement de dissemblance portait sur les timbres imaginés, des aires temporales

postérieures étaient activé, avec également une légère asymétrie du côté droit. Une étude d'IRMf n'a toutefois pas montré de supériorité hémisphérique droite mais une asymétrie d'activation des deux hémisphères pour le traitement du timbre (Menon et al., 2002). En situation d'écoute passive, un timbre X avec une attaque lente, un centre de gravité spectral aigu et dont le spectre variait dans le temps (flux spectral) engendrait plus d'activations dans des régions des lobes temporaux qu'un timbre Y avec une attaque rapide, un centre de gravité spectral plus grave et sans flux spectral. Les régions impliquées dans le traitement du timbre étaient le gyrus de Heschl postérieur et le sillon temporal supérieur jusqu'au sillon cingulaire insulaire. Si aucune région n'était significativement plus activée à l'écoute du timbre X qu'à l'écoute du timbre Y, les patterns d'activation étaient significativement plus postérieurs dans l'hémisphère gauche que dans l'hémisphère droit. Ces résultats suggéraient que les deux lobes temporaux sont impliqués dans la perception du timbre, mais que leurs rôles respectifs et le poids des différentes dimensions du timbre dans ces activations restaient à déterminer.

Nous conclurons en disant que l'étude aussi bien que le traitement cortical du timbre est un phénomène complexe. Il semblerait qu'il y ait une préférence de l'hémisphère droit pour le traitement de cet aspect, en raison du poids des composantes spectrales dans la perception du timbre. Toutefois, une contribution de l'hémisphère gauche semble exister sans que ce dernier soit indispensable. Dans une étude de potentiels évoqués, Meyer *et al.* ont mis en évidence que le traitement du timbre, impliquait différentes structures au cours de ce traitement (Meyer, Baumann, & Jancke, 2006). Il est donc raisonnable de penser que le traitement cérébral du timbre implique un réseau neuronal complexe tant les dimensions participants à la perception de timbre le sont.

2.3 PAS DE PITCH SANS TIMBRE ?

Pour résumer, on peut dire que la perception du pitch résulte en une opération de groupement des harmoniques, ou partiels, ou fréquences qui composent un son, tandis que la perception du timbre correspond plutôt à une opération de décomposition perceptive du son. Dans tous les cas le pitch et le timbre ont donc des composantes spectrales.

Mais la perception du pitch et du timbre semble liée. En effet plusieurs études on ta rapporté que la manipulation du timbre, c'est-à-dire de la répartition de l'énergie spectrale, pouvait influencer la perception du pitch (B. C. Moore & Glasberg, 1990; Pitt, 1994; Singh & Hirsh, 1992). Par exemple, dans l'étude de Singh et Hirsh en 1992, les sujets écoutaient deux sons complexes consécutifs qui pouvaient varier soient en fonction du timbre (énergie spectrale) soit en fonction du timbre. La tâche était de dire si les deux sons étaient les mêmes ou étaient différents de par leur timbre, leur pitch ou les deux. Les résultats montraient que la variation du timbre pouvait influencer la perception du pitch, c'est-à-dire que les sujets percevaient une variation de pitch même lorsque seul le timbre variait. Plus tard, Pitt a montré que cet effet était plus marqué chez les sujets non

musiciens (Pitt, 1994). Une étude en EEG a également pu montrer que la manipulation du timbre pouvait influencer la réponse en potentiels évoqués correspondant à l'onde N100 classiquement associé à la perception du pitch (Ritter, Dosch, Specht, Schneider, & Rupp, 2007).

Nous devons donc souligner la dépendance de la perception du pitch et du timbre. Cette dépendance semble plutôt s'observer du pitch en fonction du timbre.

2.4 DUREE

La durée est un attribut important du stimulus auditif. L'information temporelle est particulièrement importante dans le langage, où elle va procurer des informations phonétique et prosodiques relatives à l'enveloppe temporelle, à la périodicité ainsi qu'à la structure fine (Rosen, 1992). Certaines limites phonétiques ne sont définies que sur la durée de portions spécifique de la vocalisation. Par exemple, la différence principale entre [ta] et [da] repose sur la durée, en termes de millisecondes, de la partie non voisée (correspondant à l'attaque du son) avant la phonation de la voyelle (Fant, 1960). Sur une échelle de temps plus longue, l'organisation prosodique suprasegmentale du langage, contient également des informations subtiles concernant le sens de la phrase ou l'émotion du locuteur. Par exemple, la durée des pauses ou des syllabes dans le discours sont des cruciaux (Monrad-Krohn, 1963). La discrimination de la durée du son joue donc un rôle très important.

Sur le plan neurophysiologique, des données expérimentales chez l'homme suggèrent que le traitement de l'information temporelle implique des structures cérébrales. Au niveau du cortex auditif, plusieurs études suggèrent une implication du plan supratemporal dans le traitement de la durée du son (Jaramillo, Paavilainen, & Naatanen, 2000; Kaukoranta, Sams, Hari, Hamalainen, & Naatanen, 1989; Milner, 1962). En 2001, Rao *et al.* montraient dans une étude IRMf des traitements différents selon que le traitement était associé à l'encodage de la durée ou à la comparaison d'intervalles de temps (Rao, Mayer, & Harrington, 2001) : les activations du cortex pariétal inférieur droit et du cortex prémoteur bilatéral étaient précoces et associées avec l'encodage d'intervalles de temps, tandis que l'activation plus tardive du cortex préfrontal dorsolatéral droit apparaissait durant la comparaison d'intervalles de temps. Des structures sous-corticales et le cervelet étaient aussi impliqués dans de tels traitements.

Belin *et al*. en 2002, montraient toujours en IRMf que pendant une tâche de discrimination, un réseau fronto-pariétal droit s'activait, ainsi que le cortex préfrontal droit (BA10), les hémisphères cérébelleux et ganglion de la base bilatéraux (Belin, McAdams et al., 2002). Les auteurs suggéraient que la discrimination de la durée du son impliquait deux réseaux cérébraux ; l'un fronto-pariétal droit responsable de l'allocation de ressources attentionnelles, et l'autre, incluant le cortex préfrontal droit, le cervelet et les ganglions de la base, plus spécifiquement impliquées dans les aspects temporels de la tâche de discrimination.

2.5 INTENSITES

2.5.1 DEFINITION ET SEUILS

La pression acoustique est mesurée en micropascal (µPa), la puissance d'une source acoustique s'exprime en watts (W) tandis que l'intensité acoustique s'exprime en W/m². Comme pour grandeur physique du son la grandeur subjective associée ne correspond pas à une relation linéaire. En ce qui concerne la perception de l'intensité sonore, celle-ci est mesurée en décibel (dB) et est défini par le rapport de la puissance par unité de surface du son que l'on mesure et une puissance par unité de surface de référence. Ainsi le niveau acoustique d'intensité, exprimé en décibels (dB), est défini comme : $N_I = 10 \log (I/I_0)$, où I est l'intensité acoustique de l'onde et I_0 une intensité acoustique de référence correspondant au seuil de l'audition, fixée à 10^{-12} W/m². Une différence de niveau acoustique égale à 20 dB correspond à un rapport d'intensité acoustique égal à 100 ; une différence de 30 dB correspond à un rapport de 1000. On emploie également le niveau acoustique de pression, exprimé aussi en dB (en anglais, dB SPL pour *sound pressure level*), et défini comme : $N_P = 20 \log (P/P_0)$, où P est la pression acoustique de l'onde et P_0 une pression acoustique de référence correspondant au seuil de l'audition, fixée à 2.10^{-5} Pa.

En ce qui concerne les seuils d'intensité auditive, il faut noter également que les courbes sont des moyennes statistiques sur un grand nombre de sujets, dont on a tiré des normes. Le seuil d'audition normalisé est défini par les limites inférieure et supérieure de sensibilité de l'appareil auditif humain à des stimuli acoustiques, calculées en fonction de la fréquence. On définit ainsi une courbe de seuil d'audition normal, marquant la limite inférieure moyenne de la sensibilité auditive, et une courbe de seuil intolérable normal, qui en marque la limite supérieure. On appelle aire d'audition la plage auditive délimitée par ces deux seuils.

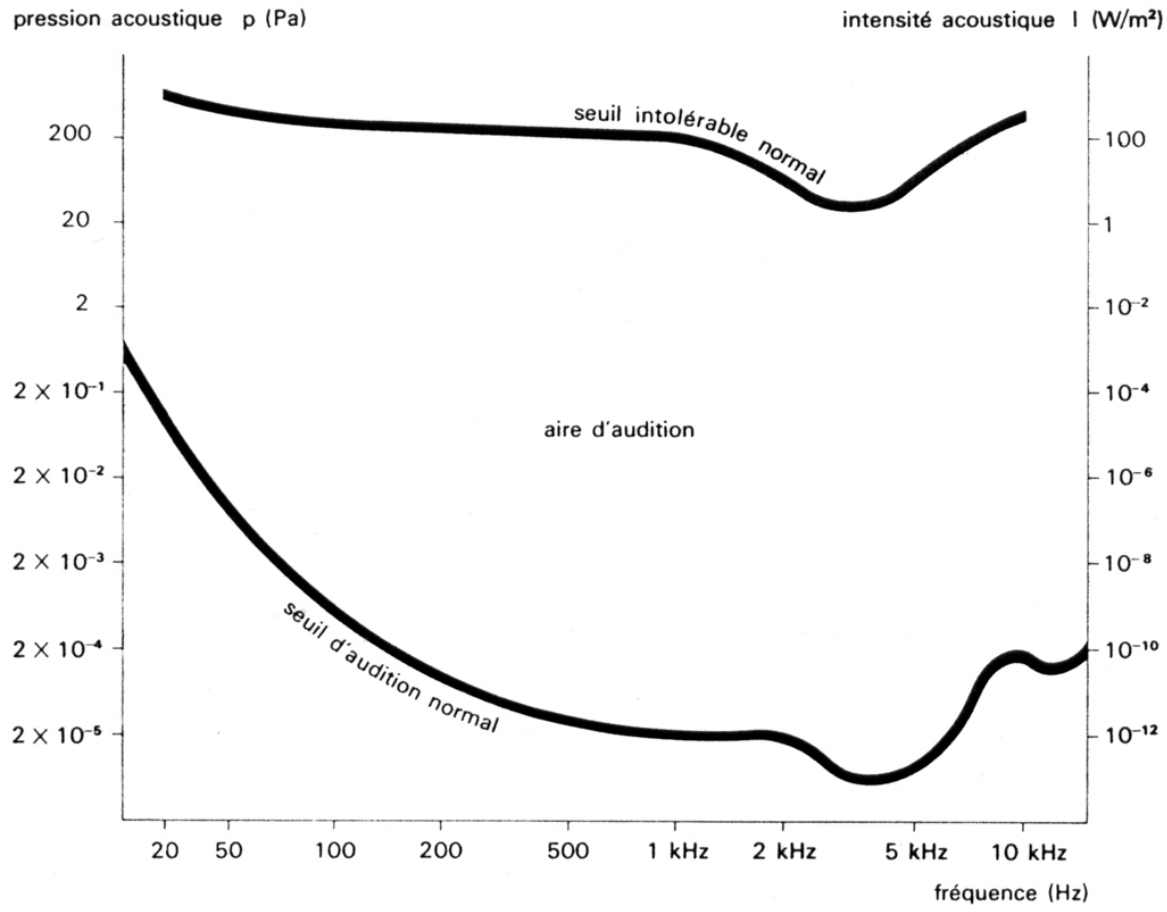


Figure 2.16 : Courbes des seuils de l'audition humaine. D'après Galifret (1991)

2.5.2 ECHELLE DES PHONES OU ECHELLE D'ISOSONIE

On appelle sonie l'intensité subjective. Depuis Fletcher et Munson, on dispose d'un ensemble de courbes appelées lignes d'isosonie(Fletcher & Munson, 1933). Chacune de ces courbes relie les coordonnées (niveau de pression acoustique et fréquence) des sons purs qui donnent à l'oreille humaine une sensation d'intensité égale (voir Figure 2.17). Le phone est l'unité qui sert à exprimer le niveau d'isosonie. Notons que l'échelle des phones correspond avec l'échelle des dB SPL pour un son de 1000Hz.

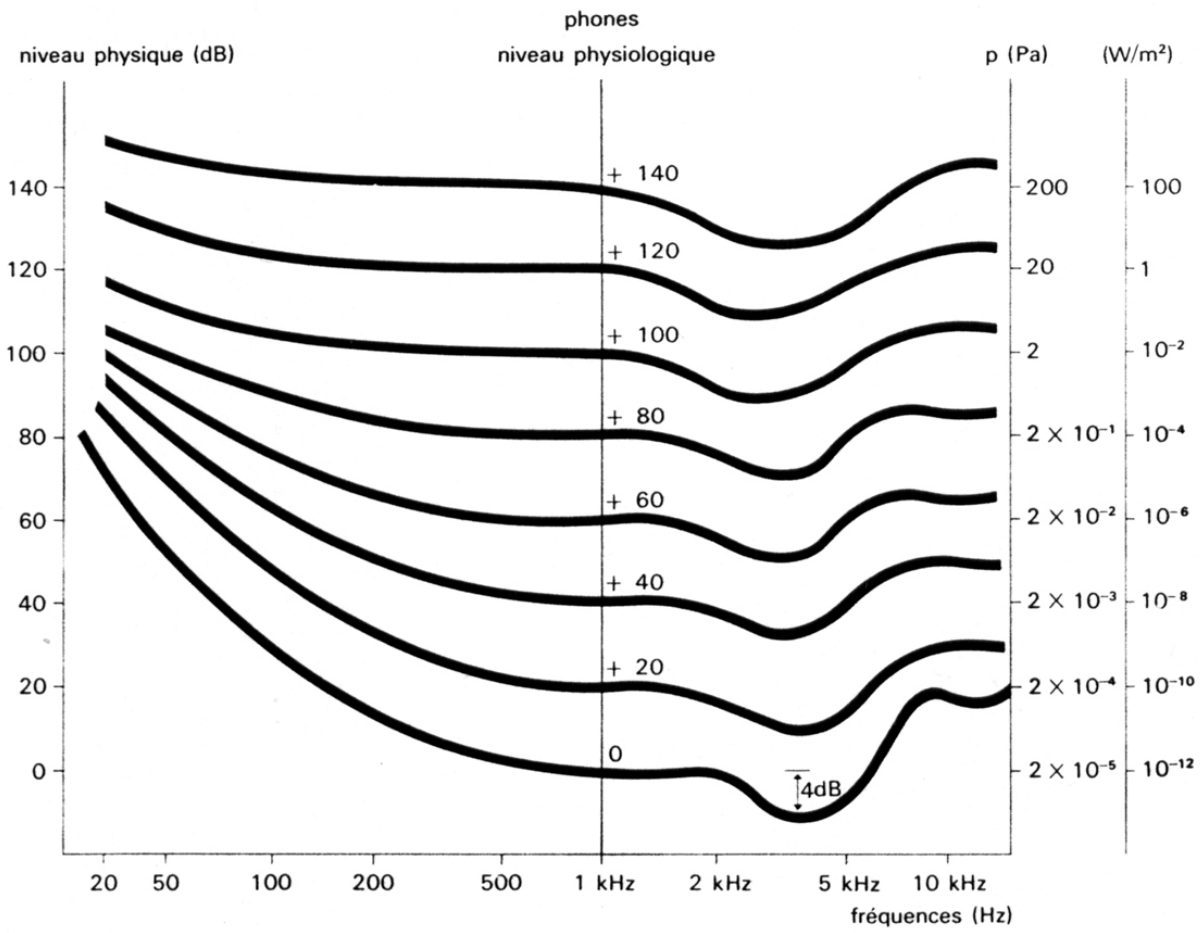


Figure 2.17 : Courbes isosoniques normalisées de l'audition humaine. D'après Galifret (1991).

On remarque une particularité intéressante de l'oreille à l'étude de ces courbes. La zone fréquentielle la plus sensible à nos oreilles est située entre 1000 et 5000 Hz.

2.5.3 FACTEUR DE VARIATION DE LA SONIE

Il semble que la sensation d'intensité soit influencée par plusieurs paramètres. Tout d'abord, le niveau de pression acoustique est, comme nous pouvons le voir sur la Figure 2.17, un facteur important. La nature du signal, a également un impact sur la sonie. Dans le cas d'un bruit large bande, les résultats expérimentaux montrent que la sonie n'obéit plus à une loi de puissance, que la sonie augmente plus rapidement aux faibles niveaux et que la sonie d'un bruit est toujours supérieure à celle d'un son pur, à intensité physique égale. De plus, suivant que les stimuli sont présentés en ordre aléatoire, croissant, ou décroissant ; dans ce dernier cas ils dépendront également du niveau initial, par un phénomène similaire à l'adaptation de la sonie. Les résultats varient encore avec la gamme des stimulations présentées (son amplitude), la première stimulation présentée, les instructions données, la symétrie de la gamme des réponses demandées. La durée du signal est un

autre facteur pouvant influencer la sonie. La sonie augmente avec la durée du signal présenté. La focalisation de l'attention sur certains paramètres améliore les performances. Par exemple, la performance de détection d'une sinusoïde dans le bruit est meilleure si la fréquence de la sinusoïde est connue a priori : le sujet focalise alors son écoute sur la fréquence qu'il cherche à entendre. Enfin l'expérience, la motivation et l'entraînement des sujets ont aussi une influence sur les résultats.

La composition spectrale a également été étudiée. On observe qu'à partir d'une certaine largeur de bande (160 Hz autour de 1 kHz), la sonie commence à augmenter Les recherches expérimentales en psychoacoustique ont permis de préciser la façon dont le spectre d'un signal sonore incident est traité par le système auditif. Un important ensemble d'études convergent pour indiquer qu'on peut assimiler le traitement acoustique effectué par le système auditif périphérique à une succession de filtres passe-bande se chevauchant, appelés filtres auditifs.

La notion de filtre auditif a été proposée pour la première fois par Fletcher (Fletcher, 1940) dans une expérience de masquage psychoacoustique. Le principe de l'expérience d'origine était le suivant : un son pur est masqué par un bruit blanc de bande dont la fréquence centrale est identique à celle du son pur. On fait varier la largeur de bande du bruit blanc tout en conservant la densité spectrale d'énergie constante, de sorte que le niveau sonore augmente avec la largeur de la bande. En exprimant le niveau sonore à partir duquel le son pur est perçu en fonction de la largeur de bande du bruit blanc, on s'aperçoit que ce niveau augmente dans un premier temps, puis se stabilise à partir d'une certaine largeur de bande. Cette largeur de bande critique correspondrait ainsi à la largeur spectrale d'un filtre auditif passe-bande rectangulaire centré sur la fréquence du son pur utilisé, de sorte que les sons dont la fréquence est située en dehors de cette bande spectrale ne jouent aucun e← et masquant pour la fréquence centrale du filtre.

La forme rectangulaire initialement proposée pour les filtres auditifs (Fletcher, 1940) n'est en réalité qu'une approximation grossière. Des méthodes fines permettant de mesurer les formes des filtres auditifs ont été proposées par la suite (Patterson, 1976), ce qui a permis de mettre en évidence une forme quasi-triangulaire dont le sommet correspond à la fréquence centrale. La sélectivité en fréquence des filtres auditifs est toutefois dépendante du niveau sonore de stimulation, et on peut observer l'apparition d'une asymétrie en faveur des basses fréquences lorsque le niveau sonore augmente (B. C. J. Moore, 1989).

2.6 LA SOURCE SONORE : NATURE ET LOCALISATION

L'environnement sonore dans lequel nous sommes plongés quotidiennement est à la fois complexe et variable, et pourtant nous évoluons avec aisance dans ce monde acoustique. En fait, deux problèmes principaux se posent à notre système auditif : d'une part, dissocier au sein du mélange acoustique qui arrive à nos oreilles les contributions des différentes sources sonores (problème de l'analyse de la scène auditive) afin reconnaître ces sources pour ensuite agir de manière

appropriée ; d'autre part et de façon indépendante du premier problème, localiser la source sonore dans l'environnement.

2.6.1 NATURE DE LA SOURCE

Les objets sonores auxquels nous sommes confrontés sont divers, d'origine humaine (bruit de pas, voix), animale (cris d'animaux), résultant de l'activité humaine (le train qui passe, un piano), ou autre (le bruit du vent dans les arbres). Les sons émis par chacune de ces sources ont un contenu spectro-temporel qui leur est propre, et qui dépend à la fois du corps entrant en vibration, et de l'évènement qui a déclenché cette vibration (un choc, un frottement, . . .). On suppose donc que cette capacité à reconnaître les sources sonores repose sur le traitement par notre système perceptif de la structure spectro-temporelle d'un stimulus auditif. La reconnaissance des sources sonores est un problème complexe que le système auditif parvient très bien à résoudre. Il est souvent présentée comme subséquente à l'extraction des attributs auditifs.

Les études sur la reconnaissance des sources sonores posent souvent la question de savoir quels indices acoustiques sont utilisés par le sujet pour reconnaître la source. Le problème est le suivant : selon le contexte, un même objet va émettre des sons assez différents. Par exemple, le spectre d'un son d'instrument de musique va dépendre de l'intensité avec laquelle on en joue ; de même, une voix est modifiée par un rhume. Il s'agit donc d'extraire les bons invariants acoustiques, c'est-à-dire les informations qui permettent d'isoler des effets du contexte et la contribution de la source. Il n'y a pas aujourd'hui de consensus sur la nature de ces invariants (Handel, 1995; McAdams, 1993). La reconnaissance des sources sonores fait également appel aux connaissances antérieures de l'auditeur, et dépend donc des informations disponibles en mémoire.

2.6.2 LOCALISATION

On s'intéresse aux capacités de localisation en direction et en distance. La direction d'incidence du son est identifiée selon les conventions suivantes ; l'azimut est l'angle que fait la direction incidente avec le demi-plan vertical situé dans l'axe du visage du sujet tandis que l'élévation est l'angle que fait cette direction incidente avec un plan horizontal.

2.6.2.1 LA LOCALISATION EN AZIMUT

Comme l'ont proposé Stevens et Newman en 1936, il semble que deux indices binauraux soient impliqués essentiellement dans la localisation. Le premier est efficace surtout pour les basses fréquences. Il s'agit de la Différence Interaurale de Temps (*Interaural Time Difference* ou ITD). L'autre indice, efficace surtout en hautes fréquences, est la Différence Interaurale d'Intensité (*Interaural Intensity Difference* ou IID).

2.6.2.2 LA LOCALISATION EN ELEVATION

L'étude des capacités de localisation en élévation donne des résultats très différents suivant le signal utilisé. Pour des sons de parole, nous avons une capacité de localisation raisonnablement bonne : de l'ordre de 10° pour des sons provenant de l'avant, un peu moins bonne pour des sons provenant de l'arrière. Pour des sons purs en revanche, les résultats sont tout à fait différents : l'élévation perçue d'un son pur n'est quasiment pas fonction de sa direction d'incidence réelle, mais uniquement de la fréquence. Ainsi, un son de basse fréquence semble provenir d'une source « basse », un son de haute fréquence d'une source « haute ». En fait on constate que c'est la largeur de bande du signal qui est déterminante. Une largeur de bande d'au moins deux tiers d'octave est nécessaire pour permettre une capacité de localisation en élévation.

2.6.2.3 LA LOCALISATION EN DISTANCE

L'étude de la distance subjective de la source montre que pour un son pur, celle-ci est en premier lieu liée à l'intensité perçue. Paradoxalement, un son pur semble provenir d'autant plus loin que son intensité est élevée. Dans le cas de sons complexes la perception de la distance utilise de multiples indices. Premièrement le niveau sonore absolu notamment pour des sources multiples, deuxièmement les variations de spectre : celui-ci change sur des distances modérées, avec l'absorption de l'air, les hautes fréquences étant plus rapidement atténuées que les basses fréquences et enfin les rapports d'intensité et les délais temporels son direct / sons réfléchis en espace clos.

Notre capacité d'exploitation de ces indices varie notablement selon notre familiarité avec la source. La précision dans tous les cas reste faible, avec des erreurs courantes de l'ordre de 20%.

3 LA VOIX

La voix humaine est probablement le son le plus important de notre environnement auditif. Un son que nous devons détecter, analyser, identifier et dont nous devons comprendre le contenu. Outre le langage, apparu très tard au cours de l'évolution de l'espèce humaine, la voix comporte beaucoup d'informations, que beaucoup d'espèces animales savent exploiter pour communiquer. En effet, si la majorité des espèces ont une communication sonore, certaines possèdent un tractus vocal qui leur permet d'émettre des sons vocaux variés qui seront alors reconnus et interprétés par leurs congénères.

Cette partie de la thèse porte précisément sur la nature et l'origine de la voix d'un point de vue physiologique ainsi que sur ce qui la caractérise et en fait un stimulus particulier d'un point de vue acoustique, et comment ce stimulus particulier est-il traité au niveau cérébral.

3.1 FORMATION DE LA VOIX

Il est courant de séparer l'appareil de production sonore humaine en trois parties. Tout d'abord l'appareil respiratoire forme avec les poumons un réservoir d'air qui apporte l'énergie nécessaire à la production sonore. Les deux autres parties sont considérées comme étant la source et le filtre sonore. La source sonore correspond à l'appareil laryngé, avec notamment les cordes vocales qui génèreront le son et ce son sera ensuite modulé par l'appareil supralaryngé, c'est-à-dire le filtre sonore.

3.1.1 LA SOUFFLERIE PULMONAIRE

La respiration humaine est constituée de deux phases : l'inspiration et l'expiration. Seule l'inspiration est sous contrôle musculaire : la contraction des muscles intercostaux et du diaphragme provoque une expansion de la cage thoracique et un appel d'air. Au cours d'une inspiration normale, les poumons font entrer environ un demi-litre d'air oxygéné. Lors de l'expiration, tous les muscles thoraciques se relâchent et le poids de la cage thoracique provoque l'expulsion passive de l'air.

La production sonore apparaît majoritairement lors de la phase d'expiration. Normalement, les phases d'inspiration et d'expiration sont de durée relativement égale. Mais on peut remarquer que lors de la production de la parole le cycle respiratoire se modifie avec seulement 15% du cycle dévolue à l'inspiration. Il existe également des productions vocales lors de l'inspiration, tel que le rire.

3.1.2 LE LARYNX

3.1.2.1 ANATOMIE

Le larynx est un organe situé au niveau de la gorge à l'intermédiaire entre le pharynx et la trachée et abrite les cordes vocales. Il est suspendu à l'os hyoïde et constitué d'un assemblage de cartilages reliés par des ligaments et recouverts de muscles et d'une muqueuse. Le larynx assure trois fonctions ; une fonction respiratoire, faisant partie intégrante des voies respiratoires, il joue également un rôle dans la déglutition, car en se fermant il protège les voies aériennes, et enfin, il est l'organe de la production de sons vocaux.

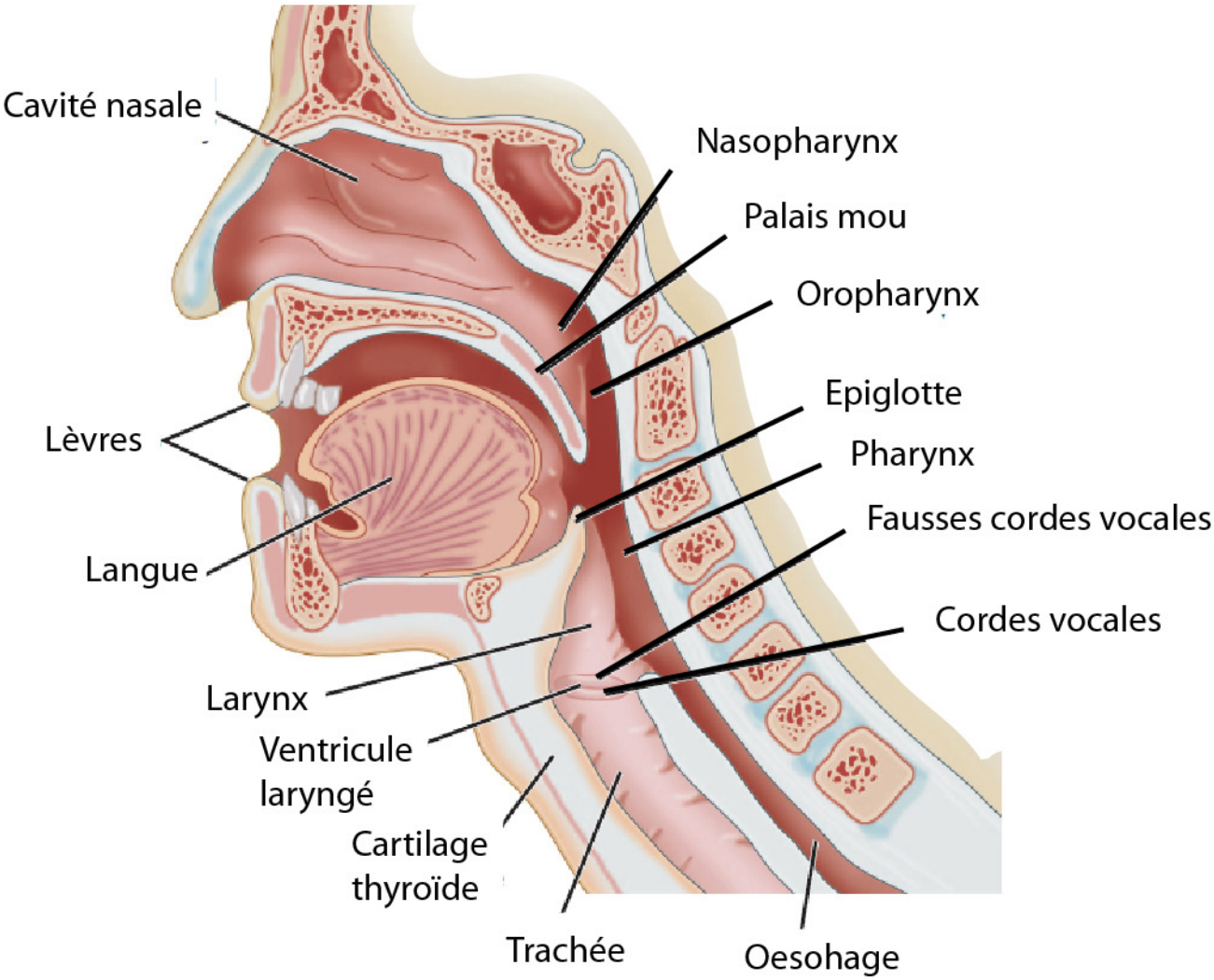


Figure 3.1 : Coupe sagittale de la bouche et du cou, comprenant le larynx et l'appareil supralaryngé.

Ses dimensions sont, chez l'homme, de 45 mm de haut et 35 mm de diamètre antéropostérieur à sa partie supérieure. Elles sont plus réduites chez la femme (respectivement 35 mm et 25 mm). Le cartilage cricoïde représente la base du larynx en forme d'anneau et fait suite à la

trachée. Sur le bord postérieur de ce cartilage, sont disposés symétriquement les cartilages aryténoïdes. Au nombre de deux, ils ont la forme de petite pyramide à leur sommet supérieur, et s'articulent de façon complexe au cricoïde au niveau de leur base inférieure. Ces cartilages, sur lesquels s'insèrent les cordes vocales, s'écartent et se rapprochent sous l'action de certains muscles, en glissant, pivotant et basculant sur la surface articulaire cricoïdienne. Les aryténoïdes peuvent donc glisser et/ou basculer, et ainsi avoir des mouvements variés et très précis. Ceci est très important car ce sont ces mouvements qui vont déterminer les degrés d'ouverture et de fermeture de la glotte, ainsi que la tension des cordes vocales.

Sur le bord antérieur du cricoïde vient se baser le cartilage thyroïde. C'est le « bouclier du larynx » et il correspond extérieurement au relief de la pomme d'Adam. Il a la forme d'un livre ouvert en arrière, et les deux lames quadrangulaires qui le constituent mesurent 1,5 à 2,5 centimètres de hauteur. Elles forment entre elles un angle de 120 degrés chez la femme et 90 degrés chez l'homme (ce qui explique que la pomme d'Adam soit plus saillante chez les hommes). Les lames sont prolongées à leur bord postérieur par des cornes (grandes cornes en haut, petites cornes en bas). De chaque côté, les petites cornes s'articulent chacune avec le cartilage cricoïde, en formant l'articulation crico-thyroïdienne qui permet un mouvement de bascule du cartilage thyroïde sur le cartilage cricoïde. Les grandes cornes thyroïdiennes, supérieures, servent de point d'insertion aux muscles qui vont arrimer le larynx dans le cou.

L'épiglotte est un cartilage en forme de « tuile de toit » à grosse extrémité supérieure, et dont la partie inférieure en forme de pointe, s'insère dans l'angle rentrant du cartilage thyroïde par un ligament, immédiatement au-dessus des cordes vocales. Pendant la déglutition, l'épiglotte se rabat sur le tube laryngé comme un couvercle, protégeant ainsi le larynx et l'arbre bronchique des pénétrations alimentaires.

La partie interne du larynx est tapissé par une muqueuse, elle-même doublée par une membrane fibroélastique tendue du ligament ary-épiglottique en haut, à l'arc cricoïdien en bas. Cette membrane comporte successivement deux replis dirigés vers l'intérieur. Les replis inférieurs sont les cordes vocales. Le relief des cordes vocales est formé de chaque coté par le muscle thyro-aryténoïdien, ou muscle vocal, auquel est accolé le ligament vocal, le tout étant recouvert de la muqueuse vocale : il s'agit de l'étage ou plan glottique. Les replis supérieurs sont appelés fausses cordes vocales ou plis vestibulaires, et sont sous-tendus par le muscle thyro-aryténoïdien latéral et le ligament thyro-aryténoïdien supérieur. Les trois étages du larynx sont délimités par le plan glottique ; on distingue donc le plan sus-glottique ou étage vestibulaire et le plan sous-glottique qui se poursuit par la trachée.

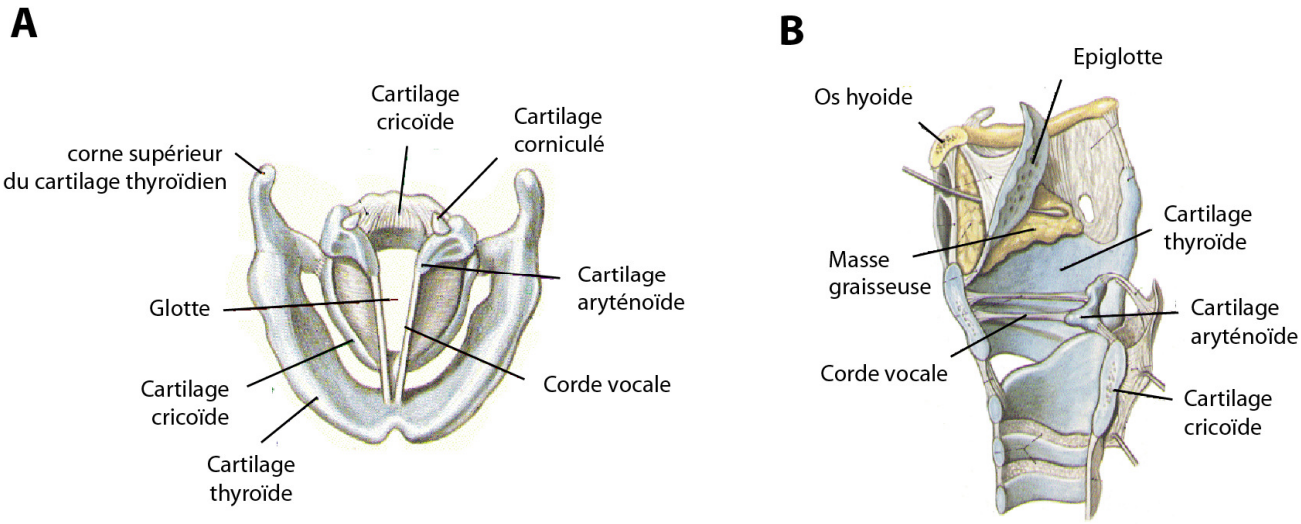


Figure 3.2 : Structure anatomique du larynx. (A) Vue du larynx par-dessous (le côté antérieur est en bas du schéma). (B) Coupe sagittale du larynx (le côté antérieur est à gauche du schéma).

Le muscle crico-thyroïdien est tendu entre le cartilage cricoïdien en bas, et le cartilage thyroïde en haut. Quand il se contracte, il attire le cartilage thyroïde vers le bas et l'avant, en le faisant basculer sur un axe horizontal passant par les deux petites cornes thyroïdiennes. Ce mouvement de bascule vers l'avant a pour effet d'étirer les cordes vocales. Ce muscle est appelé tenseur de la corde vocale. Le muscle crico-aryténoïdien postérieur s'insère à la face postérieure du cricoïdien, se dirige en haut et en dehors pour se terminer sur l'apophyse musculaire de l'aryténoïde. En se contractant, il fait pivoter vers l'extérieur le cartilage aryténoïdien sur un axe vertical passant par son sommet. Il écarte donc les cordes vocales : il est dilatateur de la glotte, ou abducteur de la corde vocale. Le muscle crico-aryténoïdien latéral se fixe à la partie latérale de l'arc cricoïdien en avant et se termine sur l'aryténoïde. Quand il se contracte, ce muscle rapproche les cordes vocales : il est constricteur de la glotte, ou adducteur de la corde vocale. Enfin, le muscle inter-aryténoïen unit les faces postérieures des deux aryténoïdes par des fibres transversales et obliques. Quand il se contracte, ce muscle rapproche les deux aryténoïdes au niveau de leur corps : il est constricteur de la glotte, ou adducteur de la corde vocale.

3.1.2.2 PRODUCTION SONORE

La production sonore correspond à la mise en vibration des cordes vocales. Pour que les cordes vocales puissent vibrer, il faut d'abord qu'elles viennent au contact de l'une de l'autre, c'est-à-dire qu'elles se ferment. Cette mise en position phonatoire, ou adduction, est produite par rapprochement des aryténoïdes sous l'action du muscle interaryténoïdien. Simultanément, les muscles crico-aryténoïdiens latéraux complètent la fermeture en rapprochant les deux apophyses vocales. La vibration peut commencer sous l'action de l'air expiré, et pendant toute la durée de la phonation, les inter-aryténoïdiens resteront contractés, assurant ce verrouillage postérieur.

Le phénomène de vibration est décrit par la théorie myo-élastique. Cette théorie fait appel à deux forces opposées : la pression sous-glottique et la pression d'accolement des cordes vocales. Lors de l'expiration, l'augmentation de la pression sous-glottique, devenant alors supérieure à la pression d'accolement des cordes vocales, permet l'écartement des cordes vocales et l'échappement d'une petite quantité d'air. Puis, lorsque le flux d'air laminaire traverse cette cavité à parois molles qu'est l'espace sus-glottique, le flux d'air dans la partie sous-glottique diminue en pression, ce qui provoque un effet d'aspiration, de dépression, tendant à rapprocher puis fermer les cordes vocales (effet ou phénomène de BERNOUILLI). Une fois les cordes vocales de nouveau accolées, la pression sous glottique augmente à nouveau et tend à pousser puis écarter les bords libres des cordes vocales, jusqu'à l'échappement d'une nouvelle petite quantité d'air. Ce phénomène va ainsi se reproduire de manière cyclique et être à l'origine de la vibration des cordes vocales et donc de la vibration de l'air qui va être expulsé via le larynx. Il y alors formation d'une onde acoustique. La fréquence de reproduction de ce cycle représente la fréquence de la voix ou hauteur tonale.

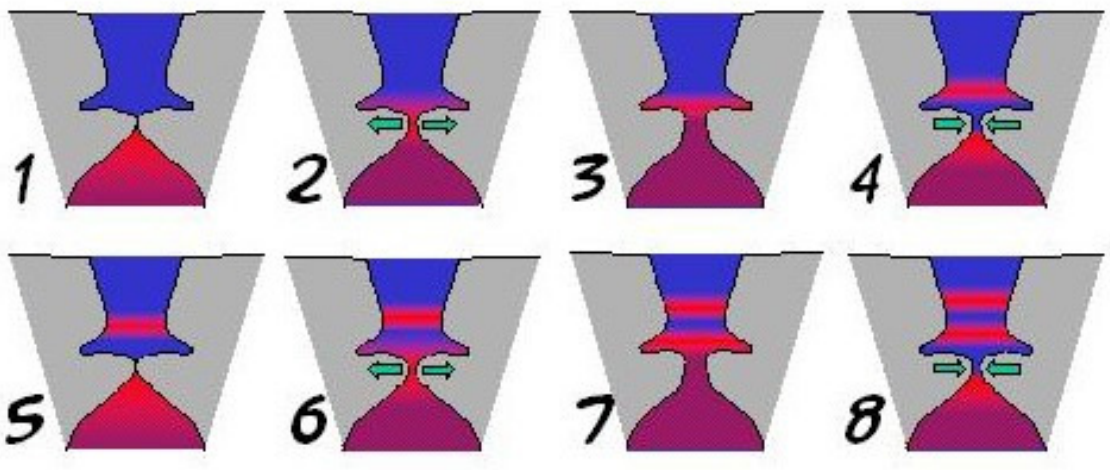


Figure 3.3 : Coupe schématisée du larynx durant deux cycles. Les couleurs indiquent les zones de pression (bleu pour des dépressions et rouge pour des surpressions).

Pour la phonation des sons graves, les cordes vocales semblent courtes et boudinées. Il existe une tension faible, à la fois dans le muscle vocal et dans les muscles qui fixent les aryténoïdes. L'amplitude vibratoire est importante, de l'ordre de 3 à 5 mm. A l'inverse, pour les sons aigus, les cordes vocales apparaissent minces et tendues et s'allongent de plusieurs mm. Elles ne s'accolent qu'à la partie haute de leur bord libre. L'amplitude vibratoire est plus faible (1/2 à un mm).

3.1.3 L'ESPACE SUPRALARYNGE

L'onde sonore provoquée par la vibration des cordes vocales se propage aux molécules de l'air occupant des cavités supralaryngées : en traversant ces cavités, l'onde sonore est propagée, mais également modifiée. En effet, on considère que cette espace supralaryngé agit comme un filtre de l'onde sonore. L'onde sonore générée va plus ou moins faire entrer en résonance les parties de l'espace supralaryngé. L'entrée en résonance ou non dépend essentiellement de la nature (taille, densité, forme, etc.) et de la position des composants du résonateur. Chaque partie du résonateur a sa fréquence de résonance propre. Ainsi, pour chaque son un certain nombre de fréquences vont être amplifiées par les composants de l'espace supralaryngé.

L'espace supralaryngé, ou conduit vocal, est le terme qui désigne la cavité remplie d'air qui s'étant de la glotte (larynx) aux lèvres. La forme de cette cavité peut-être modifiée par le mouvement. Le conduit vocal se compose du pharynx, de la cavité nasale, de la cavité orale (ou buccale), et éventuellement de la cavité labiale.

Le pharynx est un conduit musculo-membraneux à direction verticale qui aboutit en bas, d'une part au larynx, d'autre part à l'œsophage. Il est situé en avant de la colonne cervicale, en arrière des fosses nasales. On considère généralement 3 parties, le laryngo-pharynx, l'oro-pharynx et le naso-pharynx. Le pharynx a la forme d'un entonnoir irrégulier dont la longueur et le diamètre peuvent varier en fonction de l'activité des muscles qui le constituent.

La cavité nasale ou fosses nasales sont deux espaces remplis d'air au-dessus et en arrière du nez, au milieu de la face. C'est la position du palais mou (qui appartient à la cavité buccale) qui détermine si l'air expiré durant la phonation s'échappe uniquement par la bouche (palais mou relevé), par la bouche et le nez (palais mou partiellement abaissé) ou uniquement par le nez (palais mou totalement abaissé). La cavité nasale est donc peu mobile, mais le passage de l'air par cette cavité modifie considérablement la qualité du son du résonateur.

La cavité orale proprement dite est délimitée en avant et sur les côtés par les dents, et en arrière par l'isthme du gosier (qui sépare la cavité buccale de l'oropharynx). La cavité orale comprend donc les dents, les joues, la langue, la luette, les amygdales et le palais mou. La langue présente une structure extrêmement complexe avec 17 muscles. Ses mouvements sont nombreux et variés. Sa base, postérieure, est relativement peu mobile, mais sa pointe ou apex, antérieure, est très souple et peut prendre des formes très diverses. Son rôle est capital dans l'articulation des sons et des bruits de la parole. Le voile du palais est une cloison musculo-membraneuse en continuité avec le plancher des fosses nasales. Il est constitué de 5 paires de muscles, le tout recouvert d'une muqueuse, et est donc particulièrement mobile. Certains auteurs considère la cavité labiale à part tant son rôle dans la phonation est complexe et important. Il s'agit en fait de la sortie de la cavité buccale. A l'état de repos, les lèvres se situent au contact l'une de l'autre. Dans la parole, elles produisent des mouvements complexes modifiant la longueur du résonateur buccal et son degré d'aperture. Les

cavités buccale et labiale sont donc largement susceptibles de changer de forme, de par les différents muscles qui la composent. Leur rôle dans la phonation est primordial.

En résumé, on peut considérer que, si les cordes vocales vont déterminer la fréquence fondamentale de la voix, l'espace supralaryngé lui donnera son timbre.

3.2 CARACTERISTIQUES PHYSIQUES DE LA VOIX

Après avoir vu comment se formait la voix nous allons maintenant définir, d'un point de vue physique, ce qui caractérise la voix, la rend particulière et différenciable des autres sons. Comme nous l'avons vu, la fréquence fondamentale, ou le pitch de la voix, va dépendre des cordes vocales, c'est-à-dire de l'espace laryngé, et les formants, c'est-à-dire en partie le timbre de la voix, va dépendre de l'espace supralaryngé.

La production spectrale des cordes vocales est très riche en harmoniques (voir Figure 3.4). C'est d'ailleurs une propriété générale des cordes (pas seulement vocales, mais aussi guitare, piano, etc.) : lorsqu'elles subissent une excitation mécanique (le flux d'air expiré dans le cas de la voix), elles répondent en émettant leur fondamentale accompagnée de toutes ses harmoniques. Donc le spectre obtenu en sortie directe des cordes vocales ressemble à un peigne, puisque composé de tous les multiples de la fondamentale avec des amplitudes régulières (pas forcément toutes les mêmes, mais décroissant de façon régulière et progressive). Le timbre va donc être fonction des différentes cavités résonatrices (larynx, bouche, etc.) ; selon leur propre fréquence de résonance, ils vont amplifier ou atténuer certaines de ces harmoniques. Le spectre final du son émis contient toujours toutes les harmoniques initiales, mais avec des amplitudes modulées, voire quasi-nulles, en fonction des fréquences des formants (ce qui donne des creux et des bosses dans le profil final du spectre).

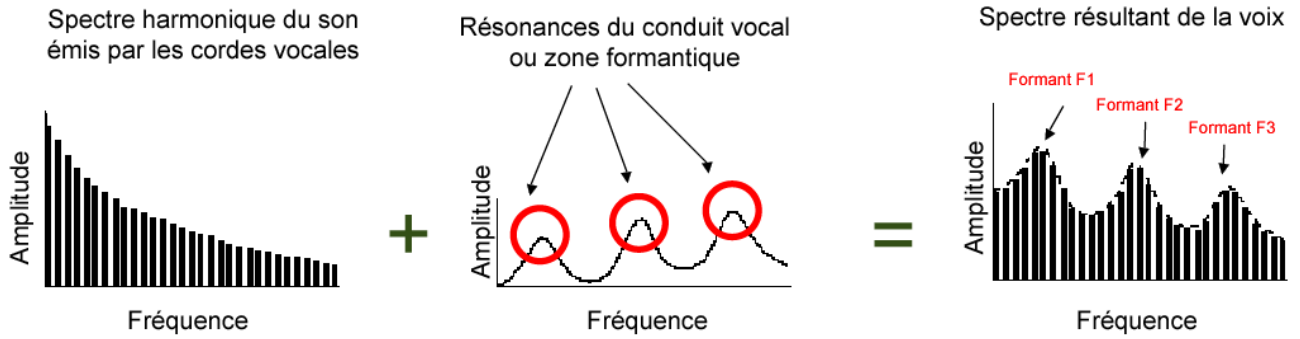


Figure 3.4 : Schéma de la formation du spectre de la voix. Les cordes vocales émettent une fréquence fondamentale avec toutes ses harmoniques, et les fréquences vont être amplifiées ou atténuées selon les fréquences de résonance du conduit vocal, ou espace supralaryngé. Il en résulte un spectre de fréquence harmonique avec une fréquence fondamentale et des formants, multiples de cette fondamentale.

Dans le cas du timbre de la voix, on considère deux aspects. Le premier est le timbre vocalique, qui est la modification de cet espace supralaryngé et qui va permettre l'articulation de phonèmes et la production de la parole. Pour cela, il y aura production de voyelles et de consonnes, ces-dernières étant plutôt apparentées à du bruit et ne nécessitent pas la vibration des cordes vocales mais simplement du passage de l'air dans le conduit vocal. En ce qui concerne les voyelles, il existe théoriquement un nombre de combinaisons infinies pour la position de chaque cavité de résonance, et donc un nombre infini de voyelles. En fait, en fonction du code linguistique utilisé, et en particulier en français, on utilise une dizaine de timbres vocaliques fondamentaux qui sont chacun caractérisé par 2 renforcements de fréquence, appelées zones formantiques ou formants. Chaque voyelle se caractérise donc, sur le plan acoustique, essentiellement par la valeur de ces deux premiers formants, F1 et F2. F1 correspondrait au renforcement fréquentiel au niveau du pharynx tandis que F2 correspondrait au renforcement fréquentiel au niveau de la cavité buccale. Pour les voyelles nasales : l'adjonction du résonateur nasal modifie le deuxième formant.

Le deuxième aspect du timbre que l'on considère au niveau de la voix est le timbre extra-vocalique. Il permet la reconnaissance des individus par le contour fréquentiel global de la voix : c'est la « couleur » de la voix. En réalité, lorsqu'un individu dit [a], on considère que l'auditeur identifie le son [a] grâce à la perception de F1 et F2, et il reconnaît l'individu par le reste des fréquences composant son enveloppe vocale. Une analyse précise de la composition fréquentielle du timbre extra-vocalique pourrait permettre l'établissement d'une véritable carte d'identité vocale. Nous essaierons dans le paragraphe suivant de développer cette notion de timbre extra-vocalique, qui fait l'identité de la voix.

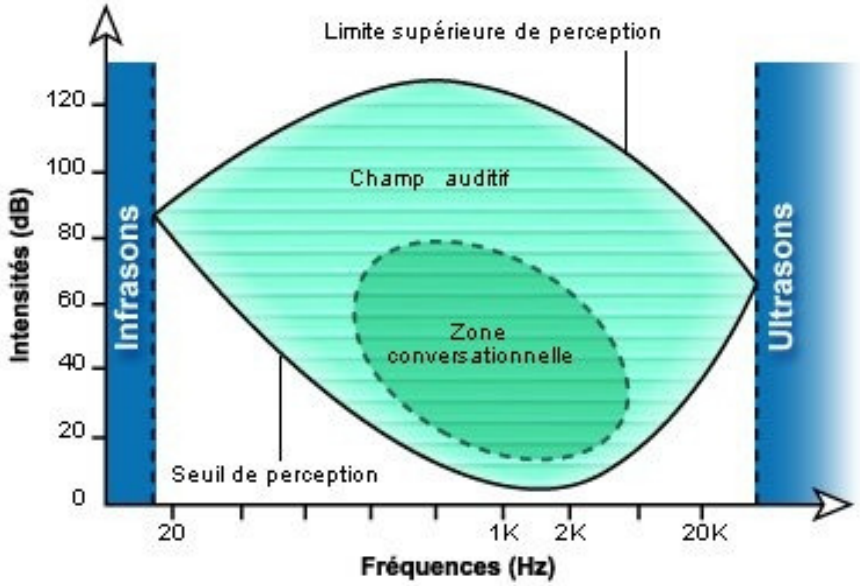


Figure 3.5 : Vue schématisée de la zone de fréquence audible par l'oreille humaine en fonction de l'amplitude et zone de fréquence correspondant aux fréquences émises par la voix humaine (« zone conversationnelle »).

Si l'on ajoute la dimension temporelle à la production de la voix, vont apparaitre certaines variations de pitch et de timbre, qui vont elle aussi être spécifiques de la nature de l'appareil phonatoire mais aussi de l'état du locuteur. Par exemple, l'intonation d'une phrase va permettre à un auditeur de reconnaître la forme de la phrase : question, affirmation, injonction, etc. Egalement, l'état émotionnel du locuteur ou encore ses origines géographiques (accent) vont être véhiculés par ces variations.

Il est intéressant de noter que la fenêtre de fréquences qui peuvent généralement être émises par la voix correspond particulièrement à la zone de fréquences où l'oreille humaine est la plus sensible (voir Figure 3.5).

3.2.1 LE PITCH DE LA VOIX

Le pitch de la voix, est déterminé par les cordes vocales. Il va donc être fonction d'une part de leur taille, mais également de leur capacité de mobilité. En effet, la taille et la masse des cordes vocales vont déterminer dans quelle gamme de fréquences la fréquence fondamentale se trouvera. La capacité à mobiliser les cordes vocales, c'est-à-dire à les tendre ou à les détendre, déterminera l'étendue des pitches qu'il leur sera possible d'émettre ; on parle de tessiture. Une étude anatomique par rayon X chez les chanteurs a montré que plus l'espace glottal, c'est-à-dire les cordes vocales, est grand, plus la gamme de fréquence de la voix, et donc la fréquence fondamentale est basse (Roers, Murbe, & Sundberg, 2009). Le pitch va donc être le reflet de nombreux traits (caractéristiques physiques) et états du locuteur, liés à la taille et aux capacités de modulation des cordes vocales.

Le genre fait partie des caractéristiques les plus étudiées dans la littérature. Une étude anatomique sur 20 cadavres a montré une différence anatomique évidente : les cordes vocales des hommes mesuraient en moyenne 13,2 mm tandis que celles des femmes mesuraient en moyenne 10,6 mm (Eckel & Sittel, 1995). Cette différence entre les genres a été confirmée par une étude plus récente chez 165 sujets anesthésiés (Su et al., 2002). Il est évident qu'une telle différence de taille, mais aussi de masse, de cordes vocales entre les genres a des conséquences sur les pitches moyens de chacun d'eux. De façon générale, le pitch moyen des hommes se situe autour de 110 Hz, et celui de la femme autour de 220 Hz (Baken & Olikoff, 2000; M. P. Gelfer & Mikos, 2005; Guimaraes & Abberton, 2005; Iseli et al., 2007; Pegaroo Krook, 1988; Russell et al., 1995).

L'âge a également une influence sur le pitch de la voix (Collins & Missing, 2003; Iseli et al., 2007; Linville, 1996). Pour les nourrissons le pitch est particulièrement élevé, et se situe aux alentours de 500Hz. Durant l'enfance, avec l'augmentation de la taille des cordes vocales, le pitch de la voix diminue et se situe aux alentours de 300Hz, sachant que cette valeur est fortement dépendante de l'âge considéré (Lee, Potamianos, & Narayanan, 1999). Chez les enfants, il y a également une plus grande variabilité du pitch au cours du temps, ce qui pourrait refléter, à la fois des différences anatomiques avec les adultes et une moins bonne habileté à contrôler les différents

muscles. A l'âge adulte, le pitch de la voix s'abaisse encore pour les femmes aux environs de 220Hz, et davantage pour les hommes, aux environs de 110 Hz (Iseli et al., 2007). En 1996, Linville a montré que le pitch de la voix reste stable à l'âge adulte jusqu'environ 45 ou 50 ans. Puis chez les hommes, le pitch moyen augmente tandis que celui des femmes diminue (Linville, 1996). Ainsi, les pitches des voix d'homme et de femme chez les personnes âgées ont tendance à se rejoindre et sont plus source de confusion pour la perception du genre. Pour les deux genres, la stabilité du pitch diminue avec l'âge (Gorham-Rowan & Laures-Gore, 2006; R. J. Morris & Brown, 1994). Ceci rejoint d'autres travaux ayant étudié la variabilité de longueur de cycle de la fréquence fondamentale (le *jitter*) et la variabilité de l'amplitude de la fréquence fondamentale (le *shimmer*) ; ces deux indices augmentaient significativement avec l'âge dans plusieurs études (Benjamin, 1981; Biever & Bless, 1989; Orlikoff, 1990).

En ce qui concerne la taille corporelle, même si on sait que, comme celle des cordes vocales, elle est plus importante chez les hommes que chez les femmes, il est rare que les études aient pu corréler le pitch de la voix à la taille corporelle (Gonzalez, 2004; Hollien, Green, & Massey, 1994; Kunzel, 1989; Rendall, Kollias, Ney, & Lloyd, 2005).

Le pitch de la voix peut aussi refléter certains comportements sociaux ayant un impact sur les cordes vocales. Par exemple, plusieurs études ont montré qu'il était significativement abaissé chez des sujets fumeurs de tabac (Gonzalez & Carpi, 2004; Murphy & Doyle, 1987; Sorensen & Horii, 1982). Spécifiquement chez les sujets féminins l'abaissement du pitch était corrélé avec une consommation quotidienne de cigarettes plus importante (Gonzalez & Carpi, 2004).

Enfin, la dimension temporelle de la voix va donner une dynamique au pitch, de par sa variation au cours du temps et sa vitesse de variation (rythme). Cet aspect dynamique va permettre de refléter d'autres aspects liés au locuteur, tel que l'état émotionnel du locuteur ainsi que la prosodie. La prosodie comprend non seulement les indices propres au langage, telle que l'intonation ou le rythme de la phrase qui vont accompagner et ainsi guider la perception du langage, mais aussi l'accent, qui est un indice propre à l'identité du locuteur et à son origine géographie.

3.2.2 LE TIMBRE DE LA VOIX

Le timbre de la voix est déterminé à la fois par la nature des composants de l'espace supralaryngé et par la façon dont le locuteur est capable de les articuler. Le timbre va donc correspondre essentiellement aux formants de la voix et à leur mise en place, ainsi qu'à la stabilité de ces formants au cours du temps. Tout comme pour le timbre et la taille des cordes vocales, le timbre est donc notamment dépendant de la taille de l'espace supralaryngé, lui-même corrélé à certains traits ou états du locuteur. Plusieurs études se sont intéressées à la relation entre le timbre de la voix et certaines caractéristiques du locuteur.

Le genre, très étudié au travers de la littérature, est corrélé avec des différences de timbres et notamment de formants. La taille du conduit vocal étant plus importante chez les hommes que chez les femmes (Bachorowski & Owren, 1999; Fitch & Giedd, 1999) le timbre s'en trouve forcément modifié. Parce que le conduit vocal masculin est environ 15% plus grand que celui de la femme, le discours des hommes a des formants plus bas, en terme de fréquences, que celui des femmes (Bachorowski & Owren, 1999; Bennett, 1981; Busby & Plant, 1995; Coleman, 1971; Huber, Stathopoulos, Curione, Ash, & Johnson, 1999). Lee *et al.* ont montré que la différence de genre au niveau des formants F2 et F3, n'était apparente qu'à partir de l'âge de 15 ans (Lee et al., 1999), tandis que Bennett *et al.* ont trouvé que cette différence de genre était déjà apparente dès l'âge de 7/8 ans pour les 4 premiers formants. L'amplitude des formants semble aussi être une différence assez récurrente entre les genres dans la littérature. En 1983, Bladon comparait le spectre des voyelles des hommes à celui des femmes et observait que chez les hommes, les bandes passantes des formants étaient plus étroites et que la diminution d'amplitude des formants le long du spectre était moins rapide (Bladon, 1983). D'autres auteurs ont montré que chez les femmes, l'amplitude du premier formant, F1, était supérieure à celle du deuxième formant, F2. L'inverse était observé chez les hommes (Henton & Bladon, 1985; Klatt & Klatt, 1990). Plus récemment, des auteurs réalisant un spectre moyen à long-terme (LTAS) sur du discours, ont montré que la répartition de l'énergie sur les fréquences de la voix était différente entre les genres, et particulièrement entre 800 et 5 000 Hz (Mendoza, Valencia, Munoz, & Trujillo, 1996). Les auteurs concluait que cette zone spectrale était donc la plus appropriée pour différencier les genres. Les mêmes auteurs ont montré que le 3ème formant, F3, présentait une différence caractéristique entre les genres : son intensité était plus élevée chez les sujets féminins, ce qui correspondrait à une plus grande aspiration nasale chez celles-ci (Klatt & Klatt, 1990; Mendoza et al., 1996).

L'âge semble également lié à la nature des formants qui compose le timbre. Tout d'abord différentes études s'accordent à dire qu'avec le passage de l'enfance à l'âge adulte, on observe une diminution des formants et notamment des 3 premiers formants (Hillenbrand, Getty, Clark, & Wheeler, 1995; Huber et al., 1999; Kent & Forner, 1979; Lee et al., 1999). Cette évolution correspond parfaitement à la maturation du conduit vocal ; les plus gros changements chez les hommes ont lieu entre 12 et 14 ans suivis de légers changements jusqu'à environ 16 ou 18 ans, tandis que chez les femmes, la maturation est quasiment complète dès l'âge de 12 ans et la taille du conduit vocal n'évolue que légèrement jusqu'à l'âge de 16 ans (Huber et al., 1999). Chez les personnes âgées, le spectre de la voix est moins étendu, ce qui se traduit par une disparition des formants dans les hautes fréquences et un centre de gravité du spectre fréquentiel de la voix décalé vers les basses fréquences (Endres, Bambach, & Flosser, 1971; Linville & Fisher, 1985; Rastatter & Jacques, 1990). Ensuite, en ce qui concerne l'amplitude des formants, il semblerait qu'elle augmente avec l'âge mais uniquement chez les sujets masculins (Huber et al., 1999).

La taille corporelle a également été corrélée à certaines variations du timbre. Par exemple des auteurs ont montré que des harmoniques élevées ainsi qu'une moins grande dispersion (distance moyenne) des formants F4 et F5 allait de pair avec une petite taille (Collins & Missing, 2003). De

façon plus artificielle, des auteurs ont montré en faisant la synthèse d'un conduit vocal, que des petites variations de taille (5 à 7%) pouvaient être détectées par les sujets (Ives, Smith, & Patterson, 2005; Smith, Patterson, Turner, Kawahara, & Irino, 2005). La taille du conduit vocal étant corrélée à la taille corporelle, les auteurs concluaient que la taille corporelle pouvait logiquement être perçue au travers du timbre.

Enfin, le timbre de part sa richesse et sa relation directe avec la mobilité des composants de l'espace supralaryngé, véhicule forcément de nombreuses informations, concernant d'autres caractéristiques, telle que la corpulence, ou des comportements, tel que le tabagisme, ou des états émotionnels.

Notons toute de même que la majorité des études sur le timbre a été réalisée à partir de production de voyelles, correspondant aux sons harmoniques de la voix, avec des formants dépassant rarement 8 000 ou 10 000 Hz. Or nous ne devons pas perdre de vue que les consonnes, et notamment les [s] et les [ʃ], ont des bandes passantes s'étendant largement jusqu'à 15 000, voir 18 000 Hz. Les variations sur les fréquences émises par ces sons sont donc peu étudiées.

3.2.3 AUTRES CARACTERISTIQUES

La voix peut également être analysée selon son rythme, son amplitude, ou encore ses contours (variation de pitch, prosodie). Tout comme pour le pitch et le timbre ces aspects sont associés à des traits physiques ou des états du sujet (Andrews & Schmidt, 1997; M. P. Gelfer & Young, 1997; Huber et al., 1999; Rodero, 2010). Des analyses acoustiques ont également été faites sur des critères physiques très précis. Nous avons déjà évoqué le *jitter* ou le *shimmer*, mais d'autres critères existent dans la littérature.

Par exemple le quotient d'ouverture (*open quotient* en anglais), ou OQ, correspond au rapport de la durée d'ouverture de la glotte sur la durée d'un cycle entier de la fréquence fondamentale. Ce critère a été corrélé à la différence de grandeur entre les deux premiers formants, il est souvent nommé dans les études « F1*-F2* ». Il semble que l'OQ soit différents entre les genres ; les femmes ont tendance à le faire varier (habitude d'utilisation de la structure vocale) dans leur voix, pour signaler des différences acoustiques, tandis que les hommes non (Iseli et al., 2007). De même, l'OQ est dépendant de l'âge, mais uniquement chez les hommes, pour qui la différence d'amplitude entre les deux premiers formants diminue avec le temps (Iseli et al., 2007). L'OQ est également un critère sensible, pouvant refléter le stress ou encore l'effort vocal (Holmberg, Hillman, & Perkell, 1988; Mooshammer, 2010).

Toujours, dans le cycle de fermeture et d'ouverture de la glotte, le *Airflow SQ* a été défini comme le rapport de la durée d'ouverture de la glotte sur sa durée de fermeture (Dromey, Stathopoulos, & Sapienza, 1992; Holmberg et al., 1988; Sapienza, Stathopoulos, & Dromey, 1998). Cet indice augmente avec l'effort vocal (Mooshammer, 2010).

Beaucoup d'études s'intéressent également à l'inclinaison spectrale (*spectral tilt* en anglais), qui correspond en pratique à la différence entre la grandeur du premier formant et le niveau corrigé du spectre au 3ème formant, souvent noté « H1*-A3* ». Il existe une différence entre les genres ; la valeur du *spectral tilt* est plus faible chez les femmes d'environ 4dB (Iseli et al., 2007). L'âge aussi peut être corrélé à certaines variations de ce critère ; il diminue avec le temps et de façon plus importante chez les hommes (Iseli et al., 2007).

Enfin, il est important de noter que s'il existe des différences physiques (pitch, timbre, rythme, etc.) liées à certains traits ou états du locuteur, ces différences sont en partie dues à des différences physiologiques et anatomiques de la structure vocale en partie dues à des habitudes d'utilisation de cette structure vocale. La part de la physiologie et du comportement dépend de ce que l'on considère ; il est facile d'imaginer que l'état émotionnel soit davantage lié à des habitudes comportementales d'intonation et donc de modulation des différents muscles et cartilages impliqués dans la phonation, alors que le genre et la taille vont majoritairement dépendre de contraintes physiologiques.

3.3 LA PERCEPTION DE LA VOIX

Comme nous l'avons vu dans la partie précédente, la voix, avec toutes ses variations acoustiques, est très riche et véhicule beaucoup d'informations propres à l'identité ou encore à certains comportements ou états émotionnels du locuteur. S'il est possible de corréler ces traits ou états du locuteur à des variations physiques de la voix, on peut se demander quelle est la perception par l'oreille humaine de ces différents paramètres acoustiques, et s'ils reflètent bien ces caractéristiques du locuteur. Nous allons donc tenter maintenant de décrire la nature des relations entre la voix produite, ce qu'elle contient, et les capacités perceptives d'un auditeur humain.

Dans le domaine de l'étude la perception de la voix on distingue deux approches. La première correspond à la corrélation d'une dimension perceptive, c'est-à-dire d'une dénomination d'un aspect de la voix, avec une (ou des) information(s) acoustique(s). Dans cette approche, comme cela a été soulevé pour l'étude de la perception du timbre en général est qu'il existe très peu de vocabulaire spécifique, et les termes utilisés ne relèvent pas uniquement du domaine sonore. La seconde approche est d'essayer de corréler des informations acoustiques avec les capacités de discrimination des sujets. Dans cette approche, l'étude de jugements de dissemblance (ou de similarité) est une méthode très utilisée. Nous allons maintenant tenter de donner un aperçu de la littérature scientifique pour ces deux approches. Nous verrons également qu'elles se complètent.

3.3.1 DIMENSIONS PERCEPTIVES ET LEURS CORRELATS ACOUSTIQUES

Dans cette partie, nous présenterons différentes dimensions perceptive de la voix et leur lien avec certaines caractéristiques de la voix. Cette approche correspond en réalité à la dénomination des attributs de la voix.

Comme nous l'avons déjà vu pour la perception du pitch, les voix peuvent également être perçues selon la dimension de **grave ou aigüe**. Le pitch de la voix de la voix est une des premières dimensions perçues permettant de différencier les genres, l'âge, ou encore l'état émotionnel (M. Brown, Perry, Cheesman, & Pring, 2000; Childers & Wu, 1991; M. P. Gelfer & Mikos, 2005; Wolfe, Ratusnik, Smith, & Northrop, 1990). Egalement les **contours de pitch**, **l'amplitude** ou **le rythme** sont des indices très importants et très bien perçus par l'oreille humaine. En revanche, les autres aspects de la voix, davantage relatifs au timbre, restent difficilement dénominés par les sujets, tant la voix est un stimulus auditif complexe. Nous allons en passer en revue quelques-uns.

Le souffle (*breathiness* en anglais) dans la voix est une notion plutôt vaste qui englobe notamment la gestion, le contrôle et la pression de l'air. D'un point de vue physiologique, cela correspondrait à une fermeture incomplète des cordes vocales (Fritzell, Hammarberg, Gauffin, Karlson, & Sundberg, 1986; Jeannon, Carding, & Wilson, 1998).Il se manifeste par la présence d'une composante bruitée dans le son. On retrouve la sensation de souffle dans la voix chuchotée. Les analyses acoustiques décrivent, entre autres, un spectre pauvre en harmoniques et présentant un rapport signal sur bruit important. Pour certains auteurs, cela s'explique par l'amplitude élevé du 3ème formant (Klatt & Klatt, 1990; Mendoza et al., 1996). Le souffle semble être un bon critère de distinction des genres ; il est plus présent chez les femmes (Klatt & Klatt, 1990; Van Borsel, Janssens, & De Bodt, 2009) .

Le caractère rauque de la voix, la raucité (*harshness* ou *roughness* en anglais) provient d'une irrégularité de l'onde glottique. Elle a été corrélée à deux paramètres acoustiques nommés *jitter* et *shimmer*. Le *jitter* est caractérisé par l'altération de la périodicité de la fréquence fondamentale et se calcule par la moyenne de la différence de fréquence fondamentale entre deux cycles de vibrations consécutifs. Le *jitter factor* correspond à la normalisation de cet indice par rapport à la fréquence fondamentale. Le *shimmer* est une variation d'amplitude entre deux périodes successives de la fréquence fondamentale. Il correspond à la moyenne des rapports d'amplitudes entre deux cycles de vibrations consécutifs. De même que pour le *jitter*, le *shimmer factor* relativise le *shimmer* moyen.

La brillance est un attribut beaucoup étudié en psychoacoustique. Sur le plan acoustique, on considère que la brillance d'une voix correspond à la présence d'harmoniques aigues dans le spectre. On la caractérise par le centre de gravité spectral (CSG), c'est-à-dire la fréquence qui comporte autant d'énergie en-deçà qu'au-delà. Il est calculé à partir des moyennes des amplitudes pondérées par les fréquences des harmoniques du son.

La nasalité est, du point de vue physiologique, liée à l'abaissement du voile du palais et aux résonances du son dans les cavités nasales. Cependant la nasalité peut être induite par deux

mécanismes différents. Sur le plan acoustique, la mise en évidence de la nasalité n'est pas simple. Certains auteurs ont associé cette dimension à la modification des formants, notamment une réduction de l'amplitude de F1 (Delattre, 1965).

D'autres dimensions ont également été proposées, mais actuellement peu d'études ont permis de corréler ces dimensions à des phénomènes acoustiques. Par exemple, Gelfer en 1988 a proposé un outil de jugement de la qualité de la voix (M.P. Gelfer, 1988). Cet outil proposait 18 échelles de jugement en 8 points. Parmi elles on retrouvait des dimensions telles que la brillance, la nasalité, la raucité, le souffle, la hauteur (grave ou aigüe) ou l'intensité. Mais on retrouvait aussi des notions telles que agréable/désagréable qui vont davantage faire appel à un jugement subjectif, ou encore des notions telles que animé/monotone, posé/chevrotant, lisse/rugueux, qui vont quand à elles être liées au rythme, à la variabilité ou à la stabilité du spectre fréquentiel ou de certains de ces composants, au cours du temps.

3.3.2 JUGEMENT DE DISSEMBLANCE ET CORRELATS ACOUSTIQUES

Dans les études utilisant une analyse multidimensionnelle de similarité ou de dissemblance, des auditeurs doivent juger de la dissemblance (ou similarité) perçue entre les sons de toutes les paires possibles d'un ensemble de sons, sur une échelle qui va de « pareil » à « différent ». Ensuite, ces dissemblances sont traitées comme des distances entre les différents sons. Ces dissemblances entre paires de sons sont analysées à l'aide d'algorithmes d'analyse multidimensionnelle de proximité (*Multidimensional Scaling* en anglais, ou MDS), qui permettent de calculer un espace avec un petit nombre de dimensions, où les distances entre les sons dans cet espace sont déterminées par les dissemblances. Cet espace reflète donc les relations de distance perceptive entre les différents stimuli. La dernière étape d'une telle étude est de proposer une interprétation à chacun des axes, généralement en les corrélant à des descripteurs issus du signal acoustique. Il existe différents modèles et techniques d'analyse multidimensionnelle permettant de répondre à des questions légèrement différentes, mais nous ne renterons pas dans les détails.

Dans le domaine spécifique de la perception de la voix, contrairement au timbre en général, peu d'études ont utilisé cette méthode. De façon générale, les études qui ont utilisé les MDS ont montré que la fréquence fondamentale pouvait expliquer en grande partie, la capacité des sujets à différencier les voix (Baumann & Belin, 2008; Hanson, 1997; Matsumoto, Hiki, Sone, & Nimura, 1973; Murry & Singh, 1980; Walden, Montgomery, Gibeily, Prosek, & Schwartz, 1978). Ensuite venait les formants, de façon différente selon les études, mais on peut parler du timbre en général (Matsumoto et al., 1973; Walden et al., 1978). Des études ont également montré que les autres paramètres acoustiques permettant de différencier les voix n'étaient pas les mêmes selon que les jugements de dissemblance étaient réalisés à partir de voix d'hommes ou de voix de femmes (Aronovitch, 1976; Coleman, 1976a). Par exemple certains auteurs ont montré que pour les hommes

la deuxième dimension permettant de différencier les voix était relative à la dispersion entre le 4ème et le 5ème formant. Pour les femmes, le jugement de dissemblance était expliqué en seconde dimension par la différence du formant F1 (Baumann & Belin, 2008).

Pour résumer, on peut dire que les études avec MDS ont montré que les voix étaient d'abord perçues selon la dimension de fréquence fondamentale, ou de pitch, et ensuite selon le timbre. De plus les paramètres acoustiques du timbre permettant de différencier les voix n'étaient pas les mêmes selon le genre.

3.4 LE TRAITEMENT CEREBRAL DE LA VOIX

Le son, qu'il soit celui d'un instrument ou celui d'une personne, se définit par les mêmes caractéristiques acoustiques. Cependant, il semble aisé de concevoir que la voix d'une personne soit unique et que nous sommes capables de reconnaître la voix d'un proche ou de quelqu'un de familier en entendant une seule syllabe. La mémoire du son d'un instrument particulier est en revanche bien plus difficile. La plupart d'entre nous sommes capables de nous représenter auditivement (et donc de reconnaître) différents types d'instruments, par exemple les instruments à vent versus les instruments à cordes. Encore une majorité de gens peut se remémorer différents types d'instruments à corde (guitare versus mandoline), mais très peu sont capables de se rappeler du son, du timbre, précis d'un instrument particulier, quand bien même le son de cet instrument aurait été entendu très fréquemment. En revanche, pour les voix il en est tout autrement. Notre faculté à mémoriser les voix, à se les remémorer et à les identifier dépasse de loin celle des autres catégories sonores. En effet, la grande majorité des gens, est capable de se réactiver en mémoire la voix de ses proches, collègues, etc., et ce même s'ils ne sont pas vus quotidiennement. Qui n'a pas été étonné d'être reconnu au téléphone (qui, qui plus est, dégrade le spectre) par une personne à qui il n'avait pas parlé depuis longtemps ? Le cerveau humain possède donc des habiletés particulières à encoder et à réactiver de façon très précise le son de la voix humaine.

Nous allons donc maintenant décrire le support neurologique sous-jacent à la perception de la voix et nous verrons ainsi que le traitement dévolu à la voix semble être de haute qualité. Pour cela nous distinguerons deux types d'arguments. Dans un premier temps nous considèrerons les études s'intéressant aux lésions ou aux dysfonctionnements cérébraux associés à une mauvaise perception de la voix. Dans un second temps, nous passerons en revue les arguments chez les sujets normo-entendants en faveur de l'existence d'un traitement cérébral spécifique à la voix.

3.4.1 LA PHONAGNOSIE : UN ARGUMENT EN FAVEUR D'UN TRAITEMENT CEREBRAL SPECIFIQUE POUR LA VOIX

Les études sur les troubles de la reconnaissance de la voix, furent les premières à émettre l'hypothèse d'un traitement spécifique.

La phonagnosie est à l'audition ce que la prosopagnosie est à la vision et ce qu'est l'anosognosie est à l'odorat. La phonagnosie se décrit comme un trouble de la reconnaissance des voix familières. Ce phénomène qui connait un regain d'intérêt actuellement avait jusque là été peu étudié. Mis en évidence en 1976 par Assal *et a.l* (Assal, Zander, Kremin, & Buttet, 1976), et nommé phonagnosie par Van Lancker et Canter en 1982(D. R. Van Lancker & Canter, 1982), la phonagnosie apporte une notion importante en termes de traitement de la voix. Les premières études montraient que la discrimination de voix était altérée chez des patients souffrant de lésions du lobe temporal. Cette phonagnosie était davantage le fait de lésions droites, mais cela ne constituait pas une différence massive (Assal, Aubert, & Buttet, 1981; D. Van Lancker & Kreiman, 1987; D. R. Van Lancker, Cummings, Kreiman, & Dobkin, 1988; D. R. Van Lancker, Kreiman, & Cummings, 1989). Cette asymétrie tendait à attribuer une supériorité hémisphérique droite dans la reconnaissance des voix. Toutefois dans la plupart des cas rapportés dans ces études, on observait également d'autres troubles auditifs, tel que l'amusie ou l'agnosie auditive, il était donc difficile de faire la part des choses. En 1981, une étude d'Assal *et al*. a permis de montré que ce déficit de la reconnaissance de la voix pouvait survenir avec une reconnaissance normale des autres sons (Assal et al., 1981). D'autres auteurs ont confirmé ce résultat en montrant qu'un déficit de reconnaissance de la voix pouvait exister alors que la reconnaissance des sons de l'environnement était préservée (Neuner & Schweinberger, 2000). A l'inverse, des auteurs ont présenté le cas de deux patients amusiques avec une reconnaissance de la voix préservée (Peretz et al., 1994). Des auteurs ont récemment rapporté un cas de phonagnosie développementale (Garrido et al., 2009).

En résumé, les études ont montré une double dissociation entre la perception du langage et de la musique et entre la reconnaissance du locuteur, ce qui constitue le premier argument en faveur d'un traitement de la voix indépendant du traitement des autres types de sons, y compris du langage.

3.4.2 UNE AIRE SPECIFIQUE A LA VOIX, LA TVA (TEMPORAL VOICE AREA)

Les études sur la phonagnosie ont permis d'orienter les recherches sur le traitement cérébral spécifique de la voix. En effet, les dissociations cliniques entre le traitement du langage et le traitement de la voix ont logiquement suscité plusieurs travaux en neuroimagerie et électrophysiologie, à la recherche d'un système anatomo-fonctionnel dédié au traitement de la voix.

Dans une étude en PET de 1997, Imaizumi *et al*. sont les premiers à retrouver une activation des pôles temporaux dans une tâche d'identification auditive d'un locuteur familier (Imaizumi et al., 1997). Mais c'est en 2000 que la première étude de neuroimagerie a identifié dans le sculus temporal

supérieur (STS) une aire qui s'activait préférentiellement lors de l'écoute passive de stimuli vocaux (langagiers et non langagiers) versus des stimuli environnementaux (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Belin *et al.* ont montré dans cette étude utilisant l'IRMf, l'existence de 3 régions neuronales localisées dans le STS, sensibles à la voix humaine : une région antérieure proche du pôle temporal, une région centrale située en regard de l'extension antérieure du gyrus de Heschl et une région plus postérieure par rapport au gyrus de Heschl, s'étendant en arrière vers le *planum temporale*. Ces régions étaient particulièrement activées au niveau de l'hémisphère droit (voir Figure 3.6). Cette étude a donc permis de démontrer un substrat neuronal spécifique au traitement des informations paralinguistiques de la voix humaine, le STS, apportant ainsi un éclairage nouveau sur le fonctionnement du cortex auditif humain. Ces aires s'activant spécifiquement aux stimuli vocaux ont été nommées par les auteurs « *temporal voice areas* », ou TVA.

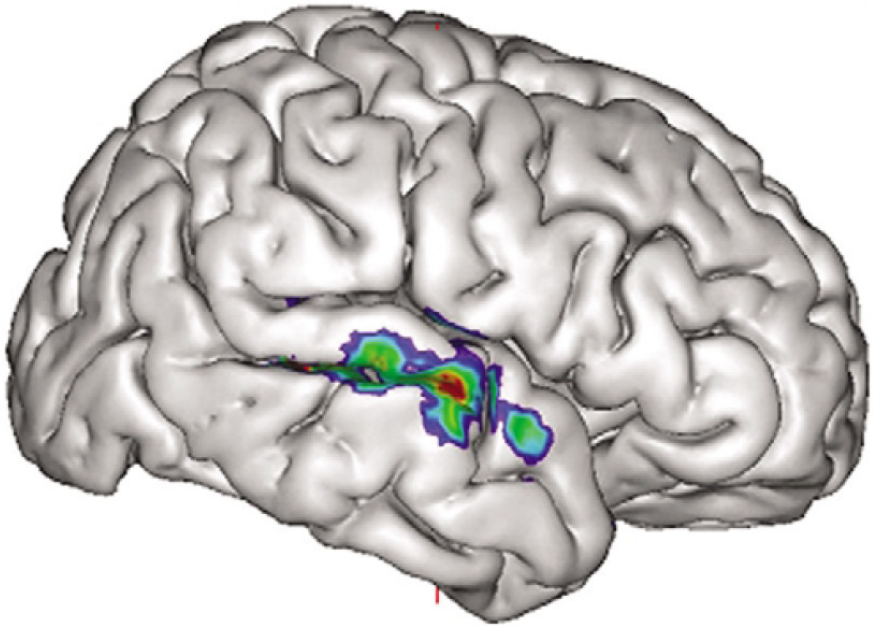


Figure 3.6 : Visualisation de TVA, mise en évidence par Pascal Belin. Les activations correspondent aux régions corticales montrant une plus grande réponse aux stimuli vocaux par rapport aux stimuli non vocaux. Extrait de Belin *et al.*, 2004.

Toutefois, en raison de la passivité de la tâche dans cette étude, la question du rôle de cette aire de la voix dans les fonctions cognitives et particulièrement de ses relations fonctionnelles avec le langage demeurait entière. C'est plus tard que des études ont permis de confirmer que cette aire de la voix était fonctionnellement distincte des autres traitements auditifs. Stevens en 2004 a montré, que lorsque les sujets étaient attentifs aux voix comparé au contenu verbal dans une tâche mnésique, on pouvait constater une augmentation dans gyrus temporal supérieur gauche et le gyrus frontal médian droit (A. A. Stevens, 2004). Deux autres études de perception de la voix ont montré l'implication du STS droit dans une tâche de reconnaissance de voix par rapport au contenu sémantique (K. von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; K. V. von Kriegstein & Giraud, 2004). Ces deux études allaient donc aussi dans le sens d'une prédominance du STS droit dans le traitement de la perception de la voix. De même, en 2003, Belin et Zatorre, en utilisant un

paradigme d'adaptation, ont révélé une plus grande activation dans le STS lorsque les sujets écoutaient plusieurs locuteurs versus un seul locuteur (Belin & Zatorre, 2003). Ce résultat a été confirmé par d'autres auteurs, dans une étude sur la perception du langage (P. C. Wong, Nusbaum, & Small, 2004). Ces auteurs ont montré une augmentation bilatérale des régions temporales supérieures quand les sujets écoutaient plusieurs locuteurs, versus un seul, suggérant ainsi une augmentation du réseau auditif pour le traitement auditif de plusieurs locuteurs. D'autres auteurs ont également montré que la TVA était sensible aux informations affectives de la voix (Ethofer, Van De Ville, Scherer, & Vuilleumier, 2009; Grandjean et al., 2005).

En EEG, des études se sont également intéressées à un traitement spécifique de la voix. Une étude de Levy *et al.* en 2001 a ainsi comparé les potentiels évoqués de moyenne latence enregistrés à l'écoute d'une voix chantée et de différents instruments de musique (Levy, Granot, & Bentin, 2001). Les auteurs observaient une réponse spécifique à la voix (« *voice specific response* »), ou VSR, survenant après N1, environ 320 ms après le début du stimulus. Cette réponse était plus marquée au niveau de l'hémisphère droit. Cette VSR n'a pas été retrouvée par des études ultérieures (Gunji et al., 2003; Levy, Granot, & Bentin, 2003). Par la suite, en utilisant des potentiels évoqué en EEG, d'autres études se sont intéressées à des réponses spécifiques à la discrimination entre sources auditives vivantes (particulièrement humaine) ou non (Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006), ou encore à des réponses spécifiques à la familiarité de la voix (Beauchemin et al., 2006), au genre de la voix (Zaske, Schweinberger, Kaufmann, & Kawahara, 2009), ou à l'identité du locuteur (Titova & Naatanen, 2001). En 2009, Charest *et al.*, ont mis en évidence, toujours avec des potentiels évoqués, une différence d'amplitude spécifique entre les réponses à des stimuli vocaux et les réponses à des stimuli non vocaux (Charest et al., 2009). Cette différence de réponse émergeait à 164 ms après le début du stimulus, avec un pic à environ 200 ms, et était positive au niveau fronto-temporal (« *fronto-temporal positivity to voices* » ou FTPV) et négative au niveau occipital.

En résumé, la spécificité du STS, préférentiellement droit, pour le traitement de la voix, dans ses aspects paralinguistiques tels que l'identité ou l'émotion, a été démontrée par Belin *et al.* en 2000 et a depuis été confirmée par de nombreuses études (Belin & Zatorre, 2003; Belin, Zatorre, & Ahad, 2002; Belin et al., 2000; Charest et al., 2009; Imaizumi et al., 1997; A. A. Stevens, 2004; K. von Kriegstein et al., 2003; K. V. von Kriegstein & Giraud, 2004). La fonctionnalité de la TVA, localisée dans le STS, a été confirmée par d'autres arguments.

3.4.3 AUTRES ARGUMENTS EN FAVEUR DE LA TVA

Tout d'abord, les enfants sont une population d'étude très intéressante pour connaître la mise en place de la perception et de la cognition. De nombreuses études ont déjà montré que les bébés, à la naissance, avant même de savoir parler donc, reconnaissaient et préféraient, d'un point de vue attentionnel, les voix appartenant à leurs père et mère (DeCasper & Fifer, 1980; Fifer & Moon, 1994; Kisilevsky et al., 2003; Ockleford, Vince, Layton, & Reader, 1988). La voix semble donc déjà

bénéficier d'un traitement particulier très tôt dans le développement humain. Des auteurs, dans une étude utilisant l'électroencéphalographie (EEG), ont montré que chez des enfants de 4 ans, en réponse à des stimuli vocaux par rapport à des stimuli non vocaux, on observait également une réponse fronto-temporale positive, spécifique à la voix, comme l'avait montré Charest *et al.* en 2009 (Rogier, Roux, Belin, Bonnet-Brilhault, & Bruneau, 2010). Cette étude suggérait donc que dès 4 ans, le traitement cérébral de la voix était similaire à celui qui avait été observé chez les adultes par Charest *et al.*. D'autre études se sont intéressé à ce qu'il ce passait chez les enfants d'âge inférieur et particulièrement chez les nouveau-nés. En 2010, Grossmann *et al.*, ont utilisé une technique non invasive, mesurant indirectement l'activité des neurones, en quantifiant la façon dont le tissu cérébral absorbait différentes ondes lumineuses, la NIRS, pour *near infrared spectrscopy* (Grossmann, Oberecker, Koch, & Friederici, 2010). Les auteurs ont pu montrer au travers d'une première expérience que des enfants de 7 mois, contrairement à des enfants de 4 mois, avaient une plus grande réponse à la voix qu'aux autres sons, dans la partie postérieure du lobe temporal bilatéral. Dans une seconde expérience chez les enfants de 7 mois, l'écoute de prosodie émotionnelle (colère ou joie) activait spécifiquement deux régions de l'hémisphère droit ; la même région du cortex temporal postérieur que lors de la première expérience, et une région du cortex préfrontal inférieur. Cette étude, permet donc de dire que dès 7 mois, un traitement cérébral spécifique à la voix existe dans les régions temporales. Le contraste avec les enfants de 4 mois laisse entendre que la période de 4 à 7 mois après la naissance est critique pour le développement du traitement de la voix. De plus, les résultats de cette étude montrent également que le pattern de sensibilité de la voix observé chez les enfants est relativement comparable à celui observé chez les adultes, bien qu'il soit plus postérieur chez les enfants, suggérant que les parties antérieures ne sont pas encore sensibles à la voix 7 mois après la naissance.

Ensuite, un autre argument en faveur de l'existence d'une aire spécifique à la voix provient des études chez l'animal. La question de savoir quel traitement cérébral, spécifique ou non, était accordé aux vocalisations conspécifiques, chez l'animal, et notamment le primate, a intéressé plusieurs auteurs. Plusieurs études se sont intéressées aux réponses du cerveau du primate aux communications verbales de ses congénères (Romanski, Averbeck, & Diltz, 2005; Romanski & Goldman-Rakic, 2002; Wang, 2000; Wang & Kadia, 2001; Winter & Funkenstein, 1973). Une première étude a suggéré l'existence de neurones « détecteur de cris » chez le singe écureuil, bien que ces résultats aient été discutés par la suite (Winter & Funkenstein, 1973). En 2008, Petkov *et al.* ont mené une étude en IRMf chez le singe rhésus et ont établi l'existence de régions neuronales dédiées au traitement de la voix de leurs congénères. Pour cela, les auteurs avaient comparé les activations des régions auditives en fonction de différents types de stimuli acoustiques (bruits environnementaux, vocalisations émises par des singes de l'espèce rhésus, vocalisations provenant d'autres espèces). Les résultats montraient que deux régions s'activaient sélectivement en réponse aux vocalisations spécifiques de l'espèce rhésus ; une première région était située dans le sillon latéral bilatéral, en continuité avec le cortex auditif primaire et une autre région plus antérieure,

latéralisée à droite. L'existence chez le singe d'une aire cérébrale impliquée sélectivement dans le traitement des vocalisations de ses congénères constitue un argument fort en faveur de son existence chez l'homme. De plus, cela plaide en faveur d'une fonctionnalité indépendante de celle du langage, qui pourrait préexister au développement du langage.

Enfin, un dernier argument venant confirmer l'idée d'une aire spécifique à la voix humaine est l'étude des troubles de la communication vocale affective, tels que l'autisme. Ces troubles peuvent aussi être caractérisés comme des déficiences du traitement cérébral de la voix. Dans une revue de la littérature de 2008, Redcay recense plusieurs arguments en faveur d'un lien entre le déficit de la perception sociale observé dans l'autisme et le dysfonctionnement du STS (Redcay, 2008). Pour l'auteur, bien que le dysfonctionnement du STS ne puisse évidemment pas expliquer à lui seul le tableau clinique de l'autisme, les arguments en faveur de ce lien sont d'ordres anatomiques d'une part et fonctionnels d'autre part. Sur le plan anatomique, différentes études ont déjà montré que chez les enfants autistes on pouvait observer un élargissement de la matière grise dans les lobes temporaux aux alentours de 2 à 4 ans (Carper, Moses, Tigue, & Courchesne, 2002), ainsi qu'une diminution de la densité de cette matière dans les STS bilatéraux (Boddaert et al., 2004). De plus, une étude a permis de corréler significativement cette diminution de densité du STS avec les symptômes autistiques (Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2006). D'autres études ont également montré une hypoperfusion bilatérale du STS et du STG (Gyrus Temporal Supérieur), qui serait corrélée, pour l'hémisphère gauche au degré de sévérité de l'autisme (Gendry Meresse et al., 2005; Ohnishi et al., 2000; Zilbovicius et al., 2000). Sur le plan fonctionnel, de nombreuses études ont pu montrer un lien entre l'activité du STS et le traitement du mouvement et particulièrement l'intention du regard (pour revue voir Redcay, 2008). Au niveau auditif, des auteurs ont montré que durant une tâche passive d'écoute de sons du langage, l'activité du lobe temporal était significativement réduite chez les patients autistes par rapport à des sujets contrôles. Plus récemment, Gervais *et al.* ont montré dans une étude utilisant l'IRMf, que des patients autistes ne pouvaient pas activer la TVA localisée dans le STS en réponse à des stimuli vocaux, alors qu'ils présentaient un pattern d'activation normal pour des sons non-vocaux (Gervais et al., 2004). Les auteurs concluaient à un traitement cortical anormal des informations auditives sociales chez ces patients.

En résumé, depuis que la TVA a été mise en évidence chez l'homme sain, des études sur d'autres thèmes sont venues corréler la fonctionnalité de cette aire. Il est maintenant bien établi qu'elle s'active spécifiquement en réponse à des stimuli vocaux, langagiers ou non, et que son rôle dans la discrimination et la reconnaissance des voix, ainsi que dans la perception d'informations paralinguistiques est primordial.

3.5 LA VOIX, UN VISAGE AUDITIF ?

L'ensemble des études sur les dissociations de traitement d'information de la voix a été cruciale pour établir un traitement hiérarchique de celui-ci. Dans l'ensemble, ces études se sont inspirées du le modèle de la perception des visages de Bruce et Young (Bruce & Young, 1986). Dans ce modèle, les auteurs décrivent plusieurs niveaux d'analyse du visage: une analyse initiale de bas niveau purement perceptuelle, reconnaissant le stimulus visage, suivie par un encodage de la structure globale du visage permettant l'analyse intégrée des éléments invariants propres à l'identité (couleur des cheveux, forme visage) et des éléments dynamiques tels que l'expressivité ou l'émotion. Sur la base de ce modèle, Ellis *et al.* en 1997, suggéraient que l'information caractéristique de la voix était d'abord encodée à un niveau très basique (*auditory structured encoding*), après quoi elle était analysée en unité de reconnaissance de la voix (*voice recognition units*) (Ellis, Jones, & Mosdell, 1997). Ce dernier module de traitement de la voix serait en lien avec un module multimodal du traitement de l'identité, *person identity nodes*, et finalement des informations sémantiques serait associées avec le stimulus.

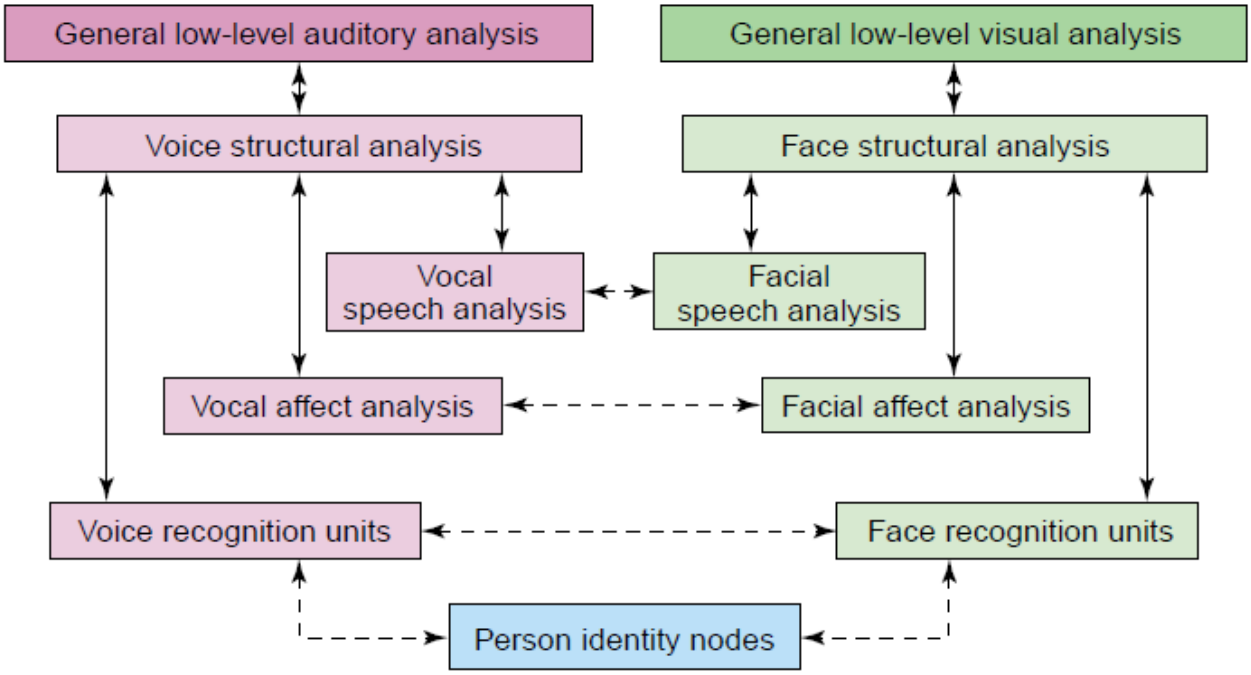


Figure 3.7 : Modèle de perception de la voix proposé par Belin *et al.* La partie de droite est adaptée d'après le modèle de perception des visages de Bruce et Young (Bruce & Young, 1986), tandis que la partie de gauche propose une organisation fonctionnelle similaire pour le traitement de la voix. Les flèches pointillées indiquent les interactions multimodales. Adapté d'après Belin *et al.* 2004.

En 2004, Belin *et al.* ont proposé un modèle de la perception de la voix plus abouti, considérant à la fois la façon dont la reconnaissance du locuteur est faite au travers de la voix, mais aussi la façon dont les autres informations vocales étaient traitées (Belin, Fecteau, & Bedard, 2004). Les auteurs parlaient donc de la voix comme d'un visage auditif. Selon ce modèle, la voix est tout d'abord analysée à un bas-niveau assez générique, commun aux autres stimuli auditifs, puis analysée

et encodée de façon structurelle. A partir de là, la voix serait ensuite analysée parallèlement par différents modules spécifiques : le premier serait sensible aux informations linguistiques, le deuxième aux informations affectives vocales, et le dernier aux informations d'identité vocale (voir Figure 3.7). Parce qu'il implique des voies de traitement séparées, ce modèle prédit des dissociations neurofonctionnelles entre les trois types de traitement. De plus pour les auteurs, les structures neuronales impliquées dans le traitement de chacune de ces informations seraient en interactions avec les traitements du visage lui correspondant. Nous allons détailler le traitement neurofonctionnel de ces 3 types d'informations vocales.

3.5.1 PERCEPTION DE L'INFORMATION LANGAGIERE

Comme nous l'avons déjà évoqué précédemment, la voix est le support de la parole. De nombreuses études de neuroimagerie dans le domaine de la voix ont étudié les aspects de l'organisation fonctionnelle impliqués dans la perception du langage.

Plusieurs revues de la littérature ont mis en avant l'implication bilatérale des aires secondaires du cortex temporal supérieur bilatéral (Hickok & Poeppel, 2000; Y. Samson et al., 2001; Scott & Johnsrude, 2003; Zatorre & Binder, 2000). Plusieurs de ces études ont également suggéré une dissociation entre les régions du STS médian, qui répondraient davantage au langage qu'au message linguistique, et les régions antérieures du STS gauche et du plan temporal supérieur, qui seraient plus impliquées dans la compréhension du langage, même sous formes de structures acoustiques dégradées. Le modèle classique d'anatomie fonctionnelle de compréhension du langage postule que le tiers postérieur du gyrus temporal supérieur gauche (aire de Wernicke) soit la zone anatomique d'interface entre signal acoustique et représentations mentales, c'est-à-dire la compréhension du langage. Toutefois, la diversité des formes cliniques d'aphasie a conduit à envisager une hétérogénéité fonctionnelle de l'aire de Wernicke (Demonet, Thierry, & Cardebat, 2005). Cette région interviendrait en effet dans des tâches distinctes, telles que la perception phonologique ou encore la récupération du mot en mémoire sémantique. Cette différenciation des fonctions de l'aire de Wernicke a favorisé l'émergence d'un nouveau modèle d'anatomie fonctionnelle du langage. Ce modèle est basé sur l'organisation du système visuel ; une voie ventrale d'identification de l'objet (*what*) et une voie dorsale de localisation (*where*). En ce qui concerne le langage, ces deux voies ont été notamment étudiées par Hicock et Poppel (Hickok & Poeppel, 2004) puis par Saur (Saur et al., 2008). La voie dorsale et la voie ventrale seraient convergentes en leur origine, la jonction temporo-occipitale gauche, et à leur terminaison, au niveau frontal inférieur. Elles seraient anatomiquement et fonctionnellement associées. Dans ce modèle, le faisceau dorsal correspondrait au traitement du « comment » du langage (*how* en anglais), c'est-à-dire aux représentations phonémiques, à la mémoire de travail verbale (par la répétition subarticulatoire) et l'interface auditivo-motrice impliquée dans la répétition orale des mots. La voie ventrale correspondrait au traitement du « quoi » du langage (*what* en anglais) et comprendrait des faisceaux

plus inférieurs (temporaux moyens et temporaux inférieurs) et permettrait l'accès aux représentations lexico sémantiques et conceptuelles, stockées dans les régions temporales moyennes.

3.5.2 PERCEPTION DE L'INFORMATION AFFECTIVE

La voix comporte beaucoup d'informations affectives. En effet, comme nous l'avons déjà évoqué, les informations acoustiques de la voix sont modulées par l'état émotionnel du locuteur. La perception de ces informations a été très étudiée dans le contexte du langage. La prosodie émotionnelle, supportée par les variations d'amplitude, la durée des pauses et la fréquence fondamentale et sa variation, permet à l'auditeur d'inférer beaucoup d'informations concernant l'état affectif du locuteur. Les signaux vocaux non langagiers, tels que les rires, les cris, les pleurs, etc., contiennent également beaucoup d'informations affectives.

En termes de support neuronal, quelques études se sont intéressées au traitement de l'information affective de la voix. Des études ayant utilisé l'IRMf ou le PET, ont permis de montrer la plus grande activation du lobe temporal droit et du cortex préfrontal inférieur droit (Buchanan et al., 2000; George et al., 1996; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002). Par la suite, les bases neuronales de la perception émotionnelle de la voix ont été étudiées en dehors du contexte du langage en utilisant des vocalisations non verbales, telles que des rires, des cris, etc. Les études PET et IRMf ont suggéré l'importance de l'amygdale et de l'insula antérieure dans le traitement de l'émotion vocale (Fecteau, Belin, Joanette, & Armony, 2007; J. S. Morris, Scott, & Dolan, 1999; Phillips et al., 1998; Sander & Scheich, 2001). Ces résultats apparaissent cohérents avec les données électrophysiologiques selon lesquelles le traitement de l'affect dans la voix serait un processus neuronal précoce, impliquant les aires auditives secondaires, intervenant en moyenne 200 ms après le stimulus (Spreckelmeyer, Kutas, Urbach, Altenmuller, & Munte, 2009). Une étude EEG de 2009, suggère que le traitement de l'information affective vocale soit plus postérieure que le traitement de l'information de l'identité vocale (Toivonen & Rama, 2009).

3.5.3 PERCEPTION DE L'INFORMATION D'IDENTITE

Nous avons déjà développé les différentes informations acoustiques de la voix qui sont modulés en fonction de l'identité (voir « Caractéristiques physiques de la voix », p77), qu'il s'agisse des informations de timbre ou de pitch, et ce dans sa dimension temporelle également (accent, intonation spécifique à une zone géographie et socioculturelle)

En termes de traitement neuronal, la perception de l'identité de la voix est une des premières informations paralinguistiques dont le traitement ait été dissocié de celui du langage. En premier lieu, furent les études sur la phonagnosie que nous avons déjà évoquées précédemment (voir « La phonagnosie : un argument en faveur d'un traitement cérébral spécifique pour la voix », p87). Plus

tard, plusieurs études ayant servie de support à la mise en évidence de la TVA dans le STS, portait sur une dissociation entre traitement verbal et traitement de l'identité du locuteur (voir « Une aire spécifique à la voix, la TVA (Temporal voice area) », p87). Pour mémoire, Imaimuzi *et al.* en 1997, avaient déjà montré que le lobe temporal antérieur bilatéral était davantage activé lors d'une tâche d'identification du locuteur que lors de la tâche d'identification de l'émotion de la voix (Imaizumi et al., 1997). En 2001, ce groupe d'auteurs a trouvé des résultats allant dans le même sens ; le cortex enthorinal et la partie antérieur du lobe temporal droit était plus activé dans une tâche de décision de la familiarité de la voix que dans une tâche de décision phonétique (Nakamura et al., 2001). Des études IRMf avaient également confirmé ce résultat. Von Kriegstein *et al.* ont montré qu'une tâche d'identification du locuteur activait davantage le STS antérieur droit et une partie du precuneus droit, qu'une tâche d'identification du contenu verbal (K. von Kriegstein et al., 2003). Puis dans un paradigme d'adaptation au locuteur, Belin et Zatorre ont montré l'implication du STS antérieur droit pour le traitement acoustique des caractéristiques de la voix liées au locuteur (Belin & Zatorre, 2003). Enfin, en 2007, une étude IRMf a également retrouvé que l'implication d'une aire auditive située sur la berge inférieure du STS était préférentiellement activée par les voix familières versus non familières (Birkett et al., 2007). Il y a donc un certain nombre d'arguments en faveur de l'implication des régions temporales antérieures de l'hémisphère droit et particulièrement du STS antérieur droit dans le traitement de l'identité. Toutefois, une étude plus récente sur la perception du changement de la taille au travers de la voix humaine a montré une activité spécifique dans gyrus temporal postérieur (STG) gauche (K. von Kriegstein, Smith, Patterson, Ives, & Griffiths, 2007). De plus cette aire n'était pas activée lors du traitement de la perception de la taille au travers d'autres sources sonores. La perception de la taille du locuteur fait partie de la perception de l'identité, le résultat de cette étude est donc surprenant, mais n'invalide pas les résultats précédents puisqu'il porte sur un attribut spécifique de l'identité.

Dans l'ensemble, ces différentes études de perception de la voix, montrent que les différentes activations du STS qui ont été observées, correspondent probablement à des régions fonctionnellement distinctes, et soutient l'idée que le traitement neuronal des informations linguistiques, émotionnelles et identitaires de la voix peut être dissocié. Ce traitement hiérarchique de l'information vocale a été confirmé dans des études plus récentes (Engel, Frum, Puce, Walker, & Lewis, 2009; Lattner, Meyer, & Friederici, 2005; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009; J. D. Warren, Scott, Price, & Griffiths, 2006; D. Wong, Miyamoto, Pisoni, Sehgal, & Hutchins, 1999), montrant notamment que différents traitements et/ou différentes catégories de sons activaient différentes régions corticales.

CHAPITRE 2 : LA SURDITÉ ET LES IMPLANTS COCHLÉAIRES

Dans ce chapitre nous allons évoquer la surdité, ses causes et ses remèdes. Plus particulièrement nous nous intéresserons aux implants cochléaires, leur technique et leur apport auditif chez les sujets sourds. Nous verrons les questions qui se posent par rapport à l'audition, et plus précisément dans le domaine de la perception des informations paralinguistiques.

4 LA SURDITE

La surdité est définie comme étant une pathologie caractérisée par une perte partielle ou totale, bilatérale ou unilatérale, de l'audition. Elle est particulièrement invalidante tant notre société fait une place importante à l'environnement sonore et à la parole en particulier. La surdité à un impact très fort sur la vie psychologique et sociale des patients. Actuellement, on estime selon les sources statistiques, que 10 à 20 % de la population mondiale est atteinte de surdité à des degrés divers. Plus la population est âgée, plus la proportion de personnes sourdes est importante.

4.1 DIAGNOSTIC DE SURDITE

4.1.1 CLASSIFICATION DES SURDITES

La classification des surdités peut s'effectuer de multiples façons. Les deux les plus couramment utilisées sont celle fonctionnelle et celle audiométrique, mais on distingue également une classification en fonction du moment d'apparition de la surdité et en fonction de l'évolution de la surdité.

4.1.1.1 CLASSIFICATION FONCTIONNELLE

On considère tout d'abord que les surdités de transmission ont pour origine des lésions de l'oreille externe et/ou de l'oreille moyenne. Elles ne créent généralement pas de perte auditive supérieure à 60 dB, ni de distorsions du message sonore, l'oreille interne étant normale. La perte auditive se trouve généralement sur les sons graves jusqu'à environ 1000 Hz, mais peut également toucher les aigus à moindre degré. L'audition par voie osseuse est normale. Le sujet n'éprouve pas

de gêne en atmosphère bruyante. Les surdités de transmission peuvent bénéficier d'un traitement médical ou médico-chirurgical adapté.

Les surdités de perception ont pour cause des lésions des voies neurosensorielles de l'audition (oreille interne, nerf auditif, voies centrales). Elles entraînent des pertes auditives qui peuvent excéder 60 dB. La perte auditive prédomine en général sur les aigus. La conduction osseuse est toujours affectée parallèlement à la conduction aérienne. Le sujet éprouve une gêne dans le bruit. Il existe des distorsions de la perception auditive qui perturbe l'intelligibilité de la parole. Les surdités de perception ne sont améliorées que par des prothèses auditives. Si l'atteinte se situe au niveau des centres nerveux auditifs, la surdité sera appelée surdité centrale. Elles constituent les surdités de loin les plus fréquentes.

Enfin, les surdités mixtes associent la surdité de transmission à celle de perception où l'une prédomine sur l'autre.

4.1.1.2 CLASSIFICATION AUDIOMETRIQUE

Pour cette classification les surdités sont différenciées selon leur degré de perte auditive. La perte auditive se calcule sur la perte en voie aérienne des fréquences conversationnelles classiques (500, 1000, 2000 Hz). On additionne les valeurs en dB et on divise la somme par trois. On compte 120 dB de perte pour toute fréquence non perçue. Le calcul est réalisé pour chaque oreille séparément. Leur somme est divisée par quatre et arrondie à l'unité supérieure. En cas de surdité asymétrique, le niveau moyen de perte en dB est multiplié par 7 pour la meilleure oreille et par 3 pour la plus mauvaise oreille. La somme est divisée par 10.

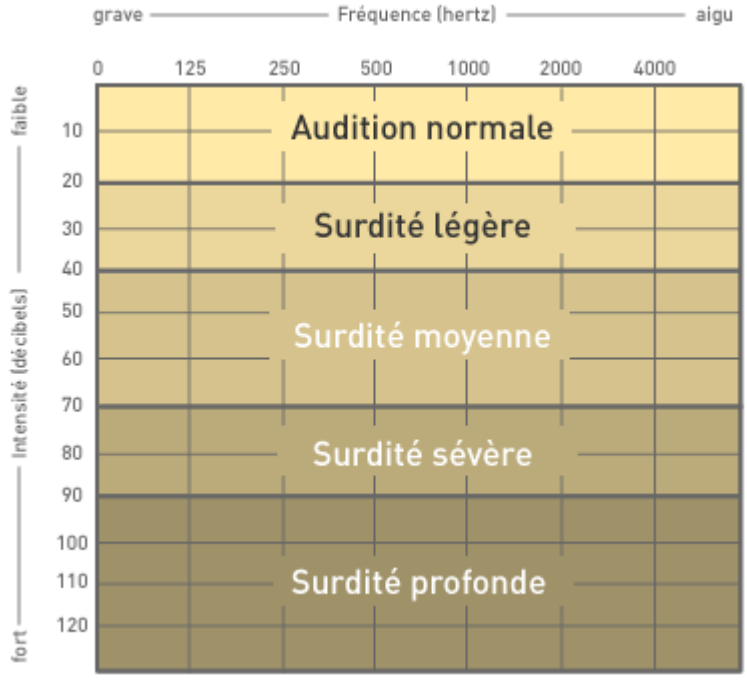


Figure 4.1 : Classification des surdités en fonction du degré de la perte auditive.

La classification s'effectue de la manière suivante :

- Audition normale et subnormale : perte auditive inférieure à 20 dB.

- Surdité légère : perte de 21 à 40 dB.

- Surdité moyenne : 41 à 70 dB de perte. Il existe deux groupes ; dans le Groupe I : la perte se situe entre 41 et 55 dB, tandis que dans le Groupe II la perte se situe entre 56 et 70 dB.

- Surdité sévère : 71 à 90 dB de perte. Il existe deux groupes : dans le Groupe I : la perte se situe entre 71 et 80 dB, tandis que dans le Groupe II la perte se situe entre 81 et 90 dB.

- Surdité profonde : au-delà de 90 dB de perte. Il existe trois groupes. Dans le Groupe profonde I, la perte est comprise entre 91 et 100 dB de perte, dans le Groupe profonde II, entre 101 et 110 dB, tandis qu'elle se situe entre 111 à 119 dB pour le Groupe profonde III.

- Surdité totale ou cophose : La perte moyenne est de 120 dB. Rien n'est perçu.

4.1.1.3 AUTRES CLASSIFICATIONS

Il existe aussi une classification en fonction du moment d'apparition de la surdité. On parle de surdité congénitale lorsque les personnes sont nées sourdes. La surdité prélinguale concerne les personnes dont la surdité est apparue lors de la petite enfance, avant l'acquisition du langage parlé. Enfin la surdité postlinguale signifie que l'audition est affectée après la période critique d'acquisition du langage.

Finalement l'évolution de la surdité est un autre critère de classification. On distingue la surdité progressive, qui évolue lentement avec une dégradation progressive de l'audition, et la surdité brusque, qui apparaît au cours d'un laps de temps très restreint, entre quelques minutes et plusieurs jours.

4.1.2 DIAGNOSTIC DE LA SURDITE

On dispose actuellement d'une grande palette d'outils diagnostiques de la surdité. Chez les nouveau-nés, le dépistage ne peut être fait qu'après la naissance. On introduit de plus en plus de dépistage systématique de la surdité dans les maternités afin de dépister et éventuellement traiter au plus tôt les surdités. Par la suite le dépistage et le diagnostic de cette pathologie se fera au moment de la demande ou de la plainte du patient.

L'interrogatoire clinique du patient doit faire apparaître :

- la date et les circonstances d'apparition de l'hypoacousie (baisse de l'acuité auditive) : traumatisme crânien, virose, prise de médicaments, etc.

- les signes d'accompagnement de la surdité : acouphènes (bourdonnements d'oreille), otorrhée (écoulement de liquide par l'oreille), otorragie (écoulement de sang par l'oreille), otalgie (douleur dans l'oreille), vertiges, paralysie faciale, céphalées, etc.

- l'allure évolutive de la surdité : progressive, fluctuante ou brutale

- le caractère uni ou bilatéral de la surdité

- les antécédents ORL oto-rhino-laryngologiques du patient
- les antécédents familiaux de surdité
- la profession du patient
- le retentissement socioprofessionnel

L'examen du tympan à l'otoscope est fondamental car il permet de distinguer les surdités à tympan normal et les surdités à tympan anormal (inflammatoire, perforé...). Il permet également d'éliminer un bouchon de cérumen ou une otite externe.

En fonction de ce premier examen et entretien clinique, le clinicien adapte les investigations afin de savoir à quel niveau se situe le problème. Pour les surdités de transmission, ils peuvent inclure une impédancemétrie ainsi qu'une radio ou un scanner des rochers, tandis que pour les surdités de perception, ils peuvent inclure une étude des potentiels évoqués auditifs, une électro-cochléographie, un test au glycérol, une électro-nystagmographie ainsi qu'un scanner par tomodensitométrie ou par résonance magnétique nucléaire.

4.1.2.1 OUTILS DIAGNOSTIQUES

ACOUMETRIE

L'Acoumétrie est un test basique de l'audition permettant de diagnostiquer et distinguer rapidement le type de surdité dont le patient est atteint, avant de pratiquer un examen audiométrique plus approfondi.

L'épreuve de Weber

Un diapason est appuyé sur le front. En cas d'audition normale ou symétrique, les vibrations sont entendues des deux côtés. Si le son est perçu par l'oreille la plus sourde, il s'agit d'une surdité de transmission ; s'il est mieux perçu par l'oreille saine, il s'agit d'une surdité de perception (lésions de l'oreille interne).

L'épreuve de Rinne

Un diapason est placé devant l'oreille, puis contre la mastoïde. A l'état normal, le son est mieux perçu et plus longtemps par voie aérienne. Dans les maladies de l'appareil de transmission, il est mieux perçu par voie osseuse : on dit que le test de Rinne est alors négatif.

AUDIOMETRIE

L'audiométrie va fournir un état précis de l'audition d'un patient.

Audiométrie tonale liminaire

Réalisée en cabine insonorisée, l'audiométrie tonale liminaire permet de rechercher les seuils de sensibilité auditive. Le patient indique dès qu'il entend le son émis, soit par un signal, soit en

faisant un signe à l'examinateur. L'exploration se fait d'abord par voie aérienne à l'aide d'écouteurs puis par voie osseuse, à l'aide d'un vibrateur appliqué sur la mastoïde. Des sons purs de fréquences connues (125 à 8000 Hertz) sont émis à des intensités choisies et variables (0 à 100 dB).

Les résultats sont notés sur un diagramme et deux courbes par oreille sont notées: l'une en conduction osseuse, l'autre en conduction aérienne. Il permet de préciser la nature de la surdité. Lors d'une surdité de transmission, la courbe de conduction osseuse est normale et la courbe de conduction aérienne abaissée. Lors d'une surdité de perception, les deux courbes sont abaissées de façon identique. Enfin, pour une surdité mixte, la courbe de conduction aérienne est plus abaissée que celle de la conduction osseuse.

Ce test est subjectif et nécessite la collaboration du patient.

Audiométrie tonale supraliminaire, ou épreuve de Fowler :

Ce test étudie la notion de recrutement. Lorsqu'on augmente progressivement l'intensité d'un son que l'on fait entendre à une oreille sourde, le sujet peut parfois percevoir ce son à partir d'un certain seuil d'intensité. On suppose que les cellules sensorielles subsistantes recrutent un plus grand nombre de terminaisons nerveuses du nerf auditif. Le recrutement n'existe que si le nerf est intact, il témoigne d'une lésion limitée à l'organe de Corti. Il n'y a pas de recrutement en cas d'atteinte du nerf auditif. A une certaine intensité, le sujet à l'impression d'entendre avec l'oreille malade au même niveau que l'oreille saine.

L'audiométrie vocale :

Elle étudie la compréhension et l'intelligibilité du langage parlé. Le patient doit répéter des mots que l'examinateur prononce. La courbe vocale obtenue détermine d'une part le seuil d'intelligibilité (niveau d'intensité nécessaire pour que 50% des mots soient compris) et d'autre part le pourcentage de discrimination (pourcentage de mots compris à une intensité de 35 dB au dessus du seuil). Ce test peut se faire en audition binaurale ou unilatérale.

EXAMENS COMPLEMENTAIRES

L'impédancemétrie permet d'étudier non seulement l'appareil de transmission par la tympanométrie, mais aussi des mécanismes réflexes tel que celui du muscle de l'étrier détection, et d'autre part, l'étude indirecte du statut de l'oreille interne par la mesure des seuils du réflexe stapédien. L'oreille moyenne avec le tympan et les osselets joue le rôle d'adaptateur d'impédance qui facilite le passage du son de l'air dans le milieu liquidien de l'oreille interne. La mesure se fait avec un impédancemètre de MADSEN. L'énergie acoustique absorbée est donnée par la différence entre l'énergie acoustique incidente (connue) et l'énergie acoustique réfléchie (mesurée). La quantité d'énergie réfléchie par l'oreille est fonction de son impédance. Des mesures de compliance (inverse de la rigidité) du tympan peuvent être effectuées. On obtient normalement un pic centré sur la pression zéro traduisant une équipression sur les deux faces du tympan. Un pic décalé traduira un

dysfonctionnement de la trompe d'eustache, alors qu'un épanchement dans la caisse du tympan correspondra à une courbe plate.

Les Potentiels Evoqués Auditifs (PEA) consistent en l'étude de l'activité bioélectrique générée par des clics de stimulation le long des voies auditives. L'activité électrique induite par un stimulus sonore recueilli par des électrodes externes est très faible et doit être extraite du bruit de fond par des méthodes électroniques pour pouvoir être enregistrée (filtrage, moyennage). Ce test permet d'obtenir un tracé fait de 7 ondes qui reflètent l'activité électrique du système nerveux auditif, chaque onde correspondant à une structure anatomique précise. Les PEA ont en grande partie remplacé l'électro-cochléogramme qui nécessitait la pose d'une électrode à travers le tympan.

Les Otoémissions Acoustiques Provoquées (OEAP) consistent à étudier indirectement les mécanismes actifs caractéristiques des cellules ciliées externes en réponse à des clics de stimulation. Les OEAP sont enregistrées par une sonde placée dans le conduit auditif externe. Il s'agit d'un test de dépistage de la surdité non invasif, rapide et indolore. Si la présence des OEAP atteste d'une audition normale sur les fréquences comprises entre 2000 et 4000 Hz (seuil maximal de 30 dB), l'absence des OEAP en revanche, ne permet pas de conclure à une surdité.

L'imagerie (radio, scanner, IRM) va permettre d'avoir une vue anatomique plus précise de différentes partie non visibles de l'appareil auditif.

4.2 CAUSES DE LA SURDITE

4.2.1 ETIOLOGIE DES SURDITES DE TRANSMISSION

La surdité de transmission est liée à une lésion de l'oreille externe ou de l'oreille moyenne. Elle touche essentiellement la conduction aérienne et est caractérisée par une conduction osseuse normale (test de Rinne négatif et test de Weber latéralisé du côté atteint) et par une absence de recrutement lors du test de Fowler. Elle se traduit par des déficiences auditives légères ou moyennes.

4.2.1.1 SURDITES DE TRANSMISSION AVEC TYMPAN NORMAL

Lorsque l'examen otologique indique que le tympan est normal, la cause la plus fréquente de surdité est l'otospongiose, ou otosclérose. Il s'agit d'une évolution pathologique des tissus osseux, et il en existe deux types. Un premier type consiste en la fixation de la platine de l'étrier dans la fenêtre ovale tandis que le second type, appelé otospongiose cochléaire, fait intervenir une calcification du labyrinthe osseux de l'oreille interne et donne une surdité de perception.

Les principales autres causes de ce type de surdité sont les bouchons de cérumen obstruant le conduit auditif externe, les ruptures de la chaîne des osselets à la suite d'un traumatisme crânien, ainsi que les malformations ossiculaires.

4.2.1.2 SURDITES DE TRANSMISSION AVEC TYMPAN ANORMAL

Lorsque l'examen otologique indique une affection du tympan, on peut distinguer trois principales causes parmi les surdités de transmissions : les affections infectieuses ou inflammatoires (notamment les otites), les traumatismes mécaniques, et les tumeurs ou malformations affectant l'oreille moyenne ou externe.

Les otites sont une cause importante de surdité. Parmi elles, on distingue en premier lieu l'otite séreuse (surtout chez les enfants). Vient ensuite l'otite moyenne aiguë, maladie fréquente, liée à une infection du conduit auditif, le plus souvent d'origine virale. Enfin, les otites chroniques et leurs séquelles représentent la cause majeure de ces surdités. Dans le cas de l'otite chronique simple non évolutive l'oreille ne coule plus depuis des années mais il persiste une surdité de transmission en rapport avec une perforation du tympan et une destruction plus ou moins importante des osselets. Dans le cas d'une otite chronique évolutive muqueuse simple, une otorrhée récidivante est souvent associée à une infection chronique des voies aériennes supérieures. Il arrive que l'otite chronique soit accompagnée de tympanosclérose qui se caractérise par des calcifications tympaniques, pouvant être considérées comme un témoin de la cicatrisation tympanique. Finalement, l'otite fibro-adhésive, autre type d'otite chronique, est la conséquence directe de l'inflammation chronique de la caisse du tympan. Elle est caractérisée par un comblement de la caisse du tympan par du tissu conjonctif épais entraînant la disparition de tout espace aérien résiduel, bloquant les osselets, ne laissant que quelques espaces remplis d'une glu épaisse.

Une autre cause de surdité de transmission est le traumatisme mécanique. L'introduction d'un corps étranger dans le conduit auditif externe peut parfois provoquer une perforation du tympan voire une luxation des osselets, et ainsi une surdité de transmission. Egalement les traumatismes crâniens affectant les os du rocher, ainsi que les barotraumatismes (plongée sous marine, avion) peuvent être à l'origine de surdité de transmission.

On observe rarement des tumeurs ou des malformations congénitales responsables des surdités de transmission. Toutefois, le cholestéatome est la présence d'épithélium pavimenteux stratifié dans l'oreille moyenne (c'est-à-dire de peau). Cet épithélium, rebelle aux traitements antibiotiques, desquame et se kératinise, envahissant à des degrés variables, entre autres, la caisse du tympan. Il peut ainsi provoquer l'érosion voire la destruction des structures contenues dans et autour de l'oreille moyenne.

4.2.2 ETIOLOGIE DES SURDITES DE PERCEPTION

Les surdités de perception sont caractérisées par une atteinte soit de la cochlée (surdité endocochléaire) soit des voies nerveuses situées en arrière de la cochlée (surdité rétrocochléaire).

4.2.2.1 LES SURDITES ENDOCOCHLEAIRES

Elles correspondent à une atteinte des cellules sensorielles de l'organe de Corti. Parmi les nombreuses causes, on distingue les causes héréditaires, congénitales, infectieuses, inflammatoires, toxiques, dégénératives ou physiopathologiques. Le mode évolutif des surdités endocochléaires doit être précisé afin de différencier deux groupes d'atteinte : les surdités d'apparition brutale et les surdités d'apparition progressive.

Parmi les surdités endocochléaires une cause courante est le vieillissement. On parle de presbyacousie il s'agit d'un phénomène plus ou moins marqué selon les individus, défini comme une perte progressive de l'audition, liée à l'âge, bilatérale et symétrique, surtout dans les fréquences élevées. Elle correspond à des mécanismes divers, et dont les effets se combinent généralement : d'une part la perte de souplesse des muscles qui gèrent le dispositif des osselets, et d'autre part la perte progressive des cellules ciliées externes de la cochlée. Les conséquences de la presbyacousie apparaissent généralement au-dessus de l'âge de cinquante ans, et évoluent progressivement et symétriquement. Par ailleurs, le vieillissement naturel de l'organisme peut parfois déboucher sur une arthérosclérose des vaisseaux cochléo-vestibulaires, également susceptible de provoquer une surdité endocochléaire.

Les traumatismes sont une autre cause de surdité endocochléaires. Les traumatismes sonores aigus, tels que des sons très forts (musique, marteaux-piqueurs, déflagration, armes à feu) ou des barotraumatismes, entraînent une baisse de l'acuité auditive, variant d'une simple encoche, unilatérale ou asymétrique, à la surdité totale. Egalement, les fractures de la base du crâne secondaires à un traumatisme crânien peuvent entraîner une atteinte auditive de type perception, lorsque le trait de fracture du rocher passe par la coque labyrinthique.

Les surdités d'origine infectieuses, virales (rougeole, oreillons, varicelle, zona, polyomyélite, grippe, cytomégalovirus) ou bactériennes, ou inflammatoires peuvent également mener à des surdités de perception. Certains cas peuvent être liés à une infection pendant la période prénatale.

L'atteinte auditive d'origine toxique est généralement bilatérale, définitive, d'où l'importance de la prévention. Le degré de l'atteinte est variable, pouvant aboutir à une cophose bilatérale. Les principales classes médicamenteuses en cause sont les antibiotiques de la famille des aminoglycosides (gentamicine, néomycine, amikacine…), les diurétiques de l'anse (furosémide, acide étacrynique…), les antimitotiques type cisplatine, les dérivés de la quinine, l'acide acétylsalicylique.

La maladie de Ménière est une maladie chronique de l'oreille interne dont la cause exacte demeure inconnue et donnant lieu à une surdité de perception endocochléaire. Cette maladie associe des symptômes témoignant de lésions endocochléaires et vestibulaires. La maladie évolue par crises associant typiquement vertiges, acouphènes, plénitude auriculaire et surdité de perception. Les crises se produisent à intervalles variables, avec tendance à l'installation d'une surdité permanente.

Les surdités auto-immunes sont une pathologie très invalidante, qui se manifeste par une surdité de perception bilatérale dans 80 % des cas, et comprend des surdités de perception

progressives mais aussi des syndromes ménièriformes, des surdités brusques et des atteintes vestibulaires bilatérales isolées (syndrome de Dandy). Aucun test biologique simple, fiable et répandu ne permet pour l'instant de porter avec certitude le diagnostic de surdité auto-immune. Celui-ci reste donc un diagnostic d'élimination posé après avoir écarté une maladie systémique par un bilan clinique et biologique adapté.

Enfin, certaines affections de type auto-immune sont susceptibles de déclencher des réactions immunitaires aboutissant à la destruction des antigènes de la cochlée et donc à une surdité neuro-sensorielle.

4.2.2.2 LES SURDITES RETROCOCHLEAIRES

Les surdités de perception rétrocochléaires correspondent à l'atteinte des voies nerveuses situées après la cochlée.

Lorsque cette atteinte porte sur le nerf auditif, elle peut être d'ordre toxique (alcool, tabac, certains antibiotiques ou composés chimiques ototoxiques), infectieux (méningo-névrites des oreillons, du zona, de la méningite ou de la syphilis) ou auto-immune (sclérose en plaques).

Une des principales étiologies de surdité rétrocochléaire sont les tumeurs de l'angle ponto-cérébelleux (85 % de neurinomes de l'acoustique). Le neurinome de l'acoustique est une tumeur bénigne, généralement développés aux dépens du nerf vestibulaire inférieur. Cette tumeur va secondairement comprimer le nerf acoustique, puis le nerf facial. Son évolution clinique commence par une surdité de perception unilatérale le plus souvent progressive, associée à des acouphènes et des vertiges.

Les surdités dont l'origine est à l'intérieur du cerveau lui-même s'associent généralement à beaucoup d'autres troubles neurologiques dont la sévérité place au second plan le désagrément d'une perte auditive qui n'est jamais complète, et qui est d'ailleurs bien plus fait d'une compréhension difficile ou mal analysée que d'une diminution de la sensibilité sonore.

L'examen de choix pour diagnostiquer les surdités d'origine rétrocochléaire est celui des potentiels évoqués auditifs (PEA).

4.3 PROTHESES AUDITIVES

Les prothèses auditives s'adressent aux patients, dont la perte auditive est stable et ne peut être traitée autrement. Par exemple dans le cas d'une otite, on pourra trouver un traitement adapté afin que se résorbe l'inflammation/infection, ou bien encore dans le cas d'une obstruction du conduit auditif, on pourra procéder à l'extraction du corps étranger.

Actuellement, il existe plusieurs types de prothèses auditives, qui se distinguent essentiellement par la nature et la cause de la surdité à palier. Leur principe général de

fonctionnement est d'adresser à l'oreille interne un message acoustique amplifié et adapté à l'audition restante.

4.3.1 LES PROTHESES AUDITIVES EXTERNES

Les prothèses auditives externes, ou audioprothèses, ont pour objectif de compenser une perte auditive par un appareillage adéquat. Ce sont les aides auditives les plus répandues, sous forme d'intra-conduit, d'intra-conque, de contours d'oreille, etc. L'appareillage est choisi, adapté et réglé par un audioprothésiste qui agit sur prescription médicale

Elles sont toutes composées des mêmes éléments dont le fonctionnement va varier selon leur forme et leur lieu de compensation. En premier lieu, le capteur (microphone) transforme les ondes de pression sonore en tension électrique. Il peut être omnidirectionnel, ou directionnel, ce qui permet de réduire les bruits latéraux et arrière par rapport à l'auditeur. Sa bande passante est limitée à la zone utile pour la compréhension de la parole (100 Hz - 8000 Hz). Le microphone peut être remplacé par une bobine magnétique pour l'écoute au téléphone. Après le capteur vient ensuite un système d'amplification, alimenté par piles. Il est caractérisé par son gain, réglable par un potentiomètre. Il possède un dispositif limiteur du niveau de sortie, et un système de compression (AGC), qui augmente l'amplification pour les signaux faibles, et la réduit pour les signaux forts. Enfin, le transducteur (écouteur ou vibrateur) transforme le signal électrique en signal acoustique, soit au moyen d'un écouteur électromagnétique, pouvant être à l'origine de distorsions en raison de sa miniaturisation, soit au moyen d'un vibrateur mastoïdien (serre-tête, branche de lunettes). La numérisation des prothèses a permis d'informatiser les réglages. La numérisation concerne presque l'ensemble des prothèses. On peut ainsi adapter parfaitement le type d'amplification aux caractéristiques du patient, en modifiant par exemple : le gain, l'écrêtage, le niveau de déclenchement de la compression, la fréquence de coupure et la pente des filtres aigus et graves, une zone d'amplification sélective, etc. Cela a également permis de s'adapter à l'environnement (salle isolée, stade, église…).

Si les prothèses sont à peu près toutes construites sur ce schéma général, il en existe de nombreux types que nous allons envisager rapidement.

4.3.1.1 LES PROTHESES A CONDUCTION AERIENNE

Ce sont les prothèses les plus répandues. Elles délivrent un signal fait de vibrations aériennes transmises au conduit auditif externe pour agir directement sur les structures de l'oreille moyenne. En général le transducteur de ces prothèses possède plusieurs propriétés. Tout d'abord il écrête le signal d'entrée, c'est-à-dire qu'il coupe tout ce qui est au-dessus de 100 décibels. Ces sonorités intenses en effet n'ont pas d'importance sur le plan de l'information et risqueraient d'être traumatisantes pour le patient. Ensuite il amplifie le signal sonore. Mais il l'amplifie en tenant compte du recrutement du malade, c'est-à-dire qu'il fait en sorte qu'à la sortie du haut-parleur le

signal maximum ne puisse pas dépasser le seuil douloureux du patient. Ce procédé, appelé compression, mettait toujours un certain temps à se déclencher sur les prothèses anciennes, qui étaient analogiques, jusqu'au milieu des années quatre-vingt-dix, ce qui expliquait que les bruits violents restaient malgré tout souvent désagréables pour le sourd. La numérisation de ces prothèses est une retombée directe de la technologie des implants cochléaires. Elle a permis d'obtenir une compression du signal presque immédiate, et surtout de voir apparaitre des prothèses présentant jusqu'à huit canaux indépendants, qui permettent d'amplifier de manière sélective les graves, les médiums, et les aigus de façon spécifique.

Ces prothèses à amplification aérienne sont ergonomiques, c'est-à-dire qu'elles ont différents aspects esthétiques. La plus séduisante, celle qui a priori plaît le plus aux patients, est le **système intra-auriculaire**, dont il existe divers modèles, plus ou moins bien cachés dans le conduit auditif externe. De face en effet la prothèse n'est pas visible. La prothèse ***intra-conduit*** est totalement située dans le conduit auditif externe. Le microphone est placé à l'entrée du conduit, et l'écouteur presque au contact du tympan. Il persiste toutefois des effets de résonance dus à la masse d'air résiduelle entre l'écouteur et le tympan. Il est particulièrement indiqué dans l'appareillage stéréophonique, car il permet d'améliorer l'intelligibilité dans le bruit. Il trouve sa limite d'application pour des surdités supérieures à 60/70% de perte auditive. La prothèse ***intra-conque*** est, comme son nom l'indique appliqué dans la conque et le conduit auditif externe. Moins esthétique que l'intra-auriculaire, il permet néanmoins d'appareiller des surdités de 70 à 80 dB. L'inconvénient de ces prothèses est que l'électronique y est forcément simplifiée, et l'amplification ne peut pas être très importante.

Tout aussi connues mais moins appréciées sont les prothèses rétro-auriculaires, les fameux **contours d'oreille**. Il s'agit de l'appareil le plus utilisé. Le corps de la prothèse ressemble aux deux premières phalanges d'un doigt. Il se trouve caché derrière le pavillon, maintenu à cheval sur le haut de l'oreille par un tube en plastique creux, qui conduit les sons issus du haut-parleur jusque dans le fond du conduit rendu étanche par un embout moulé. Ces prothèses ont l'inconvénient d'être plus visibles et de tenir un peu moins bien en place que les prothèses intra-auriculaires. Mais l'électronique qu'elles contiennent permet une sophistication bien plus poussée. Les récentes prothèses les plus perfectionnées sont des prothèses rétro-auriculaires.

Enfin, il existe également des **lunettes munies de prothèses** à conduction auditives. Il s'agit d'une paire de lunettes, sur laquelle l'une des branches contient l'amplificateur, et un tube écouteur placé dans le conduit auditif externe.

4.3.1.2 LES PROTHESES A CONDUCTION OSSEUSE

Ces prothèses transmettent l'information sonore, non pas par voie aérienne, mais par mise en route d'une vibration osseuse. Il s'agit essentiellement de vibreurs appliqués sur l'arrière de l'oreille contre la mastoïde. Ces vibreurs sont généralement fixés sur l'extrémité d'une branche de lunettes, dont les verres sont souvent factices, ou au contraire adaptés à la vue du sourd. On peut les trouver aussi sous forme de serre-tête avec un boîtier externe à la branche du serre-tête.

Ces prothèses à conduction osseuse s'adressent aux surdités inappareillables par conduction aérienne en raison d'un écoulement important dû à une otite chronique responsable de la surdité. Mais pour qu'elles soient efficaces, il faut que l'oreille interne du patient soit pratiquement indemne. Le rendement et la puissance dépendent de l'épaisseur de la peau, de la pression exercée, et de la surface de contact.

4.3.2 LES PROTHESES IMPLANTABLES

4.3.2.1 LES PROTHESES OSSICULAIRES

Elles jouent un rôle de substitution pour les osselets. Elles sont indiquées dans certains cas d'atteinte de l'oreille moyenne. Ces atteintes sont pour la plupart d'origine inflammatoire (otite chronique éventuellement accompagnée d'une lyse ossiculaire), dégénérative (otospongiose), malformative ou traumatique (luxation des articulations ossiculaires, éventuellement accompagnée d'une fracture ossiculaire). Lorsque la chaîne des osselets est détériorée, il est possible de rétablir la continuité de cette chaîne soit en remodelant les osselets (ossiculoplastie), soit en remplaçant certains d'entre eux par des prothèses ossiculaires. En cas d'otospongiose, qui provoque une ankylose de l'étrier, on remplace le corps de l'étrier par un piston millimétrique en téflon.

4.3.2.2 LES PROTHESES A ANCRAGE OSSEUX

Il s'agit d'une prothèse à conduction osseuse implantée chirurgicalement. Elle est encore quelquefois nommée BAHA, pour l'acronyme anglais de Prothèse auditive à ancrage osseux (*Bone Anchored Hearing Aid*). Pour cela, une vis en titane est vissée dans l'os mastoïdien, sous la peau. Trois mois plus tard, lorsque la vis est ostéo-intégrée, on place un pilier qui traverse le tissu cutané. Après cicatrisation, le pilier recevra la partie externe, qui transmettra directement les vibrations à l'os, en s'affranchissant de la barrière cutanée. L'intérêt de cette prothèse est qu'il n'y a pas d'absorption des sons par le tissu cutané : le son est plus clair et plus net qu'avec une prothèse à conduction osseuse par bandeau ou lunette. Toutefois, seules les surdités de moins de 50 dB en moyenne sont concernées. L'indication typique de ce type de prothèse est représentée par les malformations majeures de l'oreille, notamment en cas d'aplasie du pavillon et/ou du conduit auditif externe.

4.3.2.3 LES PROTHESES VIBRATOIRES D'OREILLE MOYENNE (VIBRANT)

Il s'agit d'un petit vibrateur placé directement sur l'enclume, qui est un des osselets de l'oreille moyenne (voir paragraphe « Oreille moyenne », p17). Ce vibrateur est relié à un processeur externe par un aimant. La partie implantée est mise en place au cours d'une intervention de type

tympanoplastie avec tympanotomie postérieure. Elle respecte totalement l'anatomie et la physiologie normale de l'oreille. Lorsque les sons arrivent au processeur externe, ils mettent en œuvre le vibrateur qui amplifie les mouvements normaux de la chaine ossiculaire. L'avantage de ce type d'appareillage est qu'il permet d'éviter les inconvénients des prothèses traditionnelles : obstruction du conduit, effet Larsen, inconfort. De plus, leur principe de fonctionnement permet de délivrer des sons de meilleure qualité que l'amplification acoustique. Cette prothèse est indiquée dans les surdités neurosensorielles d'origine endocochléaire, à condition que l'oreille interne soit normale. Leur usage est limité par leur coût (près de 10 k€), et par la nécessité d'une intervention chirurgicale, et des risques qui lui sont inhérents.

4.3.2.4 LES IMPLANTS COCHLEAIRES

Le principe de fonctionnement des implants cochléaires sera détaillée dans la partie « Les Implants cochléaires », p113. De façon générale ils sont constitués d'une partie externe, comprenant un boîtier récepteur et un processeur du signal, et d'une partie interne comprenant un récepteur et un porte-électrodes. Les sons sont captés, analysés et filtrés par le processeur externe, sont ensuite transmis au récepteur implanté qui stimule directement les cellules réceptrices de la cochlée, et ainsi le nerf auditif.

Les implants cochléaires sont réservés aux patients atteints de surdité totale bilatérale non appareillable par les moyens conventionnels, ou ne tirant aucun bénéfice des prothèses auditives classiques. L'implant cochléaire a bouleversé la prise en charge des surdités réputées jusqu'alors intraitables.

4.3.2.5 LES IMPLANTS DU TRONC CEREBRAL

Le principe de fonctionnement d'un implant du tronc cérébral est identique à celui des implants cochléaires, à la différence près que les électrodes de stimulation se trouvent dans les noyaux cochléaires du tronc cérébral.

Les implants du tronc cérébral s'adressent à la même population que les implants cochléaires, excepté qu'il est indiqué lorsque le nerf auditif est endommagé, par exemple lors d'une neurofibromatose de type 2 (NF2).

5 LES IMPLANTS COCHLEAIRES

Les implants cochléaires permettent à des patients atteints de surdité sévère ou profonde, le plus souvent bilatérale, de retrouver une audition ou une sensation auditive menant ainsi à la communication orale. L'intérêt de cette technique est de s'absoudre totalement de l'état de fonctionnement de l'oreille. Il s'agit d'une technologie particulière permettant la stimulation directe du nerf auditif. Toutefois, comme nous le verrons le signal transmis au patient au travers d'un implant cochléaire est loin d'être « naturel ». Cette limite est précisément ce qui fait l'intérêt de l'étude de ce type de technologie : comment le cerveau humain est-il capable de réentendre avec un signal dégradé ?

5.1 HISTORIQUE

En 1957, Charles Eyriès, otologiste et André Djourno, professeur de physique médicale, pratiquent la première opération assimilable à un implant cochléaire, ils redonnent de l'audition à un sourd total en stimulant grâce à une bobine d'induction les fibres nerveuses acoustiques de son oreille interne (Djourno, Eyries, & Vallancien, 1957). Si le système précaire a permis au patient implanté d'entendre certains sons, le prototype tombe en panne après quelques semaines et n'est pas réparé. En 1961 William House, otologiste américain, reprend les travaux de Djourno et Eyriès et implante un système fiable, progressivement proposé à des patients de plus en plus nombreux. Il s'agit là encore d'un système monoélectrode, ne permettant de reconnaître que les rythmes de la parole, et donc un simple complément à la lecture labiale. Bien que les premiers essais d'implants multi-électrodes remontent à 1964, le premier implant cochléaire fonctionnel remonte à septembre 1976 en France avec Chouard et MacLeod (Chouard & MacLeod, 1976; Pialoux, Chouard, & MacLeod, 1976), suivi en 1978, par l'australien Graeme Clark de l'université de Melbourne (G. M. Clark, Black, Forster, Patrick, & Tong, 1978).

Actuellement on estime à environ 120 000 le nombre de patients sourds bénéficiant d'un implant cochléaire dans le monde, dont environ 1/3 d'enfants. Plus spécifiquement en France en 2005 la Direction de l'Hospitalisation et de l'Organisation des Soins (DHOS) dénombrait 680 implantations dans l'année, dont un peu moins de la moitié sur des enfants.

5.2 TECHNIQUE

Les implants cochléaires ont, depuis la fin du XXème siècle, beaucoup évolué. Il existe actuellement 4 fabricants d'implants cochléaires, Cochlear®, MedEl®, Advenced Bionics® et Neurelec®. Chacun de ces fabricants perfectionnent sans cesse leur technologie, on comprend donc qu'il existe actuellement chez les patients un grand nombre de modèles différents. Nous allons donc décrire dans cette partie le principe général d'un implant ainsi que les différents types de stratégie de codage qu'il existe.

5.2.1 PRINCIPE ET FONCTIONNEMENT

Un implant cochléaire est un appareil destiné à rétablir partiellement l'audition chez les personnes atteintes de surdité sévère ou profonde, grâce à une stimulation électrique adéquate des fibres nerveuses auditives résiduelles de la cochlée. L'implant est indiqué dans les cas où l'organe de Corti ne s'est pas développé ou a été endommagé, rendant impossible la restitution d'une audition satisfaisante par des prothèses auditives externes. Dans tous les cas le nerf auditif doit être fonctionnel pour permettre cette technique.

Un implant cochléaire est constitué de deux parties fonctionnelles interdépendantes. La partie interne, destinée à être implantée chirurgicalement, est composée d'une antenne de réception, d'un récepteur-stimulateur et d'un porte-électrodes. La partie externe comprend un microphone permettant de capter l'information acoustique, un processeur vocal, une antenne de transmission et le système d'alimentation énergétique.

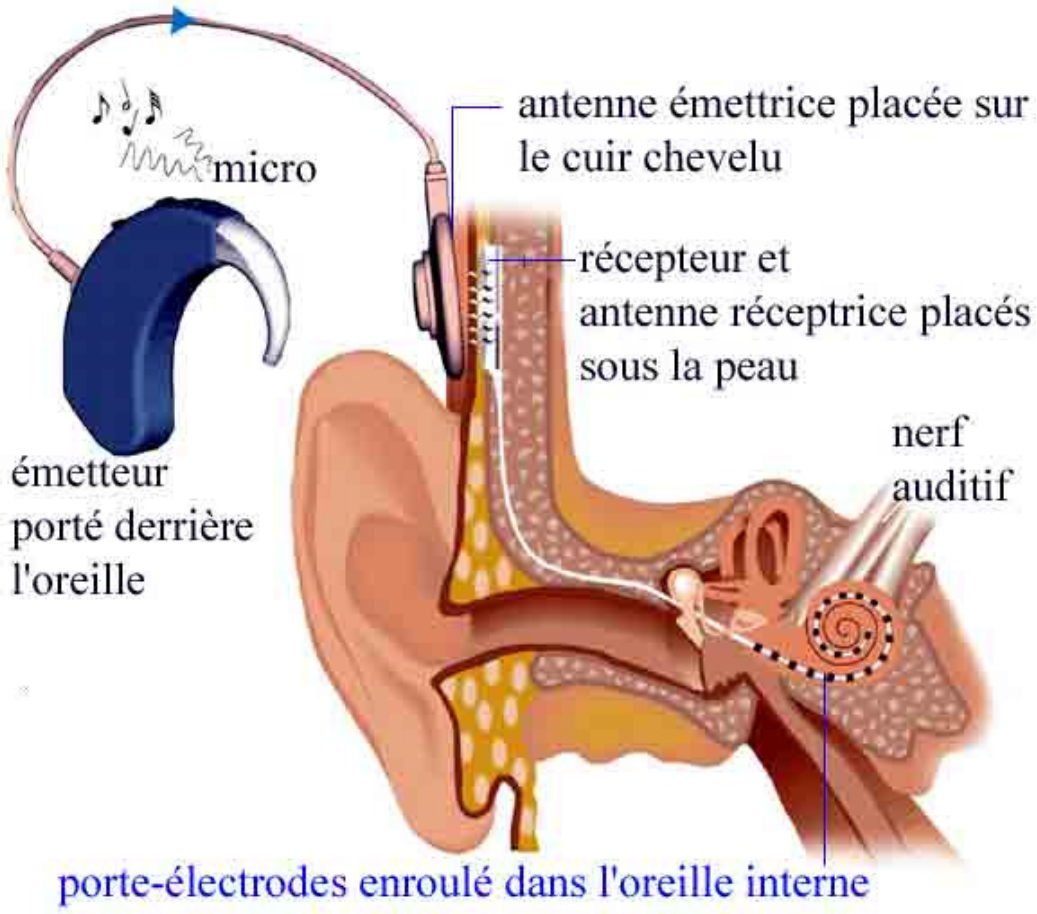


Figure 5.1 : Schéma du principe de fonctionnement d'un implant cochléaire. On peut distinguer sur ce schéma les différentes parties composant l'implant.

5.2.1.1 PARTIE EXTERNE

La partie externe analyse et code le message sonore en un signal électrique qui sera transmis aux parties internes de l'implant cochléaire.

Le microphone à pour rôle de capter les sons, tandis que le processeur vocal, contenu dans le boîtier ou dans le contour d'oreille dans le cas des implants sans boîtier, peut stocker jusqu'à trois programmes informatiques, définis par le régleur d'implant, que l'on appelle des « MAP ». Ainsi, les patients peuvent expérimenter et essayer différents programmes dans les environnements qui leur sont familiers. Ces programmes peuvent différer selon le volume, les seuils de perception ou les modes de stimulation. Pour changer de programme selon l'environnement sonore, il suffit d'actionner un petit bouton sur le processeur. De plus, le processeur est doté de deux autres commandes permettant de régler le volume et la sensibilité du microphone.

L'alimentation est soit une batterie rechargeable, soit des piles, et est contenue dans le contour d'oreille ou le boîtier suivant l'implant porté.

Enfin, l'antenne émettrice comporte un aimant et a pour rôle de transmettre l'information électrique émise par le processeur. Elle se positionne face à l'implant sous-cutané et garde cette position grâce à une attraction magnétique.

Figure 5.2 : Quelques exemples de processeurs à contour d'oreille. L'antenne émettrice n'est pas visible pour l'implant MedEl présenté mais elle est semblable aux autres. Les échelles des images ne sont pas toutes les mêmes.

5.2.1.2 PARTIE INTERNE

Le récepteur de la partie interne est une capsule électronique étanche, d'une épaisseur variant de 4 à 8mm, qui contient un aimant. Ensuite le faisceau d'électrodes est composé d'un porte-électrode (dont le nombre varie selon les fabricants) et d'une électrode de référence. Le porte-électrodes est introduit chirurgicalement dans la cochlée. Les implants cochléaires impliquent une correcte insertion dans la cochlée et permettent l'utilisation d'un courant plus faible et la possibilité d'avoir des informations plus complexes. Ils permettent l'insertion d'un implant multi-canal.

Figure 5.3 : Quelques exemples des parties implantables des implants cochléaires. Les échelles des images ne sont pas toutes les mêmes.

L'implant mono-canal n'est plus implanté depuis quelques années du fait de ses résultats limités, la discrimination fréquentielle étant considérablement appauvrie.

Les implants multi-canaux quant à eux essaient de reproduire la physiologie cochléaire en répartissant tonotopiquement l'analyse fréquentielle du message sonore le long de la cochlée : un son d'une fréquence donnée stimule une zone précise dans la cochlée, une bande de fréquence, donc certains neurones du nerf auditif. Ainsi, chaque électrode diffuse simultanément ou séquentiellement une information différente des autres électrodes. Plus il y a d'électrodes, plus le nombre de bandes de fréquences stimulées est important. L'espacement inter-électrode peut être constant ou se réduire progressivement de la base vers l'apex de la cochlée (implant Clarion). Les formes des électrodes diffèrent selon les constructeurs, tandis que le positionnement des électrodes dans la cochlée varie en

fonction du déroulement de l'intervention chirurgicale et des caractéristiques physiologiques de la cochlée. Toutefois, l'espacement minimal entre deux électrodes doit être suffisant pour qu'elles ne se perturbent pas entre elles. Si cette distance est abaissée, deux électrodes voisines vont stimuler les mêmes neurones et il y aura diaphonie entre les deux signaux.

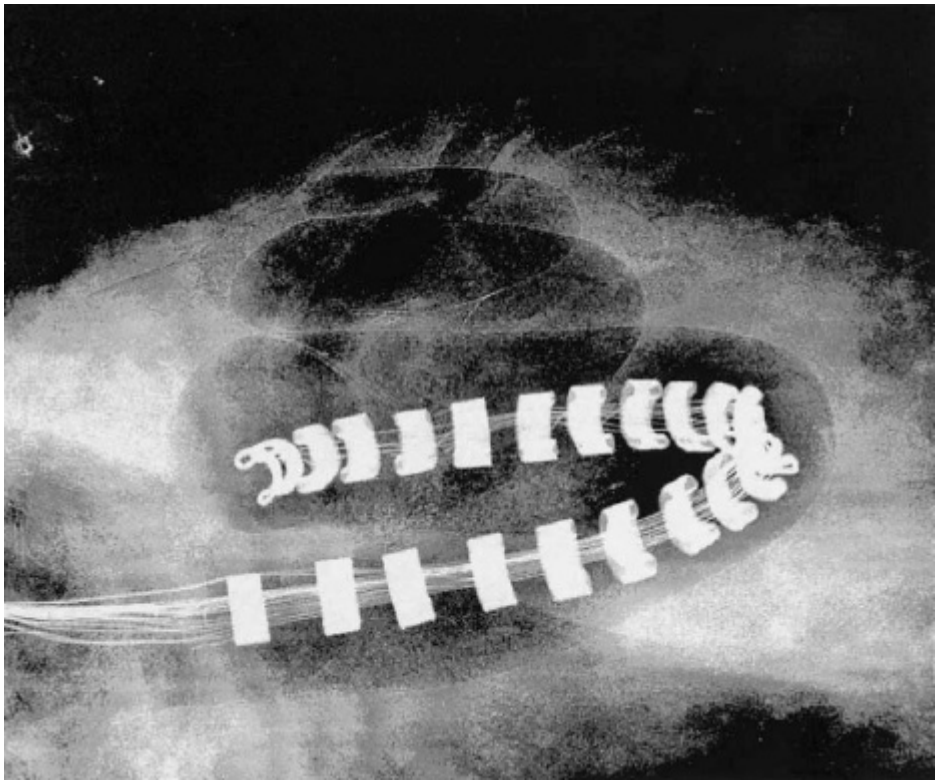


Figure 5.4 : Implant cochléaire Nucleus CI-22 inséré dans la rampe tympanique. Adapté d'après (G. Clark, 2003)

On distingue des modes de stimulation monopolaires, bipolaires, tripolaires et à masse commune. Dans les cas des modes monopolaires, une électrode est placée en dehors de la cochlée, en général dans le muscle temporal, et sert d'électrode de référence pour les électrodes de stimulation. Dans le cas des modes bipolaires, une électrode adjacente à l'électrode de stimulation est utilisée comme électrode de retour du courant électrique, ce qui offre des possibilités importantes de déclinaisons de ce mode, selon que l'électrode de stimulation est proche ou non de l'électrode de recueil. Dans le cas des modes tripolaires, ce sont les deux électrodes voisines de l'électrode de stimulation qui servent d'électrodes de retour. Enfin, dans le cas du mode à masse commune, toutes les électrodes non stimulantes à un instant donné servent de retour pour l'électrode stimulante.

5.2.2 STRATEGIES DE CODAGE

Pour l'ensemble des processeurs existants, la contrainte est le nombre limité d'électrodes au sein du porte-électrodes. Cette limite à au moins deux explications essentiellement technologiques. Comme nous l'avons vu, chaque électrode va coder une partie du signal sonore. On considère que chaque électrode va retransmettre les variations de signal d'une bande de fréquence, et que le placement de ces électrodes sur la cochlée en fonction de leur bande de fréquence respecte la tonotopie. La taille de cette bande varie d'un modèle d'implant à l'autre. On peut facilement imaginer que pour une retransmission parfaite du signal, il faudrait une électrode par fréquence, voire une électrode par neurone. Or ceci est actuellement impossible non seulement en raison de la taille réduite des neurones qui sont au-delà des capacités technologiques, mais surtout en raison des « fuites de courant » inter-électrodes, c'est-à-dire que les électrodes diffusent du courant assez largement à plusieurs neurones et que des électrodes trop rapprochées stimuleraient en partie les mêmes neurones, ce qui rendrait un signal assez confus. Actuellement pour compenser cette limite technologique, les fabricants d'implant développent des stratégies de codages différentes, en essayant d'extraire l'information pertinente pour la compréhension de l'auditeur. De plus, on sait qu'un codage trop riche en informations détériore l'intelligibilité. Il est important de noter que depuis les débuts de l'implant cochléaire, l'objectif principal, outre de rétablir une sensation sonore, était de permettre aux patients sourds de recouvrer l'usage de la communication verbale. La plupart des stratégies de codage des processeurs d'implant cochléaires sont donc axées sur la restitution des informations essentielles à la compréhension du langage. Mais on distingue tout de même 3 grandes familles de stratégie de codage. Les premières sont les stratégies utilisant une analyse temporelle du signal acoustique, ce qui permet de transmettre au mieux l'ensemble des sons ambiants sans privilégier spécifiquement la parole. Les secondes sont les stratégies spectrales, qui extraient spécifiquement les informations acoustiques significatives pour la parole humaine. Enfin, les stratégies hybrides utilisent une approche à la fois spectrale et temporelle.

La stimulation électrique délivrée aux électrodes varie en fonction de l'amplitude du signal acoustique restituée sous forme d'intensité électrique, de la vitesse de stimulation définie par électrode en impulsions par seconde (une vitesse de stimulation rapide rend mieux compte des variations acoustiques rapides accompagnant le langage) et de la nature des impulsions bi ou triphasiques. Dans les différents types de stimulations, la vitesse de stimulation constitue la variable à l'impact théorique le plus fort

5.2.2.1 LES STRATEGIES TEMPORELLES

Le but des différentes stratégies temporelles est de privilégier le traitement de l'information temporelle du message acoustique.

Pour les stratégies SAS (*simultaneous analog stimulation*) ou CA (*compressed analog*), le principe est basé sur le traitement du son par une oreille non endommagée (vers toutes les parties de

la cochlée en même temps) c'est-à-dire sur la stimulation simultanée des électrodes, on parle de codage analogique. L'avantage de ces stratégies est que davantage d'informations concernant le signal sonore peuvent être transmises avec chaque balayage de la cochlée. En revanche, l'inconvénient est le fait que la stimulation d'une électrode peut interférer avec celle d'une autre et faire apparaitre le phénomène d'interaction inter-canal (on appelle canal une bande de fréquence codée et retransmise par une électrode). Pour la stratégie SAS, une stimulation bipolaire est alors utilisée (les électrodes sont couplées lors de la stimulation) et donc 16 électrodes sont utilisées pour stimuler 8 canaux. La stratégie CA, est utilisée pour l'implant Inraid, tandis que la stratégie SAS est utilisée pour l'implant Clarion.

La stratégie CIS (*continuous interleaved sampling*) est actuellement la stratégie la plus rapide. Elle effectue un transcodage numérique afin de restituer fidèlement l'enveloppe temporelle dans chaque bande de fréquence. Les électrodes sont stimulées séquentiellement, c'est à dire les unes après les autres, ce qui permet de réduire considérablement les interactions électriques entre les différents canaux (Wilson et al., 1991). Cette stratégie a été crée pour combiner les avantages des impulsions non simultanées et ceux assurés par la préservation de la structure temporelle du signal analogue originel, tout en utilisant des impulsions brèves à un débit élevé. Actuellement cette stratégie est très utilisée par plusieurs fabricants d'implant. Par exemple, Advanced Bionics® l'utilise sous le nom de HiRes. La bande fréquentielle de codage du processeur s'étend de 250 à 8000 Hz, ce qui correspond aux fréquences les plus importantes pour la parole humaine, et est filtrée en 16 canaux. Seize électrodes peuvent donc être stimulées successivement. Pour cette stratégie il existe aussi des possibilités de stimulation simultanée partielle par paires d'électrodes. Cette stimulation multiplie théoriquement le nombre de canaux stimulables, on parle donc de canaux virtuels. MedEl® et Cochlear® utilisent également la stratégie CIS.

5.2.2.2 LES STRATEGIES SPECTRALES

Pour ces stratégies, l'approche est de considérer que les différents pics spectraux sont associés aux formants de la parole, qui traduisent la résonance acoustique des différentes cavités du conduit vocal (voir « L'espace supralaryngé », p76).

Dans la stratégie multi-peak (ou MPEAK) développé par l'implant Nucleus (Cochlear®), le processeur réalise une extraction des estimations des formants F0, F1, F2. Puis, il utilise 3 bandes de filtres pour transmettre des informations supplémentaires entre 2000 et 8000Hz. La fréquence de stimulation reste fixe.

5.2.2.3 LES STRATEGIES HYBRIDES

Les stratégies de codage les plus récentes essaient de restituer au mieux à la fois la structure spectrale et l'information temporelle.

Pour ces stratégies il s'agit de sélectionner les bandes de fréquences contenant les amplitudes maximum à chaque cycle de traitement et de les transmettre aux électrodes correspondantes. Les maxima sont variables dans le temps et peuvent donc changer à chaque cycle de traitement. Le nombre maximum de stimulations est prédéterminé par l'audiologiste ou par le technicien chargé de régler l'implant, le nombre d'électrodes réellement stimulées à un instant donné étant susceptible de varier dans la limite de ce maximum. La stratégie SPEAK a été la première basée sur ce traitement du signal et permettait l'extraction de 5 à 10 maximas, déterminés par le niveau du signal et sa composition spectrale, ainsi que le réglage propre à chaque patient. La bande passante est dépendante du nombre d'électrodes utilisables ; lorsqu'il y en a 20, la bande passante est comprise entre 250 et 10800 Hz, et s'il y en a 16 disponibles, la bande passante est comprise entre 250 et 5700 Hz. Cochlear® utilise actuellement la stratégie SMS sous le nom de ACE (*Advanced Combination Encodes*), qui est venue améliorer la stratégie SPEAK. Elle permet d'extraire jusqu'à 12 maxima et de stimuler ainsi successivement 12 électrodes. Le caractère dynamique des maximas sélectionnées limite cependant la vitesse de stimulation (en moyenne 2400 impulsions par secondes par électrode si 10 maximas sélectionnées).

5.2.2.4 D'AUTRES STRATEGIES NECESSAIRES ?

S'il est vrai que les informations spectrales, et notamment celles concernant les premiers formants, sont essentielles pour comprendre le langage, il n'en est pas moins vrai que l'information des variations temporelles (même dégradées spectralement) restent très importantes lors de la compréhension du langage (Q. J. Fu, Chinchilla, & Galvin, 2004; Q. J. Fu, Chinchilla, Nogaki, & Galvin, 2005; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). De plus l'ensemble des implants traitent une bande passante allant de 250 à 8000Hz ce qui est, comme nous l'avons vu, la zone de fréquence la plus importante pour la parole humaine. On peut donc dire que, dans l'ensemble les stratégies de codage tentent de restituer un message sonore qui permette, au minimum, la perception des informations essentielles à la compréhension du langage.

En 2009 Elliott et Theunissen (Elliott & Theunissen, 2009) suggèrent que dans le spectre fréquentiel de la parole il y ait des informations redondantes, autrement dit qu'une partie du spectre émis n'était pas pertinent pour la compréhension. Dans leur étude les auteurs ont montré que lorsqu'on supprime les modulations de fréquence supérieures à 250 Hz (et surtout celles supérieures à 1000 Hz) cela diminue considérablement les performances de reconnaissance de mots. Pour les auteurs ce sont surtout ces fréquences qui sont le support de l'information linguistique. Ils proposent alors de limiter le spectre fréquentiel des implants cochléaires en améliorant la précision du traitement du signal à l'intérieur de cette bande de fréquence. Si cette proposition de codage paraît intéressante, d'autres résultats de leur étude posent problème quant à la faisabilité de cette limitation. En effet, lorsqu'ils étudiaient la reconnaissance du genre avec les mêmes manipulations des modulations spectro-temporelles sur les même stimuli, ils montraient que la partie la plus pertinente du spectre pour avoir de bonne performance dans cette tâche se trouvait plutôt entre 3 et 7

cycles/kHz (soit 333 et 143 Hz environ). Privilégier la bande fréquentielle au-delà de 250 (ou même 1000 Hz) pourraient donc favoriser et améliorer la compréhension de la parole, mais cela reviendrait également à nuire à la perception du genre et peut-être même d'autres aspects paralinguistiques. Car si les auteurs ont montré que les fréquences au-dessous de 250 ou 1000 Hz n'apportaient pas beaucoup à la compréhension de la parole, il n'est pas exclu que leurs modulations soient, au moins en partie, le support des autres informations dont la voix humaine est le vecteur, telles que l'état émotionnel ou l'identité du locuteur. La réponse à la question d'une stratégie de codage adaptée reste donc pour l'instant relative à la nature des sons que l'on souhaite privilégier. A l'heure actuelle, les implantés possèdent généralement plusieurs « réglages » (différentes stratégies ou d'autres type de modifications) de leur implant leur permettant de s'adapter à différentes situations acoustiques. Mais la question d'une stratégie de codage permettant à la fois la restitution des informations essentielles à la compréhension de la parole et celle d'autres informations acoustiques n'est pas réglée. Doit-on améliorer la résolution spectrale (nombre d'électrodes disponibles) ou les stratégies de codages ?

5.3 L'IMPLANTATION COCHLEAIRE

L'implantation cochléaire s'adresse donc à un public de patients sourds sévères ou profond dont la surdité ne peut être traitée d'aucune autre manière et dont le nerf auditif est toujours fonctionnel. En effet, l'implantation cochléaire est un acte chirurgical qui, classiquement, va détruite l'oreille interne de manière irréversible. Toutefois, il existe de plus en plus de méthodes d'implantation, qui vont permettre de conserver l'audition résiduelle lorsqu'il y en a. Ces méthodes suscitent de plus en plus d'intérêt. Parce qu'il s'agit d'un acte chirurgical pointu, l'implantation cochléaire est réalisée dans des centres spécialisés, comportant la structure et le personnel pour accueillir et encadrer le suivi des patients implanté.

5.3.1 CRITERES D'IMPLANTATION

Les critères d'implantations sont rigoureusement définis et régulièrement réévalués par la Haute Autorité de Santé (HAS). Théoriquement dès qu'elle est reconnue, toute surdité totale ou profonde mérite d'être réhabilitée par l'implant cochléaire. Il en est de même pour les surdités sévères mal réhabilitées par les prothèses amplificatrices conventionnelles (moins de 50 % d'intelligibilité sans lecture labiale et avec prothèse). Cependant pour éviter les échecs, représentés essentiellement par un abandon de l'utilisation de l'émetteur, il faut respecter un certain nombre d'exigences avant de décider d'implanter un patient. Lorsqu'il s'agit d'une surdité prélinguale, l'implantation doit être faite le plus tôt possible. De façon consensuelle et suite aux nombreuses études sur le sujet, on considère que l'implantation pour être la plus profitable au patient sourd, doit être réalisée avant l'âge de 2 ans, si la surdité est détectée suffisamment tôt. En effet, les patients

sourds prélinguaux ayant subi une implantation cochléaire pendant leur adolescence ou à l'âge adulte rencontrent d'importantes difficultés pour accéder à une compréhension de la parole (R. S. Tyler & Lowder, 1992). Chez l'adulte sourd postlingual, il existe 3 sortes de critères.

5.3.1.1 CRITERES AUDIOMETRIQUES

Jusqu'au début des années 90, on considérait un patient comme implantable lorsqu'aucune sensation sonore au dessus de 500 Hz pour des signaux atteignant 110 dB n'était audible. L'inefficacité des prothèses auditives conventionnelles devait être évidente, et confirmée par une audiométrie vocale égale à zéro avec prothèse et sans lecture labiale. Actuellement, ces indications sont plus larges. L'aspect de la courbe d'audiométrie tonale a moins d'importance, ce qui est pris en compte est la qualité de l'audition obtenue par les prothèses amplificatrices. Lorsque, même avec l'aide de ces prothèses, le pourcentage d'intelligibilité de mots dissyllabiques sans lecture labiale est inférieur à 50 % à 65 dB, l'implant cochléaire peut être envisagé.

5.3.1.2 CRITERES OTO-RHINO-LARYNGOLOGIQUES

S'il n'existe aucun reste auditif à l'audiogramme tonal, on peut pratiquer le test de la stimulation électrique de la fenêtre ronde. La réalisation d'un implant cochléaire n'est possible que si au moins une partie de la fonctionnalité du nerf cochléaire est conservée, laquelle permettra en cas d'un implant de donner une sensation sonore, même si l'organe de Corti est totalement détruit. C'est pour cette raison qu'on réalise un test qui mesure la quantité de courant nécessaire pour provoquer la sensation sonore, et qu'ensuite on étudie la rapidité de la disparition de cette sensation, qui nous donne une idée de la fatigabilité de la future audition, en donnant une approximation du pourcentage de fibres nerveuses encore en état de fonctionnement (Shipp & Nedzelski, 1994).Ce test est réalisé sous anesthésie locale et sans hospitalisation. Il consiste à stimuler électriquement la membrane de la fenêtre ronde, ce qui provoque la diffusion du stimulus électrique dans les liquides cochléaires et permet d'atteindre les terminaisons nerveuses de la cochlée. Les résultats sont d'autant meilleurs que la quantité d'électricité nécessaire pour obtenir une réponse minimale est faible. La sensation sonore éventuellement obtenue permet d'attester la positivité du test. Les niveaux liminaires qui permettent d'obtenir une réponse sont soigneusement évalués. L'enregistrement à différentes fréquences de ces seuils permet de déterminer objectivement l'état du nerf auditif. De plus, l'enregistrement des potentiels auditifs permet l'objectivité de ce test et le pourcentage des fibres nerveuses encore utiles est alors évalué. Si le test est négatif d'un côté, on l'effectue du côté opposé. Il est rare que les réponses des deux oreilles soient négatives, mais dans ce cas, l'implantation risque d'être inutile. Si le test est positif, la réhabilitation de la surdité par l'implant est possible. On utilise aussi alternativement et plus fréquemment à celui-ci le test au promontoire. Il consiste en une stimulation électrique par une électrode placée à travers le tympan et permet de vérifier l'excitabilité du nerf cochléaire. Les seuils de détection et d'inconfort sont recherchés sur plusieurs fréquences. La

comparaison des résultats observés sur chaque oreille peut être utile dans le choix du côté à opérer. Ce test a également un certain intérêt pour apprécier l'effet de « masque » de la stimulation électrique sur des acouphènes associés à la surdité.

Le scanner permet d'analyser la structure osseuse de l'oreille. Il est utilisé pour exclure l'hypothèse de calcification de la cochlée suite à une méningite, otospongiose ou labyrinthite. Si une ossification se produit, la décision d'implanter doit être prise dans les plus brefs délais. On peut aussi détecter une dilatation du conduit auditif interne ou encore évaluer le degré de développement de la cochlée. On note aussi le degré de pneumatisation de la mastoïde qui indique si une approche chirurgicale est possible. Les variations anatomiques, malformations congénitales et lésions seront également détectées, par exemple une position anormale d'une artère qui passerait dans l'oreille moyenne, une position anormale du nerf facial, etc. Par ailleurs, le scanner est la seule méthode qui permet de voir presque en transparence la cochlée et ainsi d'évaluer très précisément la position de l'électrode dans la cochlée et sa position par rapport au modiolus.

Dans tous les cas, il faut s'assurer de la bonne perméabilité du tube cochléaire. Pour cela, le scanner peut aider, mais l'examen incontournable est l'IRM, qui permet de visualiser les liquides du labyrinthe. La présence de liquides dans la cochlée est un indice fiable de perméabilité de la cochlée. Lorsque la cochlée est ossifiée, il existe un risque majeur pour que la mise en place intra cochléaire de l'électrode soit incomplète ou impossible. Par contre, une fois l'électrode introduite, s'il y a une ossification, cela n'empêchera pas le passage des stimulations électriques qui peuvent atteindre les fibres nerveuses. L'IRM permet également de détecter d'éventuelles pathologies dans le cerveau, le canal auditif interne et la cochlée qui sont invisibles au scanner. Le diagnostic de l'absence du nerf vestibulo-cochléaire et une aplasie de la branche cochléaire du nerf vestibulo-cochléaire peut être réalisée avec une très haute résolution d'image avec l'IRM. On peut détecter les malformations congénitales de la cochlée avec la même précision que le scanner. En revanche, l'évaluation des variations anatomiques de l'oreille moyenne, les malformations et les lésions peuvent souvent rester indécelables en imagerie. Donc ces deux examens sont complémentaires.

Figure 5.5 : Visualisation par Imagerie par Résonance Magnétique (IRM) de l'ensemble du réseau liquidien du labyrinthe (canaux semi-circulaires et cochlée), confirmant leur perméabilité, et permettant de prévoir une mise en place facile du porte-électrodes.

Pour résumer, l'imagerie permet de s'assurer de l'absence de maladies ou de malformations de l'oreille qui interdisent l'implantation ainsi que de la perméabilité de la cochlée.

De façon générale, le choix du côté à implanter dépend directement du degré d'ossification cochléaire, de la durée de surdité avant implantation, des résultats de stimulation électrique de la fenêtre ronde, et indirectement de la dominance cérébrale (Deguine et al., 1995).

5.3.1.3 CRITERES PSYCHOLOGIQUES ET ENVIRONNEMENTAUX

Ce sont des critères si importants qu'ils peuvent amener à refuser un patient pour l'implantation. Le patient doit être psychologiquement stable et motivé, et doit accepter le principe d'une rééducation longue après la mise en place de l'implant. Pour cela, on évalue le degré de motivation du patient et on vérifie qu'il a bien pris connaissance et accepté les contraintes liées à l'implant. Les capacités cognitives, attentionnelles et mnésiques sont également des critères importants. Toujours sur le plan psychologique, le dépistage de pathologies psychiatriques invalidantes est aussi fait. Celles-ci peuvent être une contre-indication à l'implantation.

Particulièrement chez l'enfant mais également chez l'adulte il est essentiel que l'entourage soit patient et compréhensif avec le sujet. L'entourage est un soutien essentiel dans la réussite de l'implantation.

Le contexte social est également pris en compte car il est essentiel de voir si les contraintes liées à l'implant cochléaire sont socialement possibles (suivi lié à l'implantation).

5.3.2 L'ETAPE CHIRURGICALE

L'intervention chirurgicale permet de mettre en place la partie interne de l'implant (récepteur et porte-électrodes) et nécessite une anesthésie générale. L'intervention dure généralement 1h30 à 4h selon les circonstances.

Le récepteur-stimulateur et l'antenne réceptrice de la partie interne sont placés sous la peau par une incision derrière le pavillon de l'oreille. Un fraisage de l'os est souvent fait pour assurer leur placement convenable

On réalise ensuite une mastoïdectomie puis une tympanotomie postérieure pour accéder à la caisse du tympan. Une fois la caisse du tympan atteinte, on réalise une cochléostomie, qui permettra l'insertion du porte-électrodes. On obture ensuite de manière étanche la cochlée au moyen du fragment libre de facia temporalis et de muscle temporal de façon à éviter toute fistule labyrinthique, qui pourrait exposer le patient à un risque accru d'acouphènes ou de vertiges post-opératoires et également de méningite bactérienne en cas de survenue d'une otite moyenne aigüe.

Tous les fabricants proposent un appareil de télémétrie pour vérifier que les électrodes ont une impédance correcte en pré et per-opératoire. La télémétrie permet de vérifier le bon fonctionnement de l'implant en vérifiant notamment l'impédance et le voltage de chaque électrode,

et permet donc de détecter une éventuelle panne. La télémétrie est réalisée après l'insertion du porte-électrodes dans la cochlée.

Généralement, après la chirurgie, on prend une radio pour confirmer la position de l'appareil intérieur.

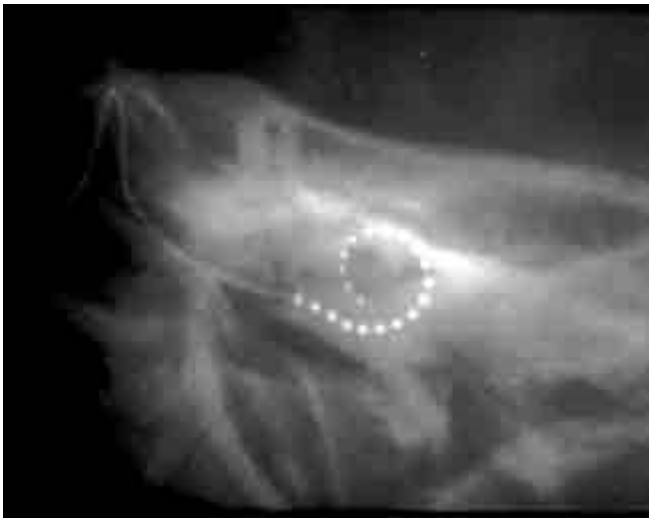


Figure 5.6 : Radiographie postopératoire montrant le porte-électrodes de l'implant Digisonic en place dans la cochlée.

En cas de cochlée ossifiée, l'intervention est beaucoup plus complexe et ses résultats plus aléatoires. Ces difficultés sont directement proportionnelles à l'importance de l'obstruction, et à la localisation de celle-ci sur la longueur du clavier cochléaire. Le cas le plus grave est l'obstruction totale, car on ne peut introduire le porte-électrode. On est obligé d'utiliser un système à électrodes séparées placées une à une dans des puits forés dans l'os cochléaire. On prend ici toute la mesure de l'importance de déceler avant l'opération cette anomalie grâce au scanner et surtout à l'IRM.

Les risques de complications associés à l'implantation cochléaires sont actuellement bien connus. Il existe des cas d'atteinte du nerf facial (<1/300 à 1/500 cas), et des possibilités restreintes d'infection. Il peut aussi survenir un tassement et une mise en court-circuit de quelques électrodes intracochléaires, ce qui est en général sans conséquence dans la mesure où un certain nombre d'électrodes restent fonctionnelles. Enfin on peut rencontrer des phénomènes d'expulsion de liquide céphalorachidien lors de la cochléostomie (souvent dans le cas d'une malformation) ou encore d'extrusion du porte-électrodes. Les autres symptômes, tels que l'engourdissement autour de l'oreille après la chirurgie, des étourdissements, etc., sont généralement transitoires.

5.3.3 SUIVI ET REEDUCATION

Généralement, en raison d'une durée de convalescence et de cicatrisation nécessaire, la mise en marche a lieu environ 1 mois après l'acte chirurgical. La première étape de cette mise en marche est le réglage de l'implant. Les réglages s'effectuent par programmation du processeur de manière informatisée. Après la mise en place de l'implant, le processeur doit être adapté à la sensibilité de chaque patient. Cette sensibilité dépend sans doute du pourcentage de fibres auditives résiduelles, et de la position des électrodes par rapport à ces fibres. Compte tenu de la technologie des appareils actuels il existe un grand nombre de paramètres réglables, qui dépendent directement du modèle d'implant et de la stratégie de codage du signal sonore utilisé. Toutefois le principe de base du réglage est de déterminer les quantités d'énergie minimales et maximales pour chacune des électrodes. Le seuil minimum sera la quantité d'énergie minimale à partir de laquelle le patient ressent une sensation auditive. Le seuil maximum ou seuil d'inconfort sera l'énergie maximale supportable par le patient : la sensation auditive doit être forte mais non douloureuse. Les réglages sont effectués directement par le rééducateur qui prendra en charge le patient au cours du suivi. Ces réglages s'effectuent en fonction des réponses du patient. Il est alors particulièrement important d'expliquer à l'implanté ce qui va évoluer grâce à la rééducation, et ce qui peut ou ne peut pas se résoudre par le réglage.

Par la suite, le suivi est réalisé par une équipe multidisciplinaire comprenant médecin ORL phoniatre, orthophoniste, audiologiste et psychologue qui ont déjà évalué le patient lors de la première étape. La prise en charge post-opératoire comporte donc un suivi médical, un entraînement au décodage de la parole, un apprentissage de l'écoute musicale, des séances de réglage de l'appareil, des tests successifs d'évaluation des performances des patients ainsi qu'un important soutien psychologique. Le premier mois, on conseille d'écouter mais sans pour autant essayer de tout comprendre, bien que certains patients arrivent à comprendre d'emblée. Les tests d'évaluation réalisés régulièrement, permettront de vérifier, de la façon la plus objective possible, les progrès des patients et d'aider le rééducateur dans les réglages de l'appareil.

Bien qu'il y ait de nombreux aspects positifs à la l'implantation cochléaire, il est important de rappeler qu'il ne procure pas une ouïe normale car l'implant cochléaire ne restitue que très partiellement la tonotopie physiologique normale. Lorsque le processeur de parole est débranché, il n'y a aucun son. De plus, la stimulation électrique appliquée au nerf auditif est encore assez éloignée de la stimulation normale produite par une cochlée saine. Le message nerveux généré par l'implant cochléaire est donc considérablement dégradé et appauvri en information, en comparaison avec un système auditif intact. Après une implantation cochléaire, il y a donc un processus d'apprentissage pour obtenir le maximum de bénéfices de l'appareil. L'adaptation à ce nouveau « son » dépend des compétences préalables à l'implantation des patients, de leurs capacités d'adaptation, et des contraintes physiologiques de leur système auditif. Ces différents paramètres expliquent ainsi qu'il y

ait une grande variabilité dans les capacités de récupération auditive des patients implantés cochléaires.

5.4 CONSEQUENCES DE L'IMPLANTATION COCHLEAIRE SUR LES PERFORMANCES AUDITIVES DES PATIENTS

Nous allons maintenant décrire l'impact qu'a l'implantation cochléaire sur les performances auditives des patients. Pour cela nous distinguerons plusieurs catégories de sons. En premier lieu la parole et plus précisément au sein de la parole nous distinguerons le langage, qui fait l'objet d'une rééducation spécifique chez les patients, et les autres informations supportées par la voix, les informations dites paralinguistiques. Nous pourrons constater le déséquilibre qui existe entre l'intérêt qui est porté aux études sur la compréhension de la parole et celui porté aux études sur la perception des informations paralinguistiques ; la récupération de la perception de la parole est toujours beaucoup plus étudiée et détaillée. Ensuite nous aborderons brièvement l'état de leurs performances dans le domaine des sons de l'environnement et enfin dans celui de la musique.

5.4.1 LE LANGAGE

En termes de perception du langage, les résultats de l'implantation cochléaire sont assez satisfaisants. Les études portant sur la perception auditive de la parole chez les personnes sourdes implantées cochléaires s'accordent pour indiquer une nette amélioration de la reconnaissance de la parole auditive après implantation cochléaire. Ceci est vérifié par des données expérimentales aussi bien chez l'enfant sourd prélingual implanté (Bergeson, Pisoni, & Davis, 2003; Calmels et al., 2004) que chez l'adulte sourd postlingual implanté (Chouard, Fugain, Meyer, & Lacombe, 1983; UKCISG, 2004; Wilson et al., 1991). Wilson et Dorman en 2008 rapportaient des performances en discrimination de phrases supérieures à 80%, et ce chez plus de 80% des patients en condition auditive seule (Wilson & Dorman, 2008). Par ailleurs les performances auditives continuent à s'améliorer progressivement en fonction de la durée d'utilisation de l'implant cochléaire, aussi bien chez les enfants sourds prélinguaux (Bergeson et al., 2003) que chez les adultes sourds postlinguaux (Hamzavi, Baumgartner, Pok, Franz, & Gstoettner, 2003; Pelizzone, Cosendai, & Tinembart, 1999).

De façon générale, les déficits perceptifs de la parole chez les patients portent notamment sur les indices auditifs de lieu d'articulation, ainsi que sur les indices de voisement et de nasalité, comme l'ont montré de nombreuses études de confusion auditives de consonnes chez les adultes sourds postlinguaux (M. F. Dorman et al., 1990; Dowell et al., 1982; Doyle et al., 1995; Kiefer, Hohl, Sturzebecher, Pfennigdorff, & Gstoettner, 2001; Parkinson, Parkinson, Tyler, Lowder, & Gantz,

1998; M. W. Skinner, Fourakis, Holden, Holden, & Demorest, 1999; R. S. Tyler, Preece, Lansing, Otto, & Gantz, 1986; Valimaa, Maatta, Lopponen, & Sorri, 2002).

Les performances des patients en perception de la parole peuvent varier en fonction de la langue considérée. En effet, les patients ayant une langue natale tonale telle que le mandarin présentent de moins bonnes performances de compréhension de la parole (Q. J. Fu, Hsu, & Horng, 2004; Xu, Tsai, & Pfingst, 2002). Cette différence s'explique essentiellement par le fait que la compréhension des langues tonales repose en partie sur la perception des variations de pitch (Q. J. Fu, Hsu et al., 2004), tandis que pour les langues occidentales la perception de la parole peut très bien se faire sur la base des variations temporelles (Shannon et al., 1995).

Les performances des patients en matière de perception de la parole font également intervenir leurs habiletés visuelles en lecture labiale et leurs capacités d'intégration bimodale visuo-auditive. Si plusieurs études ont montré que les patients sourds avaient des performances en lecture labiale supérieures à celle des normo-entendants (Grant, Walden, & Seitz, 1998; Kaiser, Kirk, Lachs, & Pisoni, 2003; Summerfield, 1992; R. S. Tyler, Parkinson, Woodworth, Lowder, & Gantz, 1997), Rouger *et al.* ont montré que ces performances ne décroissaient pas même après l'implantation cochléaire (Rouger et al., 2007). Ceci est particulièrement intéressant pour ce qui est de la reconnaissance audiovisuelle de la parole. En effet, le fait que ces habiletés visuelles perdurent après l'implantation permet une amélioration progressive des performances audiovisuelles en fonction de la durée d'utilisation de l'implant avec des scores proches de 100% dans la reconnaissance de mots dissyllabiques, dès 6 mois après l'implantation (Rouger et al., 2007). Rouger *et al.* ont également montré que les patients sourds postlinguaux avaient des habiletés d'intégration multisensorielles supérieures à celles des normo-entendants (Rouger et al., 2007).

FACTEURS PRONOSTIQUES

Les performances des patients sourds postlinguaux implantés cochléaires en discrimination et compréhension de la parole ont été largement étudiées et avec elles certains facteurs pronostiques associés à une meilleures récupération de la compréhension du langage.

Parmi ces facteurs pronostiques certains sont liés uniquement à des paramètres techniques. Notamment, des études ont pu montrer que le nombre d'électrodes ou la fréquence de stimulation du nerf auditif par les électrodes d'un implant avaient un impact sur la compréhension de la parole par les patients. Les patients implantés ont besoin d'un nombre plus grand d'électrodes stimulatrices pour la perception des voyelles (7 canaux) que pour la perception des consonnes ou des phrases (4 canaux) (Fishman, Shannon, & Slattery, 1997; Kiefer, von Ilberg, Rupprecht, Hubner-Egner, & Knecht, 2000; M. W. Skinner, Holden, & Holden, 1995). Inversement, une augmentation de la fréquence de stimulation électrique produit une plus grande amélioration de la compréhension pour les consonnes que pour les voyelles (Q. J. Fu & Shannon, 2000; Kiefer et al., 2001; Kiefer et al., 2000; Loizou, Poroy, & Dorman, 2000).

On distingue également des facteurs pronostiques chirurgicaux. En effet, la profondeur de l'insertion du porte-électrodes dans la cochlée peut jouer un rôle sur les performances de reconnaissance de la parole (M. F. Dorman & Ketten, 2003; M. F. Dorman et al., 2007; McKay & Henshall, 2002; M. W. Skinner et al., 1995). Dans l'idéal, l'implant cochléaire devrait être inséré dans la rampe tympanique de façon à ce que chaque électrode entre en contact avec la position de la cochlée codant pour la fréquence centrale représentée par cette électrode. Mais en pratique, il est difficile de contrôler précisément cette correspondance car la tonotopie cochléaire varie légèrement d'une personne à l'autre. De plus, en fonction de la géométrie des conduits cochléaires et de la façon dont l'implant est inséré dans la rampe tympanique, il n'est pas possible de maîtriser complètement le positionnement des électrodes le long de la cochlée. Actuellement, plusieurs études psychophysiques réalisées chez le patient implanté cochléaire et par simulation de l'implant chez le sujet sain ont démontré que des distorsions spectrales telles qu'un décalage vers les basses ou les hautes fréquences (M. F. Dorman, Loizou, & Rainey, 1997; Q. J. Fu & Shannon, 1999b), ou une compression ou une expansion de la carte fréquentielle (Baskent & Shannon, 2003, 2004) induisent une dégradation de la compréhension de la parole. On peut donc penser que, bien que certaines études aient démontré que la compréhension de la parole en cas de distorsion spectrale est susceptible de s'améliorer de façon significative avec l'apprentissage (M. F. Dorman & Ketten, 2003; Q. J. Fu, Nogaki, & Galvin, 2005; Rosen, Faulkner, & Wilkinson, 1999), une bonne correspondance lieu/fréquence obtenue dès le début permettra un bénéfice comportemental optimal pour les patients implantés.

Une dernière catégorie de facteurs prédictifs à la récupération perceptive de la parole est d'ordre neurofonctionnelle. En premier lieu, des études ont montré qu'une courte durée de déprivation auditive et la présence d'une audition résiduelle comptent parmi les facteurs les mieux corrélés à une bonne compréhension auditive de la parole après une implantation cochléaire (Blamey et al., 1996; Deguine et al., 1993; Kiefer et al., 1998; Rubinstein, Parkinson, Tyler, & Gantz, 1999; Shallop, Arndt, & Turnacliff, 1992; van Dijk et al., 1999). Egalement, l'âge d'implantation et la durée de stimulation par l'implant, le bilan anatomique préopératoire et le taux de survie des cellules des ganglions spiraux comptent également parmi les principaux facteurs associés à l'évolution des performances de compréhension de la parole après implantation (Blamey et al., 1996; Deguine et al., 1993; Q. J. Fu & Shannon, 1999a; R.S. Tyler, Preece, Lansing, Otto, & Grantz, 1986; Waltzman, Fisher, Niparko, & Cohen, 1995). D'autres facteurs tels que la proportion de cellules nerveuses intactes (Jyung, Miller, & Cannon, 1989; Shepherd, Clark, & Black, 1983; Walsh & Leake-Jones, 1982), les propriétés neurophysiologiques du nerf auditif (C. J. Brown, Abbas, Borland, & Bertschy, 1996; C. J. Brown, Abbas, & Gantz, 1990; Stypulkowski & van den Honert, 1984), ou encore l'intégrité des voies auditives centrales (Kraus et al., 1993; Micco et al., 1995; Oviatt & Kileny, 1991; Stypulkowski, van den Honert, & Kvistad, 1986) ont également été corrélés à une meilleure récupération de la perception de la parole.

5.4.2 LES INFORMATIONS PARALINGUISTIQUES

Si les performances de perception de la parole chez les patients implantés cochléaires ont été décrites avec minutie au cours des 30 dernières années, on pourra constater qu'en ce qui concerne la perception des informations paralinguistiques, peu d'études s'y sont intéressé.

Seules quelques études se sont intéressées récemment aux capacités d'intégration de la voix chez les patients implantés cochléaires (Cleary, Pisoni, & Kirk, 2005; Q. J. Fu, Chinchilla et al., 2004; Q. J. Fu, S. Chinchilla et al., 2005; Gfeller, Knutson, Woodworth, Witt, & DeBus, 1998; Kovacic & Balaban, 2009; Meister, Landwehr, Pyschny, Walger, & von Wedel, 2009; Xin, Fu, & Galvin, 2007). La plupart d'entre elles portaient sur la perception du genre. Les auteurs y concluent toujours que les patients présentent un déficit, qui varie selon les études, en comparaison des normo-entendants. De plus, les études sur le sujet ont souvent essayé d'expliquer ce déficit, soit en manipulant différent paramètre acoustiques, soit en comparant les patients à des sujets normo-entendant en condition de simulation d'implant (vocoder). Dans l'ensemble, les auteurs montraient que les performances des patients sont limitées par la dégradation spectrale liée au processeur de l'implant cochléaire.

On peut citer par exemple Fu *et al.*, qui ont comparé les performances d'un groupe de 11 patients implantés cochléaires et d'un groupe de 6 sujets contrôles dans la reconnaissance du genre du locuteur (Q. J. Fu, Chinchilla et al., 2004). Les sujets contrôles ont été testés dans différentes conditions de dégradation acoustique par un vocoder afin d'identifier la contribution relative de l'amplitude, du pitch et de la durée dans cette tâche. Les résultats montraient que les patients étaient capables de reconnaître le genre à 94% d'identification correcte. De plus, les auteurs identifiaient comme indices acoustiques essentiels le pitch et les variations d'amplitude rapides, regroupées sous l'appellation de structure fine. Toutefois les capacités des patients à percevoir le genre d'une voix diminuaient considérablement lorsque les pitch des voix d'homme et de femme se retrouvaient dans la même fenêtre fréquentielle (Q. J. Fu, S. Chinchilla et al., 2005; Kovacic & Balaban, 2009).

En 2007 Xin *et al.*, se sont intéressés à la perception de l'affection vocale (Xin et al., 2007). Ils ont comparé la reconnaissance de 5 émotions entre un groupe de 8 patients implantés et un groupe de 8 sujets normo-entendants. Les résultats montraient que les normo-entendants atteignaient un scores de près de 90% de reconnaissance correcte tandis que les patients ne reconnaissance les émotions correctement qu'à 45%. La différence entre les deux groupes était significative, mais les performances des patients restaient supérieures à la chance D'après les auteurs, la perception des émotions faisait intervenir la perception des variations de pitch et de l'enveloppe spectrale de manière plus importante encore que pour la reconnaissance du genre.

Dans l'ensemble on peut voir que les performances des patients concernant la perception de la voix sont très peu évaluées. Les quelques études sur le sujets se sont essentiellement intéressées à la perception du genre de la voix. Les autres aspects paralinguistiques de la voix restent mal décrit chez le patient sourd implanté cochléaire, et l'évolution de ces performances au cours du temps encore plus.

5.4.3 LES SONS DE L'ENVIRONNEMENT

La perception des sons de l'environnement, est peut-être le domaine de la perception auditive le moins étudié chez les patients implantés cochléaires. Les rares études sur le sujet montrent que les capacités des patients dans ce domaines sont très déficitaires (Inverso & Limb, 2010; Looi & Arnephy, 2009; Reed & Delhorne, 2005). Des auteurs ont évalué des patients implantés et des sujets normo-entendants dans une tâche d'identification de 45 sons de l'environnement (Looi & Arnephy, 2009). Les résultats montraient un gros déficit des patients par rapport aux normo-entendant. Les auteurs montraient même qu'il n'y avait presque pas d'amélioration de la perception des sons de l'environnement entre avant et après la chirurgie. D'autres auteurs ont confirmé ce résultat dans une étude plus récente (Inverso & Limb, 2010).

De façon générale ont peu dire que la perception et surtout l'identification de sons de l'environnement est déficitaire bien que très mal décrite chez les patients implantés cochléaires. En 2008, Shafiro montrait chez des normo-entendant en condition de simulation d'implant que la perception des sons de l'environnement pouvait être améliorée avec l'augmentation de la résolution spectrale (Shafiro, 2008).

5.4.4 LA MUSIQUE

En comparaison de la perception des informations paralinguistiques et des sons de l'environnement, la musique est un domaine assez largement étudié chez les utilisateurs d'implants cochléaires. Les diverses études se sont penchées sur plusieurs domaine musicale, comme le timbre, le rythme ou encore la mélodie.

En ce qui concerne la perception du rythme, l'habileté perceptives des patients implantés cochléaires été étudié par plusieurs études. Dans l'ensemble, les auteurs rapportent que les patient implantés sont capables de percevoir le rythme aussi bien ou presque que les patients normo-entendants (Gfeller & Lansing, 1991; Gfeller, Woodworth, Robin, Witt, & Knutson, 1997; Klajman, Koldej, & Kowalska, 1982; Kong, Cruz, Jones, & Zeng, 2004; Leal et al., 2003; Looi, McDermott, McKay, & Hickson, 2008; Szelag, Kolodziejczyk, Kanabus, Szuchnik, & Senderski, 2004).

En revanche les études qui se sont intéressées à la perception du pitch, ont toutes révélé dans l'ensemble, un déficit perceptif chez les patients implanté cochléaire. Ce déficit a été retrouvé pour des études s'intéressant aussi bien à la reconnaissance de mélodies familières (Gfeller, Turner et al., 2002; Kong et al., 2004; Leal et al., 2003; Looi et al., 2008) qu'à la reconnaissance ou discrimination de contour de pitches (M.F. Dorman, Basham, McCandless, & Dove, 1991; Gfeller & Lansing, 1991; Leal et al., 2003; Looi et al., 2008).

Enfin, les études sur la perception du timbre rapportent également un déficit perceptif (Gfeller, Christ, Knutson, Witt, & Mehr, 2003; Gfeller & Lansing, 1991; Gfeller, Witt, Woodworth, Mehr, & Knutson, 2002; Leal et al., 2003; Looi et al., 2008; M.W. Skinner, Clark, & Whitford,

1994). Les confusions de timbre était particulièrement prononcé entre les instruments appartenant à la même famille (percussion ou non-percussion).

De façon générale, on peut dire qu'en matière de perception de la musique, les patients implantés cochléaires présentent des déficits par rapport aux normo-entendants. Ce déficit est particulièrement marqué en ce qui concerne les informations spectrales (pitch et timbre), tandis que le rythme est assez bien perçu (information temporelle).

Notons que dans l'ensemble, la rééducation, de même que les stratégies de codage, est axée sur la compréhension du langage. Or, comme cela a déjà été évoqué, l'environnement sonore est infiniment plus riche que le simple message linguistique. Une telle rééducation joue certainement un rôle primordial dans la bonne récupération de la compréhension du langage. Qu'en serait-il si un entrainement des patients était aussi réalisé pour les autres informations acoustiques ?

6 REORGANISATION

Nous savons que la privation sensorielle de l'ouïe a des conséquences comportementales, comme l'absence de communication verbale ou encore l'apprentissage d'un autre mode de communication tel que la langue des signes. Cette privation sensorielle a également des conséquences sociales, telles que l'isolement, et psychologiques. De même, nous savons que l'implantation cochléaire a un impact extrêmement positif, non seulement sur la qualité de vie des patients, avec le retour à la vie sociale et un plus faible niveau de dépendance à leurs proches, mais aussi en termes de récupération de la fonction auditive. Nous allons maintenant aborder les phénomènes neuropsychologiques sous-jacents à ces deux phases que sont la déprivation auditive et l'implantation cochléaire.

6.1 EFFET DE LA DEPRIVATION AUDITIVE SUR L'ORGANISATION NEUROFONCTIONNELLE

La déprivation auditive a notamment été décrite pour ses effets sur les traitements visuels et somesthésique, au travers de la littérature scientifique.

Dans le domaine visuel, de façon générale, les patients ne semblent pas avoir de meilleures performances pour les fonctions visuelles de base, telles que la discrimination de la luminosité (Bross, 1979), le seuil temporel de discrimination visuelle (Bross & Sauerwein, 1980; Mills, 1985; Poizner & Tallal, 1987), la sensibilité au contraste (Finney, Fine, & Dobkins, 2001; C. Stevens & Neville, 2006), la sensibilité à la direction du mouvement (Bosworth & Dobkins, 1999, 2002a, 2002b) ou à la vitesse du mouvement (Brozinsky & Bavelier, 2004).

Les seuls changements que l'on peut observer pour la fonction visuelle concernent les sourds signeurs (i.e. utilisation du langage de signes). Ces changements n'ont pas été décrits comme des différences perceptives à proprement parler, mais davantage comme une modification de la répartition spatiale de l'attention visuelle (Bavelier, Dye, & Hauser, 2006). En effet plusieurs études ont décrit une amélioration perceptive dans le champ périphérique, comme dans le cas de la détection de stimuli périphériques (Loke & Song, 1991), la détection du mouvement périphérique (Bavelier et al., 2001; Bavelier & Neville, 2002; Neville & Lawson, 1987), l'orientation (Bosworth

& Dobkins, 2002a; Parasnis & Samar, 1985) et le traitement de distracteurs périphériques (Bosworth & Dobkins, 2002a; Proksch & Bavelier, 2002; Sladen, Tharpe, Ashmead, Wesley Grantham, & Chun, 2005). D'autres études ont fait remarquer que cette amélioration des performances perceptives des sourds signeurs dans leur champ visuel périphérique serait accompagnée d'une diminution de certains aspects de la perception visuelle dans le champ central, comme la capacité d'attention soutenue (Parasnis, Samar, & Berent, 2003) ou le traitement de distracteurs centraux (Proksch & Bavelier, 2002). Cependant la plupart des autres fonctions visuelles en champ central restent comparables aux personnes normo-entendantes, en particulier la détection de stimuli centraux (Loke & Song, 1991), la recherche visuelle (Rettenbach, Diller, & Sireteanu, 1999; Stivalet, Moreno, Richard, Barraud, & Raphel, 1998) et l'orientation (Parasnis, 1992; Parasnis & Samar, 1985).

Enfin, dans le domaine somesthésique, on peut également observer des capacités de détection supérieures à celles des normo-entendants chez les personnes sourdes pour les stimuli tactiles déviants (Levanen & Hamdorf, 2001).

D'un point de vue neurofonctionnel, les modifications fonctionnelles des systèmes visuel et somesthésique induites par la déprivation auditive peuvent être sous-tendues par différents mécanismes de réorganisation corticale.

Un premier mécanisme se situe au niveau des aires associatives ; ces-dernières ne recevant plus d'information auditive peuvent se réorganiser en devenant plus sensible aux modalités sensorielles préservées, comme la vision ou le toucher. Dans ce sens Bavelier *et al.* en 2001 ont montré des modifications dans les aires pariétales multisensorielles des personnes sourdes utilisant le langage des signes (Bavelier et al., 2001).

Un deuxième mécanisme est basé sur l'existence de projections nerveuses des aires auditives vers les aires visuelles de bas niveau, incluant les aires visuelles primaires. Ces connexions pourraient être à l'origine d'une réorganisation intramodale après déprivation auditive. On peut donc imaginer qu'en l'absence d'informations auditives afférentes, les aires visuelles deviennent plus disponibles pour les autres modulations intramodales, en particulier au niveau attentionnel, comme le montrent les données comportementales disponibles chez les sujets humains sourds signeurs (Bavelier et al., 2006).

Enfin, le troisième et dernier mécanisme permettant une réorganisation corticale repose sur la capacité du cortex auditif à se réorganiser afin d'assurer d'autres fonctions sensorielles comme la vision. Cette hypothèse est soutenue par plusieurs études, montrant que chez des patients sourds il existait, par rapport à des normo-entendants, une suractivation des aires auditives de bas niveau du STS, par des stimuli visuels, signés et tactiles (Bavelier et al., 2001; Fine, Finney, Boynton, & Dobkins, 2005; Finney, Clementz, Hickok, & Dobkins, 2003; Levanen, Jousmaki, & Hari, 1998; Neville et al., 1998; Nishimura et al., 1999; Petitto et al., 2000). Cette réorganisation pourrait relever d'un mécanisme de compétition similaire à celui observé dans les aires associatives. En effet, un nombre croissant d'études met en évidence le traitement d'informations non-auditives par la partie caudale du cortex auditif temporal, en particulier des informations tactiles et différentes informations

visuelles comme les mouvements biologiques et les visages (Allison, Puce, & McCarthy, 2000; Brosch, Selezneva, & Scheich, 2005; Schroeder et al., 2003). Cette hypothèse est d'autant plus cohérente que plusieurs études démontrent une corrélation positive entre l'étendue de la plasticité intermodale dans les aires auditives et le degré de perte auditive (Lambertz, Gizewski, de Greiff, & Forsting, 2005) ainsi que la durée de la surdité (Sadato et al., 2004).

6.2 EFFET DE L'IMPLANTATION COCHLEAIRE SUR L'ORGANISATION NEUROFONCTIONNELLE

Comme cela a déjà été évoqué, le message nerveux généré par l'implant cochléaire est très dégradé et appauvri par rapport au signal originel. Le cortex va donc devoir s'adapter à ces nouvelles afférences sensorielles, par l'intermédiaire des phénomènes naturels d'apprentissage et de réorganisation corticale. De par le déséquilibre survenu entre les afflux nerveux en provenance des différentes modalités sensorielles, de nouvelles associations fonctionnelles et neuronales vont se créer progressivement. Il va donc y avoir des recrutements des aires cérébrales différents d'une répartition normale. C'est cette différence que nous nous proposons de détailler dans ce chapitre, qui sert de préambule à l'exposé de nos travaux correspondants.

6.2.1 REACTIVATION DES VOIES AUDITIVES CENTRALES

Nous avons vu que la surdité modifiait considérablement l'organisation neurofonctionnelle, et ce, plus ou moins selon la durée de déprivation auditive. L'implantation cochléaire va permettre de réactiver les entrées auditives parvenant habituellement aux aires auditives. La question qui se pose est que devient le nouveau message reçu par le nerf auditif ? Dans les années 90, plusieurs études en neuroimagerie fonctionnelle ont permis de répondre à cette question. Par exemple, Naito *et al.* en 1995, ont montré chez 6 patients sourds postlinguaux implantés cochléaires que des stimuli auditifs permettaient une réactivation des centres nerveux auditifs primaires et secondaires bilatéraux, avec une activation préférentielle du côté controlatéral à l'oreille implantée. D'autres études ont retrouvé des résultats allant dans le même sens (Herzog et al., 1991; Ito, Sakakibara, Iwasaki, & Yonekura, 1993; Naito et al., 1995; Okazawa et al., 1996; Truy et al., 1995).

Dans le domaine de la perception de la parole, plusieurs études ont également retrouvé chez les sourds postlinguaux, en réponse à des stimuli langagiers, une activation de l'activité dans les aires auditives associatives et dans les aires frontales inférieures postérieures bilatérales, avec une prédominance de la suractivation dans l'hémisphère gauche (Naito et al., 1995; Okazawa et al., 1996). Plusieurs études ont suggéré une corrélation entre le degré de métabolisme dans les aires spécifiques au traitement du langage (notamment les aires de Broca et de Wernicke) et les

performances de compréhension de la parole chez les sujets sourds implantés cochléaires (Fujiki et al., 1999; Mortensen, Mirz, & Gjedde, 2006; Okazawa et al., 1996).

En ce qui concerne le traitement des informations paralinguistiques, très peu d'études se sont intéressées au sujet. En 2008, Coez *et al.* ont étudié l'activité de l'aire de la voix chez 12 patients sourds implantés cochléaires (Coez et al., 2008). Pour les patients ayant un bon niveau de compréhension (> 80%), une activité de l'aire de la voix a été retrouvée. Ceci indique donc que l'implantation cochléaires permet la réactivation du réseau impliqué dans la perception des informations paralinguistiques. Mais ce résultat demande à être confirmé par d'autres études.

6.2.2 REORGANISATION DES VOIX AUDITIVES

Si les voies auditives centrales semblent réactivées par le message auditif transmis par l'implant, on peut tout de même se demander si cette réactivation est semblable à celle que l'on observe chez le sujet normo-entendant, et si différence il y a, cela est-il dû à la nature dégradée du signal auditif transmis par l'implant ou la réorganisation consécutive à la surdité ?

En ce qui concerne le traitement de la parole, les recherches les plus récentes convergent pour indiquer que pour des niveaux comparables de performances comportementales, les sujets implantés n'utilisent pas les mêmes réseaux corticaux que les sujets normo-entendants. Tout d'abord quelques études ont rapporté des modifications fonctionnelles de certaines aires comme par exemple l'aire de Wernicke. Cette région, classiquement impliquée dans le traitement phonologique de la parole chez le sujet normo-entendant, subit après l'implantation cochléaire une forme de déspécialisation et intervient dans le traitement de toutes les catégories sonores susceptibles d'avoir un sens, comme par exemple les sons environnementaux (Giraud, Price, Graham, & Frackowiak, 2001). Egalement, des études ont rapporté des différences des niveaux d'activations corticales (augmentation ou diminution par rapport aux sujets normo-entendants) lors du traitement de la parole et du langage. Les principales sous-activations observées chez les sujets implantés se produisent dans les aires temporales ventrales et dans les régions temporo-pariétales (gyrus angulaire), consacrées à l'analyse audiovisuelle et phonologique de la parole (Giraud, Price, Graham, & Frackowiak, 2001). D'avantages d'études ont mis en avant des sur-activations chez les patients implantés cochléaires. Ces sur-activation, ont quant à elles été observées dans des régions impliquées dans les processus attentionnels et mnésiques, notamment les régions pariétales supérieures et le gyrus parahippocampique (Giraud et al., 2000), mais également dans l'aire de Broca, la partie postérieure du gyrus temporal supérieur droit, l'aire prémotrice droite, ou encore des régions impliquées dans la mémoire comme l'hippocampe gauche (Giraud, Price, Graham, Truy, & Frackowiak, 2001; Giraud & Truy, 2002; Giraud et al., 2000). De manière générale, cette réorganisation a été interprétée par les auteurs comme le témoin des stratégies développées par les patients pour optimiser leur compréhension du langage et compenser la perception auditive dégradée. Les patients compenseraient donc leur difficultés de traitement acoustique et phonologique par des stratégies

cognitives de haut niveau, notamment grâce au contexte sémantique et syntaxique encodé en mémoire et aux ressources attentionnelles.

Au niveau des interactions multimodales, on observe chez les patients implantés cochléaires, des connexions intermodalitaires plus fortes que chez les sujets normo-entendants (Giraud, Price, Graham, Truy et al., 2001; Zatorre, 2001). Ces connexions intermodalitaires, sont probablement la conséquence de la surdité ; nous avons vu que lors d'une déprivation auditive, les aires habituellement consacrées au traitement auditif sont impliquées dans d'autres types de tâches dépendantes d'autres modalités (voir « Effet de la déprivation auditive sur l'organisation neurofonctionnelle », p135). Mais ces connexions sont probablement aussi le support des stratégies compensatrices adoptées par les patients.

Dans le domaine du traitement des informations paralinguistiques, peu d'études se sont intéressées à la réorganisation fonctionnelle du réseau cortical impliqué dans la perception des informations paralinguistiques, après implantations. L'étude déjà citée de Coez *et al.* en 2008 a comparé l'activité de l'aire de la voix, mise en évidence par Pascal Belin (voir « La perception de la voix », p83), entre 2 groupes de patients et des normo-entendants (voir Figure 6.1)

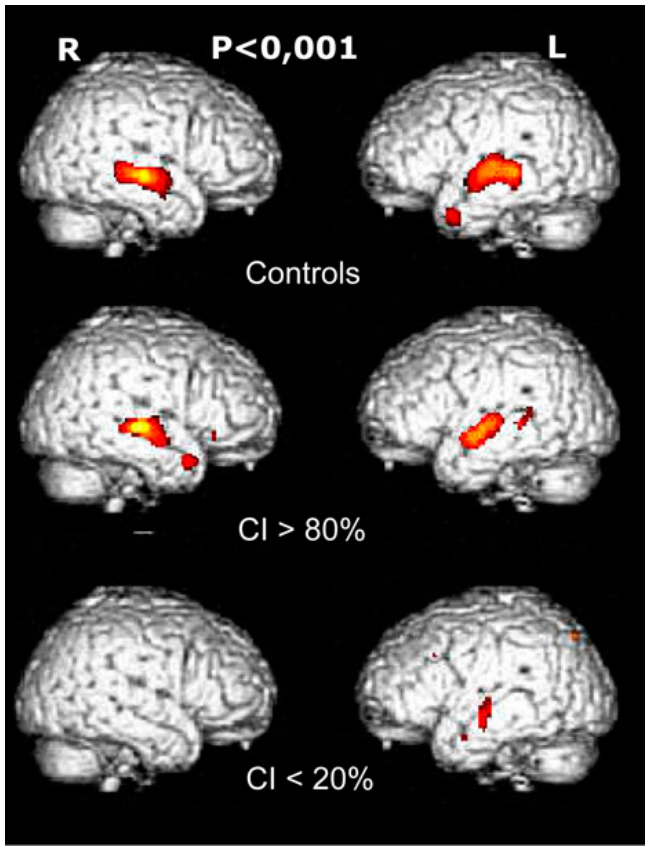


Figure 6.1 : Résultats du contraste entre l'activation lors de la présentation des stimuli « voix » et l'activation lors de la présentation des stimuli « non voix ». Lors de ce contraste, l'activité de l'aire de la voix était retrouvée pour le groupe de normo-entendant (« Controls ») et le groupe de patients ayant un bon niveau de compréhension du langage (« CI < 80%), mais pas pour le groupe de patients ayant un faible niveau de compréhension du langage (« CI < 20%), Adapté de Coez *et al.* (2008)

Les patients étaient séparés en deux groupes : ceux qui avaient un bon niveau (> 80%) de compréhension de la parole (N = 6) et ceux qui avaient un faible (< 20%) niveau de compréhension de la parole. La tâche dans le scan était une écoute passive de stimuli vocaux et de stimuli non vocaux. Le contraste d'activité cérébrale en réponse à ces deux groupes de stimuli a montré qu'en comparaison des normo-entendants, les patients ayant une bonne compréhension du langage avaient un niveau d'activité de l'aire de la voix similaire. En revanche aucune activité significative n'était retrouvée chez les patients ayant une faible compréhension du langage. Cette étude suggère donc que la réactivation de l'aire sensible à la voix humaine, soit liée à la compréhension du langage. Toutefois aucune information concernant les performances de détection de la voix n'étant données, on peut se demander si l'activité de cette aire de la voix chez les patients reflète bien les performances en matière de perception de la voix, ou bien est simplement impliqué à un traitement de la parole.

En dehors de cette étude, nous n'avons pas retrouvé d'autres travaux portant sur la réorganisation des voix auditives impliquées dans le traitement des informations paralinguistiques suite à l'implantation cochléaire. Il semble donc que cet aspect de l'audition et des réorganisations corticales qui y sont liées soit très peu étudiées dans la littérature scientifique, bien qu'elle suscite de plus en plus l'intérêt des patients et, par conséquent des fabriquant d'implants cochléaires.

CHAPITRE 3 : ÉTUDE EXPÉRIMENTALE DE LA PERCEPTION DE LA VOIX APRES IMPLANTATION COCHLÉAIRE

7 PROBLEMATIQUE

La voix est un outil de communication très puissant au delà-même de ses aspects langagiers, tant elle est un stimulus sonore particulier. Comme nous l'avons décrit précédemment, le langage, sa perception, sa reconnaissance et sa compréhension ont été très largement étudiés. Aujourd'hui, les questions qui se posent en matière de perception de la voix chez les sujets sourds implantés cochléaires sont de plus en plus nombreuses. Plusieurs limites ont déjà été mises en avant, tant du point de vue de la technique de l'implant que des capacités corticales des patients implantés. Par exemple, pour la plupart d'entre elles, les stratégies de codage du signal sonore des processeurs d'implant, se concentre donc essentiellement sur les informations essentielles à la compréhension du langage. Comme nous l'avons vu, en matière de compréhension de la parole, la récupération chez les patients sourd implantés cochléaire est retrouvé au moins partiellement et souvent de façon très satisfaisante dans la majorité des cas (proportion variables selon les études). Mais si ces stratégies se concentrent sur le langage, il est évident qu'il n'est pas la seule information auditive de l'environnement. D'une part la voix humaine contient, beaucoup d'autres informations, telles que l'identité du locuteur ou son état émotionnel, et d'autre part il existe il existe une forte demande des patients en termes d'amélioration de leur perception des informations paralinguistiques, des bruits de l'environnement ou de la musique. Or, les informations acoustiques caractéristiques de chacun de ces stimuli auditifs ne sont pas les mêmes. La question se pose donc de savoir si en privilégiant la parole au travers des stratégies d'implant ou de la rééducation post-implantation, on ne limite pas la perception des autres informations auditives. Enfin, comme nous avons pu le voir, très peu d'étude se sont intéressées de près à la récupération de la perception auditive des informations paralinguistique lors de l'implantation cochléaires.

L'objectif de cette thèse est donc de faire un état des lieux de la perception de la voix chez le sujet sourd implanté cochléaire et son évolution au cours du temps, ainsi que l'évaluation des effets liés au processeur et ceux liés à des facteurs neurofonctionnels sur cette perception. Pour cela, nous avions comme projet de départ d'explorer la fonctionnalité de l'aire de la voix chez les patients implantés cochléaires, en utilisant un protocole TEP. Nous souhaitions ainsi évaluer l'évolution de cette fonctionnalité au cours du temps et la corréler avec les performances comportementales. Cependant, pour des raisons techniques, cette partie du projet a due être laissée de côté au cours de ce travail de thèse. Pour répondre à notre objectif de départ, et afin de compenser l'absence de ces données de neuroimagerie, nous avons donc eu recours à l'IRM chez des sujets normo-entendants en condition de simulation d'implant. Ce travail a donc été structuré en 3 parties :

1) Une première partie de l'étude a été menée en **psychophysique chez des sujets sains**, espérant ainsi mieux comprendre les aspects psychophysiques et neurofonctionnels qui les caractérisent, et en faire une population contrôle pour les patients. Les sujets normo-entendants ont également été placés en conditions de simulation d'implant, ce qui en fait un modèle d'étude intéressant de l'implantation cochléaire en termes de perception auditive.

2) Ensuite, une autre partie, toujours en **psychophysique**, a été conduite chez des **patients sourds implantés cochléaires** afin de rendre compte de leurs performances dans ce domaine et de comparer leurs données aux sujets contrôles. Les implantés ont été abordées de deux façons différentes. Une partie des patients à été suivie sur plusieurs mois à partir de l'activation de leur implant et l'autre partie n'a été vue qu'une seule fois à des délais d'implantation variables.

3) Enfin, la troisième et dernière partie de ce travail était une **étude IRMf** chez les normo-entendants en condition de simulation d'implant, ayant pour but d'observer l'effet de la simulation d'implant (ou vocoder) sur l'aire de la voix mise en évidence par Pascal Belin (Belin et al., 2000).

A la fin de chacune de ces parties, les résultats seront discutés. Puis une discussion générale en fin d'ouvrage essayera de synthétiser l'ensemble des résultats obtenus et de discuter leur interprétation et implication dans le domaine de la perception de la voix chez les sujets sourds implantés cochléaires.

8 METHODES

Dans cette partie nous décrivons le matériel et les méthodes communs aux tests utilisés pour les études comportementales, soit les deux premières parties de ce travail. Nous décrivons notamment la méthode de simulation d'implant et les analyses statistiques.

8.1 TRAITEMENT DES STIMULI COMPORTEMENTAUX

8.1.1 GENERATION DE STIMULI

Pour les tests auxquels nous avons procédés, l'ensemble des stimuli originaux qui ont été utilisés pour les tests de détection de la voix et le test de catégorisation du genre provenait de la base de données de Pascal BELIN. Tous ces sons avaient les mêmes caractéristiques acoustiques (mono, 16-bits, 22050 Hz d'échantillonnage) et avaient été égalisés en amplitudes.

Les stimuli utilisés dans le test de reconnaissance de l'âge ont été enregistrés par nos soins, au sein des populations concernées.

Les détails de tous ces stimuli seront donnés plus bas, avec la description du test les concernant.

8.1.2 SIMULATION ACOUSTIQUE DE L'IMPLANT COCHLEAIRE OU VOCODER

Les méthodes numériques décrites ici sont celles que nous avons utilisées pour réaliser les simulations acoustiques d'implant cochléaire employées dans nos études chez le sujet normo-entendant. Ces méthodes sont celles développées et utilisées par Julien Rouger durant sa thèse au sein du laboratoire du CerCo (Rouger, 2007). Ces simulations sont souvent désignées dans la littérature par les appellations vocoder (Dudley, 1939a, 1939b) ou SSSN, pour *speech-spectrum shaped noise* (Shannon et al., 1995).

L'implant cochléaire contemporain est constitué de plusieurs électrodes, dont chacune va traiter une bande spectrale acoustique spécifique. L'étendue de celle-ci va dépendre essentiellement du nombre d'électrodes de l'implant, des réglages du processeur, du positionnement de l'électrode le

long de la rampe tympanique. Dans le cas de notre simulation, les fréquences de coupure qui délimitent les différents canaux fréquentiels ont été calculé grâce à la carte tonotopique de Greenwood (Greenwood, 1990). Celle-ci propose une relation sous forme de formules numériques entre la position sur la cochlée et la fréquence codée. Ensuite les fréquences extrémales choisies étaient 125 à 8000 Hz, ce qui correspond à ce que l'on peut retrouver dans le processeur des implants cochléaires Ainsi, les fréquences de coupures obtenues pour les différentes conditions de vocoder (64, 32, 16, 8, 4 et 2 canaux) sont présentées en annexes (voir Tableau 14.1, p338). La décomposition spectrale consistait à procéder au sein de chaque bande de fréquence découpée dans le spectre, à une uniformisation de l'énergie spectrale, c'est-à-dire à une moyenne pour toutes les fréquences appartenant à ce canal fréquentiel. Cela revient à créer un bruit blanc dans chaque bande spectrale en gardant la modulation temporelle du son originel. Ensuite, les bruits modulés par les enveloppes de chaque bande sont recombinés additivement. Enfin, un réajustement du niveau sonore moyen est effectué pour assurer l'équivalence des intensités sonores entre le son original et le son filtré. Les détails techniques de cette simulation sont disponibles dans la thèse de Julien Rouger (Rouger, 2007).

En clinique, le nombre d'électrode maximum d'un implant cochléaire est de 22 électrodes. Ce nombre est actuellement limité pour des raisons techniques et physiologiques. En raison de cette limitation les fabricants d'implant cochléaire cherchent davantage à optimiser la transmission du signal sonore au travers des stratégies de codages. La question se pose alors du choix de nos conditions expérimentales en ce qui concerne la simulation d'implant : pourquoi choisir des conditions de vocoder ayant un nombre de canaux supérieur à celui techniquement possible des implants cochléaires (i.e. 32 et 64 canaux) ? L'idée de la prise en compte de telles conditions est d'étudier l'effet de l'amélioration de la résolution spectrale sur les compétences des sujets. En effet, on sait que des raisons techniques et physiologique limite le nombre d'électrodes à 22 sur un porte-électrode d'implant cochléaire, or la question se pose de savoir si en augmentant le nombre d'électrodes dans la cochlée, autrement dit en améliorant la résolution spectrale, on augmenterait significativement la perception du signal auditif ? Les conditions expérimentales 32 et 64 canaux de vocoder vont nous permettre d'apporter une réponse à cette question.

8.2 ANALYSES

8.2.1 RAPPELS SUR LA THEORIE DE DETECTION DU SIGNAL

Dans notre étude, plutôt que d'analyser classiquement les performances correctes, nous avons décidé d'analyser les données psychophysiques des sujets, lorsque celles-ci nous le permettaient, avec la Théorie de détection du signal (*Signal Detection Theory* en anglais ou SDT). Cette méthode

nous a permis de dissocier deux processus mis en œuvre dans les réponses comportementales des sujets, à savoir les processus décisionnels et les processus perceptifs. La méthode que nous avons utilisée est celle qui a été développée essentiellement par Tanner et Swets (Tanner & Swets, 1954).

Avant tout, il faut savoir que le paradigme qui est utilisé (voir Tableau 8.1) est bien particulier : à chaque essai, soit le stimulus est présent soit il ne l'est pas, et la tâche du sujet est de détecter sa présence. Il existe donc quatre types de réponses possibles : « Hit » (détection correcte), « Miss » (manqué), « False Alarm » (Fausse Alarme) et « Correct Rejection » (rejet correct).

	Réponses du sujet	
	« Oui »	« Non »
Signal présent	Taux de Détection Correcte (Hit Rate)	Taux de Manqué (Miss Rate)
Signal absent	Taux de Fausse alarmes (False Alarm Rate)	Taux de Rejet Correct (Correct Rejection Rate)

Tableau 8.1 : Paradigme expérimentale pour appliquer la théorie de détection du signal (SDT), avec les types de réponses possibles du sujet selon que le signal est présent ou absent

La proportion de détection correcte (Hit) d'un stimulus (surtout proche du seuil de sensibilité du sujet) dépend non seulement de la capacité discriminative du sujet, c'est à dire sa capacité à percevoir ou non le stimulus, mais aussi de facteurs non sensoriels qui sont les critères de décision de la présence ou non du stimulus. Grâce à la SDT, il nous est possible d'estimer la capacité discriminative (ou sensibilité) indépendamment du critère de décision du sujet (biais de réponse).

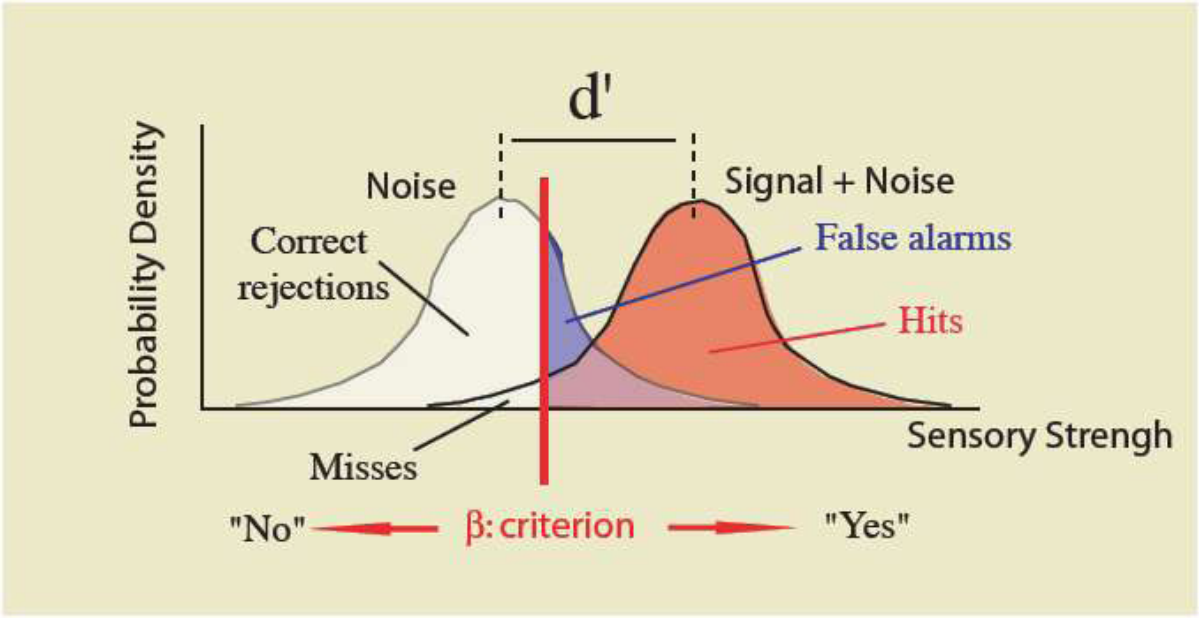


Figure 8.1 : Illustration de la représentation interne au sujet du « bruit » et du « signal + bruit » : l'intensité de la sensation du sujet correspond à la densité de probabilité. La sensibilité *d'*, est égale à la distance entre les deux moyennes des distributions « bruit » et « signal + bruit ». Le critère de décision, est représenté par un trait rouge entre les deux distributions et varie indépendamment de la sensibilité *d'*.

La Figure 8.1, illustre la représentation interne du sujet pour le « bruit » et un « signal + bruit ». En abscisse on peut lire l'intensité sensorielle perçue par le sujet pour chaque essai, qu'il y ait eu signal ou pas. Ce qui nous donne, en multipliant les essais, les distributions (représentées par les deux courbes en cloches) de l'intensité sensorielle perçue par le sujet en ordonnée, c'est-à-dire la densité de probabilité. Le bruit est conçu comme une variable aléatoire car il est attribué aux fluctuations aléatoires du système détecteur (récepteurs périphériques et structures de traitement de l'information) et à celles de l'environnement physique. Le signal ayant des caractéristiques précises et constantes, les deux courbes (« bruit » et « signal + bruit ») vont donc avoir une distribution normale. Etant donné qu'il y a plus d'information sensorielle lorsque le signal est présent, on comprend que les valeurs de la courbe « signal + bruit » (Signal + Noise) soit supérieure à celle du « bruit » (Noise). La sensibilité (*d'*) correspond aux capacités physiologiques du sujet à détecter un stimulus. Elle est calculée en fonction de la distance entre les deux courbes. Plus le signal est fort plus les deux courbes seront espacés et plus la sensibilité *d'* sera élevé. Un *d'* égal à 0, signifie un recouvrement complet des deux distributions, les moyennes étant ainsi les mêmes ; le sujet n'est alors pas capable de différencier le signal du bruit. Dans ce modèle, la zone de recouvrement des deux courbes est une situation ambiguë pour le sujet. Sa décision repose alors sur un critère de décision (ß) qu'il se serait donné (fig. 8), c'est à dire un état d'observation en-deçà duquel il attribuerait l'observation à Noise (il répondrait alors que le signal est absent), et au-delà duquel il l'attribuerait à SN (il répondrait que le signal est présent). Avec ce critère de décision on peut clairement déterminer sur les courbes les aires de probabilités correspondant aux différentes réponses possibles (hit, false alarm, correct rejection et miss).

Les valeurs de la sensibilité *d'* et du critère de décision ß sont calculés à partir des valeurs de détections correctes (Hit Rate) et de Fausses alarmes (False alarm Rate). Dans un premier temps ces valeurs sont normalisées, puis, en considérant le rapport des variances « bruit » et « signal sur bruit » égal à 1, on peut calculer la sensibilité et le critère de décision avec les équations suivantes :

$$d' = \mathbf{z(HR)} - \mathbf{z(FAR)}$$

$$\beta = e^{(-\frac{1}{2})[\mathbf{z^2(HR)} - \mathbf{z^2(FAR)}]}$$

où **z(HR)** est le z-score correspondant à la probabilité de *hits* dans la table de la loi normale et **z(FAR)** celui correspondant à la probabilité de *false alarm*.

On comprend avec ce modèle que les performances de reconnaissances correctes (Hit Rate) ne reflètent pas uniquement les capacités perceptives du sujet, mais sont aussi largement influencées par le critère de décision de celui-ci. En effet, lorsqu'on regarde la Figure 8.1, on voit clairement que les variations de *Hits* peuvent être indépendamment dues à des variations de la sensibilité ou à des variations du critère de décision, ou bien encore à des variations de ces deux aspects. Pour prendre un exemple extrême, imaginons un sujet qui répond toujours que le signal est présent (« oui »), sont

taux de *Hits* sera donc au maximum, 100%. On pourrait alors penser que ses performances sont très bonnes, or il est clair aussi que le taux de *False alarm* sera lui aussi au maximum. Dans ce cas on ne saurait dire si le sujet n'est pas capable de faire la différence entre « signal » et « signal + bruit », ayant donc des performances réelles plus proche du hasard, ou bien si son critère de décision est tellement permissif que dans le moindre doute il acceptait de répondre que le signal était présent. Appliquer la théorie de détection du signal permet donc de tenir compte de toutes les réponses du sujet et reflètent bien plus la sensibilité perceptive du sujet que les simples détections correctes, dans le cas d'un paradigme expérimental approprié. C'est pourquoi, nous nous attacherons à utiliser cette méthode d'analyse dans les tests où le paradigme le permet, c'est-à-dire dans les tests de détection de la voix humaine versus sons de l'environnement binaural et monaural, ainsi que dans le test de détection de voix humaine versus voix animale.

8.2.2 STATISTIQUES

L'ensemble de nos échantillons et/ou groupe de sujets étant de taille inférieure ou égale à 14, nous avons opté pour des analyses non paramétriques.

Pour comparer des groupes en mesure répétée, nous avons utilisé le test de Friedman. En test post-hoc et pour des comparaisons 2 à 2 nous avons utilisé le test de rang de Wilcoxon

Pour comparer des groupes lors de mesures indépendantes, le test de Kruskal-Wallis a été utilisé ainsi que le test de Mann-Whitney pour les analyses post-hoc 2 à 2.

Le seuil de significativité était fixé à $p = 0.05$.

L'article « *Voice detection in cochlear-implanted deaf subjects* », accepté sous révision dans **Hearing Recearch** se base sur les résultats obtenus pour le test de détection de la voix chez le sujet sain et le patient sourd implanté cochléaire (voir Publication à la fin du manuscrit). Toutefois, à la demande des reviewers, les analyses statistiques ont été faites avec des tests paramétriques. On pourra remarquer que, dans l'ensemble, les résultats ne changent pas beaucoup, et l'interprétation qui peut en être faite est largement similaire.

9 TESTS COMPORTEMENTAUX DE LA RECONNAISSANCE DE LA VOIX HUMANE CHEZ LE SUJET SAIN

En se basant sur le modèle de la perception de la voix de Pascal Belin, nous avons souhaité aborder plus particulièrement la perception de l'identité de la voix. Pour cela nous avons développé plusieurs tests. Une première série cherche à mesurer particulièrement la détection de la voix parmi des sons divers (sons de l'environnement ou voix animales). Puis nous avons souhaité mesurer les capacités perceptives des sujets en ce qui concerne le genre, ainsi que leur capacité à discriminer deux voix et l'âge du locuteur.

9.1 TEST DE DETECTION DE LA VOIX HUMAINE

Selon le modèle proposé par Pascal Belin que nous avons présenté dans partie théorique de ce manuscrit (voir partie « La voix, un visage auditif ? », p92), l'analyse structurale de la voix serait le module qui supporterait la détection de la voix (Belin et al., 2004). Elle serait donc le plus bas niveau d'analyse de ce modèle, celle qui permettrait d'accéder aux autres niveaux d'analyse, qui sont l'analyse du langage, du contenu affectif et de l'identité. Il nous a donc semblé indispensable pour commencer ce travail de thèse d'étudier le fonctionnement de ce module au travers de la simulation d'implant grâce au vocoder (voir « Simulation acoustique de l'implant cochléaire ou vocoder », p145).

Pour cela nous avons mis en place 3 tests, dont la tâche était de détecter de la voix humaine. Le premier a consisté à détecter la voix humaine parmi des bruits de l'environnement. Dans un second temps nous nous sommes posé la question de savoir si la latéralité, plus marquée à droite, de l'activité de l'aire de la voix mise en évidence par Pascal Belin avait un impact sur les performances, notamment au niveau des temps de réaction, selon que l'on stimulait l'oreille droite ou l'oreille gauche. Enfin nous avons voulu étudier les performances de détection de la voix parmi d'autres voix, des voix animales.

9.1.1 VOIX HUMAINE VERSUS SONS ENVIRONNEMENTAUX, STIMULATION BINAURALE

9.1.1.1 MATERIEL ET METHODES

SUJETS

Vingt-huit sujets normo-entendants (NHS), dont 14 hommes, ont été inclus dans cette étude, avec une moyenne d'âge de 23,8 ans (+/-2,4). Tous les sujets étaient de langue maternelle française, et aucun d'entre eux n'avait rapporté de problème auditif, neurologique ou psychiatrique. Le groupe de sujets a été divisé en deux groupes de 14 sujets (7 hommes). Le groupe A a passé le test avec les sons originaux, ainsi que dans 4 autres conditions de vocoder (64, 32, 16 et 8 canaux). Ce premier groupe de sujet passait également le test de reconnaissance « Voix humaine-Voix animale », décrit dans la partie le concernant (voir « Voix humaine versus voix animale » p167). Le groupe B a été testé en utilisant les conditions de vocoder à 16, 8, 4 et 2 bandes. Ces sujets, à l'inverse de ceux du groupe A, étaient naïfs vis-à-vis du protocole, n'avaient jamais été exposés à de tels stimuli.

STIMULI

Les 110 stimuli utilisés pour le test « Voix-Non Voix » provenaient tous de la base de donnée de Pascal Belin (Belin, Zatorre et al., 2002; Belin et al., 2000) et avaient une durée de 500 ms. Deux groupes de stimuli avaient été constitués. Le premier consistait en 55 sons « Non-Voix », issus de l'environnement et incluant des sons provenant de voitures (moteur, klaxons, etc.) de téléphones, de cloches, d'eau qui coule. Le deuxième groupe comprenait 55 sons de voix humaine qui avaient été enregistrés auprès d'une grande variété de locuteurs d'âges et de genres variés. Vingt-six de ces stimuli étaient des stimuli non langagiers, c'est-à-dire ne contenant pas de syllabes compréhensible ou identifiable (toux, bâillement, rire, etc.). Les vingt-neuf autres étaient des stimuli dits langagiers (e.g., des phonèmes présentés dans un contexte /h/voyelle/d/, des mots issus de langues étrangères, des syllabes), qui contenaient donc des syllabes compréhensibles mais n'avaient pas de contenu sémantique.

Une première analyse des stimuli a consisté à comparer le spectre de puissance des stimuli des groupes « Non-Voix » et « Voix » par bootstrap (voir Figure 9.1). Aucune différence significative n'a été révélée entre les deux groupes de stimuli. A priori la répartition des fréquences et leur amplitude était équivalente dans les deux groupes de stimuli. De même, la comparaison du spectre de puissance des stimuli des groupes vocaux langagiers et vocaux non langagiers par bootstrap n'a révélée aucune différence (voir Figure 9.2).

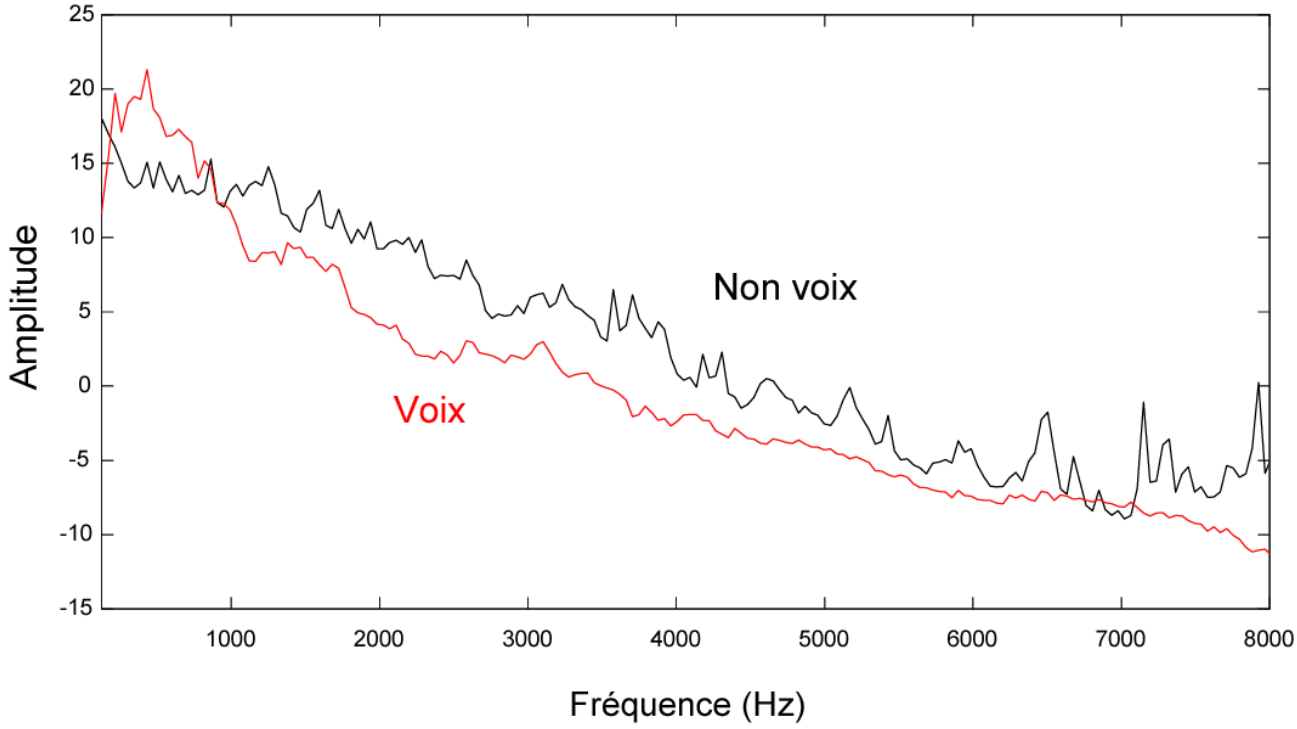


Figure 9.1 : Spectre de puissance moyen sur l'ensemble des stimuli "Non-Voix" (noir) et sur l'ensemble des stimuli "Voix" (rouge). Aucune différence significative n'a été retrouvée entre les deux groupes de stimuli.

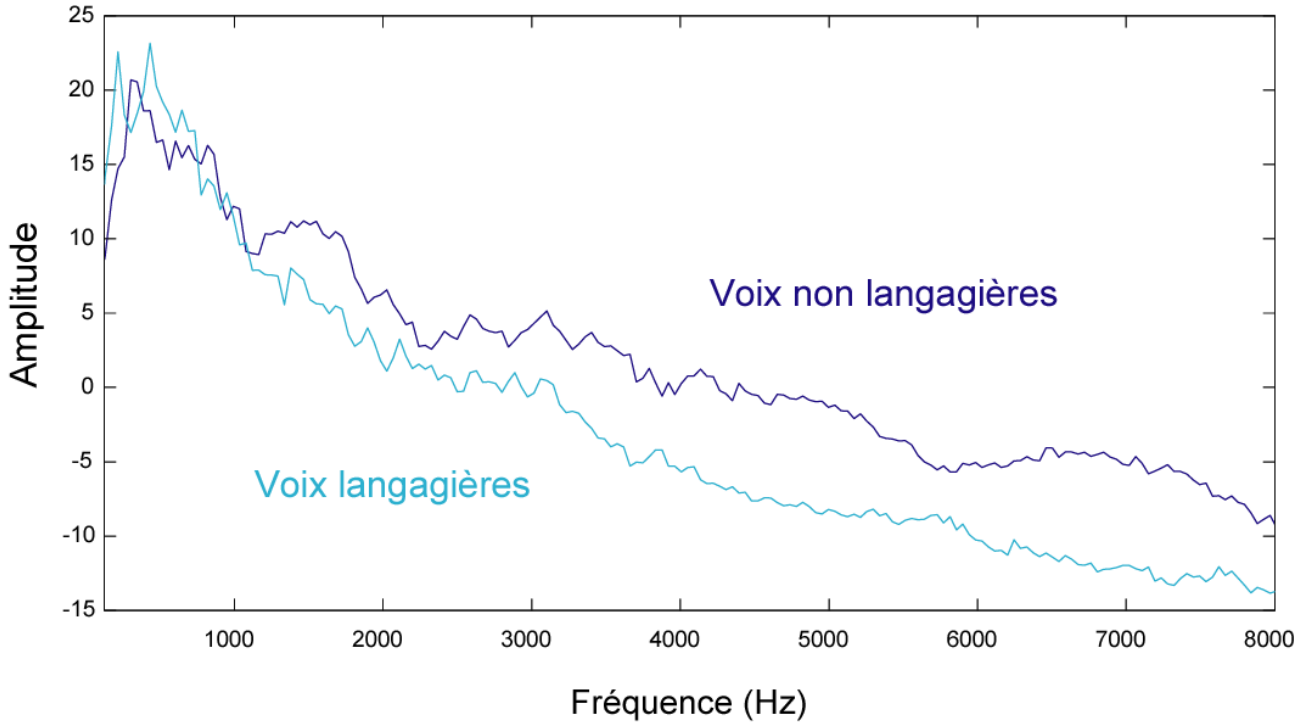


Figure 9.2 : Spectre de puissance moyen sur l'ensemble des stimuli vocaux langagiers (cyan) et sur l'ensemble des stimuli vocaux non langagiers (bleu foncé). Aucune différence significative n'a été retrouvée entre les deux groupes de stimuli.

PROCEDURE

Les deux groupes de sujets étaient testés dans une pièce calme du laboratoire du CerCo. Les stimuli étaient présentés binauralement au travers d'un casque (modèle Eh 250, Sennheiser), dans un ordre aléatoire à un volume sonore ajusté à 72 dB SPL. Le groupe A était testé avec les stimuli originaux ainsi que dans 4 autres conditions de vocoder : 64, 32, 16 et 8 canaux. Le groupe B était testé seulement en conditions vocodées à 16, 8, 4 et 2 canaux. Pour les deux groupes les conditions étaient présentées par block dans un ordre aléatoire. Toutefois pour le groupe A, la condition Voix Originale était toujours passée en dernière, ceci dans le but d'éviter un effet de « reconnaissance du stimuli ». Les sujets étaient placés devant un clavier avec le casque. La tâche donnée au sujet était un choix forcé à 2 alternatives (2AFC) : ils devaient dire si le son qu'ils avaient entendu était de la Voix humaine ou un son provenant de l'environnement avec pour consigne d'être le plus précis et le plus rapide possible dans leur réponse. Pour répondre, ils devaient appuyer sur une des deux touches « contrôle » correspondant à leur réponse. Le côté des réponses « Voix » et « Non Voix » sur le clavier étaient contrebalancée entre les sujets. La durée qui séparait la réponse des sujets et le début su stimulus suivant était d'1 seconde. Chaque bloc de condition durait approximativement 4 minutes, et les sujets étaient libres de faire une courte pause entre chaque condition. La présentation des stimuli et l'enregistrement des réponses se faisaient via le logiciel Matlab 6.5®.

ANALYSE

Pour analyser les réponses des sujets nous avons appliqué la théorie de détection du signal (SDT, (Green & Swets, 1966; Tanner & Swets, 1954)) afin de calculer l'indice de sensibilité *d'* indépendamment des biais liés aux critères de décision. Pour cela, les Hits (détection correcte) et les Fausses Alarmes (FA) étaient calculés en plaçant les stimuli « Voix » comme cible à détecter. Le *d'* a aussi été calculé selon que les stimuli « Voix » étaient « langagiers » ou « non langagiers ». Les analyses statistiques ont testé l'évolution des valeurs de *d'* entre les différentes conditions de vocoder. Ces analyses ont été faites sur les *d'* Voix ainsi que sur les *d'* « Voix langagière » et « Voix non langagière ». Nous avons également comparé les valeurs de *d'* entre les stimuli Voix langagière et Voix non langagière pour chaque condition de vocoder.

Les temps de réactions pour les Hits et les FA étant similaires, nous avons choisis de présenter les temps de réactions pour l'ensemble des essais, toujours en distinguant, Voix, Non-Voix, Voix langagière et Voix non langagière. Les analyses statistiques ont tout d'abord testé l'évolution des temps de réaction mesurés pour chaque groupe de stimuli en fonction de la dégradation du signal par le vocoder. Dans un deuxième temps, nous avons comparé les temps de réaction entre les stimuli Voix et Non-Voix ainsi qu'entre les Voix langagières et Voix non langagière.

Les tests précédents ont été réalisés au sein de chaque groupe de sujets, A et B. Des tests ont également été pratiqués pour comparer les performances du groupe A avec celles du groupe B, dans les mêmes conditions de vocoder.

9.1.1.2 RESULTATS

EFFET DU VOCODER

Lorsqu'on présentait les sons originaux, les sujets normo-entendants avaient de bonnes performances pour détecter les sons issus de la voix humaine parmi les sons issus de l'environnement, avec un *d'* de 4,3, soit un Hit de 98% (Voir Figure 9.3). Pour les groupe A et B, la diminution du nombre de canaux du vocoder diminuait significativement la valeur de *d'* ($p < 0,0001$). Les analyses post-hoc ont montré que les performances diminuaient dès 64 canaux pour le groupe A (Wilcoxon entre Voix Originale et 64 canaux : $p = 0,0036$). De plus, toutes les comparaisons 2 à 2 au sein de chaque groupe étaient significatives, excepté entre 64 et 32 canaux et entre 16 et 8 canaux, où les performances étaient statistiquement les mêmes. A partir de ces résultats nous pouvons dire que la dégradation du signal acoustique par la diminution du nombre de bande du vocoder augmente la difficulté de détection de la voix parmi des sons de l'environnement.

Lorsque l'on a comparé les deux groupes entre eux sur les conditions 16 et 8 canaux qu'ils avaient tous les deux passés, on retrouvait une différence significative pour ces deux conditions ($p < 0,01$). Autrement dit le groupe B avait des performances significativement moins bonnes que celles du groupe A pour les mêmes conditions de vocoder. Cette différence peut s'expliquer par le fait que, globalement, le groupe B avait accès à moins d'information acoustique que le groupe A. En effet, même si les conditions de vocoder étaient passées dans un ordre aléatoire, les conditions 8 et 16 canaux étaient les plus pauvres et donc les plus « difficiles » pour le groupe A alors qu'elles étaient les plus riches et les plus « faciles » pour le groupe B. On peut donc supposer que les sujets du groupe A, au moins pour ceux qui ne passaient pas 16 et 8 canaux en premiers pouvaient reconnaître, ou mieux détecter, les sons issus de la voix humaine. Rappelons qu'il s'agissait du même set de stimuli d'une condition à l'autre de vocoder, et que seul variait le nombre de bandes. Notre hypothèse semble se confirmer lorsqu'on regarde les données individuelles. Les 5 sujets du groupe A qui ont commencé par les conditions 8 ou 16 canaux, n'ont un *d'* moyen que de 2,28 pour 8 canaux et 2, 59 pour 16 canaux. Ces valeurs se rapprochent beaucoup plus de celles du groupe B.

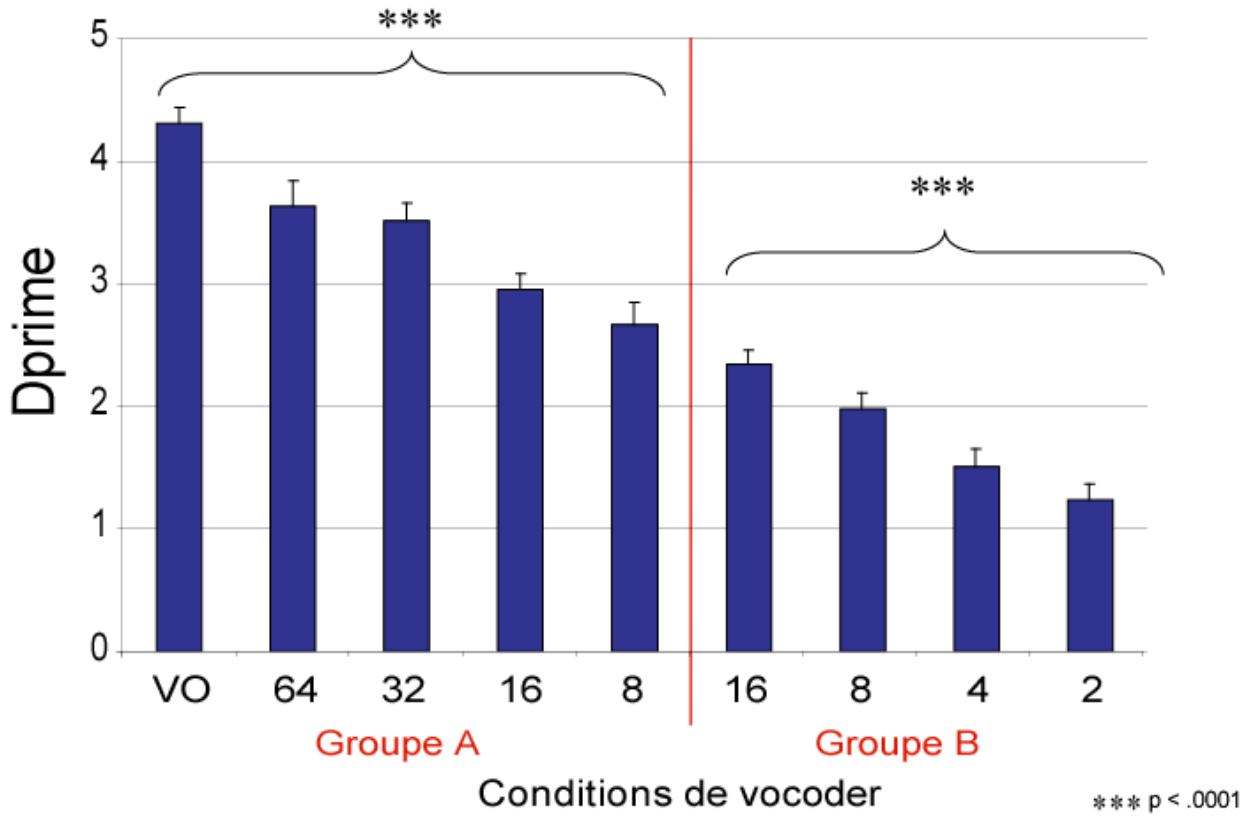


Figure 9.3 : Mesures de *d'* pour les groupes A (N = 14) en condition de vocoder 64, 32, 16 et 8 canaux ainsi qu'en condition Voix Originales (VO), et pour le groupe B (N = 14) en condition 16, 8, 4 et 2 canaux. L'effet du vocoder est significatif pour les deux groupes

COMPARAISON ENTRE STIMULI LANGAGIERS ET NON-LANGAGIERS

Nous avons analysé les performances de détection de la voix en distinguant les sons issus de la voix humaine, selon qu'ils étaient langagiers (mais non sémantiques) ou non langagiers (cf. ch. Stimuli p 152). On peut voir sur la Figure 9.4 que les résultats confirment la diminution des performances des sujets à détecter la voix parmi des sons issus de l'environnement à mesure que le nombre de bande du vocoder diminuait, et ce, que la voix soit langagière ou non ($p < 0{,}01$). Toutefois, le vocoder n'affectait pas les performances de détection de la voix langagières et non langagière de façon similaire. Pour commencer on peut remarquer que déjà en condition Voix Originales (VO), la détection de voix langagières est significativement meilleure que pour les voix non langagières, avec des *d'* de 4,48 et 4,13 respectivement (Hit 97,4% et 96,8%) et un $p = 0{,}0135$. Cette supériorité des capacités à détecter les voix langagières s'observait dans toutes les conditions de vocoder et pour les deux groupes. De plus, le taux de décroissance des performances était différent pour les deux groupes de voix. Pour les voix non langagières on observait une diminution forte et continue de 64 à 2 canaux de vocoder. Cette même pente de diminution ne s'observait pour

les stimuli langagiers, qu'à partir de 8 canaux ; les comparaisons 2 à 2 entre 64, 32 et 16 canaux étaient non significatives, bien que les *d'* de ces 3 conditions soient significativement plus faibles qu'en condition originale (VO).

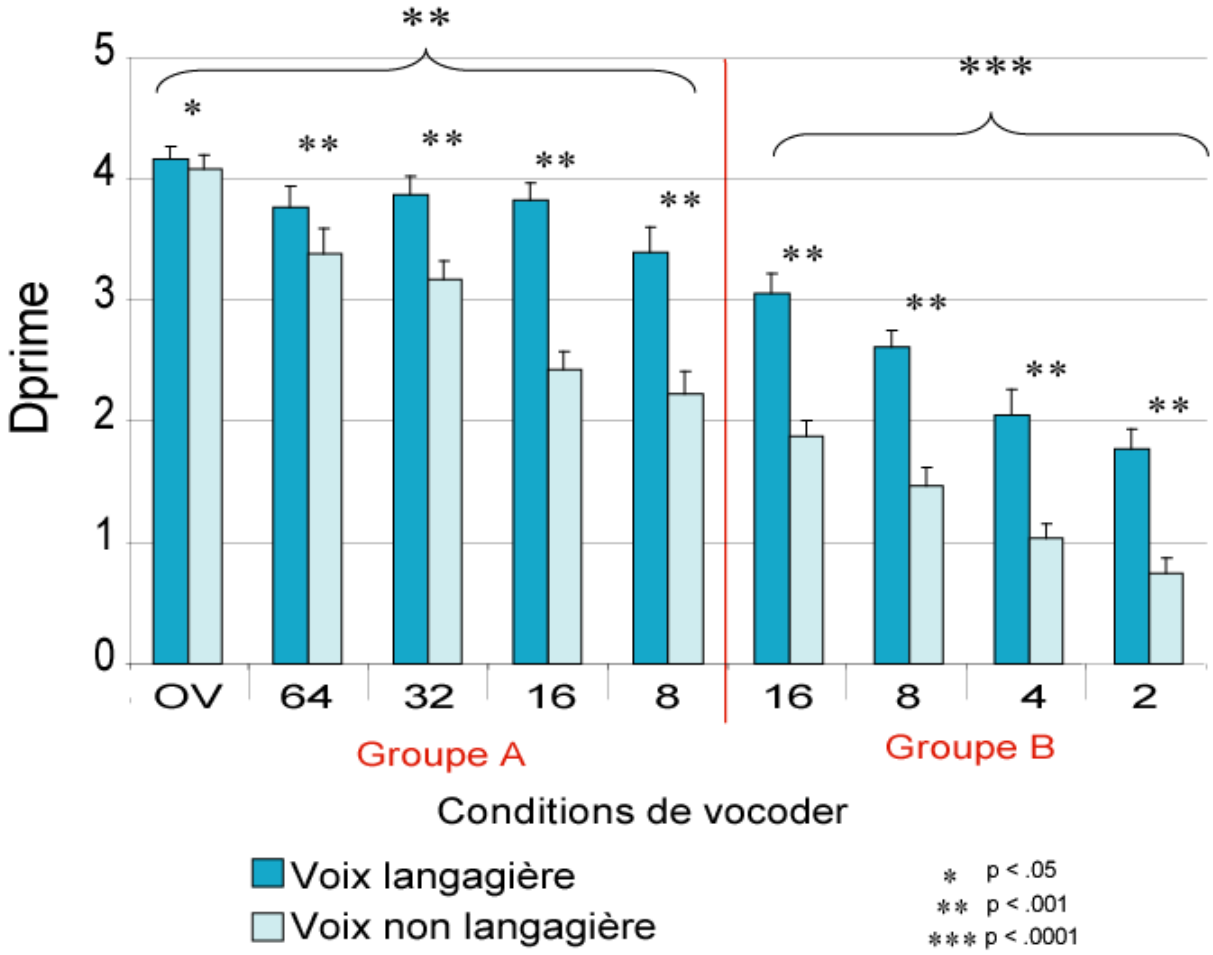


Figure 9.4 : Mesures de *d'* chez les normo-entendant, pour les groupes A (N = 14) en condition de vocoder 64, 32, 16 et 8 canaux ainsi qu'en condition Voix Originales (VO), et pour le groupe B (N = 14) en condition 16, 8, 4 et 2 canaux. L'effet du vocoder est significatif pour les deux groupes

L'ensemble de ces données prouvent que les sons vocaux non langagiers, tels que le bâillement, la toux, etc., étaient beaucoup plus affectés par la dégradation spectrale que les sons vocaux langagiers qui sont pourtant non sémantiques. Nous avons donc cherché à savoir quels paramètres acoustiques particuliers pouvaient expliquer ces résultats.

Dans un premier temps nous nous sommes intéressés au pitch de chacun de ces groupes de sons (voir Tableau 9.1). Le pitch moyen pour chaque stimulus a été mesuré avec le logiciel PRAAT. Pour le groupe de stimuli Non Voix, nous n'avons pas pu calculer de pitch pour 12 sons sur 55, car il

s'agissait de bruit tel que de l'eau qui coule, un moteur de voiture, etc., avec donc des fréquences aléatoires. Pour les autres groupes de stimuli, tous les pitch ont pu être calculés.

Types de Stimuli	Non Voix	Voix	
		Voix Langagière	Voix Non Langagière
Pitch moyen (Hz)	257,6	252,7	
		193,4	318,9

Tableau 9.1 : Pitch moyen pour les types de stimuli. Voix langagière et Voix non langagière sont des distinctions des stilui Voix. Seules les moyennes entre les stimuli Voix langagière et Voix non langagière sont significativement différentes ($p < 0,0001$)

On peut voir sur le Tableau 9.1 que les pitch des stimuli Voix et Non-Voix sont très similaires. Il semble donc évident que les sujets ne se basent pas sur cette différence de pitch pour réaliser la tâche. Toutefois lorsqu'on compare les pitch moyen des stimuli langagiers et non langagiers, une forte différence significative apparaît ($p < 0,0001$) ; les pitch des stimuli langagiers sont donc, de manière générale, plus bas que ceux des stimuli non langagiers. Cependant, il paraît peu probable que les sujets se soient basé sur une différence de pitch par rapport aux stimuli Non Voix pour mieux détecter les stimuli langagiers. En effet, on pourrait penser que cette différence explique la meilleure capacité des sujets à détecter des stimuli vocaux langagier, mais rappelons que la tâche était de détecter des stimuli vocaux parmi des stimuli non vocaux. Or, lorsqu'on compare les pitch des stimuli langagiers et non Langagiers avec les pitch des stimuli Non Voix, il n'apparait aucune différence significative. La différence entre les pitch Non Voix et Voix langagière est de 59,3 Hz et celle entre Non Voix et Voix non langagière est 61,3 Hz. Les Voix langagières et non langagières sont sensiblement à la même distance des stimuli Non Voix.

Dans un deuxième temps, nous nous sommes demandé si la meilleure perception pour les sons vocaux langagiers pouvait reposer sur une information temporelle plus riche. Nous avons réalisé une autre analyse des stimuli, qui a porté sur le nombre de « composants temporels » qui composaient les stimuli. Nous voulions savoir s'il y avait de meilleures performances lorsque la variation temporelle est plus riche. Nous avons donc distingué pour les sons vocaux langagiers le nombre de syllabes qui les composaient et pour les sons vocaux non langagiers le nombre d'unités temporelles élémentaires, c'est-à-dire le nombre d'élément que l'on distingue à l'écoute de la voix. Les stimuli ont donc été répartis en 3 groupes ; 1, 2 ou 3 syllabes ou unités temporelles élémentaires (voir Tableau 9.2). Il n'existait aucune différence de répartition entre les stimuli vocaux langagiers et non langagiers ($Chi^2 = 0,061$, $p > 0,05$). Il n'y a donc pas de raisons de penser que les sujets se basaient sur une information temporelle macroscopique plus riche pour les sons vocaux langagiers pour mieux les percevoir.

Type de stimuli / Nombre de syllabes/unités	Sons vocaux langagiers	Sons vocaux non langagiers
1	16	14
2	6	5
3	7	7

Tableau 9.2 : Répartition des différents stimuli vocaux langagiers et non langagiers en fonction du nombre de syllabes ou d'unités temporelles élémentaires.

Nous avons voulu vérifier si la richesse de ces informations temporelles macroscopiques améliorait la perception des stimuli et était plus robuste à l'effet du vocoder. Dans ce cas alors on s'attendait à ce que les stimuli les plus riches en syllabes et unités temporelle élémentaires furent les mieux reconnus. Nos analyses montrent que le vocodeur n'affecte pas plus les stimuli contenant une syllabe ou unité temporelle élémentaire que ceux en contenant plus (voir Figure 9.5). On peut donc dire que les sujets ne se basaient pas sur des informations temporelles macroscopiques, car on ne remarque pas d'effet du nombre de syllabe ou d'unité temporelle élémentaire sur la détection de voix. Cependant on sait que la structure temporelle fine est plus riche pour le langage que pour n'importe quel autre son, même issu de la voix. Il semble donc probable que les caractéristiques acoustiques permettant aux sujets de mieux détecter les sons vocaux langagiers que les non langagiers se situaient à une échelle beaucoup plus petite que celle que nous avions définis. Ces explications n'excluent pas la place du traitement cortical qui est très performant lorsqu'il s'agit de parole. Deux explications non exclusives peuvent donc être avancées pour expliquer la meilleure détection des voix langagières (non linguistiques) ; d'une part une plus grande richesse de la structure temporelle fine, autrement dit davantage de modulations temporelles, et d'autre part une meilleure capacité corticale de traitement des sons langagiers que pour tout autre son.

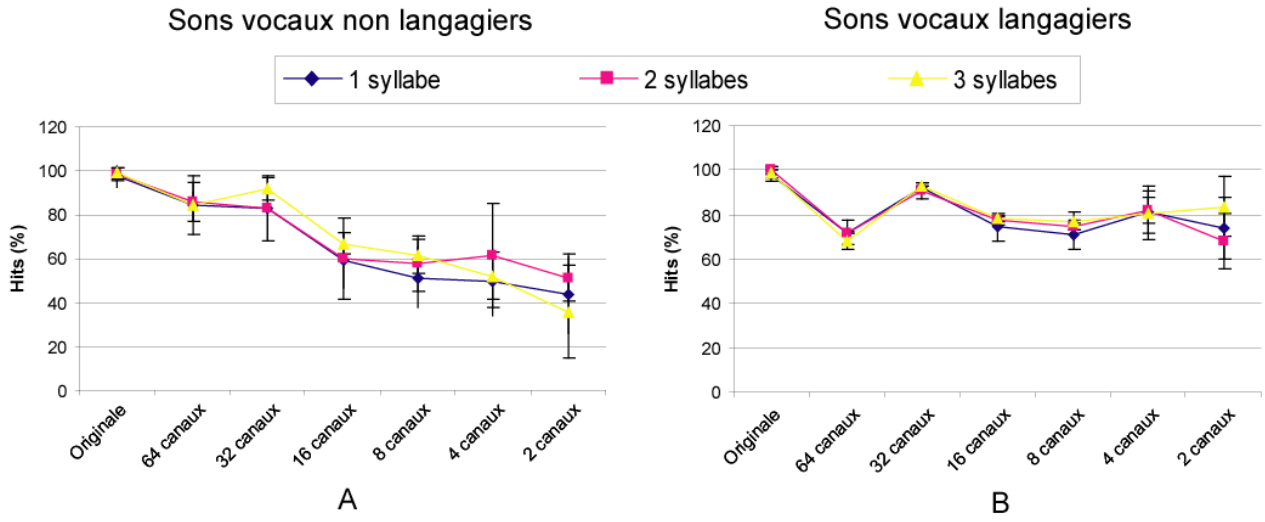


Figure 9.5 : Performances des sujets en fonction du nombre des syllabes et du nombre de canaux de vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée), pour les stimuli vocaux langagiers (B) et non langagiers (A).

MESURE DES TEMPS DE REACTION

Globalement l'analyse des temps de réaction (TR) a montré un effet significatif du vocoder qui se traduisait par une augmentation des RT avec la diminution de l'information spectrale par le vocoder (voir Figure 9.6). Pour le groupe A, un effet significatif a été retrouvé pour les sons Non Voix (sons environnementaux), les sons vocaux langagiers ainsi que pour les sons vocaux non langagiers ($p < 0.0001$). Pour le groupe B, qui a été testé avec un nombre de canaux moins élevé, un effet statistiquement significatif a été retrouvé uniquement pour les stimuli vocaux langagiers et Non Voix ($p < 0,03$). Aucun effet significatif n'a été retrouvé pour les sons vocaux non langagiers avec ce deuxième groupe, suggérant qu'avec ces conditions et instructions de taches, les sujets avaient atteint leurs performances les plus faibles.

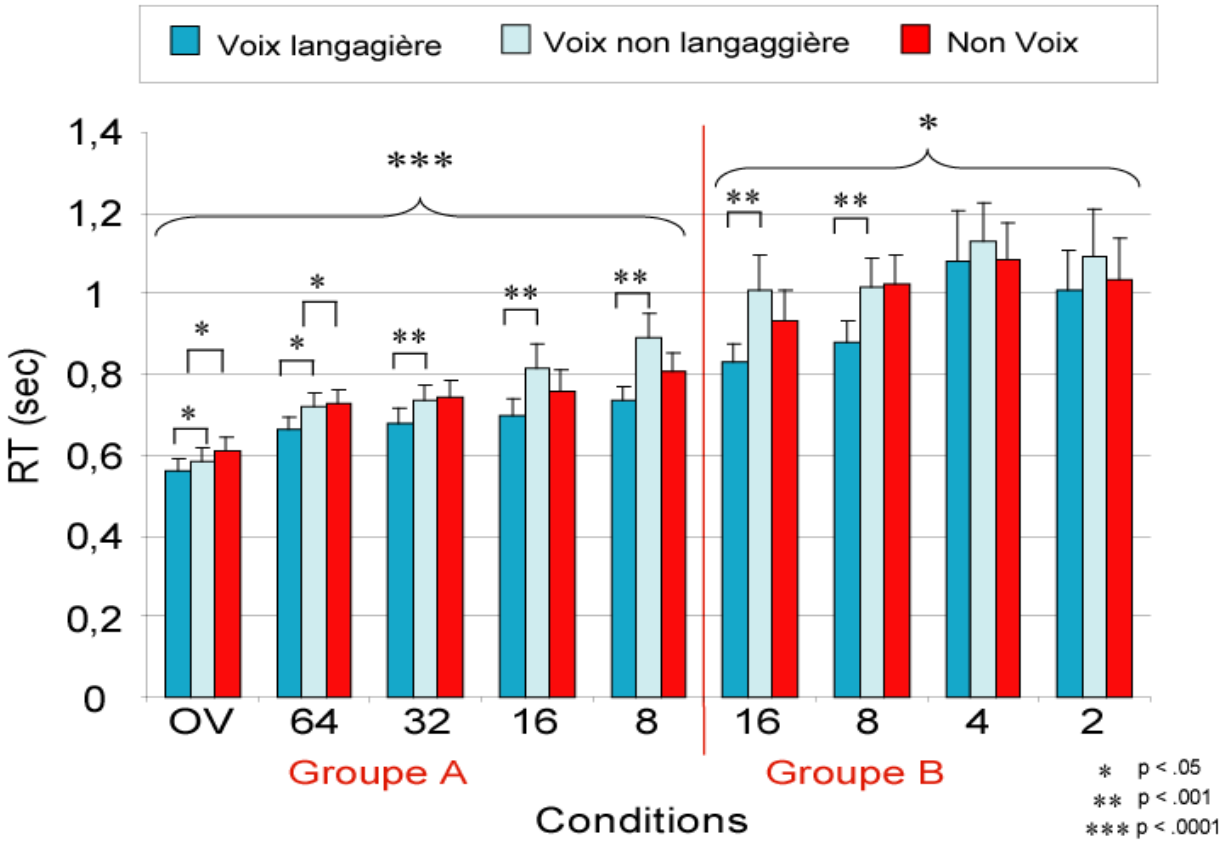


Figure 9.6 : Représentation des temps de réaction, pour les détections correctes (hits) des voix langagières, non langagières et les non voix (sons de l'environnement). Les temps de réaction sont représentés pour les groupes A et B, dans les différentes conditions de vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée). Les différences significatives sont indiquées

Une première comparaison entre les stimuli Vocaux et non Vocaux (non représentée sur la Figure 9.6) a montré que ces valeurs n'étaient statistiquement pas différentes, excepté pour la condition originale et la condition 64 canaux ($p < 0{,}011$). Pour ces deux derniers cas, on observait que les sujets étaient plus rapides pour les stimuli Vocaux. Ensuite, la comparaison entre les TR observés pour les Voix langagières et les Voix non langagières a montré que les sujets étaient statistiquement toujours plus rapides pour détecter les stimuli vocaux langagiers que les autres types de stimuli, excepté dans les conditions extrêmes de dégradation (4 et 2 canaux).

9.1.1.3 RESUME ET DISCUSSION PRELIMINAIRE

En résumé, l'analyse de l'indice de sensibilité *d'* et des temps de réaction a montré plusieurs choses. D'une part la difficulté de détection de la voix parmi des sons de l'environnement augmente avec la dégradation des sons par le vocoder, et d'autre part, la détection de la voix était facilitée pour les sons vocaux langagiers part rapport aux non langagiers. Les stimuli langagiers semblent donc avoir un statut particulier, du point de vue acoustique et perceptif.

9.1.2 VOIX HUMAINE VERSUS SONS DE L'ENVIRONNEMENT, STIMULATION MONAURALE

Suite à la mise en évidence de l'aire de la voix, dont l'activité serait plus marquée à droite (Belin et al., 2000), l'objectif de ce test était de mettre en évidence une éventuelle facilitation du traitement de la voix par l'hémisphère droit, notamment en terme de temps de réaction. On s'attendait donc à des temps de réaction plus court, voire à des différences de performances entre stimulation droite et gauche.

SUJETS

Pour ce test, ont été inclus 14 sujets naïfs, c'est à dire n'ayant pas, auparavant, été exposés à des sons simulant un implant cochléaire. Pour être inclus les sujets devaient également être de langue maternelle française et n'avoir rapporté aucun trouble auditif, neurologique ou psychiatrique. Les sujets avaient été recrutés sur la base du volontariat. Le groupe était âgé en moyenne de 31 ans (+/- 10,1) et composé de 7 femmes et 7 hommes.

STIMULI

Pour ce test, les stimuli étaient les mêmes que pour le test « Voix-Non Voix » en binaural. Il y avait donc 55 stimuli Non Voix et 55 stimuli Voix, dont 29 langagiers et 26 non langagiers. Les sujets ont passé le test avec les sons originaux et dans 3 simulations d'implant : 64, 32 et 16 canaux. Afin de limiter un effet attentionnel pour l'oreille stimulée, nous avions généré pour chaque stimulus, un bruit blanc en conservant la même enveloppe temporelle. Le bruit blanc correspondant au stimulus était présenté à l'oreille « non stimulée », qui recevrait alors la même énergie acoustique mais sans information relative au contenu du son.

PROCEDURE

Le matériel et le logiciel utilisés, la procédure expérimentale et les consignes transmises au sujet étaient exactement les mêmes que lors du test « Voix-Non Voix » en binaural. Toutefois, rappelons que les 110 stimuli étaient présentés monauralement à une oreille tandis que l'autre oreille recevait un bruit blanc ayant la même enveloppe temporelle que le stimulus. Les 110 stimuli étaient présentés dans un ordre aléatoire à gauche et à droite, soit 220 stimuli par condition de vocoder. La durée d'une condition de test était environ de 7 minutes. Les sujets répondaient à l'aide des deux touches « contrôle » d'un clavier correspondant à de leur réponse. Le côté des réponses « voix » ou « non voix », pour les touches « contrôle » était contrebalancé d'un sujet à l'autre. La présentation des stimuli et l'enregistrement des réponses et temps de réaction du sujet à chaque essai se faisait à l'aide du logiciel Matlab 6.5®.

ANALYSE

Pour analyser les réponses des sujets nous avons également appliqué la théorie de détection du signal afin de calculer l'indice de sensibilité *d'*, et ce, de la même manière que pour le test de

détection de la voix en binaural (voir «Matériel et méthodes » dans la partie « Voix humaine versus sons environnementaux, stimulation binaurale », p152). Nous avons donc calculé un *d'* pour l'ensemble des stimuli « Voix » pour l'oreille droite et pour l'oreille gauche, puis avons différencié les performances selon que les stimuli « Voix » étaient « langagiers » ou « non langagiers », toujours pour les deux oreilles. Les analyses statistiques ont testé l'évolution des valeurs de *d'* entre les différentes conditions de vocoder pour chaque côté de stimulation. Ces analyses ont été faites sur les *d'* Voix ainsi que sur les *d'* « Voix langagière » et « Voix non langagière ». Nous avons également comparé les valeurs de *d'* entre les stimuli Voix langagière et Voix non langagière pour chaque condition de vocoder. Bien sûr nous avons comparé les valeurs de *d'* selon que les stimuli étaient présentés dans l'oreille droite ou l'oreille gauche.

Les temps de réactions pour les Hits ont été analysés, toujours en distinguant, Voix, Non-Voix, Voix langagière et Voix non langagière. Les analyses statistiques ont tout d'abord testé l'effet du vocoder sur les temps de réaction pour chaque groupe de stimuli en distinguant le côté de stimulation, droite ou gauche. Dans un deuxième temps, nous avons comparé les temps de réaction entre les présentations droites et gauches ainsi qu'entre les stimuli Voix et Non-Voix et les Voix langagières et Voix non langagière.

9.1.2.1 RESULTATS

EFFET DU VOCODER

Lorsqu'on présentait les signaux originaux on notait de bonnes performances des sujets avec un *d'* pour les stimuli vocaux, égale à 3,7 lorsqu'ils étaient présentés à droite et 3,59 lorsqu'ils étaient présentés à gauche (voir Figure 9.7). Il n'y avait aucune différence entre les stimulations droite et gauche. Par la suite, les performances se dégradaient de façon significative avec la diminution du nombre de canaux du vocoder lorsqu'on considère l'ensemble des stimuli vocaux. Cette diminution avait lieu aussi bien pour la stimulation droite que la stimulation gauche (Friedman ; Droite $p < 0,0001$ et Gauche $p = 0,0009$) et aucune différence n'était retrouvée entre les deux. Les analyses post-hoc 2 à 2, nous permettent de dire que cet effet du vocoder est uniquement dû à une meilleure détection en condition originale, avec un *d'* significativement meilleur aux conditions vocoder (Wilcoxon ; Droite et Gauche $p < 0,007$). Aucune différence entre les conditions de vocoder n'a été retrouvé que ce soit à droite ou à gauche.

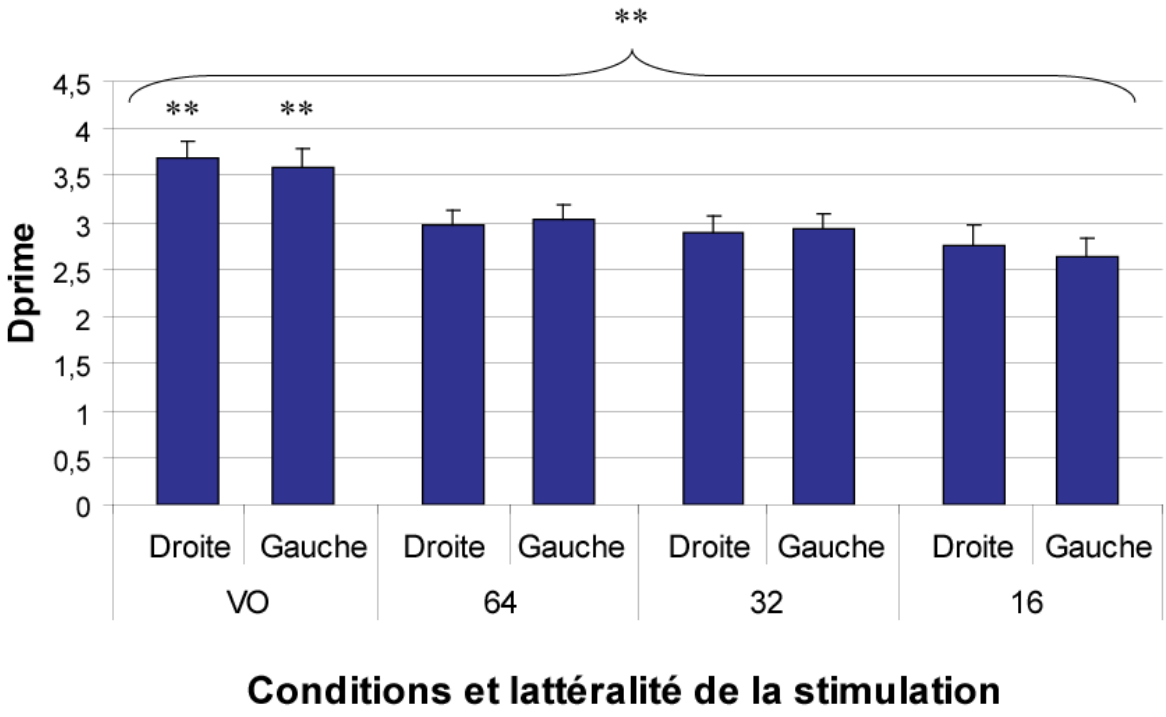


Figure 9.7 : Effet du vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée) selon la latéralité de la stimulation, droite ou gauche. Il n'existe pas de différences lorsqu'on stimule l'oreille droite ou gauche. L'effet global du vocoder est significatif, mais les différences ne se trouvent qu'entre la condition originale et les conditions de vocoder. Ces dernières ne sont pas différentes entre elles.

Nous avons par ailleurs comparé ces données à celles obtenues en conditions de simulation binaurale. Il est apparu que, globalement, les performances étaient meilleures en stimulation binaurale qu'en condition monaurale droite et gauche. Tout d'abord, il est possible que la stimulation d'une seule oreille limite la capacité perceptive des sujets, les empêchant ainsi de percevoir efficacement le signal émis. Notons aussi que l'oreille controlatérale recevait un bruit blanc de même durée et même amplitude ce qui a pu interférer ou parasiter la perception du stimulus cible. Enfin, cela s'explique aussi par l'addition et la confrontation des informations en conditions binaurales qui améliorent la perception et le traitement du signal sonore. Ainsi, si on ne retrouve pas de différences significatives entre les différentes conditions de vocoder dans cette stimulation monaurale, on peut noter une tendance des données à diminuer avec l'augmentation de la dégradation par le vocoder, comme on avait pu le retrouver en condition binaurale. Cette absence de significativité peut s'expliquer, d'une part par une plus grande difficulté en condition monaurale, gommant ainsi certaines différences, et d'autre part par un effet d'échantillon ; rappelons que nous disposions dans les tests binaural et monaural de deux échantillons différents, apportant ainsi, une certaine variabilité aux données.

DISSOCIATION DES STIMULI LANGAGIERS ET NON LANGAGIERS

Une fois de plus, la comparaison des stimuli vocaux langagiers et non langagiers n'a révélé aucune différence entre les capacités de détection de la voix, entre la stimulation droite et gauche. De même que pour le test binaural, la même dissociation a été mise en évidence. A droite comme à gauche on retrouvait un indice de perception *d'* significativement meilleur pour les voix langagières que pour les voix non langagières (Wilcoxon : $p < 0{,}022$) excepté pour la comparaison à droite en condition originale (voir Figure 9.8). De plus on a pu constater que l'effet du vocoder n'était significatif que sur les sons non langagiers, montrant ainsi que l'effet retrouvé sur l'ensemble des stimuli vocaux s'explique par l'effet sur le groupe de stimuli vocaux non langagiers (Friedman : $p < 0{,}0002$ à droite et à gauche). Egalement, pour les tests post-hoc sur les stimuli non langagiers, on observe la même chose que pour l'ensemble des stimuli vocaux, à savoir, des meilleures performances en conditions originale, avec *d'* significativement supérieur aux conditions de vocoder, et une absence de différence entre les conditions de vocoder.

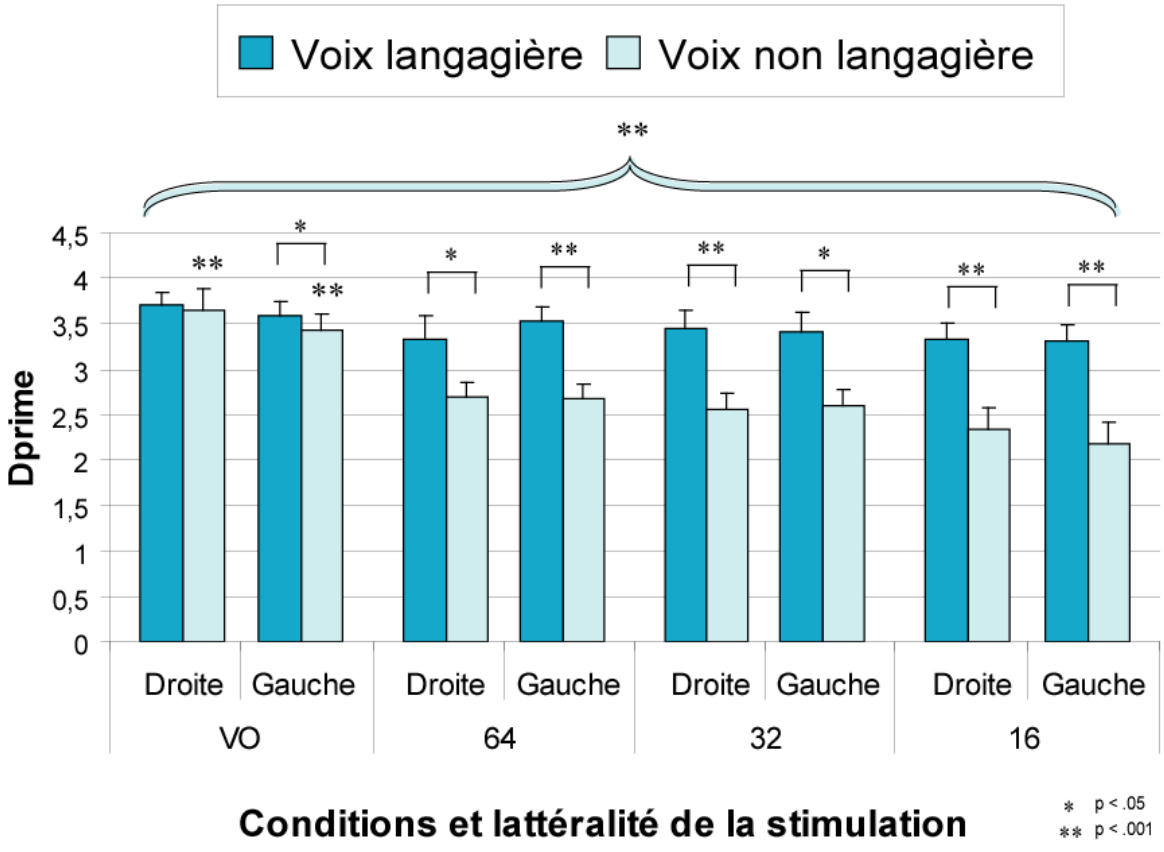


Figure 9.8 : Effet du vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée) selon la latéralité de la stimulation, droite ou gauche et selon le type de stimuli, langagier ou non langagier. Il n'existe pas de différences lorsqu'on stimule l'oreille droite ou gauche. L'effet global du vocoder est significatif uniquement pour les stimuli non langagiers, mais les différences ne se trouvent qu'entre la condition originale et les conditions de vocoder. Ces dernières ne sont pas différentes entre elles. Toutes les comparaisons Voix langagières versus Voix non langagières montrent un *d'* significativement supérieur pour les Voix langagières excepté pour la comparaison en condition originale à droite.

La comparaison avec les performances obtenues en stimulation binaural vont dans le même sens que lorsqu'on compare l'ensemble des stimuli vocaux, à savoir, globalement, une diminution des performances en stimulation monaurale. Les mêmes conclusions peuvent en être tirées.

MESURE DES TEMPS DE REACTION

L'analyse des temps de réaction a confirmé les résultats observés pour l'indice de perception *d'*, ainsi que ceux obtenus au test binaural. D'une part aucune différence significative entre les stimulations droite et gauche n'a pu être mise en évidence. D'autre part, le vocoder semblait affecter significativement les temps de réaction (Friedman : $p < 0{,}0001$ pour les stimulations droite et gauche et pour tous les types de stimuli, vocaux et non vocaux) ; on observait un allongement de ces derniers à mesure que le signal était dégradé par le vocoder (voir Figure 9.9). Toutefois, même s'il semble que les temps de réaction se soient allongés à mesure que le nombre de canaux du vocoder diminue, seuls les temps de réaction en condition originale étaient différents des conditions vocodées. Ceci était vrai pour tous les stimuli vocaux (langagiers et non langagiers) et non vocaux.

Lorsqu'on a comparé les stimuli Vocaux et non Vocaux (non représentée sur la Figure 9.9) on a pu constater que ces valeurs n'étaient statistiquement pas différentes, excepté pour la condition originale à gauche (Wilcoxon ; $p < 0{,}011$). Pour ce cas, on observait que les sujets étaient plus rapides pour les stimuli Vocaux. Cette tendance s'observait aussi pour les autres conditions de vocoder même si elle n'était pas significative.

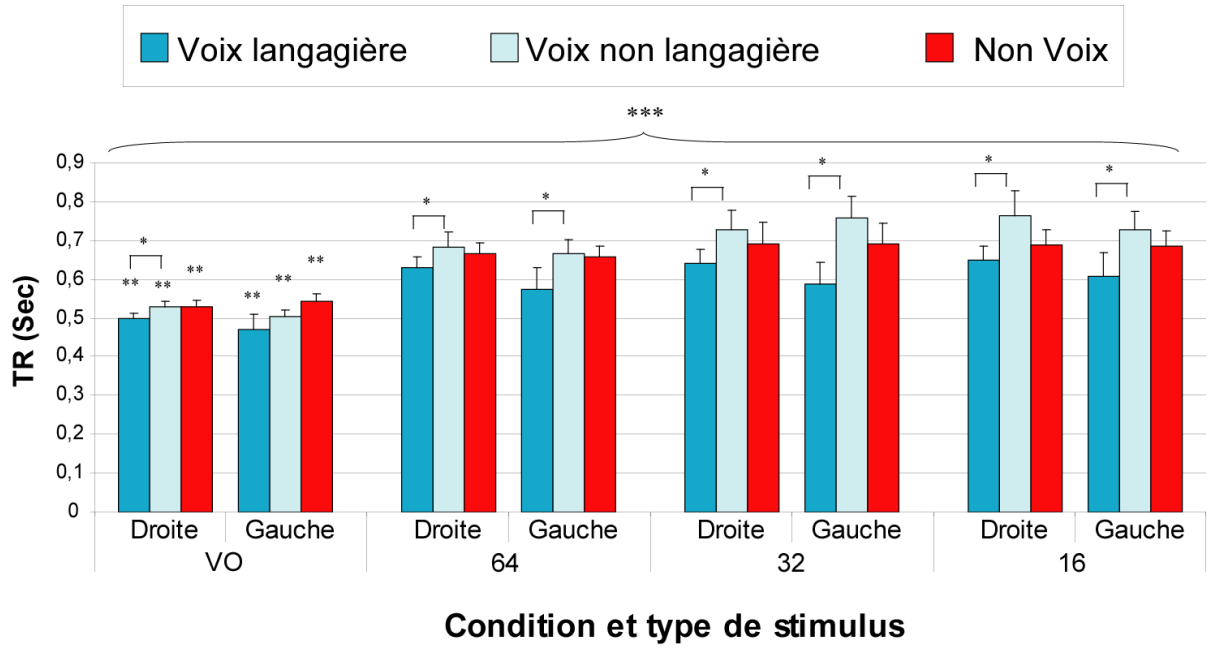


Figure 9.9 : Effet du vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée) sur les temps de réaction. On a distingué la latéralité de la stimulation, droite ou gauche et le type de stimulus, langagier ou non langagier. Il n'existe pas de différences lorsqu'on stimule l'oreille droite ou gauche. L'effet global du vocoder est significatif pour les stimuli vocaux langagiers et non langagiers, ainsi que pour les stimuli non vocaux. Seules, les conditions originales sont différentes des conditions de vocoder. Toutes les comparaisons Voix langagières versus Voix non langagières montrent un temps de réaction significativement plus court pour les Voix langagières excepté pour la comparaison en condition originale à droite.

Ensuite, la comparaison des stimuli vocaux langagiers et non langagiers, révèle que, de façon générale les temps de réaction étaient plus courts pour les stimuli langagiers que pour les stimuli non langagiers (Wilcoxon : $p < 0,02$) excepté pour la comparaison à gauche en condition originale.

Enfin, lorsqu'on compare les temps de réaction avec ceux que nous avons obtenus au test binaural, on remarque que les temps de réaction sont beaucoup plus courts en condition monaurale. Ceci est probablement du au fait qu'il ne s'agit pas du même échantillon de sujets, puisqu'on observe dans les deux cas les mêmes effets du vocoder sur les temps de réactions.

9.1.2.2 RESUME ET DISCUSSION PRELIMINAIRE

En résumé, le test de détection de la voix en stimulation monaural, ne nous permet pas de révéler des différences de performances ou de temps de réaction, selon que l'on stimule l'oreille droite ou gauche. De plus les mêmes variations ont été mises en évidence à droite et à gauche lors de ce test. Nos résultats ne nous permettent donc pas de conclure à l'existence d'une facilitation de la perception des stimuli vocaux par l'hémisphère droit, c'est-à-dire par l'oreille gauche, qui serait liée à une plus grande activité de l'aire de la voix mise en évidence par Pascal Belin (Belin et al., 2000). Toutefois, les résultats ont confirmé les résultats obtenus lors du test de détection de la voix en binaural, à savoir, premièrement la diminution des performances avec la dégradation du signal auditif par le vocoder, et deuxièmement la dissociation des performances entre les stimuli vocaux langagiers et non langagiers. Enfin, si nous n'avons pas pu mettre en évidence un meilleur traitement de la voix lorsque le stimulus est émis à l'oreille gauche, la comparaison avec le test en stimulation binaurale, nous permet en revanche d'affirmer que le traitement binaural améliore la perception des stimuli vocaux. Toutefois, notons que le groupe de sujet ayant passé le test binaural était plus rapide. Ceci peut refléter le fait que les sujet aient privilégié la rapidité de la réponse à la précision et ainsi expliquer les moins bonnes performances mesurées avec le *d'*.

9.1.3 VOIX HUMAINE VERSUS VOIX ANIMALE

9.1.3.1 MATERIEL ET METHODES

SUJETS

Les sujets inclus pour cette expérimentation étaient ceux du groupe A du test de détection de voix parmi des sons de l'environnement (test « Voix-Non Voix ») décrit plus haut. Pour ce dernier test ils avaient passé les conditions originale et 64, 32 et 16 canaux de vocoder. Le groupe était âgé en moyenne de 31 ans (+/-10,1) et composé de 7 femmes et 7 hommes.

STIMULI

Les 110 stimuli de 500 ms utilisés pour le test Voix-humaine-Voix animale provenaient également de la base de données de Pascal Belin. Le groupe de stimuli Voix humaine était constitué sur le modèle du test « Voix-Non Voix » (voir Stimuli p152), avec la même distribution des sons langagiers et non langagiers et des voix d'homme, de femme et d'enfant. Les stimuli Voix animale comprenaient des voix d'animaux de toute sorte, tels que des mammifères, des oiseaux, etc.

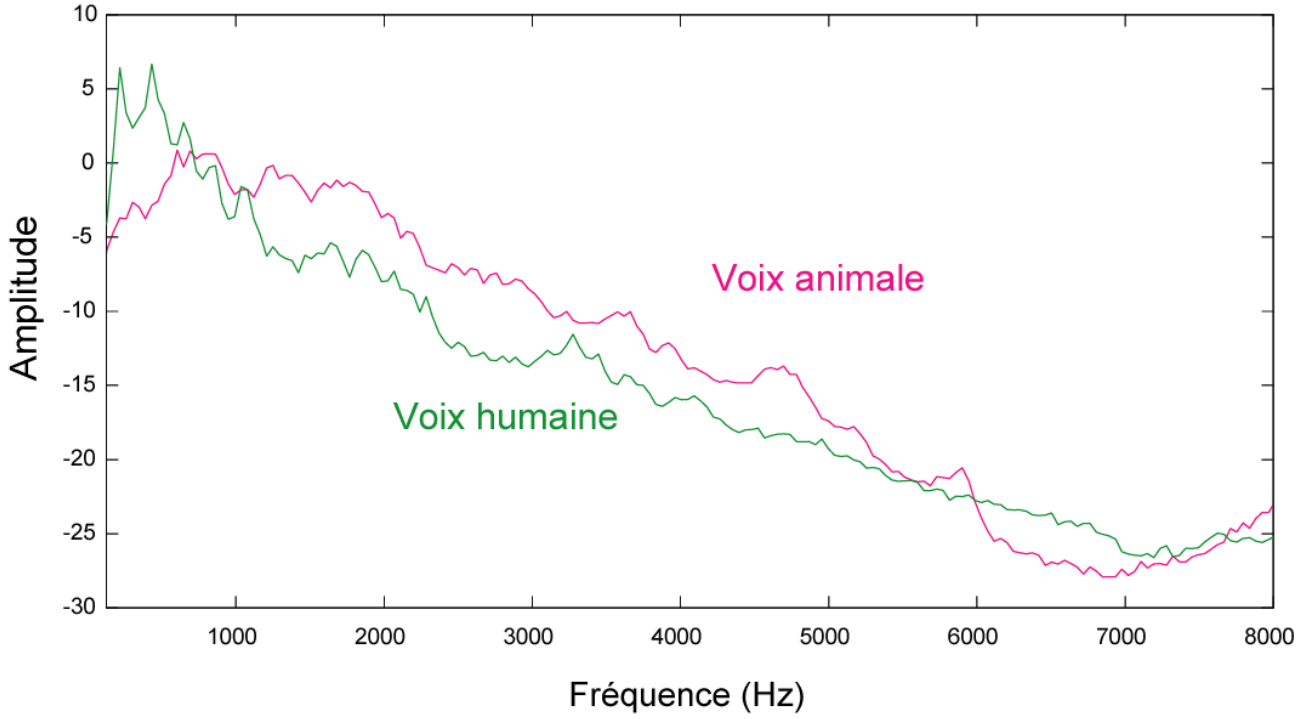


Figure 9.10 : Spectre de puissance moyen sur l'ensemble des stimuli "voix humaine" (vert) et sur l'ensemble des stimuli "Voix animale" (magenta). Aucune différence significative n'a été retrouvée entre les deux groupes de stimuli.

L'analyse par bootstrap comparant le spectre de puissance des stimuli des groupes « Voix humaine » et « Voix animale » n'a révélé aucune différence significative (voir Figure 9.10). Aucune différence non plus n'a été retrouvée entre le spectre de puissance des stimuli vocaux langagiers et celui des stimuli vocaux non langagiers (voir Figure 9.11). La répartition des fréquences et leur amplitude était équivalente dans les deux groupes de stimuli.

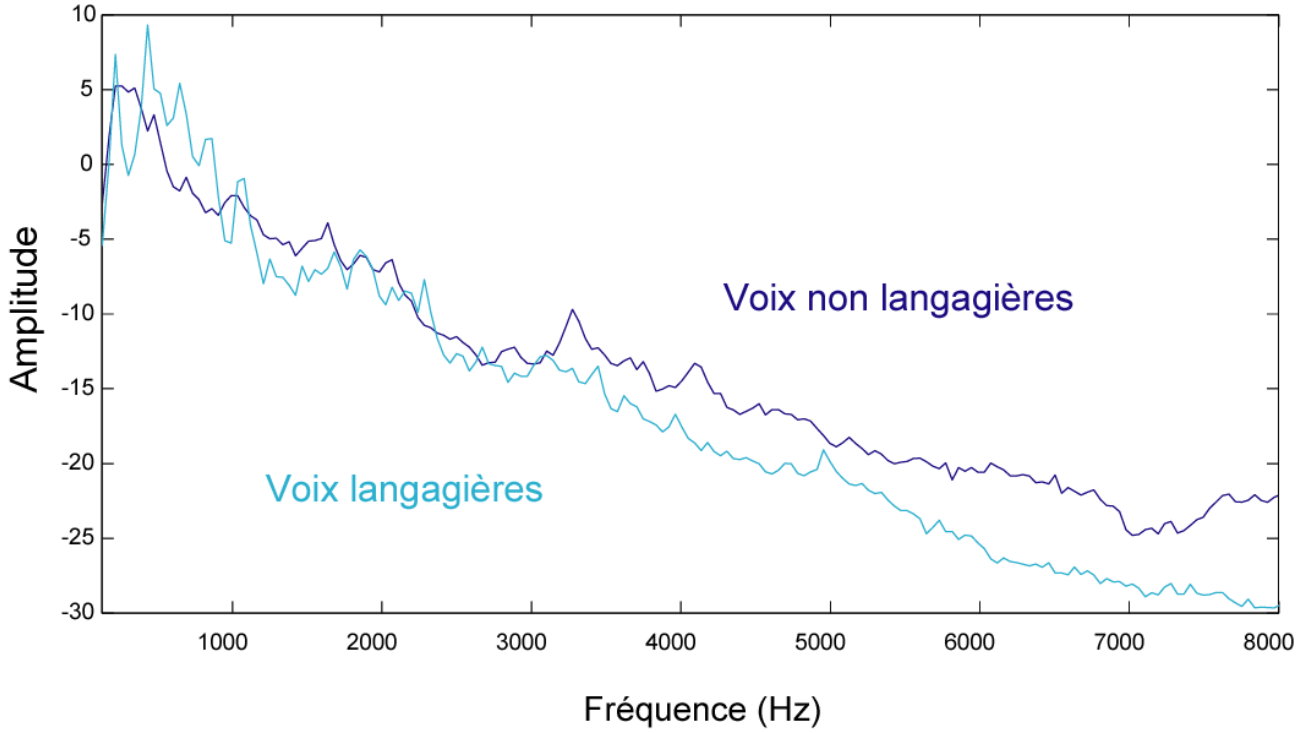


Figure 9.11 : Spectre de puissance moyen sur l'ensemble des stimuli vocaux langagiers (cyan) et sur l'ensemble des stimuli vocaux non langagiers (bleu foncé). Aucune différence significative n'a été retrouvée entre les deux groupes de stimuli.

PROCEDURE

Le matériel, le logiciel utilisés et la procédure expérimentale étaient exactement les mêmes que lors du test « Voix-Non Voix » en binaural et monaural. Les 110 stimuli étaient présentés binauralement et dans un ordre aléatoire et par bloc de condition de vocoder ; Voix originales (VO), 64, 32 et 16 canaux. Ces conditions ont été choisies, car nous nous sommes particulièrement intéressés à l'effet de l'amélioration de la résolution spectrale sur les performances, au-delà de ce qui existe actuellement pour les implants cochléaires. L'ordre des différentes conditions était contrebalancé entre les sujets. Le sujet avait pour consigne de dire si le son qu'il avait entendu était une voix humaine ou une voix animale. Pour répondre, ils devaient appuyer sur une des deux touches « contrôle » correspondant à leur réponse. Le côté des réponses « Voix » et « Non Voix » sur le clavier était contrebalancée entre les sujets. La durée qui séparait la réponse des sujets et le début du stimulus suivant était d'1 seconde. Chaque block de condition durait approximativement 4 minutes, et les sujets étaient libres de faire une courte pause entre chaque condition. La présentation des stimuli et l'enregistrement des réponses se faisaient via le logiciel Matlab 6.5®.

ANALYSE

Pour analyser les réponses des sujets nous avons également appliqué la théorie de détection du signal afin de calculer l'indice de sensibilité *d'*, et ce, de la même manière que pour le test de détection de la voix en binaural (voir Matériel et méthodes, p152). Nous avons donc calculé un *d'* pour l'ensemble des stimuli « Voix », puis avons différencié les performances selon que les stimuli « Voix » étaient « langagiers » ou « non langagiers ». Les analyses statistiques ont testé l'évolution

des valeurs de *d'* entre les différentes conditions de vocoder. Ces analyses ont été faites sur les *d'* Voix ainsi que sur les *d'* « Voix langagière » et « Voix non langagière ». Nous avons également comparé les valeurs de *d'* entre les stimuli Voix langagière et Voix non langagière pour chaque condition de vocoder.

Les temps de réaction pour les Hits ont été analysés, toujours en distinguant, Voix, Non-Voix, Voix langagière et Voix non langagière. Les analyses statistiques ont tout d'abord testé l'effet du vocoder sur les temps de réaction pour chaque groupe de stimuli. Dans un deuxième temps, nous avons comparé les temps de réaction entre les stimuli Voix et Non-Voix et les Voix langagières et Voix non langagière.

9.1.3.2 RESULTATS

EFFET DU VOCODER SUR L'INDICE DE PERCEPTION DPRIME

Lors de la présentation des stimuli originaux, on obtenait de bonnes performances en valeurs de *d'*, bien qu'inférieures à celles obtenues aux tests de détection de voix versus sons environnementaux en stimulation binaurale et monaurale (voir Figure 9.12A). Ces performances sont bonnes quelque soit le type de stimulus vocal ; les différences entre stimuli langagiers et non langagiers n'était pas significative dans la condition originale (voir Figure 9.12B). L'effet du vocoder était significatif que l'on considère l'ensemble des stimuli vocaux ou que l'on distingue les stimuli langagiers et non langagiers (Friedman : $p < 0{,}0001$). La détection des voix humaines parmi des voix animales décroissait à mesure que le signal était dégradé par le vocoder. Lorsqu'on considère l'ensemble des stimuli vocaux (voir Figure 9.12A), on peut voir que les valeurs de *d'* en condition originale étaient significativement meilleures que pour les conditions de vocoder. A 16 canaux les performances étaient significativement plus faibles que toutes les autres conditions. En revanche on ne retrouvait pas de différences entre les conditions de vocoder 32 et 64 canaux.

Lorsque nous avons analysé les valeurs en distinguant les stimuli langagiers (Voix langagières) et les stimuli non langagiers (Voix non langagières), nous avons pu observer que l'effet du vocoder était significatif sur les deux types de stimuli (voir Figure 9.12B). Pour les stimuli langagiers, on retrouvait une meilleure détection des voix humaines en condition originale (VO), avec un *d'* significativement supérieur à ceux obtenus pour les conditions de vocoder (Wilcoxon ; $p < 0{,}03$). Les différences entre les valeurs des 3 conditions de vocoder, même si elles avaient tendance à diminuer avec la diminution du nombre de canaux du vocoder n'étaient pas significatives. Pour les stimuli non langagiers, on retrouvait aussi un effet du vocoder, plus fort que pour les stimuli langagiers. En effet, les valeurs de l'indice de perception *d'* chutaient plus rapidement. Toutes les conditions de vocoder avaient des valeurs de *d'* inférieures à celle observées en conditions originales (Wilcoxon ; $p < 0{,}001$). Pour la condition 16 canaux du vocoder, les performances étaient significativement plus faibles que toutes les autres conditions (Wilcoxon ; $p < 0{,}006$).

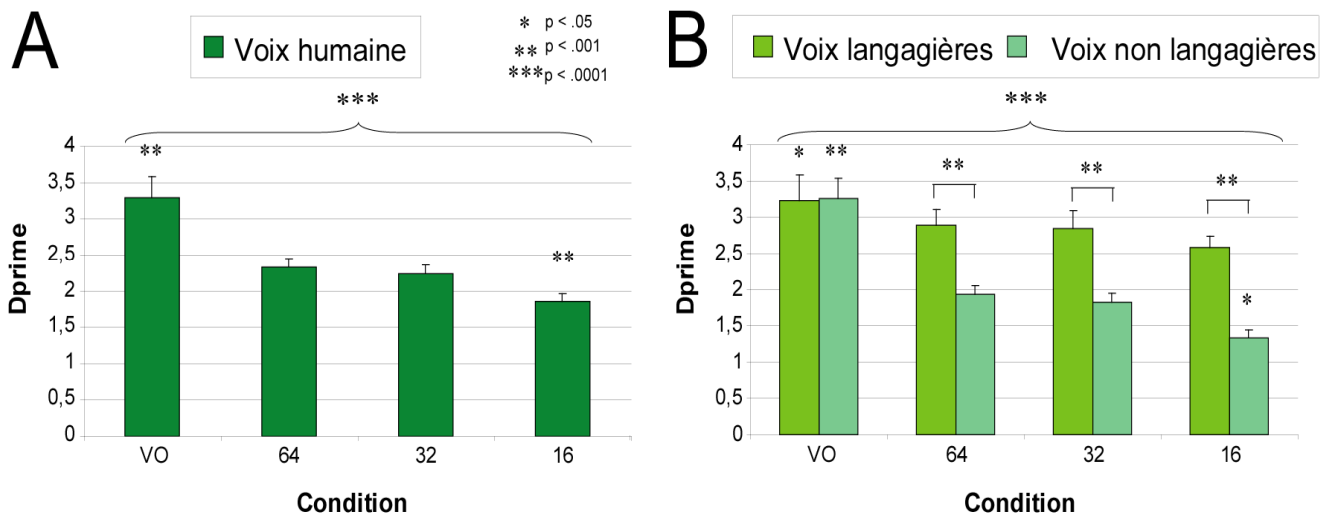


Figure 9.12 : Représentation des valeurs de *d'* obtenues au test de détection de la voix humaines parmi des voix animales, en fonction des différentes conditions de vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée). (A) Les valeurs sont représentées pour l'ensemble des stimuli vocaux. L'effet du vocoder était significatif, avec des valeurs de *d'* qui décroissaient à mesures que le nombre de canaux diminuait. (B) Les valeurs sont représentées en distinguant les stimuli langagiers (Voix langagières) et les stimuli non langagiers (Voix non langagières). L'effet du vocoder était significatif sur les deux types de stimuli, mais plus important pour les stimuli non lagagiers. Dans toutes les conditions de vococder (64, 32 et 16 canaux), les performances obtenues pour les stimuli langagiers étaient significativement suppérieures à celles obtenues pour les non langagiers.

Afin de savoir si les différences de performances observées pour ce test pouvaient être liées au pitch des stimuli, nous avons mesuré ce dernier pour chaque stimulus à l'aide du logiciel PRAAT (voir Tableau 9.3). Pour le groupe de stimuli « Voix Animale » nous avons pu calculer le pitch de 50 stimuli sur 55, tandis que le pitch d'un seul stimulus « Voix non langagière » était manquant pour le groupe « Voix humaine ». L'analyse de pitch moyen des stimuli a révélé que le pitch moyen des stimuli « Voix humaine » était significativement plus grave que celui des stimuli « Voix Animale » (Mann-Whitney : $p < 0{,}0001$). Aucune différence n'a été retrouvée entre les stimuli vocaux langagiers et non langagiers. En revanche les pitches moyens de ces deux distinctions du groupe de stimuli « Voix humaine » étaient aussi significativement plus grave que celui des voix animales.

Types de Stimuli	Animal	Humain	
		Voix Langagière	Voix Non Langagière
Pitch moyen (Hz)	322	246	
		231,4	263

Tableau 9.3 : Pitch moyen pour les différents types de stimuli. Voix langagière et Voix non langagière sont des distinctions des stimuli Voix humaine. Le pitch moyen de l'ensemble des stimuli voix humaine, ainsi que les pitches moyens des voix langagières et non langagières, étaient significativement inférieur ($p < 0{,}0001$) à celui des voix animales (Animal). Aucune différence sigbificative n'a été retrouvée entre les pitches des stimuli Voix langagière et Voix non langagière.

De cette analyse on peut donc dire que la différence moyenne de pitch entre les stimuli humain et les stimuli animaux pouvait être un indice de détection pour les sujets. Dans le test

« Voix-non Voix » on retrouvait une différence significative entre les pitches moyens des stimuli humains langagiers et non langagiers. En revanche ici, bien qu'il existe une différence allant dans le même sens, entre les pitches moyen des deux groupes, mais cette différence n'était pas significative. Or tout comme dans le test de détection de la voix versus les sons de l'environnement, nous avons observé de meilleures performances des sujets pour détecter les stimuli humains vocaux que les stimuli humains non-vocaux. Si on ne peut pas totalement exclure l'idée que le pitch participe à cette meilleure aptitude, on peut dire que d'autres indices que le pitch participent à la meilleure reconnaissance des stimuli humain vocaux, notamment comme cela à déjà été évoqué, des modulations temporelles plus importantes.

EFFET DU VOCODER SUR LES TEMPS DE REACTION

L'analyse des temps de réaction nous montre que l'effet du vocoder était significatif pour tous les types de stimuli ; voix humaines, voix animales, voix langagières et non langagières (Friedman ; $p < 0{,}0001$). Cela se traduisait par une augmentation des RT avec la diminution de l'information spectrale par le vocoder (voir Figure 9.13). Les comparaisons entre les différentes conditions de dégradation ont montré que seuls les temps de réactions en condition originale étaient plus courts que ceux dans les conditions de vocoder (Wilcoxon : $p < 0{,}019$). Ceci était vrai pour tous les types de stimuli, voix humaines, langagières et non langagières et voix animales.

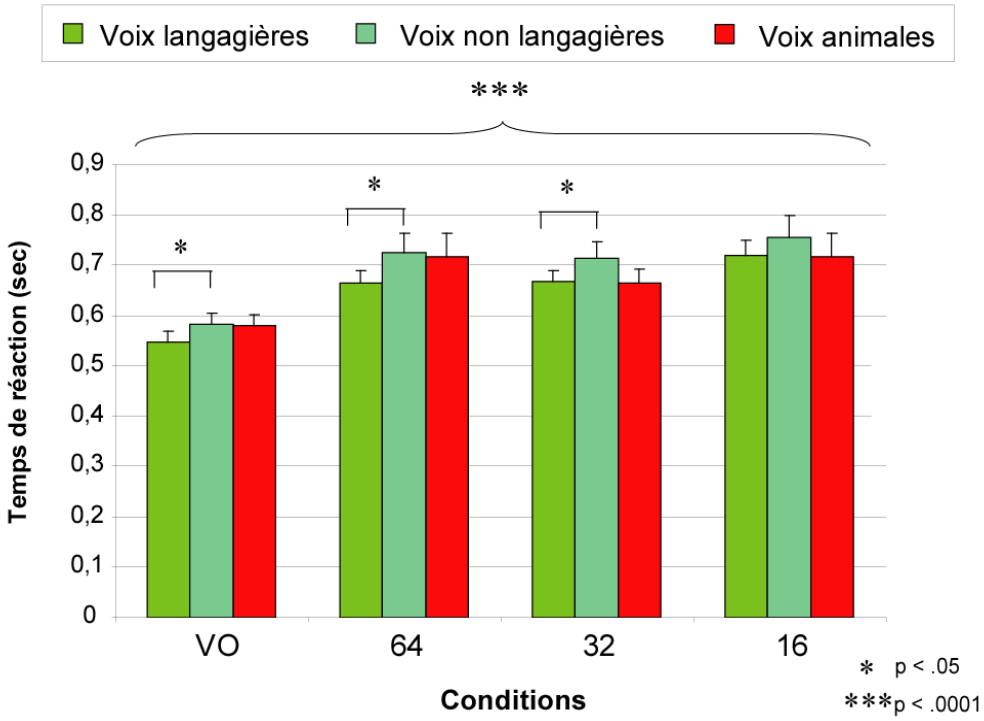


Figure 9.13 : Effet du vocoder (VO correspondait à la condition Voix originale, 64, 32 et 16 correspondaient respectivement à 64, 32 et 16 canaux en condition vocodée) sur les temps de réaction. Le type de stimulus, langagier, non langagier et voix animale. L'effet global du vocoder est significatif pour les stimuli vocaux langagiers et non langagiers, ainsi que pour les stimuli voix animales. Seules, les conditions originales sont différentes des conditions de vocoder. Toutes les comparaisons Voix langagières versus Voix non langagières montrent un temps de réaction significativement plus court pour les Voix langagières excepté pour la comparaison en condition originale à droite

Une première comparaison entre les stimuli Voix humaine et Voix animale (non représentée sur la Figure 9.13) a montré que ces valeurs n'étaient statistiquement pas différentes. Ensuite, la comparaison entre les temps de réaction observés pour la Voix humaine langagière et la Voix humaine non langagière a montré que les sujets étaient statistiquement toujours plus rapides pour détecter les stimuli vocaux langagiers que les autres types de stimuli ($p < 0,03$), excepté dans la plus forte condition de dégradation, 16 canaux.

COMPARAISON AVEC LE TEST « VOIX-NON VOIX »

Nous avons voulu comparer directement les performances entre le test de détection de la voix versus des sons de l'environnement (test « Voix-Non voix ») et l'actuel test de détection de la voix humaine versus la voix animale (test « Voix humaine-Voix animale »). Dans un premier temps, lorsqu'on compare les performances en condition originale, on peut déjà remarquer qu'elles étaient inférieures pour le test « Voix humaine-Voix animale » par rapport au test « Voix-Non voix » (binaural et monaural). Ceci signifie donc qu'il est plus difficile pour les sujets de distinguer la voix humaine parmi des voix animales que parmi des sons de l'environnement. D'autant plus que, rappelons-le, les sujets du test « Voix humaine-Voix animale » étaient les mêmes que pour le test « Voix-Non voix » monaural, ce qui rend donc l'effet de la variation interindividuelle nulle pour expliquer cette différence. Toutefois, toujours en conditions originales, les sujets du test « Voix-Non voix » monaural étaient moins bons que ceux en conditions binaural, mais meilleurs que ceux du test « Voix humaine-Voix animale ».

Afin de comparer les variations induites par le vocoder entre le test « Voix-Non voix » binaural et le test « Voix humaine-Voix animale », nous avons normalisé pour chaque sujet, les variations de performances en condition de vocoder par rapport à la condition originale (voir Figure 9.14).

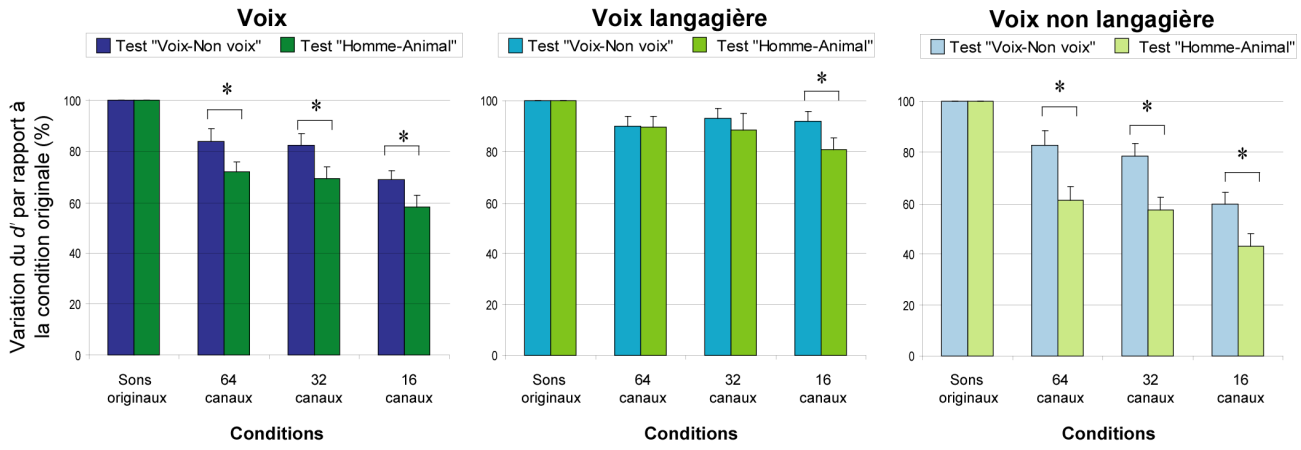


Figure 9.14 : Comparaison des variations de performances en fonction du vocoder pour les tests « Voix-non Voix » et « Voix humaine-Voix animale ». Les données des conditions de vocoder des deux tests sont normalisées par rapport à la condition originale. La comparaison entre les deux tests est faite sur la détection des stimuli vocaux (Voix) ainsi que lorsqu'on distingue les stimuli vocaux langagiers (Voix langagière) et non langagiers (Voix non langagière).

Cette analyse confirme que l'effet du vocoder dégrade plus les performances lorsque la détection de la voix se fait parmi des voix animales que parmi des sons de l'environnement. En effet, on peut voir sur la Figure 9.14 que la diminution en condition de vocoder par rapport à la condition originale était plus importante pour le test « Voix humaine-Voix animale » que pour le test « Voix-Non voix ». Ceci était vrai (Mann-Whitney ; $p < 0,05$) lorsqu'on considérait l'ensemble des stimuli vocaux (« Voix » sur la Figure 9.14), ainsi que lorsqu'on distinguait les stimuli vocaux non langagiers (« Voix non langagière » sur la Figure 9.14) et langagiers (« Voix langagière » sur la Figure 9.14), excepté pour ces derniers en condition de vocoder à 64 canaux. Cette analyse de l'effet du vocoder confirme donc la comparaison des performances en condition originale. Les voix animales comme distracteurs des voix humaines augmentent davantage la difficulté de détection de la voix humaine que lorsqu'on on utilise des sons de l'environnement et rendent ces performances plus sensible à la dégradation.

Pour les temps de réaction, la fenêtre des valeurs des ces temps de réaction était davantage semblable à celle des temps de réaction du test en stimulation monaural qu'à ceux en stimulation binaurale. Ceci s'explique par le fait qu'il s'agissait des mêmes sujets pour les deux premiers tests.

9.1.3.3 RESUME ET DISCUSSION PRELIMINAIRE

Les résultats de ce test confirment les résultats obtenus au test de détection de la voix versus des sons de l'environnement, avec une diminution de l'indice de perception *d'* et une augmentation des temps de réaction à mesure que l'on dégrade le signal. On retrouve le statut particulier des stimuli vocaux langagiers, qui sont toujours mieux détecté.

Lorsqu'on a comparé les résultats du test « Voix humaine-Voix animale » avec ceux des tests « Voix-Non voix » binaural et monaural, on a pu remarquer que, pour des mêmes conditions de stimulation, la tâche était plus difficile pour le test « Voix humaine-Voix animale ». Cela signifie donc qu'il est plus difficile de distinguer la voix humaines parmi des voix animales que parmi des sons de l'environnement. Cette observation semble cohérente. En effet, parmi les voix d'animaux certaines étaient aussi produites par des cordes vocales et un tractus vocal, il semble logique que les modulations spectro-temporelles soient plus proches de celles de la voix humaine, que ne l'étaient celles des sons de l'environnement. Il est donc plus évident que les voix humaines soient plus facilement confondues avec des voix animales qu'avec des sons de l'environnement, dont les sources physiques sonores sont très éloignées de la voix. De plus cette comparaison apporte un point de discussion supplémentaire en ce qui concerne le pitch. Nous avons déjà discuté du fait que les différences de pitch entre les stimuli Voix humaine et les distracteurs, c'est-à-dire les sons de l'environnement dans un cas et les voix animales dans un autre, pouvaient être un indice qui permettrait aux sujets de réaliser la tâche. Or nous avons vu que les performances était meilleures pour le test « Voix-Non voix » alors que la différence entre les stimuli vocaux et les stimuli non vocaux n'était pas significative, contrairement au test « Voix humaine-Voix animale ». Ceci indique

donc que les sujets ont de meilleures performances alors que les indices de pitch sont réduits et ces performances meilleures se maintiennent avec la dégradation du signal par le vocoder. On peut donc clairement avancer le fait que le pitch n'est pas nécessaire pour réaliser la tâche, même si bien entendu il doit apporter de l'information quand il est disponible.

9.2 TESTS COMPORTEMENTAUX DE LA RECONNAISSANCE DE L'IDENTITE DE LA VOIX

Plusieurs attributs de l'identité peuvent être perçus au travers de la voix. Par exemple, comme cela a déjà été évoqué, nous pouvons percevoir le genre, l'âge, l'origine géographique (accent), éventuellement la corpulence, etc. Parmi ces différents attributs nous avons choisi d'étudier plus particulièrement le genre, la discrimination de voix et l'âge.

9.2.1 TEST DE CATEGORISATION DU GENRE

9.2.1.1 MATERIEL ET METHODES

STIMULI

Les sons utilisés pour ce test ont été générés par Pascal Belin et consistaient en un continuum réalisé par morphing entre une voix d'homme et une voix de femme. La voix de l'homme et de la femme correspondaient chacune à une voix moyennée au niveau de la fréquence fondamentale, des formants et de la durée à partir de 30 voix du même genre. Les fréquences fondamentales des voix de l'homme et de la femme étaient respectivement 126 et 212Hz. A partir de ces deux voix moyennées, 3 continuums ont été générés par morphing avec le logiciel Matlab 6.5®. Le premier a consisté en un glissement progressif et linéaire des fréquences fondamentales (pitch) et des formants (timbre) de la voix d'homme vers ceux de la voix de femme (voir Figure 9.15, test 1). Pour le deuxième continuum, un pitch moyen (169 Hz) était maintenu constant et seuls les formants variaient de la voix d'homme vers la voix de femme (voir Figure 9.15, test 2). Pour le dernier continuum, des formants moyennés (timbre) entre la voix d'homme et de femme, étaient maintenus constants, et seule la fréquence fondamentale variait (voir Figure 9.15, test 3).

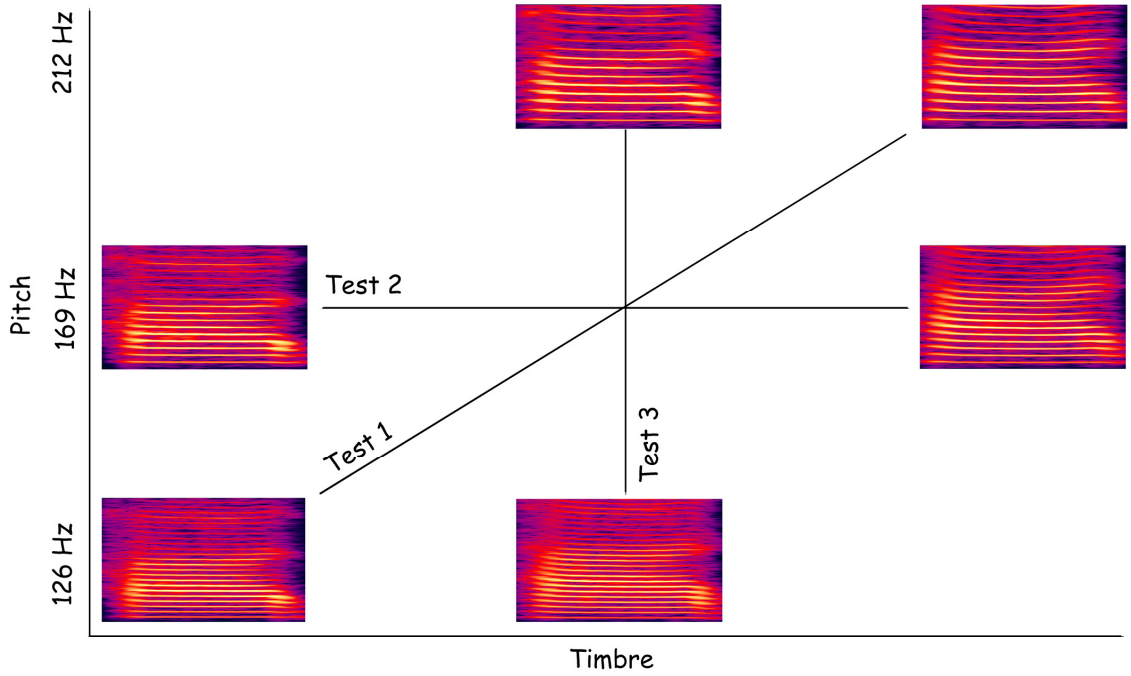


Figure 9.15 : Illustration des différentes continua réalisés par morphing. Le test 1 fait varier à la fois le pitch et timbre de la voix original « homme » vers la voix originale « femme ». Le test 2 maintient constant un pitch intermédiaire en les 2 voix, et fait varier uniquement le pitch. Enfin, le test 3 maintient un timbre moyen constant en faisant varier le pitch

Mille pas pour chaque continuum ont été générés avec ce morphing et nous disposions donc de 1001 voix pour chacun. Pour le test de la catégorisation du genre, nous avons utilisé 11 voix à égale distance physique pour chaque continuum. Pour le test 1, les extrêmes étaient donc les voix d'homme et de femme et la voix du milieu correspondait au stimulus intermédiaire du continuum (50% homme et 50% femme). Pour le test 2, les extrêmes étaient les voix avec les timbres d'homme et de femme et un pitch constant. Pour le test 3, les extrêmes étaient les voix avec les pitch d'homme et de femme avec un timbre moyen constant. La voix du milieu pour les tests 2 et 3 correspondait au même stimulus intermédiaire du continuum que pour le test 1 (50% homme et 50% femme).

SUJETS

Pour ce test, les 14 sujets inclus étaient tous volontaires, naïfs et n'avaient rapporté aucun trouble auditif, neurologique ou psychiatrique. Le groupe était composé de 7 hommes et la moyenne d'âge était 24.63 (+/-2,91). Seul un sujet n'a pas pu passer les conditions 8, 4 et 2 canaux.

PROCEDURE

Tous les sujets ont passés le test avec les sons originaux ainsi que dans 6 conditions de vocoder, 64, 32, 16, 8, 4 et 2 canaux, et dans les conditions pitch constant et timbre constant. Les conditions ont été passées sur deux jours. Le premier jour les sujets passaient les conditions de vocoder 64, 32 et 16 canaux et les conditions pitch constant et timbre constant par block dans un ordre aléatoire et terminait toujours pas la condition originale. Puis, afin de compléter les données et de permettre des analyses statistiques plus robustes, les sujets on été revus une secondes fois pour

passer les conditions de vocoder 8, 4 et 2 canaux, toujours dans un ordre aléatoire. Les sujets étaient testés dans une pièce calme du laboratoire du CerCo. Les stimuli étaient présentés binauralement au travers d'un casque (modèle Eh 250, Sennheiser), dans un ordre aléatoire à un volume sonore ajusté à 72 dB SPL. Les sujets étaient placés devant un clavier avec le casque et avaient reçu pour consigne de dire le plus précisément et le plus rapidement possible, si la voix qu'ils avaient entendue était celle d'un homme ou d'une femme. Ils étaient également prévenus que parfois la voix leur paraîtrait ambigüe et qu'il faudrait répondre ce qui leur paraîtrait être le plus probable. Pour répondre, les sujets devaient appuyer sur une des deux touches contrôles du clavier correspondant à sa réponse « homme » ou « femme ». Le côté des réponses « homme » et « femme » était contrebalancé entre les sujets. Avant chaque test les sujets entendaient 3 fois la voix la plus masculine et la voix la plus féminine. La présentation des stimuli et l'enregistrement des réponses et temps de réaction des sujets étaient fait grâce au logiciel Matlab 6.5®. Les sujets étaient libres de faire une courte pause entre chaque block qui durait approximativement 5 minutes.

ANALYSES

Pour chaque condition, le nombre de réponses « femme » et « homme » et les temps de réactions étaient relevé pour chacune des 11 voix du continuum (10 présentations). Ainsi, pour chaque condition étaient disponible le pourcentage de réponses « femme » pour les 11 voix du continuum, ce qui nous permettait d'estimer une courbe sigmoïde (tangente hyperbolique), qui est la réponse psychophysique classique, en minimisant les distances entre les points observés et la courbe, c'est-à-dire en minimisant l'erreur de la courbe. Un exemple d'estimation à partir de résultats réels est représenté en Figure 9.16. A partir de là, nous nous sommes particulièrement intéressé au point d'inflexion de la courbe, c'est-à-dire l'endroit où les sujets répondait 50% « homme » et 50% « femme », point nommé C50. De ce point, deux indices ont été analysés. Le premier était la pente, qui est un indice très représentatif de la catégorisation ; plus elle est élevée, meilleure est la catégorisation, avec des réponses proches de 100% à l'extrême féminin du continuum et de 0% de l'autre côté. Le deuxième indice est le seuil, c'est-à-dire le pourcentage de féminité du continuum pour lequel les sujets répondaient 50% « homme » et 50% « femme ». Plus ce point se trouvait sur la droite du continuum, plus les sujets avaient tendance à entendre les voix se trouvant à gauche de ce point comme masculines, et inversement. En résumé, la pente va refléter la qualité de la catégorisation, et le seuil va refléter à partir de quel moment du continuum le sujet vont passer d'une catégorie à l'autre.

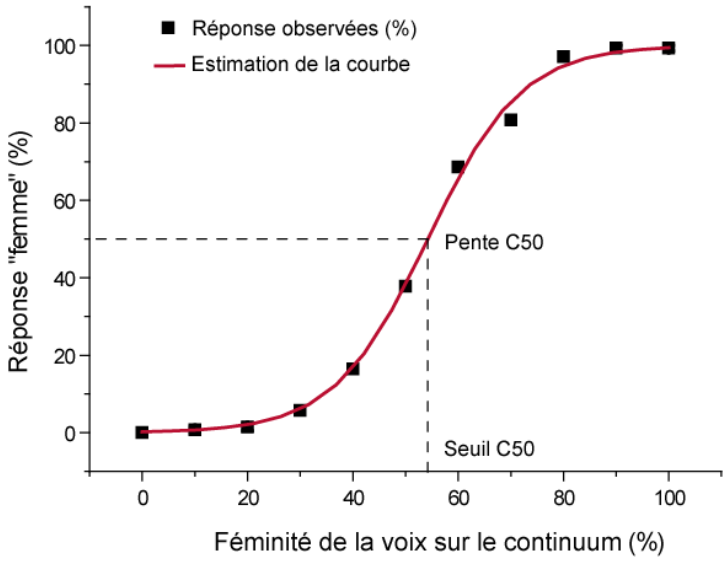


Figure 9.16 : Exemple de réponses d'un sujet et estimation d'une courbe psychophysique (tangente hyperbolique) avec Matlab 6.5 ® en minimisant la distance des points observés avec la courbe estimée.

Nous avons également mesuré les réponses aux 2 extrêmes du continuum. Ces deux stimuli correspondaient respectivement à une voix moyenne d'homme et une voix moyenne de femme. Nous avons regardé le pourcentage de reconnaissance correcte pour ces deux voix. Pour les voix intermédiaires, nous avons estimé que le biais physique introduit par la manipulation liée au morphing ne nous permettait pas d'appréhender correctement cet indice de reconnaissance correcte. Autrement dit il était difficile d'estimer quel impact avait le morphing sur le percept et de dire ce qui était correct ou non comme réponse. Le pourcentage de réponses correctes n'a donc pas été étudié pour les voix intermédiaires.

Les analyses statistiques ont donc porté sur la reconnaissance correcte de la voix d'homme et de la voix de femme (extrêmes du continuum), les pentes et les seuils au point C50, ainsi que l'adéquation du modèle sigmoïde aux réponses comportementales (ou l'erreur de la courbe). Ce dernier paramètre a été mesuré grâce à la distance moyenne des points (réponses comportementales) par rapport à la courbe estimée. Enfin, nous avons également analysé les temps de réactions pour chaque voix du continuum.

Une première analyse a consisté à comparer ces différents critères entre les différentes conditions manipulées ; originales (VO), vocoder (64, 32, 16, 8, 4 et 2 canaux) et pitch constant et timbre constant. Dans un second temps nous avons également regardé s'il existait un effet du genre du sujet sur la catégorisation. Nous avons comparé les réponses aux extrêmes, le seuil et la pente du point C50, l'adéquation du modèle ainsi que les temps de réaction entre les sujets féminins et masculins.

9.2.1.2 RESULTATS

LES COURBES

Une première description des données concerne les courbes sigmoïdes (voir Figure 9.17, A). Cette représentation nous permet de voir que les sujets en condition originale étaient capables de très bien catégoriser les voix d'hommes et de femmes avec des réponses très stables et catégoriques aux extrêmes. De plus on peut voir que la pente de la courbe est plus élevée pour cette condition que pour toutes les autres. A mesure que l'on dégrade le son par le vocoder, on peut voir ces deux indices varier : les réponses au extrêmes sont plus variables et moins catégoriques, et la pente s'aplatie. Pour plus de clarté, sur la Figure 9.17 A, seules les conditions originales et 64, 16 et 2 canaux de vocoder ont été représentées, mais les autres conditions variaient régulièrement entre celles qui sont représentées.

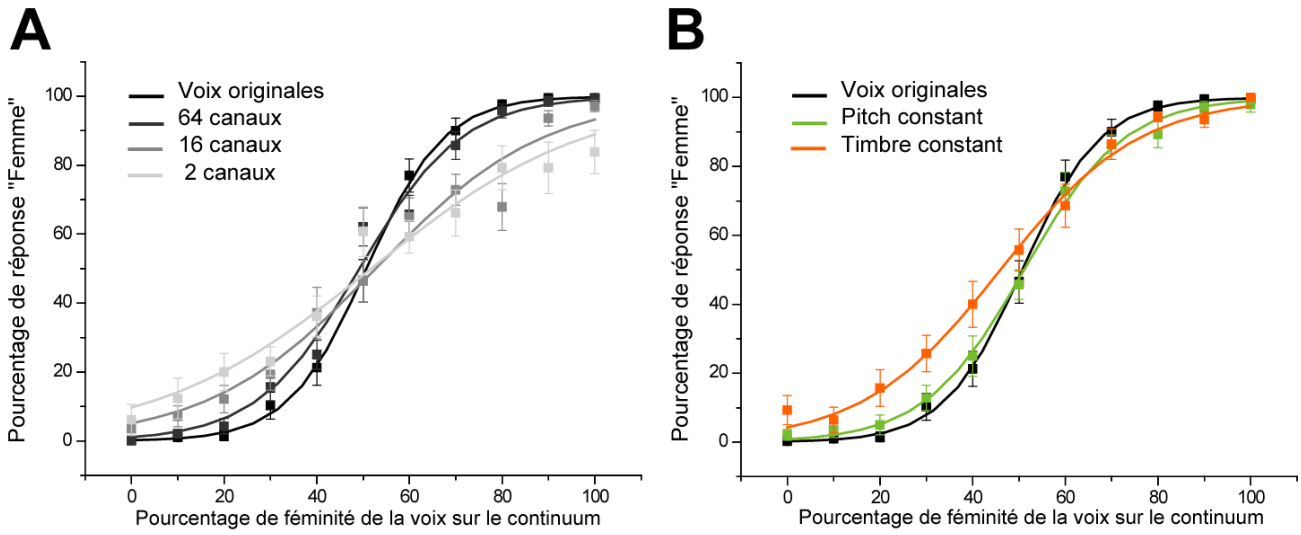


Figure 9.17 : Représentations des réponses (±SE) moyennes des sujets (les points) et des courbes sigmoïdes moyennes calculées à partir des réponses des sujets. (A) Pour plus de clarté, seules les conditions originales et 64, 16 et 2 canaux de vocoder ont été représentées. Les conditions manquantes variaient régulièrement entre celles qui sont représentées. (B) Les conditions Pitch constant, Timbre constant et originales sont représentées.

Lorsqu'on manipulait le pitch ou le timbre (voir Figure 9.17, B), on peut voir avec les courbes sigmoïdes que la catégorisation du genre sur la seule information du pitch (condition timbre constant) s'avérait plus difficile que lorsque seule l'information de timbre était présente (condition pitch constant).

Les analyses du seuil et de la pente au point d'inflexion (C50), de l'adéquation du modèle, des réponses aux extrêmes ainsi que l'adéquation du modèle sigmoïde aux données, vont nous permettre de mieux quantifier les différences sur les courbes, entre les conditions que nous avons manipulées.

L'ANALYSE DU SEUIL

L'analyse du seuil au point d'inflexion de la courbe nous montre que tous les seuils sauf celui pour la condition 4 canaux, oscillaient à plus ou moins 5% autour du milieu du continuum. Il semble donc que dans l'ensemble, les sujets trouvaient que la voix la plus ambigüe se situait à peu près au milieu du continuum (voir Figure 9.18).

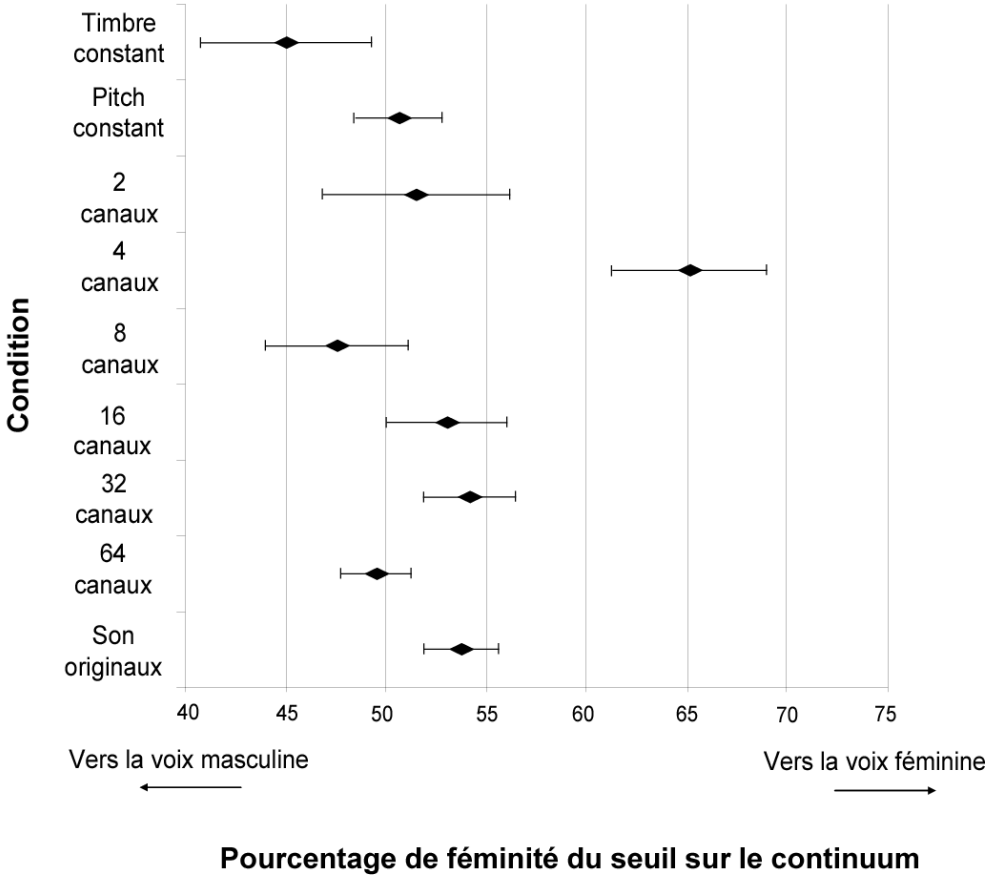


Figure 9.18 : Valeurs (±SE) du seuil au point d'inflexion C50 en fonction des différentes conditions auditives (originales, vocoder, timbre et pitch). Le seuil de la condition 4 canaux du vocoder est significativement décalée du coté de la voix de femme par rapport à toutes les autres. Les conditions timbre constant, pitch constant et 8 canaux de vocoder sont significativement plus du côté homme par rapport à la condition originale.

L'analyse du seuil révèle que, globalement, les manipulations des stimuli, que ce soit le vocoder ou encore le pitch ou le timbre constant, influence la capacité des sujets à catégoriser le genre de la voix (Friedman : $p < 0{,}0005$). Lorsqu'on s'intéresse aux comparaisons entre les conditions, on se rend compte que cet effet s'expliquait notamment par un seuil dans la condition de vocoder à 4 canaux, qui était significativement décalé du côté droit, c'est-à-dire du côté de la voix de femme, par rapport à toutes les autres conditions (Wilcoxon : $p < 0{,}04$). Cela signifie que pour la condition 4 canaux, en partant du côté de la voix d'homme, les voix du continuum étaient davantage perçues comme masculines. Cet effet n'était pas retrouvé pour la condition la plus dégradée par le vocoder (2 canaux). Il apparaissait également, que les seuils en conditions pitch constant, timbre

constant et 16 canaux de vocoder étaient significativement décalés du côté de la voix de l'homme. Par rapport à la condition originale, les conditions timbre constant, pitch constant et 8 canaux du vocoder étaient significativement plus du côté de la voix d'homme qu'en condition originale (Wilcoxon : $p = 0,0258$). De même pour la condition timbre constant, par rapport à la condition 32 canaux (Wilcoxon : $p = 0,011$).

Ce que nous pouvons dire à partir de cette analyse du seuil, c'est que, bien que l'effet des différentes manipulations des stimuli soit significatif, il semblerait que cet effet soit essentiellement dû à une condition ; la condition 4 canaux du vocoder voit son seuil s'écarter davantage de la condition originale. Lorsqu'on enlevait cette condition pour l'analyse globale, les variations devenaient non significatives. L'écart observé pour la condition 4 canaux, est difficile à expliquer étant donné qu'on ne retrouve pas cet effet pour la condition la plus dégradée par le vocoder (2 canaux). Nous expliquons donc cet effet, probablement par un effet de l'échantillon, ce qui mériterait d'être vérifié auprès d'un échantillon plus grand. Nous concluons donc que les seuils semblaient peu affectés par les différentes modulations.

Lorsque nous avons comparé pour chaque condition les seuils des sujets masculins et féminins, il n'est apparu aucune différence significative. Toutefois, on peut noter que, pour les conditions les plus dégradées par le vocoder (8, 4 et 2 canaux), les sujets masculins avaient tendance à catégoriser les voix du continuum plus comme masculine que les sujets féminins. Cela s'illustre par des seuils du point C50 qui étaient davantage du côté « voix de femme » pour les sujets masculins que pour les sujets féminins. Cet effet était probablement dû aux conditions particulièrement dégradées du stimulus auditif, qui amenaient les sujet à adopter une stratégie ; les sujets étant plus hésitants ils avaient tendance, semble-t-il, à choisir plutôt homme pour les sujet masculins et plutôt femme pour les sujets féminins. Mais rappelons que cet effet n'était pas significatif.

L'ANALYSE DE LA PENTE

Les résultats concernant la pente de la courbe psychophysique, montrent que la dégradation du signal auditif par le vocoder diminuait les capacités de catégorisation des sujets, en diminuant progressivement les valeurs de la pente (voir Figure 9.19). L'effet des différentes manipulations de vocoder, pitch et timbre était significatif (Friedman : $p < 0,0001$). On remarque que la pente en condition originale (Sons originaux) avait une valeur maximale de 0,43. Les comparaisons 2 à 2 montrent que la pente, qui reflète la qualité de la catégorisation, n'était pas différente entre les conditions originales, 64 canaux, 32 canaux et pitch constant. Pour les 3 premières conditions les valeurs de pente de la catégorisation étaient significativement supérieures aux conditions 16, 4, 8 (sauf pour 32 canaux) et 2 canaux (Wilcoxon : $p < 0,0445$). Autrement dit, la catégorisation du genre n'était significativement affectée par le vocoder qu'à partir de 16 canaux. Toutefois, en-deçà de 16 canaux de vocoder, les différences entre les conditions n'étaient pas significatives. Les analyses ont également montré que la condition pitch constant avait une valeur de pente supérieure à 16, 8 et 2

canaux. Il n'existait pas de différences entre les conditions pitch constant et timbre constant, bien que pour cette dernière la pente était significativement inférieure à celle en condition originale.

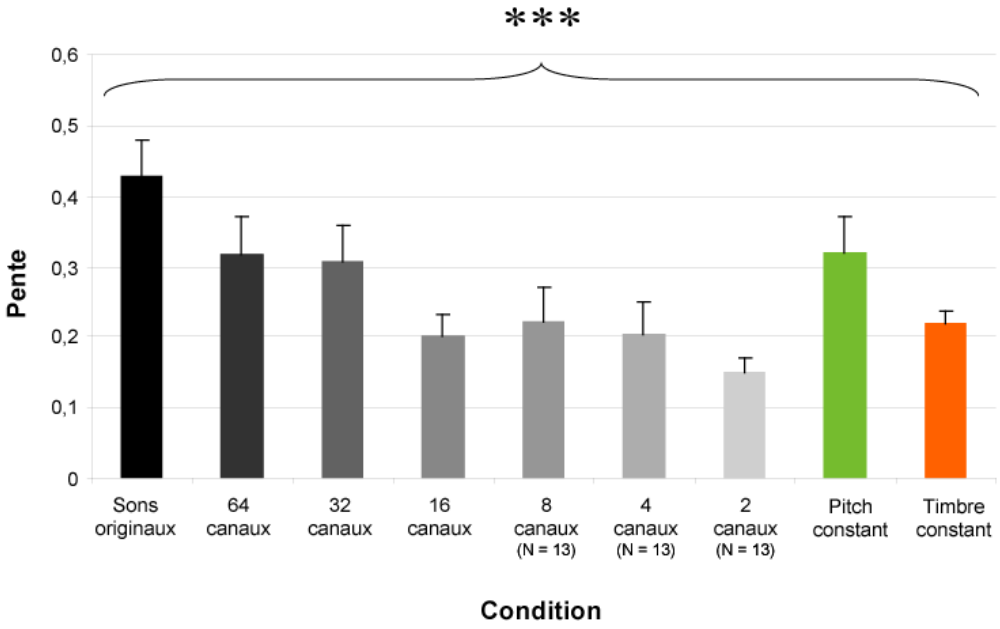


Figure 9.19 : Valeurs (±SE) de la pente au point d'inflexion C50 en fonction des différentes conditions auditives (originales, vocoder, timbre et pitch). Les résultats sont présentés pour 14 sujets, bien qu'un sujet soit manquant pour les conditions de vocoder 8, 4 et 2 canaux.

Nos résultats semblent donc montrer que la catégorisation du genre est plus facile lorsque seules les variations de timbre étaient présentes que lorsqu'il n'y avait que les variations de pitch. Le timbre semble donc être une information importante et un indice plus fort que le pitch dans la perception du genre de la voix. Toutefois, comme nous allons le voir, cet effet est expliqué uniquement par une habileté particulière du groupe de sujets féminin.

La distinction des valeurs de la pente en fonction du genre du sujet, nous indique que globalement, les femmes ont tendance à mieux catégoriser le genre de la voix que les hommes, mais la différence entre les deux groupes de sujets n'était significative que pour la condition pitch constant (voir Figure 9.20). Pour cette condition on remarque que la pente moyenne pour les femmes était similaire à la pente en condition originale. Autrement dit, les femmes arrivaient à catégoriser le genre de la voix uniquement avec l'information de timbre, et ce aussi bien qu'en condition originale. Il semble donc que les femmes aient des habiletés particulières en ce qui concerne la perception du genre de la voix et de façon significative pour l'information de timbre. A l'inverse, les sujets masculins en condition pitch constant avaient des performances similaires à celles qu'ils avaient en condition de vocoder 16 et 8 canaux ainsi qu'en condition timbre constant. Pour les hommes, il semble donc que les variations de pitch (timbre constant) ou les variations de timbre (pitch constant)

seules, ne soient pas suffisantes pour une bonne catégorisation. Cette différence entre les sujets masculins et féminins est intéressante, et nous amène à conclure, qu'aucun des deux groupes ne semble être meilleur pour catégoriser le genre de la voix sur la base du pitch (timbre constant). En revanche, les femmes semblent mieux percevoir ou mieux traiter l'information du timbre (pitch constant) et ceci pourrait expliquer leur tendance à mieux catégoriser le genre de la voix au travers du vocoder ; il est probable que les femmes ayant une meilleure perception ou un meilleur traitement de cette information, soit capable de mieux percevoir les informations de timbre même dégradées au travers du vocoder.

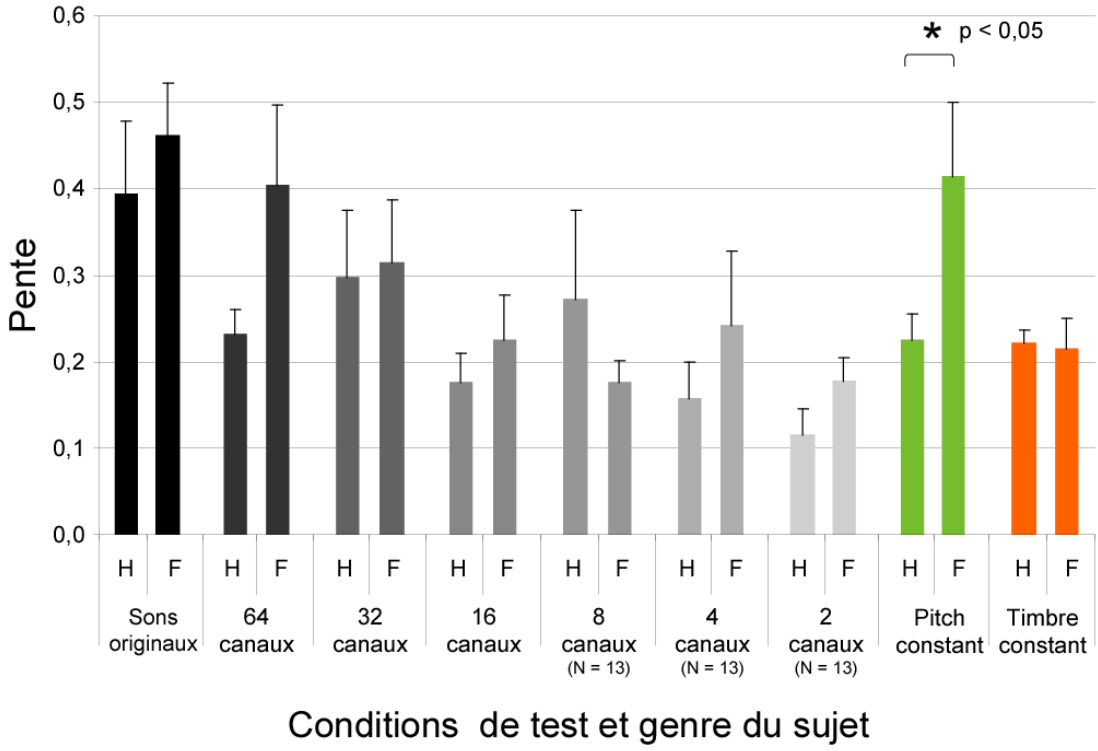


Figure 9.20 : Distinction en fonction du genre du sujet (H pour homme et F pour femme), des valeurs (±SE) de la pente au point d'inflexion C50 pour les différentes conditions auditives (originales, vocoder, timbre et pitch). Les femmes (N = 7) avaient tendance à mieux catégoriser et la différence était significative pour la condition pitch constant. Un sujet masculin était manquant pour les mesures à 8, 4 et 2 canaux de vocoder.

L'ADEQUATION DU MODELE SIGMOÏDE

Au travers de cette analyse nous avons cherché à savoir à quel point l'estimation du modèle sigmoïde, et donc d'une réponse psychophysique de catégorisation, était pertinente par rapport aux données de chaque condition.

La manipulation des différentes conditions faisait varier significativement l'erreur (Friedman ; $p < 0{,}0001$) : l'erreur augmentait à mesure que l'on dégradait les sons par le vocoder, signifiant ainsi que plus on dégradait le signal, moins les sujets arrivaient à catégoriser, car le modèle sigmoïde était moins adapté (voir Figure 9.21). Cela correspond bien à ce que l'on peut observer sur

les courbes sigmoïdes, sur la Figure 9.17, avec une variabilité des points qui augmentait. Les analyses statistiques ont montré que l'erreur pour la condition originale était significativement plus faible en comparaison de toutes les autres conditions (Wilcoxon ; $p < 0,01$), excepté les conditions 64 et 32 canaux de vocoder et pitch constant. On peut donc dire que le modèle sigmoïde était aussi bien adapté pour ces 4 conditions (originale, 64 et 32 canaux de vocoder et pitch constant). Pour les conditions de vocoder 16, 8, 4 et 2 canaux, l'erreur était statistiquement similaire et pour chacune de ces 4 conditions l'erreur était significativement supérieure aux conditions de vocoder 64 et 32 canaux, ainsi qu'aux conditions originales, pitch constant et timbre constant (Wilcoxon ; $p < 0,04$). On peut voir sur la Figure 9.21 qu'à 2 canaux de vocoder, l'erreur moyenne doublait presque par rapport aux 3 conditions précédentes. Cet effet était dû à un seul sujet qui avait une erreur particulièrement importante de 3,2. Ceci explique la variabilité importante de cette dernière condition de vocoder. Lorsqu'on retirait la valeur de l'erreur de ce sujet pour le calcul de la moyenne de l'erreur à 2 canaux, on retrouvait une erreur moyenne de 0,17, ce qui était tout à fait comparable aux 3 conditions précédentes de vocoder (16, 8 et 4 canaux). Notons toutefois que, pour les analyses présentées dans ce paragraphe, les mêmes différences entres les conditions étaient retrouvées avec ou sans ce sujet, aussi, nous avons conservé cette donnée.

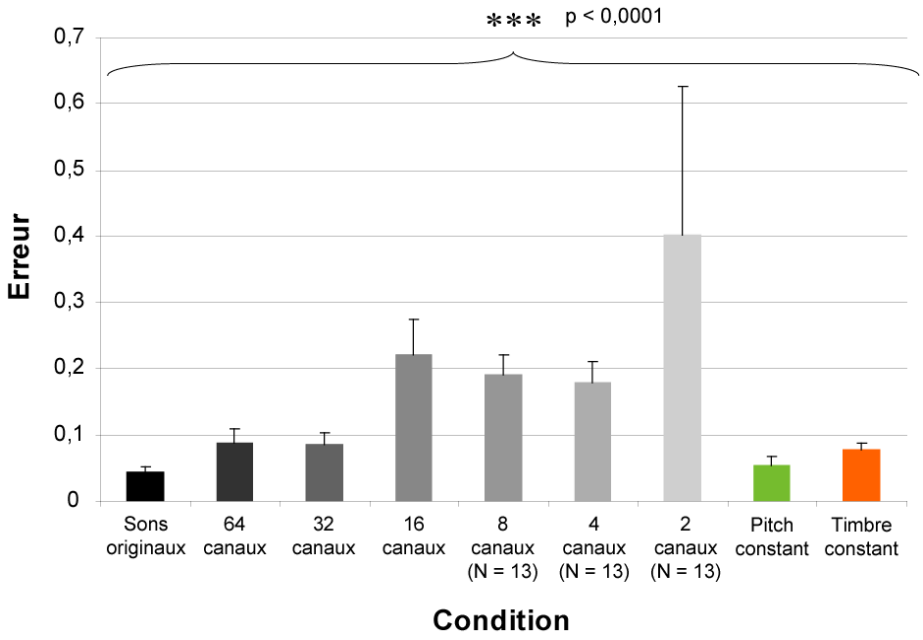


Figure 9.21 : Valeur de l'erreur (±SE) de l'estimation de la courbe sigmoïde par rapport aux données des sujets (i.e., la distance moyenne des points par rapport à la courbe). La manipulation des différentes conditions de stimulation était significative. Les conditions originale et 64 et 32 canaux de vocoder avaient des valeurs similaires et significativement plus faibles que les autres conditions de vocoder (16, 8, 4 et 2 canaux). Les conditions pitch constant et timbre constant étaient similaires à 64 et 32 canaux de vocoder.

A partir de ces analyses on peut dire que le modèle sigmoïde, typique d'un réponse psychophysique de catégorisation, était significativement moins adapté aux données des sujets en condition de vocoder à 16, 8, 4 et 2 canaux, qu'aux conditions originale, pitch constant, timbre constant et 64 et 32 canaux de vocoder. Ces résultats corroborent ceux obtenus lors de l'analyse de la

pente, à savoir une similarité entre les conditions originale et 64 et 32 canaux de vocoder, ainsi qu'une différence entre ces 3 conditions et les autres conditions de vocoder. De même, les analyses de la pente révélaient également que pour la condition pitch constant, c'est-à-dire où seule l'information du timbre était présente, les sujets montraient de meilleures performances que pour la condition timbre constant.

Les analyses en fonction du genre du sujet, n'ont révélé aucune différence significative entre les sujets masculins et les sujets féminins en ce qui concerne l'adéquation du modèle sigmoïde aux données comportementales.

ANALYSE DES REPONSES AUX EXTREMES

Au travers de cette analyse, nous avons souhaité mesurer l'effet des différentes conditions de stimulation sur la reconnaissance correcte des voix extrêmes du continuum, à savoir la voix d'homme et celle de femme. Rappelons que, sur les graphiques précédents, la voix d'homme était la première sur le continuum, tandis que la voix de femme était la dernière.

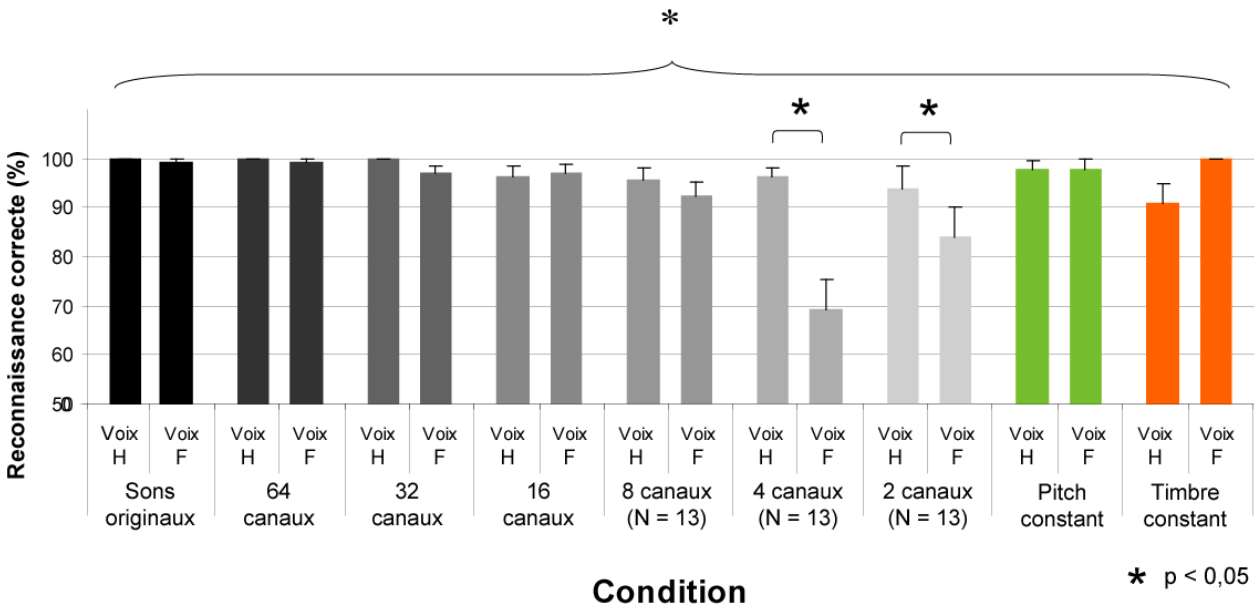


Figure 9.22 : Pourcentage de reconnaissance correcte (±SE) pour la voix d'homme et la voix de femme dans chaque condition de vocoder.

Les analyses ont montré que les différentes conditions de stimulation diminuait significativement la reconnaissance correcte de la voix d'homme (Friedman ; $p = 0,0031$) et de la voix de femme (Friedman ; $p < 0,0001$). Du côté de la voix d'homme, les analyses post-hoc ont montré que l'effet était spécifiquement dû à la condition timbre constant. En effet comme on peut le voir sur la Figure 9.22, les performances étaient particulièrement diminuées pour la reconnaissance de la voix d'homme dans cette condition et étaient significativement inférieures aux performances en conditions originales et 64 et 32 canaux de vocoder. Aucune autre différence significative n'a été retrouvée du côté de la voix d'homme. On peut donc dire que les performances de reconnaissances correctes de la voix d'homme n'étaient pas significativement affectées par le vocoder. Du côté de la

voix de femme, les performances de reconnaissance correctes étaient plus vite affectées par la dégradation du signal par le vocoder que du côté de la voix d'homme. En effet, on peut voir une légère diminution des performances dès la condition 32 canaux, mais cette diminution n'était significative qu'à partir de 8 canaux de vocoder (Wilcoxon ; $p < 0,04$). Les performances de reconnaissance de la voix de femme pour les conditions de vocoder 4 et 2 canaux étaient similaires entre elles et toutes deux significativement inférieures aux conditions de vocoder 64 et 32 canaux ainsi qu'à la condition originale (Wilcoxon ; $p < 0,05$). En revanche, et contrairement à la voix d'homme, la reconnaissance correcte de la voix de femme n'était pas significativement affectée par les conditions pitch constant et timbre constant. De plus, lorsqu'on comparait directement les performances entre la voix d'homme et la voix de femme, on observait une différence significative pour les conditions 4 et 2 canaux (Wilcoxon ; $p < 0,005$).

En résumé, on peut dire que la reconnaissance de la voix d'homme était très peu affectée par le vocoder tandis que la reconnaissance de la voix de femme l'était beaucoup plus, surtout pour les conditions 8, 4 et 2 canaux. A l'inverse, la manipulation du pitch et du timbre n'affectait pas la reconnaissance de la voix de femme, tandis que l'absence de l'information de timbre (condition timbre constant) affectait la reconnaissance de la voix d'homme.

Les analyses en fonction du genre du sujet, n'ont révélé aucune différence significative entre les sujets masculins et les sujets féminins en ce qui concerne la reconnaissance correcte des voix extrêmes d'homme et de femme.

ANALYSE DES TEMPS DE REACTION

L'analyse des temps de réaction nous montre qu'en condition originale, ils formaient une courbe en cloche le long du continuum (voir Figure 9.23). Ceci évoque une réponse psychophysique classique. Sur les parties du continuum où les sujets étaient sûrs d'eux, c'est-à-dire aux extrémités, les temps de réaction étaient courts. Là où la voix devenait plus ambigüe, autour de 50% du continuum (seuil moyen en condition originale = 53,8%, voir paragraphe « L'analyse du seuil » p180), les sujets étaient plus hésitants, incertains, et mettaient donc un peu plus de temps à répondre. Les différences entre les TR aux deux extrémités et le TR moyen du pic (à 50% du continuum) étaient significatives ($p < 0,001$). Avec le vocoder, la distribution des TR moyens le long du continuum avait tendance à s'aplatir, comme on peut le voir sur la Figure 9.23. Les temps de réactions aux extrémités semblaient s'allonger. Cependant, la variabilité interindividuelle était telle qu'aucune différence significative n'a été retrouvée entre les différentes conditions de vocoder.

Lors de la tâche de catégorisation en conditions de pitch constant et timbre constant, les TR étaient particulièrement semblables à ceux observés en condition originale. Aucune différence significative n'a été retrouvée.

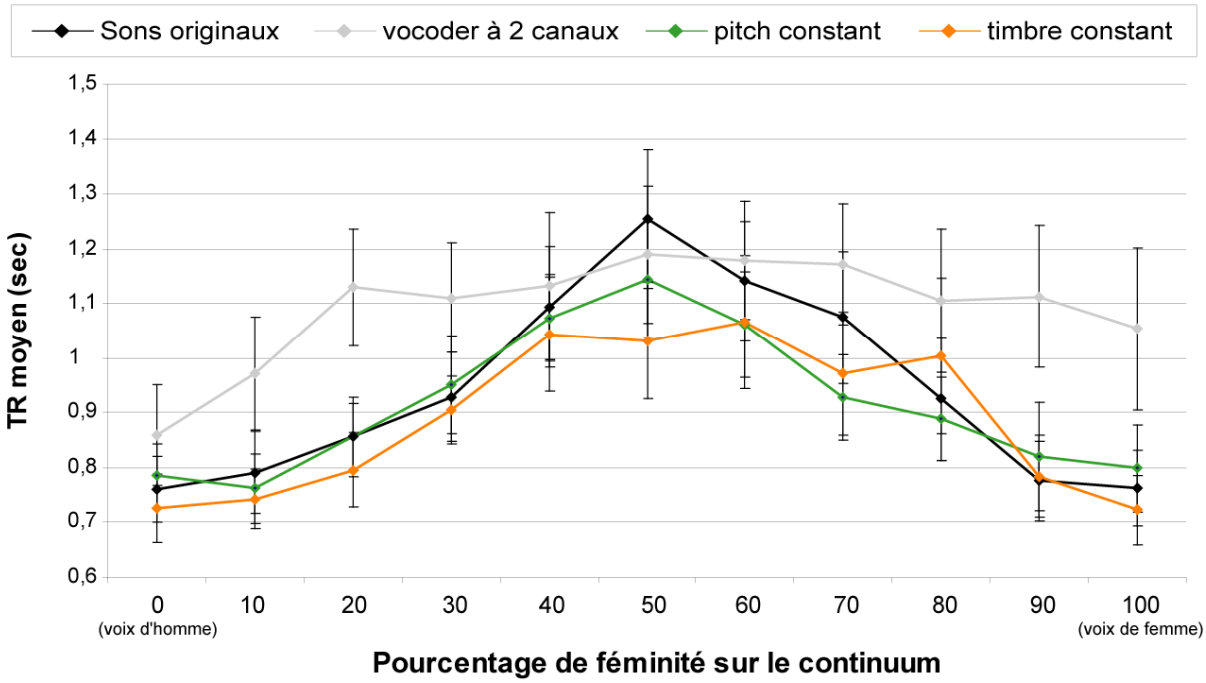


Figure 9.23 : Temps de réaction moyen (±SE) pour les conditions originale, vocoder à 2 canaux, timbre constant et pitch constant. Les résultats sont présentés pour 14 sujets, bien que les données d'un sujet soient manquantes pour les conditions de vocoder 8, 4 et 2 canaux. On observe une courbe en cloche pour la condition originale, avec des TR courts aux extrémités, et plus longs au milieu du continuum, près du point d'inflexion. Bien qu'il y ait une tendance à l'aplatissement de la courbe avec le vocoder, aucune différence significative n'a été retrouvée. Les conditions pitch constant et timbre constant ont des TR très similaires à la condition originale.

9.2.1.3 RESUME ET DISCUSSION PRELIMINAIRE

Avant tout dans pour ce test, on peut dire l'ensemble des critères pris comme indicateurs des performances étaient très cohérents. La pente, qui reflète la qualité de la catégorisation, et le seuil, qui reflète le percept ambigu du continuum, peuvent donc tout à fait bien représenter et résumer l'ensemble des variations.

Nos résultats ont montré que la tâche de catégorisation du genre était affectée par le vocoder surtout à partir de 16 canaux, avec une pente au point d'inflexion de la courbe significativement plus plate. En condition originale, les sujets étaient très sûrs d'eux pour les 3 voix, voire 4 voix, du côté de la voix féminine ainsi que du côté de la voix masculine, avec des réponses à 100% « femme » pour les unes et 100% « homme » pour les autres. Lorsqu'on dégradait la voix avec le vocoder, les sujets étaient beaucoup moins sûrs d'eux, et étaient rarement catégoriques dans la réponse, ce qui avait pour conséquence l'aplatissement de la courbe sigmoïde et une diminution de l'adéquation du modèle aux données. De plus on a pu remarquer que les voix d'hommes étaient moins affectées par la dégradation spectrale. Cette tendance à mieux reconnaître les voix d'hommes avait déjà été retrouvée dans différentes études (Coleman, 1971, 1976a; M. P. Gelfer & Mikos, 2005; Owren, Berkowitz, & Bachorowski, 2007; Whiteside, 1998) sans pour autant être davantage expliqué. Elliott

et Theunissen (Elliott & Theunissen, 2009), ont trouvé une région de modulation spectro-temporelle qui serait importante pour l'identification du genre. En effet, les filtres de modulation de fréquences comprises entre 3 et 7 cycles par kHz (soit entre 335 et 140 Hz environ) entraînaient une confusion des genres de la voix et particulièrement pour les voix féminines. Les auteurs concluent que les modulations comprises dans cette bande de fréquences permettent particulièrement de reconnaître le genre féminin. Ces résultats paraissent cohérents puisque c'est à peu près la bande de fréquence dans laquelle on retrouve en générale la fréquence fondamentale de la voix. Toutefois les auteurs se posent également la question de savoir pourquoi la reconnaissance des voix d'homme est moins affectée par les filtres de fréquences correspondants aux fréquences fondamentales de voix d'hommes.

Il ressort également de nos données que le pitch et le timbre sont tous deux des indices essentiels pour la catégorisation du genre. Ceci corrobore un certains nombres d'études portant sur les indices essentiels permettant la catégorisation du genre (Andrews & Schmidt, 1997; Aronovitch, 1976; Baumann & Belin, 2008; Coleman, 1976a, 1976b; M. P. Gelfer & Mikos, 2005; Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995). Toutefois dans l'ensemble, ces études rapportent que l'indice acoustique sur lequel se basent le plus les sujets est le pitch, et qu'en second lieu, les autres indices importants appartenaient au timbre, avec notamment la répartition des formants. Or, nos résultats montrent que, particulièrement pour les sujets féminins, le timbre permet une très bonne catégorisation du genre, aussi bonne que lorsque le pitch est présent et meilleure que lorsque qu'il n'y a que le pitch. On peut donc dire que si le pitch est, d'après la littérature, un très bon indicateur du genre du locuteur, le timbre soit un indicateur au moins aussi important, si ce n'est plus pour les sujets féminins.

Cette différence entre les sujets masculins et féminins est également un point intéressant. Il semblerait en effet que les femmes aient tendance de façon générale à mieux catégoriser le genre de la voix. Cette différence avec les hommes était très marquée et devenait significative lorsque seule l'information du timbre était présente. Les femmes étaient capables de catégoriser aussi bien qu'en condition originale. En revanche les variations de pitch seules, c'est-à-dire de fréquences fondamentales, ne permettaient pas aux hommes ni aux femmes de bien catégoriser le genre de la voix, puisqu'ils avaient tous deux des performances inférieures à la condition originale et similaires à la condition de vocoder à 16 canaux. Il semble donc que les hommes ne bénéficiaient pas de l'information du pitch ou du timbre seule, mais que la présence des deux informations soit nécessaires pour leur permettre de catégoriser correctement. L'habileté des femmes à percevoir ou traiter le timbre est peut-être due à une meilleure perception ou bien un meilleur traitement de l'information. Ce résultat semble être assez nouveau ; dans la littérature, la différence du genre sur la perception du genre n'a pas été étudiée. Toutefois de nombreuses études se sont intéressées aux différences de perception auditive et au traitement cérébral sous-jacent entre hommes et femmes. Dans l'ensemble il semblerait que le traitement cérébral auditif soit davantage bilatéral chez les femmes tandis qu'il serait plutôt latéralisé à gauche chez les hommes (Kansaku & Kitazawa, 2001; Kansaku, Yamaura, & Kitazawa, 2000; Kocak et al., 2005; Ruytjens et al., 2007). Or, nous savons

que le traitement du timbre dépendrait davantage de l'hémisphère droit (voir paragraphe « La neuropsychologie du timbre », p60). Il se pourrait donc que cette activation davantage bilatérale chez les femmes puisse leur permettre un meilleur traitement du timbre, tandis que les hommes, activant moins facilement l'hémisphère droit lors du traitement auditif, pourraient moins bien analyser ces indices, surtout si la tâche n'est pas explicitement indiquée comme un traitement du timbre. Toutefois, cette hypothèse nécessite d'être vérifié et étudié de manière plus approfondi.

9.2.2 TEST DE DISCRIMINATION ENTRE DEUX VOIX (DU CONTINUUM)

Ce test de discrimination entre deux voix a été fait à partir des stimuli générés avec les morphing entre la voix d'homme et la voix de femme, utilisé pour le test de catégorisation du genre. Ces stimuli nous ont permis de tester la discrimination de deux voix, dans 3 parties différentes du continuum ; la première concernant les voix d'hommes, la seconde les voix de femme et la dernière concernant les voix intermédiaires.

9.2.2.1 MATERIEL ET METHODES

SUJETS

Les sujets ayant participé à ce test étaient les mêmes que ceux qui avaient participé au test de catégorisation du genre. Il y avait donc 14 sujets volontaires, naïfs et n'ayant rapporté aucun trouble auditif, neurologique ou psychiatrique. Le groupe était composé de 7 hommes et la moyenne d'âge était 24.63 (+/-2,91). Un sujet n'a pas pu passer les conditions 8, 4 et 2 canaux, ainsi que timbre constant et pitch constant.

STIMULI

Pour ce test nous nous sommes servi des mêmes continua que pour le test de catégorisation du genre ; un continuum faisant varier le pitch et le timbre (voir Figure 9.15, test 1, p176), un autre faisant varier uniquement le timbre (voir Figure 9.15, test 2, p176) et le dernier faisant varier seulement le pitch (voir Figure 9.15, test 1, p176). Le premier continuum a été utilisé en condition originale, et en 6 conditions de vocoder, 64, 32, 16, 8, 4 et 2 canaux. Nous avions donc au total, 9 conditions de stimuli ; une condition originale ou le pitch et le timbre variaient, puis 6 conditions de vocoder ou le pitch et le timbre variaient, ensuite, une condition où seul l'indice de timbre variait et enfin une condition où il n'y avait que l'indice du pitch qui variait. Pour chaque continuum généré nous avons considéré le continuum sur 3 parties (voir Figure 9.24). La première (partie « Homme ») concernait le côté des voix d'homme, de la première à la 250ème voix. La seconde concernait les voix intermédiaires (partie « Intermédiaire »), de la 376ème à la 625ème voix. Enfin la dernière partie concernait les voix du côté des voix de femme (partie « Femme »), de la 750ème à la dernière voix.

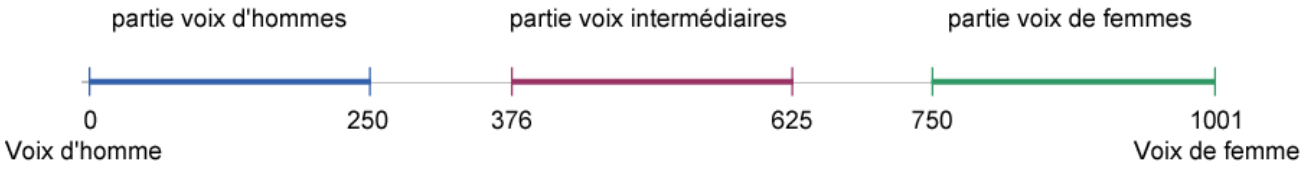


Figure 9.24 : Représentation des 3 parties du continuum utilisées pour le test de discrimination entre deux voix. 3 parties distinctes ont été considérées : les voix d'hommes de la 1ère à la 250ème voix, les voix intermédiaires de 376ème à 625ème voix et les voix de femmes de 750ème à dernière voix.

PROCEDURE

Les sujets étaient testés dans une pièce calme du laboratoire du CerCo. Les stimuli étaient présentés binauralement au travers d'un casque (modèle Eh 250, Sennheiser), dans un ordre aléatoire à un volume sonore ajusté à 72 dB SPL. Tous les sujets ont passé le test avec les sons originaux, dans 6 conditions de vocoder, 64, 32, 16, 8, 4 et 2 canaux ainsi qu'en conditions de pitch constant et de timbre constant. Les conditions ont été passées sur deux jours. Le premier jour les sujets passaient les conditions de vocoder 64, 32 et 16 canaux par block dans un ordre aléatoire et terminait toujours pas la condition originale. Puis, afin de compléter les données et de permettre des analyses statistiques plus robustes, les sujets on été revus une seconde fois pour passer les conditions de vocoder 8, 4 et 2 canaux et les conditions pitch constant et timbre constant, toujours dans un ordre aléatoire. Les sujets étaient placés devant un clavier avec le casque (modèle Eh 250, Sennheiser). La distance entre les deux sons était déterminée par une méthode de procédure adaptative ; lorsque pour 2 présentations de voix où la distance était la même, le sujet répondait correctement (i.e. « différent »), la distance diminuait. Si les sujets se trompaient une fois, la distance entre les voix de l'essai suivant (3 présentations de deux voix) augmentait. Lorsque la distance entre les deux voix présentées à chaque essai augmentait et diminuait (en fonction des réponses des sujets) 7 fois consécutives, la mesure de la distance moyenne pour que le sujet entende une différence, était calculée en moyennant les distances des 7 derniers essais. Cette procédure adaptative était la même pour les 3 parties du continuum (homme, intermédiaire et femme) et les essais concernant chacune de ces parties étaient présentés aléatoirement. La consigne pour les sujets était de dire le plus précisément et le plus rapidement possible, si les deux voix qu'ils avaient entendues étaient identiques ou différentes. Pour cela ils répondaient à l'aide des deux touches « contrôle » du clavier, « Pareil » ou « Différent », dont l'ordre était contrebalancé entre les sujets. S'ils le désiraient, les sujets pouvaient faire une petite pause entre chaque block, qui durait environ 5 minutes. La présentation des stimuli ainsi que l'enregistrement des réponses et des temps de réactions de sujets étaient réalisés à l'aide du logiciel Matlab 6.5®.

ANALYSES

Nous avons analysé l'effet des différentes manipulations de vocoder, pitch et timbre sur la distance minimum et sur les temps de réactions. Cet effet a été analysé sur les 3 parties du continuum. Dans un second temps nous avons également regardé s'il existait un effet du genre du sujet sur la discrimination entre 2 voix. Pour cela, nous avons comparé la distance minimum et les temps de réaction entre les sujets féminins et masculins.

9.2.2.2 RESULTATS

LA DISTANCE NECESSAIRE POUR PERCEVOIR UNE DIFFERENCE ENTRE 2 VOIX

L'analyse de la distance minimum sur le continuum pour que les sujets entendent une différence est présentée sur la Figure 9.25. La représentation de ces distances par rapport au continuum est présentée sur la Figure 9.26. L'analyse révèle que les manipulations de vocoder, de pitch et de timbre affectaient significativement les performances des sujets, mais uniquement pour les voix intermédiaires et les voix de femmes (voir la Figure 9.25 et la Figure 9.26 ; Friedman : $p < 0{,}0004$). Pour les voix d'homme, on ne retrouvait aucune différence significative au niveau des performances entre les différentes conditions de stimulation. En revanche, les voix intermédiaires et les voix de femmes étaient significativement affectées par ces différentes conditions (Friedman ; $p < 0{,}0004$). Les analyses 2 à 2 montraient pour ces deux types de voix que les distances de discrimination étaient significativement plus petites dans les condition originale, 64 et 32 canaux de vocoder que pour les conditions de stimulation 8, 4 et 2 canaux de vocoder (Wilcoxon ; $p < 0{,}05$). A 16 canaux, pour les voix de femme, la distance était significativement plus courte que pour les 3 premières conditions (Wilcoxon ; $p < 0{,}003$).

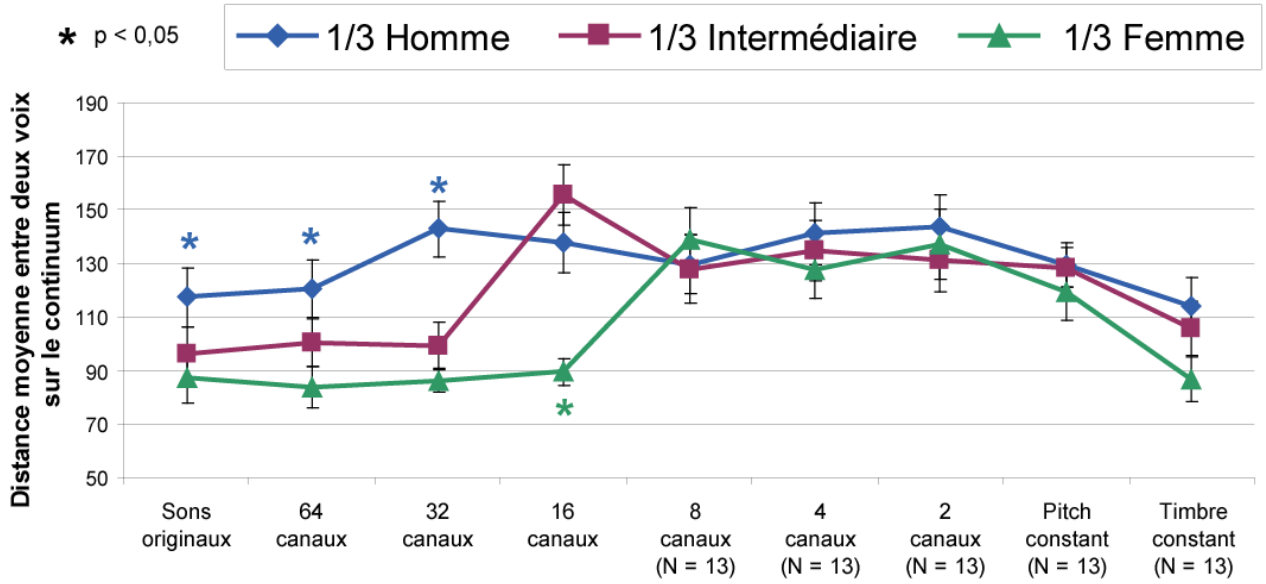


Figure 9.25 : Distances minimum moyenne (±SE) pour discriminer deux voix du continuum, pour les 3 différentes parties du continuum (voies d'hommes, voix intermédiaires et voix de femmes) et pour toutes les conditions de stimulation (sons originaux, 64, 32, 16 8, 4 et 2 canaux de vocoder, pitch constant et timbre constant). Aucun effet n'a été retrouvé sur les voix d'hommes. Pour les 3 premières conditions, la distance pour les voix de femmes et intermédiaires étaient significativement plus courtes que pour les voix d'hommes. La distance de discrimination des voix intermédiaires augmentait à 16 canaux pour devenir similaire à la distance des voix d'hommes, tandis que pour les voix de femmes cette distance n'augmentait qu'à partir de 8 canaux. Aucune différence entre les 3 types de voix n'a été retrouvée pour les conditions 8, 4 et 2 canaux de vocoder, ainsi que pour les conditions pitch constant et timbre constant. La condition pitch constant était similaire aux 3 dernières conditions de vocoder tandis que la condition timbre constant était similaire à la condition originale.

La comparaison de la distance de discrimination entre les différente parties du continuum (homme, intermédiaire et femme) nous montrait que dans la condition originale, les performances de discrimination pour les voix intermédiaires et les voix de femmes étaient similaires entre les deux conditions et significativement plus courtes que la distance minimum pour les voix d'hommes (Wilcoxon ; $p < 0{,}0052$). On retrouvait des résultats similaires pour les conditions 64 et 32 canaux de vocoder. Autrement dit, pour ces 3 premières conditions les sujets percevaient mieux la différence entre deux voix lorsqu'elles se situaient sur la partie intermédiaire et la partie femme du continuum que sur la partie homme. A la condition 16 canaux de vocoder, la distance de discrimination pour les voix intermédiaires augmentait pour devenir similaire à celle des voix d'hommes et significativement supérieure à celle des voix de femmes (Wilcoxon ; $p < 0{,}001$). Puis, de 8 à 2 canaux, la distance de discrimination pour les voix de femmes augmentait pour devenir similaire à celle des voix d'hommes et intermédiaires. Pour ces 3 dernières conditions de vocoder, la distance minimum entre deux voix pour les discriminer semblait atteindre un plafond et était similaire entre les 3 types de voix.

Dans la condition où seule l'information de timbre était présente (pitch constant), on peut voir sur la Figure 9.25 et la Figure 9.26 que les performances étaient similaires pour les 3 types de voix (femme, homme et intermédiaire) et que ces performances étaient similaires à celles observées à 8, 4 et 2 canaux de vocoder. En revanche lorsque seul le pitch était présent (timbre constant), bien qu'aucune différence n'ait été révélée entre les 3 types de voix par les analyses statistiques, les performances étaient similaires à celles observées pour les conditions originales, 64 et 32 canaux de vocoder. De plus pour les voix de femmes et intermédiaires, les performances étaient meilleures que pour les conditions de vocoder 8, 4 et 2 canaux de vocoder (Wilcoxon ; $p < 0{,}05$).

Enfin, aucune différence n'a été retrouvée entre les sujets masculins et féminins.

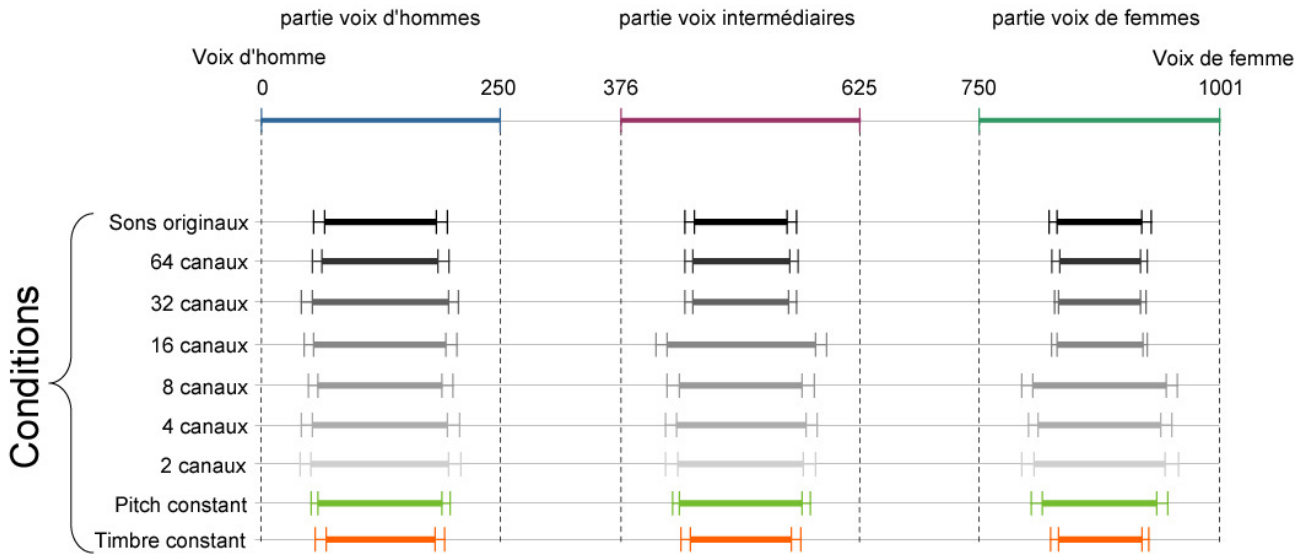


Figure 9.26 : Représentation par rapport au continuum des distances minimum moyenne (±SE) pour discriminer deux voix du continuum, pour les 3 différentes parties du continuum (voix d'hommes, voix intermédiaire et voix de femmes) et pour toutes les conditions de stimulation (sons originaux, 64, 32, 16 8, 4 et 2 canaux de vocoder, pitch constant et timbre constant.

En résumé, on peut dire qu'en condition originale, les performances de discrimination entre deux voix sont meilleures en ce qui concerne la dégradation du son par le vocoder, mais seulement pour les voix de femmes et intermédiaires. Il semble que les performances ne soient affectées qu'à partir de 16 canaux pour les voix intermédiaires et 8 canaux pour les voix de femmes puis restent stables jusqu'à la condition la plus dégradée (2 canaux de vocoder). Comme pour les tests de catégorisation du genre et de détection de la voix, les conditions 64 et 32 canaux semblaient assez bien préserver les performances des sujets, excepté pour les voix d'hommes ; au-delà on voyait les performances se dégrader. Par ailleurs les résultats nous ont montré que pour ce test de discrimination de voix, les sujets semblaient bénéficier davantage des variations de pitch (timbre constant), que de l'information du timbre. En effet dans la condition timbre constant, les performances étaient similaires à la condition originale.

L'interprétation de ces résultats en fonction du pitch paraît intéressante. Bien que, comme cela a déjà été expliqué, ce morphing faisait varier à la fois les informations de pitch et de timbre, les informations de timbre restent difficilement quantifiables. En revanche, étant donné que le pitch variait de façon linéaire entre 126 et 212 Hz, on peut clairement quantifier entre quelles valeurs de pitch se trouvaient les parties des voix d'hommes de femmes et intermédiaires en condition originale. Le pitch des voix d'hommes variaient dans une gamme de 126 à 147,5 Hz, tandis que les voix intermédiaires variaient entre 158,25 et 179,75Hz et les voix de femmes entre 190,5 et 212 Hz. A partir de là nous avons vu qu'en condition originale les sujets était meilleurs pour discriminer deux voix de femmes ou intermédiaire alors que la taille de la gamme de fréquence dans laquelle variait les pitches des différents types de voix était la même. Si la discrimination entre deux voix était faite uniquement sur la base du pitch, on observerait un résultat inverse. En effet, on sait que plus la fréquence est élevée, plus l'écart entre deux fréquences doit être important pour que l'on entende une différence (Attneave & Olson, 1971). Toutefois il est vrai que la gamme de pitch du continuum était assez restreinte à l'échelle de ce que peut entendre l'oreille humaine. Il est probable que cet effet soit minime dans la gamme que nous utilisions. Nos résultats en condition originale confirment donc que d'autres indices que le pitch étaient utilisés pour discriminer deux voix. Au fur et à mesure que l'on dégrade le signal par le vocoder en réduisant le nombre de canaux, on dégrade aussi les performances pour les voix de femmes et intermédiaires. Le profil de dégradation des performances s'explique essentiellement par la dégradation du pitch. En effet, comme nous l'avons vu, les performances pour les voix intermédiaires diminuaient significativement à partir de 16 canaux. Or, c'est précisément pour cette condition de vocoder que la gamme de pitch des voix intermédiaires se retrouvait dans la même bande de fréquences du vocoder que celle des voix d'hommes (voir Tableau 14.1). A 16 canaux les pitches des voix de femmes étaient dans une bande de fréquence du vocoder distincte de celle des pitches des deux autres types de voix. Ceci explique donc pourquoi les performances à 16 canaux restaient stables pour les voix de femmes et chutaient pour les voix intermédiaires ; leur pitch se confond avec celui des voix d'hommes. A 8 canaux, la même explication peut être avancée pour les voix de femmes. Pour cette condition, les pitches de tous les types de voix, celles de femmes y compris, se retrouvaient dans la même bande de

fréquence. L'indice de pitch qui était disponible jusqu'alors ne l'était donc plus. L'indice de pitch semble d'autant plus pertinent pour cette tâche de discrimination de voix lorsqu'on s'intéresse aux conditions pitch constant et timbre constant. On peut voir que lorsque seul l'indice de pitch était présent (condition timbre constant) les performances étaient similaires à celles obtenues en condition originale. Ceci confirme donc bien l'importance de cet indice pour cette tâche. A l'inverse lorsque le pitch était constant et que la tâche pouvait être réalisée seulement sur les variations de timbre les performances étaient aussi mauvaises qu'à 8, 4 et 2 canaux de vocoder. Autrement dit l'indice qui prédomine pour réaliser cette tâche de discrimination de voix est celui du pitch.

LES TEMPS DE REACTION

En ce qui concerne l'analyse des temps de réaction, nous observons une légère augmentation des temps de réaction avec la dégradation du signal par le vocoder mais cette tendance n'était pas significative (voir Figure 9.27). Les analyses n'ont révélé aucune différence significative entre les différentes conditions ni entre les différents types de voix.

De même, aucun effet du genre du sujet sur les temps de réaction n'a été retrouvé.

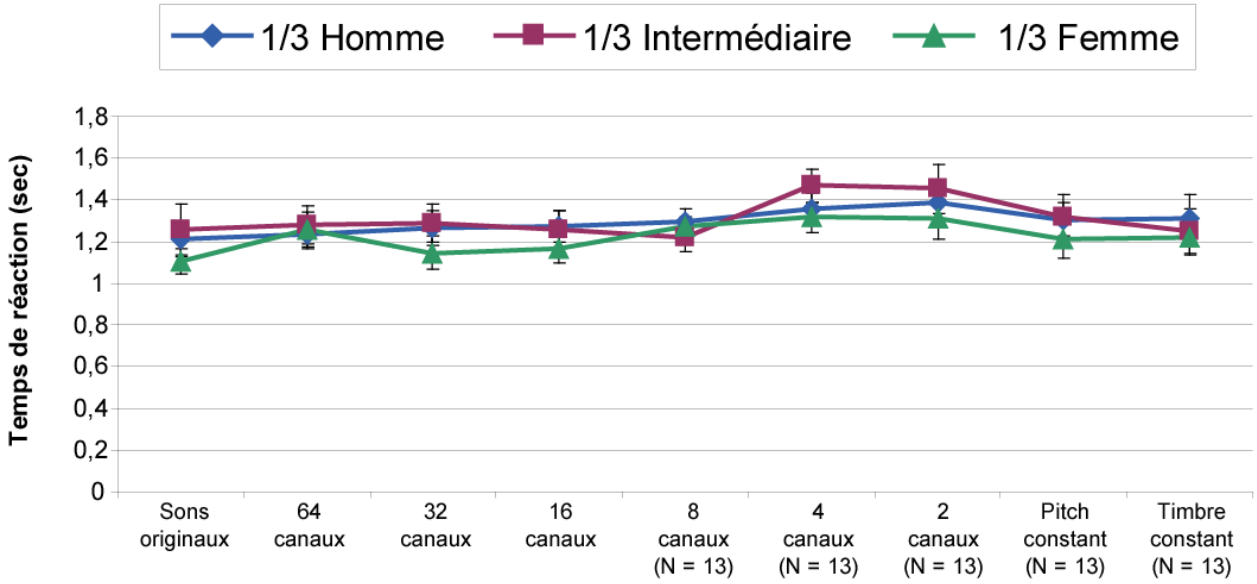


Figure 9.27 : Temps de réaction moyens (±SE) pour discriminer deux voix du continuum, pour les 3 différents tiers du continuum (« Homme », « Intermédiaire » et « Femme ») et pour toutes les conditions de stimulation (sons originaux, 64, 32, 16 8, 4 et 2 canaux de vocoder, pitch constant et timbre constant. Aucun effet significatif n'a été retrouvé entre les différentes conditions, ni entre les différentes parties du continuum.

9.2.2.3 RESUME ET DISCUSSION PRELIMINAIRE

Les résultats de ce test de discrimination entre deux voix, nous montrent plusieurs choses. D'une part le vocoder affecte les capacités des sujets à discriminer deux voix comme étant identiques ou différentes, et ce, uniquement à partir de 16 canaux. Ce résultat semble cohérent avec celui du test de catégorisation du genre où, finalement, il semble que le vocoder n'affecte pas les performances avant 16 canaux. D'autre part, l'information de pitch semblait être un meilleur indice

pour la tache de discrimination, puisque les performances étaient meilleures en condition timbre constant. Dans le test de catégorisation du genre de la voix nous avons observé des résultats inverses, où le timbre semblait apporter davantage d'information. Il est intéressant de voir que pour ces deux types de tâches, les sujets ne vont pas tirer autant d'information des mêmes indices acoustiques pour les réaliser. Pourtant, il s'agit bien des mêmes stimuli. On peut donc penser que ces différentes tâches en s'appuyant sur des informations différentes, impliquent différents mécanismes de traitement cérébraux et par conséquent, différents supports neuronaux. Cette idée est appuyée par le fait que déjà en 1987, Van Lancker et Freiman soutenaient que les tâches de reconnaissance (familiarité) de la voix et de discrimination étaient deux tâches bien distinctes et impliquaient même des substrats corticaux différents (D. Van Lancker & Kreiman, 1987).

En ce qui concerne l'absence de différences entre les sujets féminins et masculins, cela peut s'expliquer par le fait que dans ce test, l'information du timbre ne semble pas un élément déterminant pour réaliser la tâche. Or nous avons vu que les femmes ont des facilités lorsque le timbre est impliqué comme dans le test de catégorisation. Ce test nous montre que les sujets féminins ne semblent pas avoir d'habiletés particulières concernant le traitement du pitch.

9.2.3 TEST DE LA RECONNAISSANCE DE L'AGE

Ce test de reconnaissance de l'âge était inclus dans le projet de stage de première année de master de Sophie OUTH-GAUER, étudiante en 3ème année de médecine. Ce stage à eu lieu au CerCo, de septembre 2009 à janvier 2010 et a été encadré par Mathieu MARX et moi-même, doctorant. Le projet de Sophie OUTH-GAUER appartenait au projet plus large de l'étude de la perception de la voix chez les patients sourds implantés cochléaires, aussi les résultats de ce test ont été inclus dans ce travail de thèse. L'ensemble de ce test à donc été mené par Sophie OUTH-GAUER me permettant ainsi de procéder à l'analyse des résultats.

9.2.3.1 MATERIEL ET METHODES

SUJETS

Dans cette étude un groupe de 22 sujets, dont 9 hommes, a été inclus. La moyenne d'âge du groupe était 22,2 (+/-7,2). Le groupe a été scindé en deux afin de limiter les biais liés à l'habituation et la fatigue. Un premier groupe de 11 sujets (Groupe 1), composé de 5 hommes, a passé les conditions voix originales ainsi que 32, 16 et 8 canaux et le second groupe (Groupe 2) a passé les conditions originales et 8, 4 et 2 canaux de vocoder.

STIMULI

Les stimuli originaux pour le test de la reconnaissance de l'âge étaient des phrases fournies par Cochlear® et servant dans le test d'audiométrie vocale *Hearing In Noise Test*, le HINT (Nilsson,

Soli, & Sullivan, 1994). Ces phrases étaient équilibrées en longueur et en fréquence de mots. Elles ont été enregistrées auprès des 3 populations concernées. Ces enregistrements ont été faits à l'aide d'un micro et du logiciel PRAAT sur un PC (système d'exploitation Windows XP). Les voix d'enfants ont été enregistrées chez des enfants âgés de 5 à 9 ans, alors qu'ils étaient présents dans le centre de loisirs de l'Ecole Nationale d'Aviation Civile. Les enregistrements ont été faits sur une seule journée dans une pièce calme. L'accord écrit des parents avait été donné pour chaque enfant. Les voix de seniors ont été enregistrées auprès de personnes de plus de 65 ans. Les volontaires ont été trouvés auprès de personnes accompagnant un patient dans le service de « Neurologie vasculaire, pathologie neuro-dégénérative et explorations fonctionnelles du système nerveux » de Purpan pour une consultation avec le Dr Jérémie Pariente, ou venant pour une consultation dans le service d'ophtalmologie de l'hôpital de Purpan. Les enregistrements ont eu lieu dans une pièce insonorisée du service ORL ou du service de Neurologie de Purpan. Enfin, le groupe d'adulte, qui a été enregistré dans la pièce insonorisée du service ORL, consistait en des volontaires sains.

Pour tous les stimuli, les bruits de fond ont été filtrés, et les amplitudes ont été égalisées à l'aide du logiciel Matlab 6.5® (Mathworks, Inc.).

Au total 4 sets de phrases avaient été constitués, chaque groupe de sujets ne passant que 4 conditions. Chaque set comprenait 18 phrases ; 6 étaient prononcées par un adulte, 6 par un enfant, et 6 par un senior. 5 phrases par conditions étaient prononcées par un locuteur féminin, avec au moins une voix féminine pour chaque type de voix (i.e., enfant, adulte et senior). Un même locuteur ne prononçait pas plus de 2 phrases dans un set de stimuli. Le pitch moyen pour chaque set de stimuli était équivalent (voir Tableau 9.4). Afin de s'assurer que le contenu des sets de stimuli était équivalent, nous avions préalablement vérifié que le niveau de difficulté de reconnaissance de phrase pour chaque set était similaire ($p < 0.05$). La répartition des sets pour les différentes conditions est présentée dans le Tableau 9.4.

	Pitch moyen	Groupe 1	Groupe 2
Set 1	192,3 Hz	Condition originale	Condition originale
Set 2	194,7 Hz	32 canaux	4 canaux
Set 3	192,8 Hz	16 canaux	2 canaux
Set 4	193,7 Hz	8 canaux	8 canaux

Tableau 9.4 : Répartition des différents sets de phases en fonction des différentes conditions présentées aux sujets.

PROCEDURE

Tous les sujets ont passé le test dans une pièce calme, en présence de l'expérimentateur. La passation de toutes les conditions durait environ 20 minutes mais les sujets pouvaient faire une petite pause entre chaque condition s'ils le désiraient. Les conditions de vocoder étaient passées dans un ordre aléatoire et la condition originale était toujours présentée en dernier. Pour chaque condition

(originale, 2, 4, 8, 16 ou 32 canaux de vocoder), le test se déroulait de la manière suivante : le sujet était placé face à un écran avec le casque (modèle HD 280 Pro, Sennheiser) et une souris et devait fixer une croix de fixation jusqu'à la présentation du stimulus auditif (voir Figure 9.28). Les stimuli étaient présentés dans un ordre aléatoire. Une fois le stimulus présenté, le sujet avait pour consigne de ne prêter aucune attention au contenu de la phrase et de dire si la voix qu'il avait entendu était celle d'un enfant, d'un adulte ou d'un sénior. Pour répondre le sujet cliquait sur la partie de l'écran correspondant à sa réponse, adulte, enfant ou senior. Nous enregistrions le nombre de reconnaissances correctes pour chaque type de voix dans chaque condition.

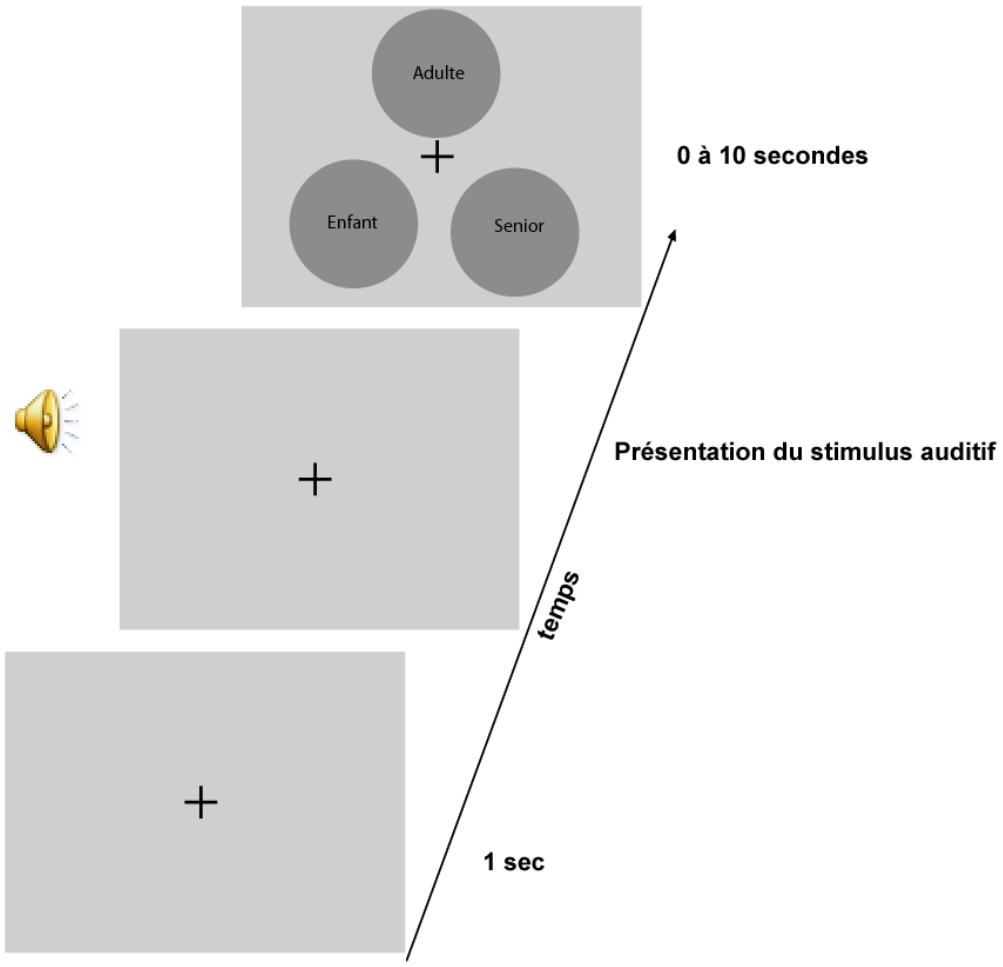


Figure 9.28 : Design expérimental dans le test de reconnaissance de l'âge pour un essai. Chaque condition comprend 18 phrases (6 prononcées par un adulte, 6 par un enfant et 6 par un senior)

ANALYSES

Nous avons relevé pour ce test les taux de reconnaissance correcte pour chaque catégorie de stimuli ; adulte, enfant et sénior. Nos analyses statistiques ont donc concerné l'effet du vocoder sur les taux de reconnaissances pour les stimuli Adulte, Enfant et Senior. Ces tests ont été réalisés indépendamment pour les groupes 1 et 2. Des tests ont également été pratiqués pour comparer les

performances du groupe 1 avec celles du groupe 2, dans les mêmes conditions (VO et 8 canaux). Enfin, nous avons analysé les différences des taux de reconnaissance entre chaque condition pour les voix Adulte, Enfant et Senior. Dans le but de mieux comprendre les résultats précédents nous avons regardé si le genre du locuteur avait une influence sur le taux de reconnaissance.

9.2.3.2 RESULTATS

EFFET DU VOCODER SUR LA RECONNAISSANCE DE L'AGE

Pour le groupe 1, qui est passé en condition originale et condition de vocoder à 32, 16 et 8 canaux, nous avons retrouvé un effet significatif ($p < 0,005$) du vocoder sur la reconnaissance des voix d'adulte et d'enfant (voir Figure 9.29). Pour la reconnaissance des voix Senior, on remarque, même si elle n'est pas significative, une tendance à la diminution des performances avec la diminution du nombre de canaux pour le vocoder. Cette absence de significativité peut s'expliquer par un échantillon insuffisant. Les tests post hoc montraient que l'effet significatif du vocoder pour les voix Enfant et Adulte s'expliquait essentiellement par la différence entre la condition originale et les conditions vocodées, surtout avec 8 canaux. En effet, pour les voix Adulte, les différences se situent entre les conditions originales et les conditions de vocoder à 32, 16 et 8 canaux, ainsi qu'entre 32 et 8 canaux ($p < 0,05$). Pour les voix Enfant, les performances en condition vocodée à 8 canaux sont significativement plus faibles que pour les 3 autres conditions. Les performances en condition originales sont également meilleures qu'à 32 canaux. La condition 16 canaux n'est pas différente des conditions originale et 32 canaux.

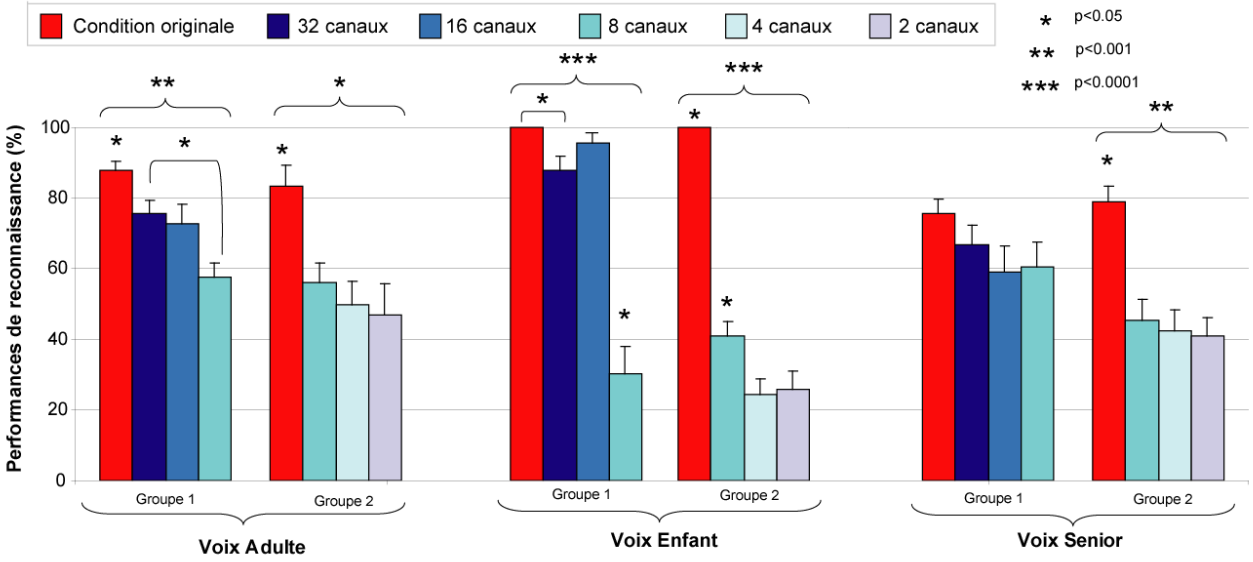


Figure 9.29 : Effet du vocoder sur la reconnaissance de l'âge, pour les voix Adulte, Enfant et Senior. Les résultats sont présentés séparéments pour les deux groupes de sujets. L'effet du vocoder diminue significativement la reconnaissance de la voix, pour les deux groupes, sauf les voix Senior pour le groupe 1.

Pour le groupe 2 (condition originale, 8, 4 et 2 canaux de vocoder), nous avons également observé un effet significatif du vocoder ($p < 0,0072$) sur les performances de reconnaissance de l'âge (voir Figure 9.29). Les test post-hoc ont montré que les performances dans les différentes conditions de vocoder ne sont pas significativement différentes entre elles, excepté pour les voix Enfant, où les performances à 8 canaux étaient significativement supérieures à 4 et 2 canaux. Ceci dit on note tout de même une tendance à la diminution des performances avec la diminution du nombre de canaux pour les voix Adulte et Senior.

Lorsque nous avons comparé les deux groupes sur les conditions qu'ils avaient tous les deux passé (condition originale et 8 canaux), nous n'avons retrouvé aucune différence significative. Ceci nous a permis, dans les résultats suivants, de comparer les performances pour les voix Adulte, Enfant et Senior en fusionnant les données des groupes 1 et 2 pour les analyses statistiques.

Pour les conditions originale, 32 et 16 canaux, les voix d'enfant étaient mieux reconnus que les voix d'adultes et de senior. Cette différence était significative ($p < 0,0008$) sauf à 32 canaux, probablement en raison de la variabilité. Toutefois on note que les voix Enfant restaient supérieurement reconnues pour cette condition aussi. A 8 canaux les tendances s'inversaient complètement. Alors que les performances pour les Voix Adulte et Senior continuaient de diminuer progressivement, les performances pour les voix Enfant chutaient brutalement pour devenir significativement très inférieur aux deux autres ($p = 0,018$). Cette infériorité se maintenait à 4 et 2 canaux, bien qu'elle ne soit significative que pour la condition 4 canaux ($p = 0,014$). Notons qu'aucune différence n'a été retrouvée entre les performances des voix Adultes et Senior.

On peut donc dire, à partir de ces résultats, qu'en condition originale, les voix d'enfants sont mieux reconnues que les voix d'adultes et de personnes âgées. Nous avons également pu montrer que la diminution du nombre de canaux augmentait significativement la difficulté de reconnaissance de l'âge. Toutefois, cette diminution était progressive pour les voix Adulte et Senior, mais les performances restaient au dessus du niveau de la chance même à 2 canaux. En revanche pour les voix Enfant, les performances étaient meilleures que pour les voix Adulte et Senior et la diminution des performances était plus tardive, à partir de 8 canaux, et abrupte, chutant au niveau de chance pour 2 et 4 canaux.

EFFET DU PITCH DE LA VOIX

Pour mieux comprendre ces résultats, qui semblent montrer que les voix d'enfants ont un statut particulier, nous nous sommes intéressés aux pitches de ces différentes voix. Il apparaît que, pour les échantillons de voix que nous avons utilisés les voix Enfant avaient un pitch moyen particulièrement plus élevé (264,5 Hz) par rapport aux voix Adulte et Senior (respectivement 156,7 et 159 Hz), et ce dans les 4 sets de stimuli utilisés (voir Figure 9.30). Ces résultats correspondent aux pitches moyens rapportés dans la littérature pour les mêmes types de voix (voir partie « Le pitch de la voix », p79).

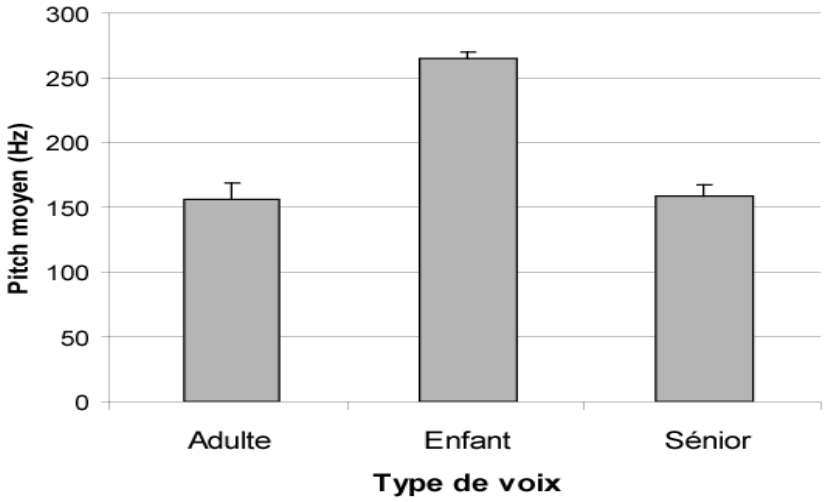


Figure 9.30 : Pitch moyens pour les voix Adulte, Enfant et Senior. Les 4 sets de phrases sont confondus car il n'existe pas de différences.

Ceci peut expliquer deux choses. La première, c'est la meilleure reconnaissance des voix Enfant en condition originale ; en se basant sur le pitch, les sujets pouvaient plus facilement distinguer les voix d'enfants qui étaient beaucoup plus aigües que les autres qui se ressemblaient plus en termes de pitch. La deuxième, c'est pourquoi les performances pour les voix Enfant chutent soudainement à partir de 8 canaux. En effet, pour 32 et 16 canaux, les bandes spectrales du vocoder contenant les pitches autour de 150 ou 160 Hz pour les voix Adulte et Senior sont bien distinctes des bandes spectrales contenant les pitches de 264 Hz des voix d'enfants (voir Tableau 14.1 en Annexes p315). Ainsi, les voix Enfants exciteraient des bandes spectrales bien distinctes jusqu'à 16 canaux, et il est toujours possible pour le sujet de se baser sur la hauteur (d'une bande spectrale) pour distinguer les voix d'enfants des autres. Ensuite, pour 8, 4 et 2 canaux, les pitch moyens de toutes les voix se retrouvaient dans la même bande spectrale du vocoder. Les sujets ne pouvaient plus différencier, du moins par le pitch, les voix Enfant des deux autres types de voix. Reste néanmoins que les performances chutent jusqu'au niveau de chance, il semblerait que les sujets de 8 à 4 canaux ne soient plus capables de reconnaitre les voix d'enfants.

En revanche, on peut s'étonner de la relative robustesse des voix Adulte et Senior à être reconnues. On peut comprendre qu'elles soient moins bien reconnues que les voix Enfant car leur pitches sont très proches, mais lorsque leur pitches, à partir de 8 canaux se retrouve dans la même bande spectrale que celui des voix d'enfants, on s'attendrait à voir aussi chuter les performances avec la soudaine difficulté. Or à 2, 4 et 8 canaux, ces voix ont toujours de meilleurs taux de reconnaissance que les voix d'enfants. Ces différences sont significatives pour 8 et 4 canaux mais la tendance est la même à 2 canaux. Il est donc probable que les sujets se basent sur d'autres indices, tels que les modulations temporelles de l'énergie spectrale ou encore le rythme et les variations d'amplitude, pour reconnaitre ces deux types de voix. En effet, nous avons vu dans la partie « Caractéristiques physiques de la voix » (voir p77), que le rythme, le timbre et les variations d'amplitude et de pitch, étaient des indices variant de manière significative en fonction de l'âge du

locuteur. De façon générale, l'écart de pitch est plus important entre les voix d'enfants et les voix d'adultes et de personnes âgées, ce qui fait du pitch un bon indicateur pour détecter les voix d'enfants. En revanche, les pitches de voix d'adultes et de personnes âgées étant plus proches, d'autres indices tels que l'amplitude, le timbre (notamment la répartition de formants) ou le rythme sont utiles voire nécessaires pour les discriminer. Nos données vont donc bien dans le sens de la littérature.

9.2.3.3 RESUME ET DISCUSSION PRELIMINAIRE

Ce test de reconnaissance de l'âge nous montre une fois de plus que le vocoder affecte les performances, ici dès 32 canaux pour les voix Adulte et Senior et à partir de 8 canaux pour les voix Enfant. Toutefois, on remarque que lorsqu'il est disponible, et notamment pour les voix d'enfants, le pitch est un indice important pour réaliser la tâche. Mais, si le pitch est sensible à la dégradation du vocoder, expliquant notamment la diminution pour les voix d'enfants, les sujets s'appuient également sur d'autres indices, probablement des indices de timbre, de rythme, d'amplitude et/ou de prosodie, qui sont plus robustes à la dégradation spectrale consécutive au vocoder et sont moins affectés par cette manipulation.

9.3 DISCUSSION

En premier lieu, il semble important de rappeler l'objectif concernant l'ensemble de ces tests chez les sujets normo-entendants. Le but était double : d'une part connaître l'effet de la dégradation spectrale sur la perception de la voix, et d'autre part savoir si l'amélioration de la résolution spectrale au-delà de ce qui est actuellement et techniquement possible, c'est-à-dire environ 22 électrodes (canaux), peut avoir un effet bénéfique sur la perception de la voix, lors de l'écoute de stimuli au travers d'un processeur d'implant cochléaire. Nous allons maintenant reprendre les résultats importants et les discuter.

9.3.1 SYNTHESE DES RESULTATS

Tout d'abord, l'ensemble des tests réalisés chez le sujet normo-entendant en condition de simulation d'implant révèlent plusieurs choses. En premier lieu, quelques soient les compétences testées - détection de la voix, catégorisation du genre, discrimination entre deux voix - on peut dire que le vocoder affectait les performances des sujets, qui diminuaient de façon graduelle. Plus le son était dégradé spectralement, c'est-à-dire, moins il y avait de canaux, plus les performances diminuaient. Toutefois, il convient de nuancer ces résultats. En effet, pour le test de détection de la voix (monaural ou binaural), il semble que la diminution soit plus progressive que pour le test de

catégorisation du genre où de façon générale on observe une diminution des performances jusqu'à 16 canaux, après quoi les différences significatives entre les conditions sont rares. Une conclusion importante doit être tirée de ce constat : à 16 canaux de résolution spectrale les résultats sont significativement plus faibles qu'en condition originale. A 64 et 32 canaux, les résultats sont le plus souvent meilleurs que toutes les autres conditions de vocoder et parfois similaires à la condition originale. On peut donc dire que l'amélioration de la résolution spectrale améliore les performances des sujets dans différentes tâches de perception de la voix.

On peut voir dans la littérature scientifique que la détection de la voix a été peu étudiée jusqu'à présent. Toutefois, au cours de ce test, nous avons cherché à évaluer les capacités des sujets au niveau voix le plus bas, du modèle de la perception de la voix, l'analyse structurelle de la voix. Nos résultats montrent que ce premier niveau d'analyse est déjà affecté par le vocoder. Toutefois nous devons noter que la tâche que nous utilisons est particulièrement difficile, dans la mesure où les stimuli sont très courts et très variables.

La perception du genre a en revanche été plus étudiée. S'il est difficile de comparer, parmi les critères que nous avons évalués, la pente, le seuil ou l'erreur, on peut en revanche s'intéresser aux performances de reconnaissances correctes de la voix d'homme et de la voix de femme (extrêmes du continuum). D'autres études ont également étudié ces performances de reconnaissance correcte du genre de la voix. Notamment, Fu *et al.* ont étudié la reconnaissance du genre de la voix à partir de voix prononçant des syllabes similaires à celle que nous avons utilisé pour le morphing (Q. J. Fu, Chinchilla et al., 2004; Q. J. Fu, S. Chinchilla et al., 2005). Nos résultats en matière de reconnaissance la voix sont très similaires ; les performances de reconnaissances correctes sont proches de 100 % à 32 canaux de vocoder et de 90 % à 16 canaux. Les performances que nous retrouvons à 8 et 4 canaux semblent être un peu meilleures, mais cela peu s'expliquer par le fait que nous utilisons un échantillon de chaque genre pour mesurer les performances.

Dans nos résultats, nous avons montré que l'amélioration de résolution spectrale améliorait la détection de la voix parmi les sons de l'environnement, de même que la catégorisation du genre ou de l'âge. Une récente étude chez les sujets normo-entendants a aussi étudié l'impact de la résolution spectrale sur l'indentification de stimuli environnementaux, et a permis de montrer que l'augmentation du nombre de canaux jusqu'à 32 canaux pouvait considérablement améliorer les performances des sujets (Shafiro, 2008). Il semble donc que, bien que les techniques actuelles ne permettent pas encore de dépasser environ 22 électrodes pour stimuler la cochlée sans créer d'interférence entre électrodes, cette piste ne semble pas vaine, en termes de bénéfices perceptifs pour les utilisateurs d'implant cochléaire. Ce bénéfice toucherait particulièrement la perception de la voix ainsi que les sons environnementaux.

9.3.2 LATERALITE DU TRAITEMENT DE LA VOIX

Un autre point important de ces résultats est que la détection de la voix ne semble pas facilitée par une oreille comme on aurait pu s'y attendre en fonction des résultats des études neurofonctionnelles, qui montraient une aire de la voix préférentiellement activée à droite. Ce résultat est assez nouveau au niveau comportemental. D'après les résultats présentés dans ce travail de thèse, cette tendance à la latéralisation à droite ne semble pas faciliter les détections de la voix, une analyse relativement bas niveau. Mais, la nature des tâches utilisées dans la plus part des études qui montraient un tendance à la latéralité de l'aire la voix, étaient assez diverses (voir partie « La voix, un visage auditif ? », p92),. Dans beaucoup d'études, l'aire de la voix, ou TVA, était activée lors d'une tâche portant sur l'identité ou la familiarité du locuteur. Ceci constitue donc une piste intéressante. Il se pourrait donc qu'on retrouve davantage un effet de cette tendance à la latéralité sur des tâches ayant trait à l'identité du locuteur, comme la tâche du genre. Cette hypothèse mérite d'être tester dans des expérimentations ultérieures.

De plus, nous pourrons rediscuter de cette latéralité dans la partie d'imagerie cérébrale. Dans cette partie nous étudierons la fonctionnalité de l'aire de la voix en condition de simulation d'implant.

9.3.3 LE STATUT PATICULIER DU LANGAGE

Nos résultats ont également montré le statut particulier du langage dans la voix. Nous avons pu voir ceci, à la fois dans le test de détection de la voix, monaural ou binaural, que les distracteurs soient des sons de l'environnement ou des voix animales. Ce résultat va dans le sens de ce que plusieurs auteurs ont déjà montré, à savoir que le langage avait un statut très particulier en termes perceptif. Tout d'abord, il semble que déjà chez les nouveau-nés il existe un biais attentionnel pour le langage (Vouloumanos, Hauser, Werker, & Martin, 2010; Vouloumanos & Werker, 2004, 2007). Ces résultats, nous amènent à penser que le langage bénéficie d'un traitement particulier. Les mêmes auteurs ont également montré chez l'adulte, que le signal vocal activait un réseau cortical particulier, et que lorsque ce signal vocal était langagier, une aire supplémentaire s'activait dans le gyrus frontal supérieur droit (Vouloumanos, Kiehl, Werker, & Liddle, 2001). On peut donc penser que le langage, en termes perceptifs, a un statut particulier. Ceci est probablement lié au fait que la source du langage (et de la voix) est unique (les cordes vocales et l'espace supralaryngé). Egalement, la structure du langage est elle aussi la même entre les individus. Il y a donc moins de variabilité dans la structure acoustique du langage, que dans toute autre source sonore.

Nous pouvons probablement considérer que les sons langagiers et non langagiers sont différemment affectés par le vocoder en raison de leurs différents contenus spectro-temporels. Les sons langagiers sont caractérisés par des changements rapides propres à l'articulation, et sont donc riches en modulations temporelles. Or nous avons déjà vu que les modulations temporelles sont

relativement bien préservées par le vocoder. On peut donc penser que, étant typiques du langage, lui-même étant particulièrement bien perçu, ces modulations temporelles sont un indice fort pour détecter la voix humaine. Au contraire, les sons vocaux non langagiers sont beaucoup plus hétérogènes en structure que les sons du langage, donc moins facilement reconnaissables sur la base d'une structure typique. De plus, la typicité de leur structure vocale est davantage basée sur le contenu fréquentiel qui est beaucoup plus affecté par la dégradation spectrale du vocoder ; leurs indices acoustiques typiques de la voix seront alors moins saillants.

En résumé, la différence de performances en faveur des sons langagiers en comparaison des sons non langagiers, s'explique probablement par une structure typique des sons langagiers qui résiste mieux à la dégradation par le vocoder, que celle des sons non langagiers et par un statut perceptif particulier de ces sons langagiers, procurant ainsi un indice saillant pour détecter la voix

Cette analyse est intéressante, car la voix ne véhicule pas toujours du langage, et les stimuli non langagiers sont plus difficiles à détecter, et par conséquent à analyser. Or, le test de catégorisation du genre ainsi que le test de catégorisation de l'âge ont été faits en utilisant des stimuli langagiers ; la syllabe « *had* » pour le test de catégorisation du genre et le test de discrimination, et des phrases pour le test de perception de l'âge. Pour ces tests nous utilisons donc des stimuli relativement riches du point de vue de la structure temporelle fine. II serait intéressant de voir si des résultats similaires seraient obtenus avec des stimuli non langagiers, tels que des rires, de la toux, des cris, etc.

9.3.4 DIFFERENTS INDICES ACOUSTIQUES POUR DIFFERENTES TACHES

Enfin, il semble que dans l'ensemble, les résultats montrent que, dans les différentes tâches que nous avons évaluées, les sujets n'utilisaient pas toujours le même type d'information comme indice principal. Par exemple, pour la détection de la voix, il semble que le pitch ne suffise pas à lui seul à expliquer les performances des sujets ; la structure temporelle fine, bien que la durée des stimuli fût très réduite (500 ms), aidait probablement beaucoup les sujets à réaliser la tâche. Dans le test de catégorisation il semble que le timbre ait été un indice plus important pour réaliser la tâche, notamment pour les sujets féminins, tandis que pour le test de discrimination de voix, le pitch semblait être un meilleur indicateur que le timbre. Enfin pour le test de l'âge, il semble que plusieurs indices acoustiques soient utilisés pour réaliser la tâche ; le pitch mais aussi le timbre, les variations d'amplitudes, etc. On peut donc dire à partir de ces résultats, que les différentes tâches réalisées par les sujets dans le cadre de la perception de la voix, reposent sur la perception et l'analyse de différentes informations acoustiques. On comprend donc ainsi la variabilité d'effet du vocoder sur les performances pour chacune de ces tâches. En outre, si ces habiletés reposent sur des indices acoustiques différents, cela implique également qu'ils reposent probablement sur des réseaux neurofonctionnels différents, comme le suggère le modèle de perception de la voix proposé par Pascal Belin (Belin et al., 2004).

9.3.5 CONCLUSION

En conclusion, l'ensemble de ces résultats nous amène à dire que le processeur de l'implant cochléaire, de part la dégradation spectrale du signal acoustique qu'il induit, est effectivement source de diminution des performances par rapport à une stimulation acoustique avec le signal original. La question qui se pose maintenant est de savoir ce qu'il en est chez les patients implantés cochléaires. En comparaison aux sujets normo-entendants, les patients bénéficient pour certains de l'effet de l'expérience et de l'entrainement avec un tel signal, qui peut mener à une habituation et une adaptation des stratégies de perception de la voix. Une autre différence avec les normo-entendants est la surdité en tant que telle, qui comme nous l'avons vu dans la partie théorique implique des réorganisations corticales plus ou moins importantes, de même que l'implantation cochléaire. La partie qui suit va nous permettre d'explorer les performances des patients dans le même type de tâche et de déduire l'effet de l'expérience et de la surdité sur les performances.

10 TESTS COMPORTEMENTAUX DE LA RECONNAISSANCE DE LA VOIX CHEZ LE PATIENT IMPLANTE COCHLEAIRE

Avant de commencer cette partie, nous souhaitons rappeler qu'un des objectifs de cette étude était de pouvoir comparer toutes les données obtenues chez le sujet sain à celles des patients. Ceci dans le but de pouvoir différencier, dans la perception de la voix, le déficit qui dépendait de la nature du signal transmis, c'est-à-dire au processeur, à la stratégie de l'implant, de celui qui dépendait du déficit lié à la surdité, c'est-à-dire à la réorganisation qu'elle implique. Toutefois l'acquisition des données chez les patients étaient plus difficile que chez le sujet normo-entendant, en termes de temps notamment. Sur l'ensemble des tests passés par les NHS, nous n'avons pas actuellement recueillis toutes les données pour tous ces tests chez les patients, et nous en poursuivons l'acquisition. Nous sommes actuellement en mesure de présenter les résultats chez les patients pour le test de détection de la voix parmi des sons environnementaux et pour le test de catégorisation du genre en condition original et en conditions de pitch constant et timbre constant. Des résultats préliminaires seront présentés pour le test de discrimination entre deux voix.

Les patients implantés cochléaires inclus dans cette étude, ont été recruté auprès du service ORL de Purpan. La réalisation de l'étude a exigé la rédaction d'un dossier exposant le déroulement et les buts du protocole. Ce dossier a été examiné par la Direction de la Recherche Clinique des Hôpitaux de Toulouse puis par le Comité de Protection des Personnes du Sud-Ouest et d'Outre-Mer qui a donné son autorisation pour inclure les sujets (n°08 161 03). Tous les patients inclus dans cette étude avaient été opérés par un chirurgien du service du service ORL de l'Hôpital de Purpan, et tous étaient suivis pour la rééducation et les contrôles dans l'Unité d'Implant Auditif (UIA) du même hôpital, avec l'orthophoniste Marie-Laurence Laborde. L'étiologie de la surdité parmi les patients était très variable, de même que la durée de surdité qui précédait l'implantation. Toutefois, tous ces patients étaient implantés unilatéralement, à droite ou à gauche. Certains patients portaient une prothèse auditive controlatérale à l'implant, mais étaient toujours testés sans cette prothèse. Un tableau détaillé des patients est présenté en annexe (voir Tableau 14.2, p345), en précisant à quels tests ils ont participés.

Chez les patients sourds implantés cochléaires unilatéraux, il nous a paru intéressant d'utiliser deux approches différentes. Avec la première, nous avons inclus un groupe de patients en mesure transversale : les patients n'étaient vu qu'une seule fois et avaient des durées d'activation d'implant variables. La deuxième approche était le suivi d'un groupe de patients ; ils étaient inclus

au premier jour de leur activation et suivis au cours du temps. L'objectif de ce suivi était de permettre d'obtenir des mesures avec moins de variabilité interindividuelle. Egalement, en comparant ces données à celles du groupe transversal, nous pourrions valider ou non l'intérêt des mesures en suivi. En effet, le suivi nécessitant des contraintes de temps plus importantes, il allait de soit que nous ne poursuivrions ces méthodes que dans la mesure où elles nous permettaient de mettre en évidence des variations plus fines.

Pour l'ensemble des patients nous avons souhaité savoir si les performances étaient reliées à d'autres paramètres ou d'autres habiletés. Dans un premier temps, étant donné que la compréhension du langage est un indice d'évolution important chez les patients implantés cochléaires, nous avons souhaité voir comment évoluaient les performances de détection de la voix en comparaison des performances de compréhension de mots. Ensuite, nous avons déjà évoqué l'importance de la lecture labiale chez les patients et l'impact que cela peut avoir sur la réorganisation corticale. En effet, Rouger *et al.* ont montré que la lecture labiale avait une importance primordiale chez les patients sourds, et ce même après implantation cochléaire (Rouger et al., 2007). Les auteurs ont également montré, chez des patients implantés cochléaires, une activation de l'aire de la voix par la lecture labiale seule (Rouger, Fraysse, Deguine, & Barone, 2008), montrant ainsi une possible colonisation de l'aire de la voix par des fonctions en lien avec la lecture labiale. Nous nous sommes par conséquent intéressés aux performances de lecture labiale et leurs relations avec celles de la détection de la voix. L'objectif était de voir si de bonnes performances en lecture labiale pouvaient être corrélées à de moins bonnes performances de détection de la voix. Enfin, certains patients portaient, en plus de l'implant cochléaire, une prothèse auditive controlatérale à l'implant. Nos avons souhaité savoir si cette prothèse avait un impact sur les performances auditives au travers de l'implant, et notamment en termes de détection de la voix. En effet la prothèse est portée par des patients ayant une audition résiduelle, particulièrement dans les basses fréquences La question que nous nous sommes posée était de savoir si cette audition résiduelle pouvait guider ou accélérer, en termes d'évolution, la perception du son au travers de l'implant seul. L'idée étant que cette audition résiduelle pourrait agir comme un feed-back auditif sur ce qui est perçu via l'implant, et permettre au patient d'associer les sons perçus avec l'implant avec des sons connus et perçus via l'audition résiduelle. Plus particulièrement, l'analyse à consisté à rechercher si le port de cette prothèse auditive, permettait une meilleure perception au travers de l'implant seul.

10.1 TEST DE DETECTION DE LA VOIX HUMAINE

10.1.1 VOIX HUMAINE VERSUS SONS ENVIRONNEMENTAUX

10.1.1.1 MATERIEL ET METHODES

SUJETS

Le groupe transversal était constitué de 31 patients, dont 13 hommes et 18 femmes, avec une moyenne d'âge de 65,5 ans (+/- 14.9). Les sujets avaient des délais d'activation variables allant de 1 jour à 131 mois et ont été répartis en 4 groupes : 1er jour d'activation (N = 9), 1 à 6 mois d'activation (N =6), 6 à 18 mois d'activation (N = 8) et plus de 18 mois d'activation (N =8).

Le groupe de suivi était constitué de 14 patients, dont 5 hommes et 9 femmes, avec une moyenne d'âge à l'inclusion de 51,5 ans (+/- 14,5). Tous les patients de ce groupe ont été inclus au premier jour de l'activation de leur implant.

Certains patients portaient une prothèse auditive controlatérale au quotidien. Toutefois tous les patients passaient le test avec l'implant seul. L'effet du port de cette prothèse sera analysé à part. Un tableau descriptif pour l'ensemble des patients est donné en Annexes (voir Tableau 14.2, p345).

STIMULI

Les stimuli utilisés étaient les mêmes que pour les sujets sain passant le test « Voix-Non Voix » en binaural (voir Voix humaine versus sons environnementaux, stimulation binaurale, Stimuli, p152). Soient 55 stimuli Non-Voix (cloche, eau qui coule, sonnerie, bruit de voiture, vent, etc.) et 55 stimuli Voix, dont 29 étaient langagiers mais sans information sémantique et 26 non langagiers. Tous les stimuli duraient 500 ms et aucun traitement supplémentaire n'avait été apporté à ces stimuli.

PROCEDURE

Les sujets des deux groupes étaient vus au cours de leur venue à l'hôpital dans le cadre de la prise en charge par l'Unité d'Implant Auditif (UIA) de Purpan, et étaient testés dans une pièce insonorisée du service ORL. Les stimuli étaient présentés en champ libre via deux haut-parleurs (KINYO, modèle PS-240) et le volume était ajusté pour être environ à 72 dB à 60 cm des enceintes. Les patients qui portaient une prothèse auditive controlatérale à l'implant, la retiraient et mettaient un bouchon d'oreille, afin qu'ils soient testés avec l'implant seul, et que les conditions de test puissent être comparables entre tous les patients. Les sujets étaient placés à cette même distance des haut-parleurs et face à un clavier, avec lequel ils avaient pour consigne de répondre si le stimulus présenté était une voix humaine ou un son de l'environnement, en appuyant sur une des deux touches « contrôle » correspondant à leur réponse. Pour des raisons de difficulté évidente chez les

patients, aucune consigne de rapidité n'était donnée, et l'expérimentateur indiquait également qu'il leur était possible de ne pas répondre, mais qu'ils devaient essayer de répondre le plus possible. Après 10 secondes sans réponse du sujet, le programme de stimulation passait à l'essai suivant. Les stimuli étaient présentés aléatoirement et la durée entre deux essais était de 1,5 seconde. Le test durait entre 5 et 10 minutes, selon les temps de réponse des sujets et le nombre de non réponse. La présentation des stimuli et l'enregistrement des réponses des patients se faisaient à l'aide du logiciel Matlab 6.5®.

Les 31 sujets du groupe transversal n'étaient vus qu'une seule fois. Pour ce groupe, le sous-groupe de patients vus au premier jour de l'activation, étaient des sujets qui furent inclus dans le groupe suivi. Pour ce deuxième groupe, les sujets en question, inclus au premier jour de l'activation et revus à des délais d'activation fixe qui correspondaient aux évaluations de la prise en charge par l'UIA : 1, 3, 6, 9, 12 et 18 mois post-activation d'implant. Le suivi impliquant des contraintes de temps particulières, tous les sujets n'ont pas pu être vus à tous les temps du suivi.

ANALYSES

Pour les analyses, seules les données des patients qui avaient fait moins de 40% de non réponses étaient analysées, et tous les essais où les patients n'avaient pas répondus n'étaient pas pris en compte. Ensuite la même analyse que pour les sujets normo-entendants était appliquée. On calculait ainsi à partir du nombre de détections correctes des stimuli vocaux et du nombre de fausse alarme des stimuli issus de l'environnement, l'indice de sensibilité *d'* (voir « Rappels sur la théorie de détection du signal », p146 et « Analyse » du test « Voix-Non Voix », p154).

10.1.1.2 RESULTATS DU GROUPE TRANSVERSAL

L'EFFET DE L'EXPERIENCE AVEC L'IMPLANT SUR LA DETECTION DE LA VOIX

Nous avons analysé les performances de détection de la voix en fonction du délai d'activation de l'implant. Les sujets étaient répartis selon les groupes suivant : 1 jour, 1 à 6 mois, 6 à 18 mois et au-delà de 18 mois d'activation.

Nos résultats montrent que les patients implantés cochléaires avaient un déficit important lors de l'activation de l'implant, avec un *d'* égal à 0,85 comparé aux normo-entendants qui avaient un *d'* de 4,3 (voir Figure 10.1). Ces performances semblaient s'améliorer à partir du groupe de 1 à 6 mois puis rester stables pour les deux autres groupes ; en effet les moyennes des 3 groupes étaient très proches de 1,5. L'évolution des performances entre les 4 groupes de patients n'était pas significative, on peut donc dire que même s'il y avait une tendance, leurs performances ne s'amélioraient pas. De plus elles restaient très inférieures aux NHS en condition originale (Mann-Whitney : $p < 0,001$). En fait en comparant les performances des patients à celles des NHS en condition de vocoder, on remarquait qu'elles étaient similaires à celles des NHS en condition de vocoder à 4 canaux. De plus les performances des patients étaient supérieures à la condition vocoder 2 canaux (sauf pour le

premier groupe qui était significativement équivalent à 2 et 4 canaux), et inférieures à toutes les autres conditions de vocoder.

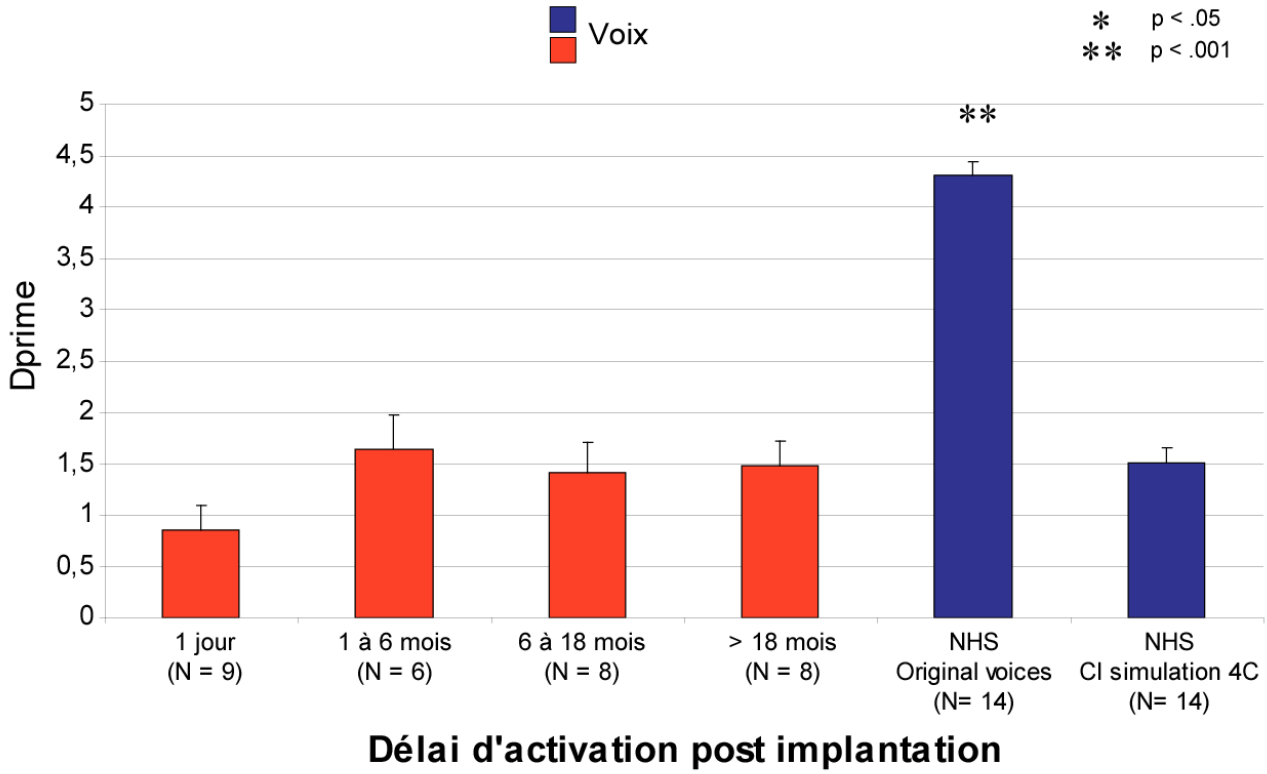


Figure 10.1 : Performance des patients (en rouge) pour le test de détection de la voix, effet du délai d'implantation et comparaison avec des sujets normo-entendants (en bleu) en condition originale et en simulation d'implant par le vocoder. Chez les patients, la faible évolution des performances avec l'expérience de l'implant n'est pas significative, et leurs performances, même après 18 mois d'implantation sont similaire à celles des NHS en condition de simulation d'implant à 4 canaux et significativement très inférieures aux NHS en condition originale.

Nos résultats indiquent donc que les patients avaient un très gros déficit pour la tâche de détection de la voix. Ce déficit peut être expliqué en partie par la difficulté de la tâche, en effet les stimuli sont très courts (500 ms). Mais lorsqu'on compare des patients implantés expérimentés, même de plus de 18 mois, avec des sujets normo-entendants naïf vis-à-vis de ce type de signal acoustique, on remarque qu'à nombre de canaux égaux, les NHS étaient bien meilleurs. Au-delà de 18 mois les patients, qui avaient entre 12 et 22 électrodes, avaient un *d'* égal à 1,48 et les NHS à 16 canaux de vocoder avaient un *d'* compris entre 2,3 et 3, soit près du double de celui des patients. Il semble donc que les NHS bénéficiaient d'une meilleure perception ou d'un meilleur traitement du signal acoustique. Toutefois, notons qu'une comparaison directe du signal transmis par le vocoder et celui par les processeurs des implants cochléaires n'est pas possible étant donnée que les patients bénéficiaient de modèles d'implants, de processeurs et de stratégies de codage différents. Mais de manière générale, les stratégies de codages développées par les différents fabriquant d'implant ont pour but d'améliorer la transmission et la perception du signal auditif par rapport au traitement basique du vocoder (voir « Stratégies de codage », p119). On peut donc dire que le déficit en termes

de détection de la voix chez les patients n'est pas uniquement dû à un signal acoustique dégradé, et que probablement, il y a chez ces patients des déficits au niveau du traitement neuronal du signal acoustique de la voix.

LA DISSOCIATION DES STIMULI LANGAGIERS ET NON LANGAGIERS

Lorsqu'on a distingué les performances de détection de la voix selon que les stimuli était langagiers ou non, nous avons obtenu la même dissociation au niveau des performances (voir Figure 10.2). Les patients étaient meilleurs pour la détection de la voix langagière, et ceci dès le premier jour de l'activation (Mann-Whitney : p = 0,0077). Pour les deux groupes suivant, « 1 à 6 mois » et « 6 à 18 mois », on ne retrouvait pas cette différence, même si la tendance était là pour le groupe « 6 à 18 mois ». En revanche, pour les patients dont l'activation d'implant était supérieur à 18 mois, on retrouvait de nouveau une habileté significativement supérieure pour la détection des sons vocaux langagiers par rapport aux non langagiers (Mann-Whitney : p = 0,0117). Lors de cette distinction des stimuli vocaux langagiers ou non, on retrouvait également des performances égales à celles des NHS en condition de simulation d'implant à 4 canaux. Le premier groupe, dont l'implant était activé depuis 1 jour avait également des performances similaires aux NHS en condition de vocoder à 2 canaux.

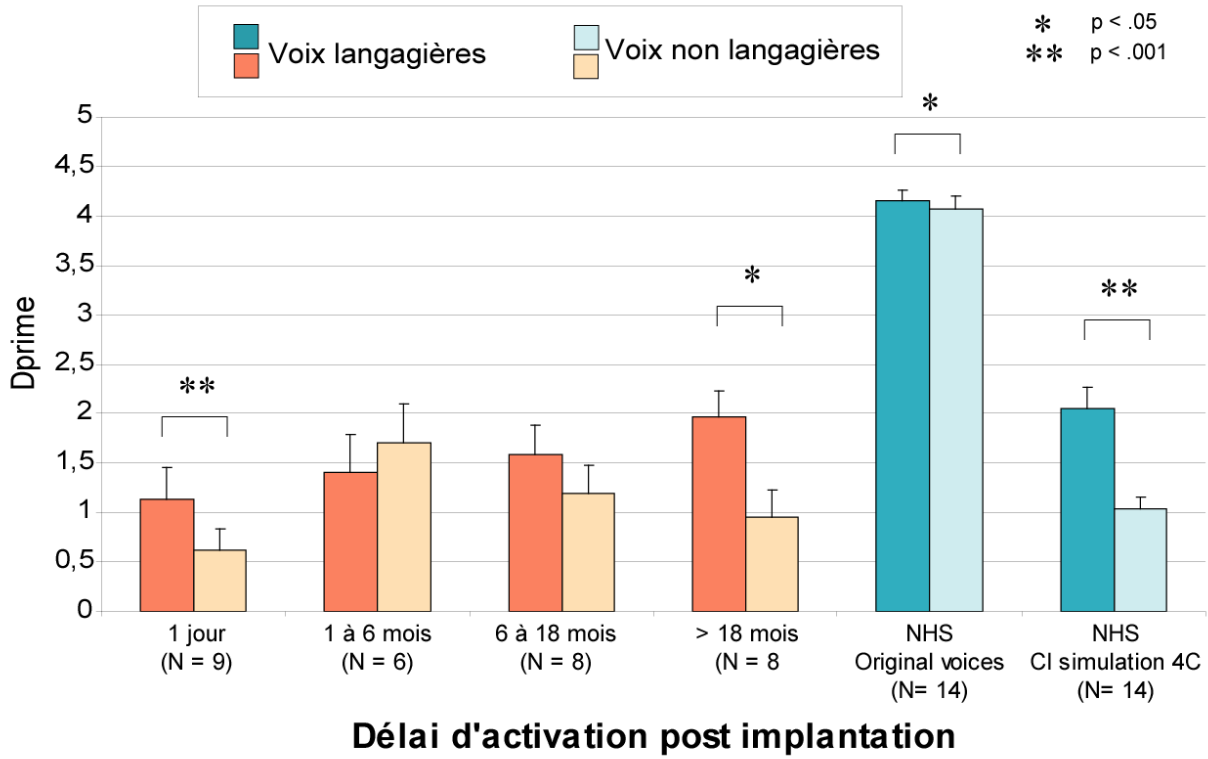


Figure 10.2 : Performance des patients (en rouge) pour le test de détection de la voix en distinguant les stimuli langagiers et non langagiers en fonction du délai d'implantation et comparaison avec des sujets normo-entendants (en bleu) en condition originale et en simulation d'implant par le vocoder. Chez les patients, la faible évolution des performances avec l'expérience de l'implant n'est pas significative, et leur performances, même après 18 mois d'implantation sont similaire à celles des NHS en condition de simulation d'implant à 4 canaux et significativement très inférieures aux NHS en condition originale.

Il semble donc que chez les patients, à l'instar des NHS, il existe aussi une différence entre la détection des stimuli vocaux langagiers et non langagiers. Tout comme pour les sujets en condition de vocoder, on peut expliquer cette différence par une plus grande richesse des modulations temporelles des sons langagiers. On sait que ces modulations sont préservées au travers du vocoder et des processeurs d'implant cochléaires, contrairement aux modulations spectrales. Les stimuli langagiers, bénéficiaient donc de cette information préservée et étaient mieux détectés lors de la tâche.

LES PERFORMANCES DE COMPREHENSION DU LANGAGE

Pour notre échantillon de patients, nous nous sommes intéressés à leurs performances de reconnaissance de mots dissyllabiques en auditif, avec l'implant seul. Les valeurs présentées ici, sont celles mesurées par l'orthophoniste (Marie-Laurence Laborde) de l'UIA de Purpan, lors de la prise en charge de suivi des patients (voir Figure 10.3). Les valeurs indiquées ici ont été mesurées le même jour que le test de détection de la voix. En analysant les données, on peut remarquer que, contrairement au test de détection de la voix, les performances de reconnaissance de mot progressaient de manière significative, jusqu'à atteindre une moyenne de 73,5 % de mots correctement reconnus pour le groupe de patients dont l'implant était activé depuis plus de 18 mois. Toutefois le niveau de reconnaissance des patients restait significativement inférieur à celle des normo-entendants en condition originale, qui eux reconnaissaient tous les mots sans erreurs. Nous présentons également sur la Figure 10.3 des données issues de travaux antérieurs au sein du laboratoire, réalisés et publiés par Julien Rouger (Rouger et al., 2007). Ces données correspondent aux performances de reconnaissance correcte de mots dissyllabiques par des normo-entendants en condition de vocoder à 4 canaux. Les mots dissyllabiques étaient les mêmes et prononcés par la même locutrice (l'orthophoniste de l'UIA de Purpan) que pour les patients implantés cochléaires de notre étude. Ainsi, on peut remarquer que les patients, dès 1 à 6 mois d'activation d'implant étaient nettement meilleurs que des normo-entendants en condition de simulation d'implant à 4 canaux. Ceci met en évidence une dissociation de récupération entre la détection de la voix, qui semble ne pas s'améliorer avec l'usage de l'implant, et la compréhension de mots, qui elle s'améliore très nettement et très vite.

Nos résultats montrent donc une différence très nette chez les patients entre leurs capacités de compréhension le langage et celles de détection de la voix. En effet, leur performances de détection de la voix, sont faibles et n'évoluent pas, elles sont comparables aux sujets NHS en condition de vocoder à 4 canaux. En revanche, en compréhension de mots, les performances évoluaient rapidement pour devenir très bonnes et largement au dessus d'un sujets NHS en condition de vocoder à 4 canaux. Trois explications peuvent être avancées. La première concerne le signal acoustique de la tâche de compréhension de mots, qui, étant plus riche du point de vue temporel, va pouvoir être mieux reconnu par les patients. Mais on a pu voir dans la tâche de détection de la voix que la richesse en termes de modulation temporelle des stimuli vocaux langagiers, même si elle

permettait une meilleure détection par rapport aux stimuli non langagiers, ne permettait pas d'amélioration des performances au cours du temps. Il semble donc que des mécanismes supplémentaires viennent expliquer cette évolution. On peut notamment évoquer l'effet du contexte lexical chez les patients. Celui-ci pourrait expliquer en partie la reconnaissance des mots. Par exemple dans le cas d'une compréhension partielle du mot, les connaissances lexicales du patient lui permettraient de compléter et de retrouver l'information manquante. Un tel effet semble très probable d'autant plus qu'il a été mis en évidence récemment dans d'autres études (Chatterjee, Peredo, Nelson, & Baskent, 2010). Ceci n'exclu pas l'idée de deux processus différents. Il est probable que ces deux types de tâches impliquent des processus différents ayant des substrats neuronaux différents, qui ne bénéficieraient des mêmes récupérations fonctionnelles au cours du port de l'implant, et dépendant peut-être de facteurs de récupération différents. Notamment, l'effet de la rééducation axée sur la compréhension du langage pourrait accentuer cette différence de récupération.

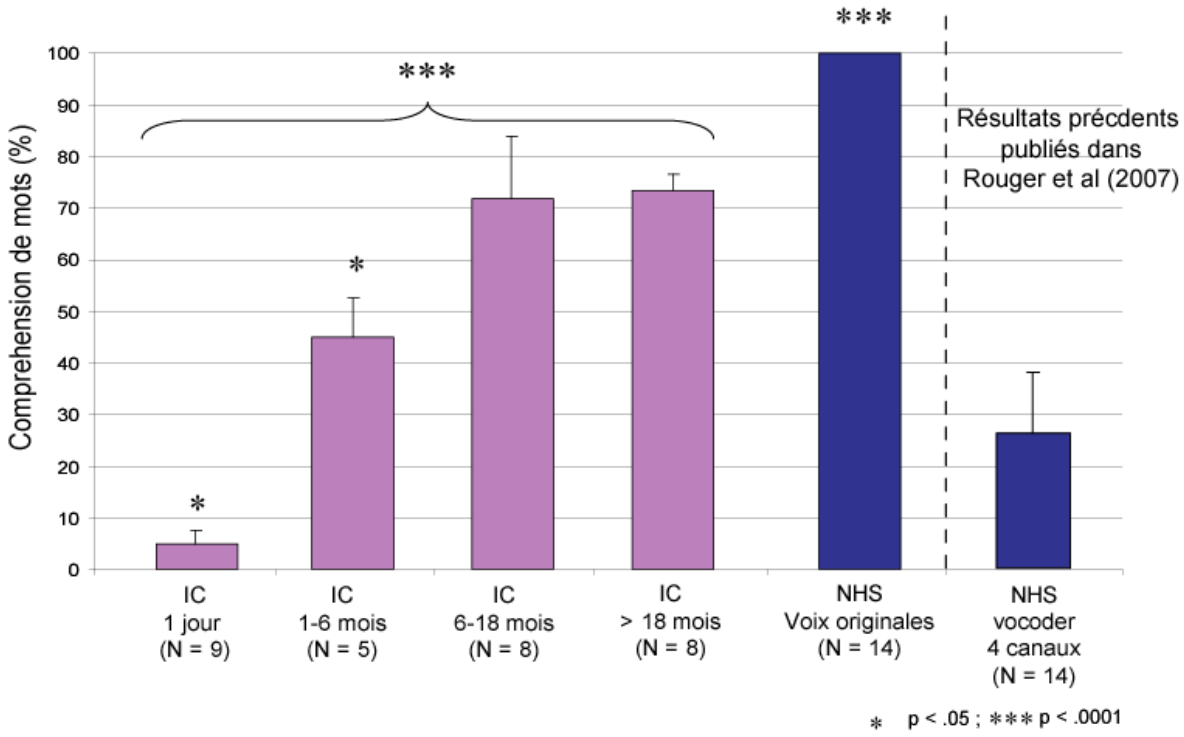


Figure 10.3 : Performance de reconnaissance de mot avec l'implant seul pour les patients implantés cochléaires du groupe transversal, et comparaison avec les sujets normo-entendants (NHS) de cette même étude, en condition originale, et à des sujets NHS d'une étude précédente (Rouger et al., 2007) en condition de vocoder à 4 canaux. Le niveau de reconnaissance de mots dissyllabiques augmente de manière significative au cours du temps pour les patients implantés cochléaires (IC) pour atteindre des performances très correctes. Toutefois, même pour les patients dont l'implant a été activé depuis plus de 18 mois, les performances restent inférieures aux normo-entendants en condition originale.

Bien qu'il semble exister une dissociation de récupération entre les deux tâches, la corrélation entre les performances de détection de la voix et celles de compréhension de mots était positive (voir Figure 10.4, données en rouge). Le coefficient de corrélation R^2 était de 0,194, et cette corrélation avait un seuil de significativité égal à 0,0123. Lorsqu'on considérait les sujets par groupes, on pouvait constater que la corrélation était également positive (voir Figure 10.4, données à 1 jour en bleu et à plus de 18 mois en vert), mais elle n'était pas significative, probablement en raison du petit nombre d'individu par groupe (respectivement, N = 9 et N = 8).

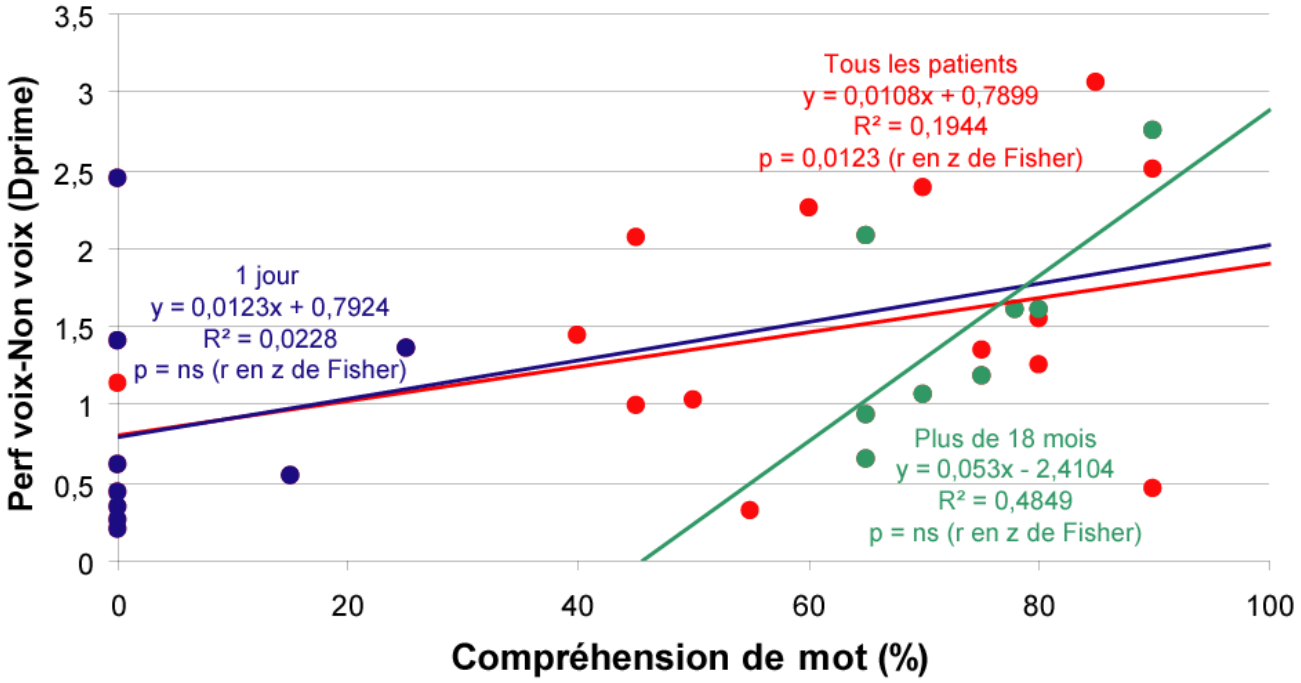


Figure 10.4 : Corrélation positive pour les patients, entre les performances de détection de la voix en ordonnées (*d'*) et celles de compréhension de mots en abscisses (%). La corrélation était positive que l'on considère les sujets à 1 jour post-activation (bleu) ou expérimenté à plus de 18 mois post-activation (vert). mais elle n'était significative que lorsqu'on prenait en compte l'ensemble des sujets (rouge). Les sujets ayants les meilleures performances au test de compréhension de mots étaient donc aussi les meilleurs pour la détection de la voix. Toutefois, les perfromances de détection de la voix restaient très faibles pour tous les patients en comparaison aux sujets normo-entendants.

Cette corrélation positive sur l'ensemble des patients indique que ceux qui avaient les meilleures performances au test de compréhension de mots étaient aussi ceux qui avaient les meilleures performances au test de détection de la voix. Cependant, reste que, pour la détection de la voix, les performances étaient très faibles pour tous les patients en comparaison des NHS, indépendamment de leurs performances en compréhension de mots. Ce résultat reflète donc essentiellement l'effet de l'expérience de l'implant. En effet les nouveaux implantés (1 jour post-activation) avaient des performances faibles dans les deux types de tâches et les patients plus expérimentés avec l'implant (plus de 18 mois post-activation) avaient de meilleures performances dans les deux tâches, bien qu'elles restent faibles pour la détection de la voix en comparaison aux NHS.

LES PERFORMANCES DE LECTURE LABIALE

Nous nous sommes également intéressés aux performances de lectures labiales et leurs relations avec celles de détection de la voix. Rappelons que l'idée était de voir si de bonnes performances en lecture labiales pouvaient être corrélées à de moins bonnes performances de détection de la voix.

Tout comme pour la compréhension de mots, le test de lecture était réalisé le même jour que le test de détection de la voix par l'orthophoniste (Marie-Laurence Laborde) de l'UIA de Purpan, lors de la prise en charge de suivi des patients (voir Figure 10.5). Toutefois ce test n'était pas réalisé systématiquement, ce qui explique le plus grand nombre de données manquantes.

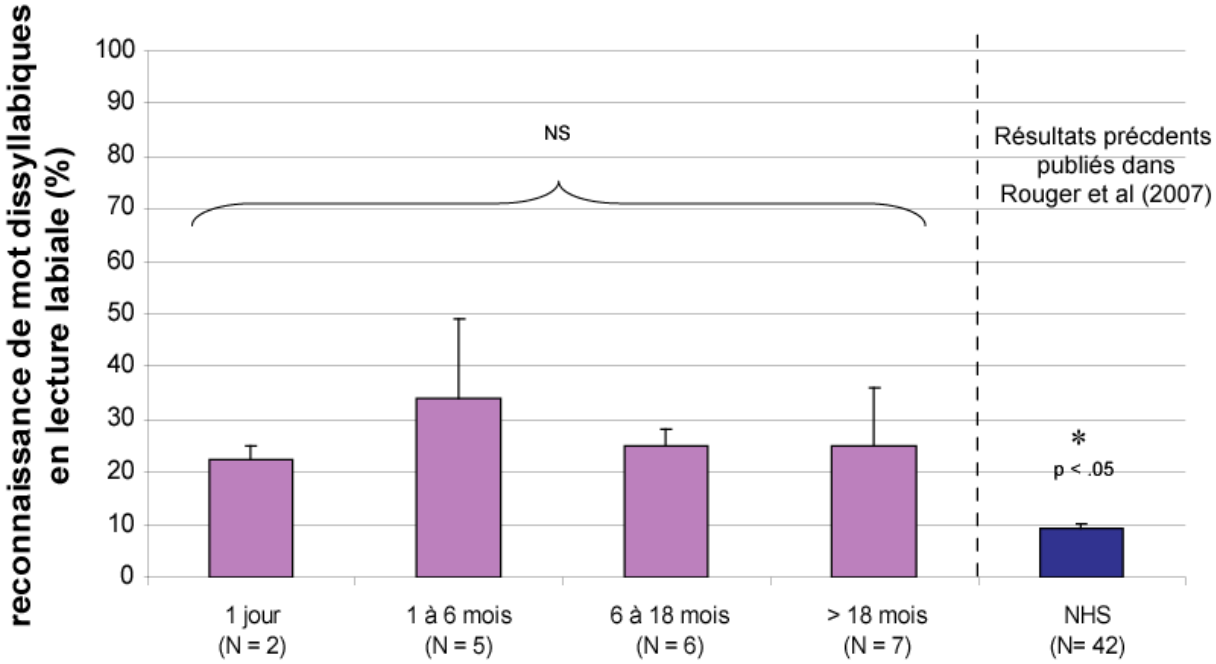


Figure 10.5 : Performance de lecture labiale pour les patients implantés cochléaires du groupe transversal, et comparaison avec des NHS d'une étude précédente (Rouger et al., 2007). Le niveau de reconnaissance de mots dissyllabiques restait stable au cours du temps pour les patients implantés cochléaires (IC) et restait toujours supérieur à celui des NHS.

Nos résultats montrent que les patients avaient de bien meilleures performances de lecture labiale (entre 22,5 et 34 %) que les normo-entendants (9,4%), soit jusqu'à plus de 3 fois leur score. Ces performances, ne décroissaient pas au cours de l'expérience avec l'implant ($p > 0,99$). Nos données correspondent à ce qui a déjà été publié par Rouger *et al.* (Rouger et al., 2007). Il est intéressant de voir que les patients, bien qu'ils aient retrouvé un très bon niveau de compréhension de la parole, conservent des performances de lecture labiale très supérieures à celles des sujets normo-entendants.

De plus nous pouvons voir sur la Figure 10.6 que ces performances de lecture labiale sont très fortement corrélées ($R^2 = 0,31$ et $p = 0,0096$). Ceci nous indique que les patients qui avaient

tendance à mieux détecter la voix étaient aussi ceux qui avaient les meilleures performances de lecture labiale. Cette corrélation était vraie pour les différents groupes de patients (voir données pour le groupe activé depuis plus de 18 mois en vert sur la Figure 10.6), bien qu'elle n'ait pas pu être réalisée pour le groupe de patients à 1 jour post-activation d'implant, en raison du nombre trop faible de données (N = 2).

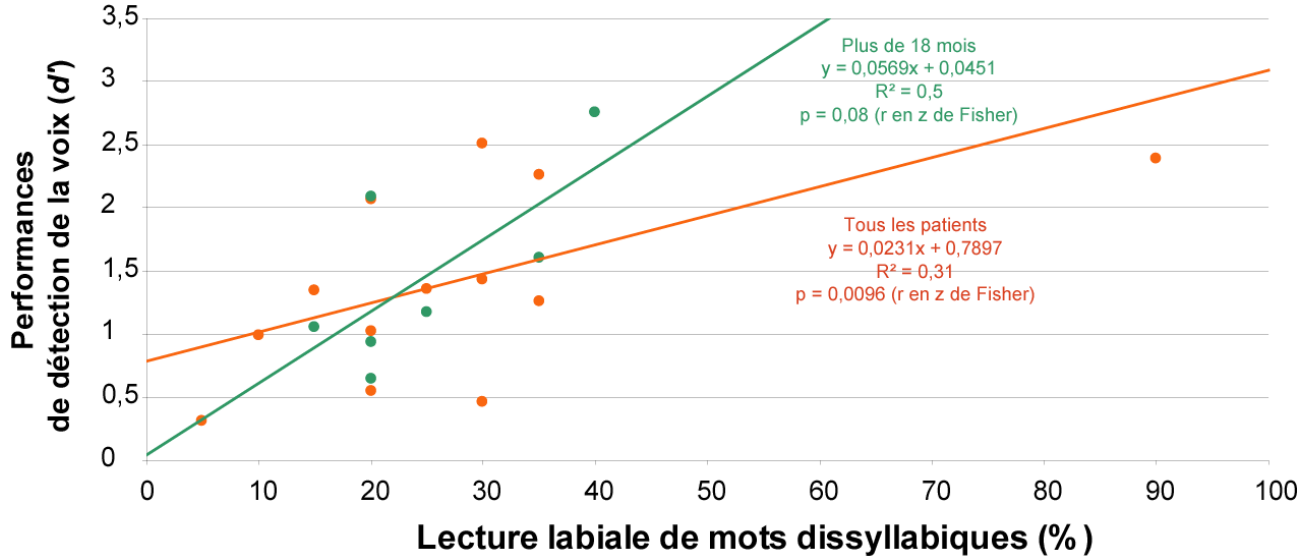


Figure 10.6 : Corrélation positive et significative entre les performances de détection de la voix (*d'*) et celles de lecture labiale (%). La corrélation n'a pas pu être réalisée pour les sujets à 1 jour post-activation en raison du faible nombre de données (N = 2). A plus de 18 mois post-activation d'implant, la corrélation était positive mais pas significative là encore probablement en raison d'un nombre de données insuffisantes.

Les résultats de cette analyse ne confirment pas du tout l'idée que de meilleures performances en lecture labiale soient associées à de moins bonnes performances en détection de voix, nous aurions dû pour cela constater une corrélation négative. Les résultats vont donc à l'opposé de ce à quoi on s'attendait. Toutefois on peut noter que les données de lecture labiale manquent pour 1/3 des patients (11 sur 31) et que cela introduit un biais important dans l'interprétation des résultats.

10.1.1.3 RESULTATS DU GROUPE EN SUIVI

Dans ce groupe de 14 patients, 4 sujets n'ont pas pu être vu au premier jour de l'activation, en raison d'un manque de temps ou de la fatigue du sujet après de nombreux réglages, et un sujet n'a pas réussi à faire le test du tout. Par la suite le nombre de sujets manquants est dû à une absence du sujet (voir Tableau 10.1) excepté à 6 et 12 mois où un problème expérimental à invalidé les données d'un patient. Toutefois aucun patient n'a été absent du suivi plus de deux fois.

Délai d'activation	1 jour	1 mois	3 mois	6 mois	9 mois	12 mois	18 mois
Nombre de patients manquants	5	0	3	4	1	3	3

Tableau 10.1 : Nombre de sujet manquant selon le délai. Les sujets n'ont pu être vus, soit à cause de trop grande fatigue, soit en raison d'une absence lors du RDV de suivi. Un patient au premier jour d'activation n'avait pas réussi à faire la tâche et un problème expérimental a invalidé les données d'un patient à 6 et 12 mois. Aucun patient n'a été absent du suivi plus de deux fois.

L'EFFET DE L'EXPERIENCE AVEC L'IMPLANT SUR LA DETECTION DE LA VOIX

La Figure 10.7 représente les données individuelles moyennées pour chaque visite de suivi (en rouge) et comparées à celles des NHS (en bleu). On peut voir sur cette figure que les performances, qui étaient très faibles au premier jour d'activation de l'implant (*d'* = 0,847)) n'évoluaient que très peu avec le temps ; à 18 mois la valeur du *d'* était de 1,35. Pour les analyses, en raison d'un nombre insuffisant d'observations, nous avons réalisé les analyses (Friedman) sans les mesures faites à 6 mois et à 12 mois. Nous avons choisi d'exclure ces mesures car on a remarqué qu'entre 1 mois et 18 mois les valeurs étaient relativement stables, et les mesures à 6 mois ne paraissaient pas particulièrement atypiques pour que leur exclusion nous empêche de voir un effet du délai d'implantation sur la détection de la voix. De plus ces deux mesures n'étaient ni successives ni aux extrémités du suivi, ce qui permettait de conserver la tendance générale de la variation. Les résultats ont montré que la valeur du test était nettement non significative (Friedman : $p = 0,78$). Le résultat de ce test nous indique d'une part que, étant donné le seuil de significativité du test, l'absence des mesures à 6 et 12 mois ne changerait probablement pas la non significativité du test. D'autre part nos résultats montrent que les patients avaient un déficit important en ce qui concerne la détection de la voix, et que ce déficit ne s'améliorait pas avec l'expérience de l'implant. Ceci confirme les résultats obtenus pour le groupe transversal. Les valeurs de *d'* pour les mesures prisent entre 1 et 18 mois pour le suivi oscillaient entre 1,26 et 1,51 ce qui semble légèrement plus faible que celle du groupe transversal qui étaient 1,7 pour les groups 1 à 6 mois, 1,4 pour le groupe 6 à 18 mois et 1,48 pour le groupe dont l'implant était activé depuis plus de 18 mois. Le groupe dont l'implant était activé depuis 1 jour était constitué des mêmes patients pour l'approche transversale et le suivi ; les valeurs sont donc exactement les mêmes.

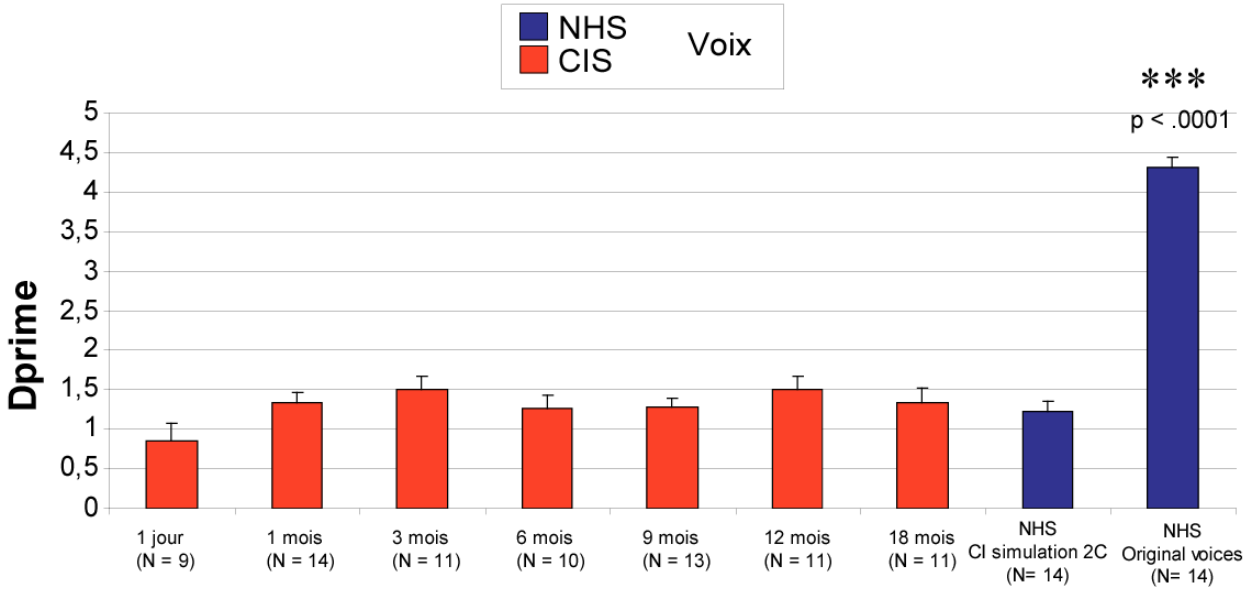


Figure 10.7 : Performance (±SE) des patients du groupe suivi (en rouge) pour le test de détection de la voix, effet du délai d'implantation et comparaison avec des sujets normo-entendants (en bleu) en condition originale et en simulation d'implant par le vocoder. Chez les patients, la faible évolution des performances avec l'expérience de l'implant n'était pas significative (les mesures à 6 et 12 mois ont été exclues pour les statistiques en raison d'un nombre de mesures insuffisant), et leurs performances, même à 18 mois d'activation d'implant étaient similaires à celles des NHS en condition de simulation d'implant à 2 canaux et significativement très inférieures aux NHS en condition originale.

On peut voir sur la Figure 10.8 les données individuelles de 4 patients ; 2 d'entre eux ont été vus à tous les temps et les deux autres n'ont qu'une mesure manquante. On constate que les données étaient variables au cours du temps, et ce même pour un patient ; les sujets pouvaient s'améliorer par rapport à une mesure précédente ou au contraire avoir de moins bonnes performances. Sur l'évolution générale des performances, on peut remarquer que pour les 4 sujets présentés, les valeurs étaient assez proches des valeurs moyennes présentées en Figure 10.7. Si le sujet 2 semblait s'améliorer légèrement, les 3 autres n'amélioraient pas du tout leur capacité à détecter la voix. On peut noter aussi l'exemple du sujet 4, qui avait de très bonnes performances lors de l'activation de l'implant (d' = 2,4), puis avait pour les autres mesures des performances moyennes (d' autour de 1,5), ce qui montre là encore les variabilités intra-individuelles.

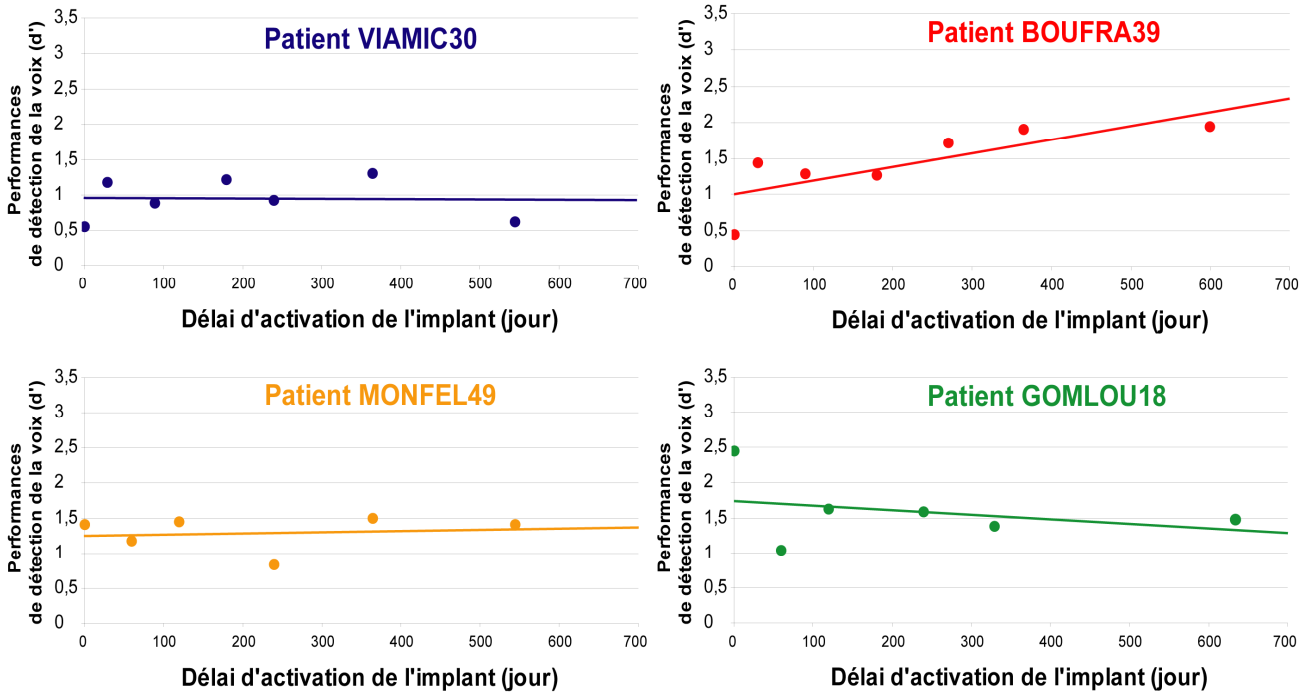


Figure 10.8 : Données individuelles de 4 patients. Les patients VIAMIC30 et BOUFRA39 ont été vus à tous les temps et les patients MONFEL49 et GOMLOU18 n'ont que la mesure à 6 mois manquante. On peut voir que les performances étaient variables au cours du temps pour un même patient. Si le sujet BOUFRA39 semblait s'améliorer légèrement avec le temps les 3 autres avaient des performances qui stagnaient.

Nous avons tenté de corréler les données de performances de détection de la voix des patients en fonction du délai d'implantation. Pour cela, à l'instar de la Figure 10.8, les délais réels en jour ont été pris pour chaque valeur de *d'*, étant donné que tous les patients n'ont pas été vu exactement au jour près. Avec cette analyse, nous n'avons pas non plus mis en évidence d'amélioration des performances au cours du temps (voir Figure 10.9). En effet on pouvait se rendre compte que cette corrélation n'existait pas ($R^2 = 0{,}021$; $p > 0{,}05$). Lorsqu'on regardait les données individuelles, seuls 4 patients sur 14 s'amélioraient légèrement, comme le sujet 4 de la Figure 10.8, ou de façon moindre. Les 10 autres patients, avaient des performances qui stagnaient parfaitement au cours temps ou qui diminuaient très légèrement. On peut donc en conclure que même d'un point de vue individuel, on ne peut pas vraiment constater une amélioration des performances pour ce qui est la détection de la voix dans l'environnement.

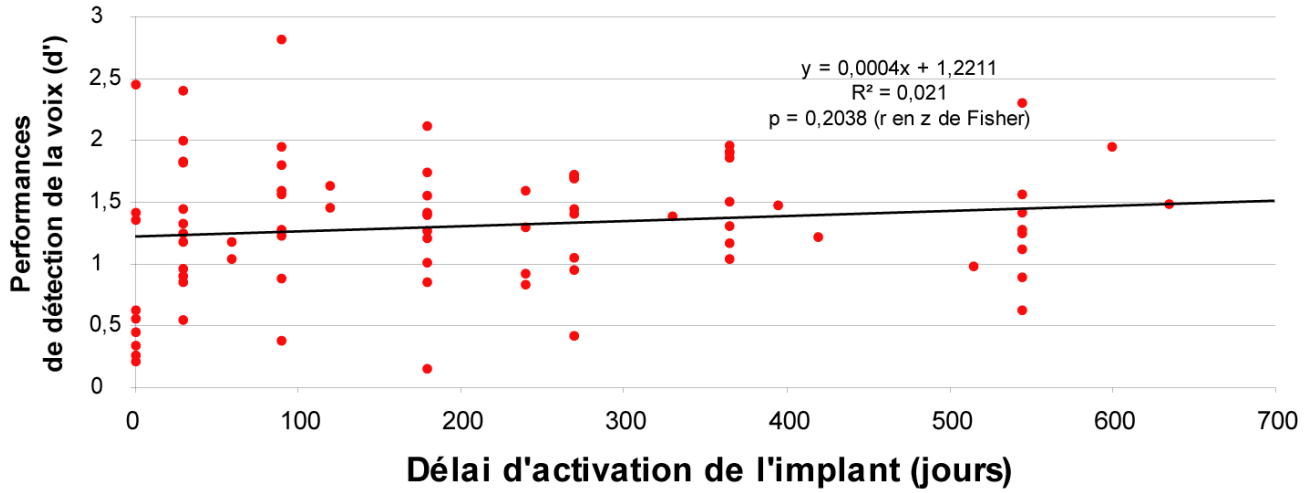


Figure 10.9 : Corrélation des valeurs de *d'* en fonction du délai d'activation. Toutes les mesures des 14 patients sont représentées et la corrélation pour l'ensemble de ces valeurs était quasi-nulle et non significative.

Lorsqu'on comparait le groupe de patients suivis aux sujets normo-entendants en condition originale, on retrouvait un déficit important, comme pour le groupe transversal. En revanche, la comparaison avec les NHS en condition de vocoder, nous indique que les performances des sujets en suivi étaient légèrement inférieures à celles des sujets de l'approche transversale. En effet, les performances des sujets en suivi étaient similaires à celles des NHS en condition de vocoder à 2 canaux et significativement inférieures à celles des NHS en condition de vocoder à 4 canaux, tandis que le groupe transversal avait des performances similaires à celles des NHS en condition de vocoder à 4 canaux (pour les groupes 1 à 6 mois, 6 à 18 mois et plus de 18 mois d'activation d'implant). Ceci nous permet de dire que les performances des sujets suivis étaient légèrement inférieures à celles des sujets de l'approche transversale. Nous voyons donc ici un intérêt à la méthode du suivi des patients, à savoir une mesure plus réelle de l'évolution de leurs performances. Toutefois on peut noter que dans les deux approches, en suivi ou transversale, nous arrivons aux mêmes conclusions, c'est-à-dire à un déficit important de la détection de la voix et pas d'amélioration avec l'expérience du port de l'implant. Il semble donc que, si l'évolution des performances est mieux appréhendée avec le suivi, la variabilité interindividuelle de cette évolution ne semble pas primordiale.

LA DISSOCIATION DES STIMULI LANGAGIERS ET NON LANGAGIERS

Pour les sujets en suivi nous nous sommes également intéressés à la distinction entre les stimuli vocaux langagiers et non langagiers (voir Figure 10.10). Comme pour le groupe transversal, nous n'avons pas non plus retrouvé d'amélioration significative au cours du temps, que l'on considère les stimuli vocaux langagiers ou non langagiers En revanche nous avons confirmé la dissociation entre les performances de détection de stimuli langagiers et non langagiers. Les patients étaient significativement meilleurs pour détecter les stimuli langagiers que les stimuli non langagiers, et ce quelque soit le délai d'activation de leur implant (Wilcoxon : $p < 0{,}016$).

L'approche en suivi nous a permis de confirmer cette différence de performances entre stimuli langagiers et non langagiers, alors qu'elle n'était pas présente pour tous les groupes lors de l'approche transversale.

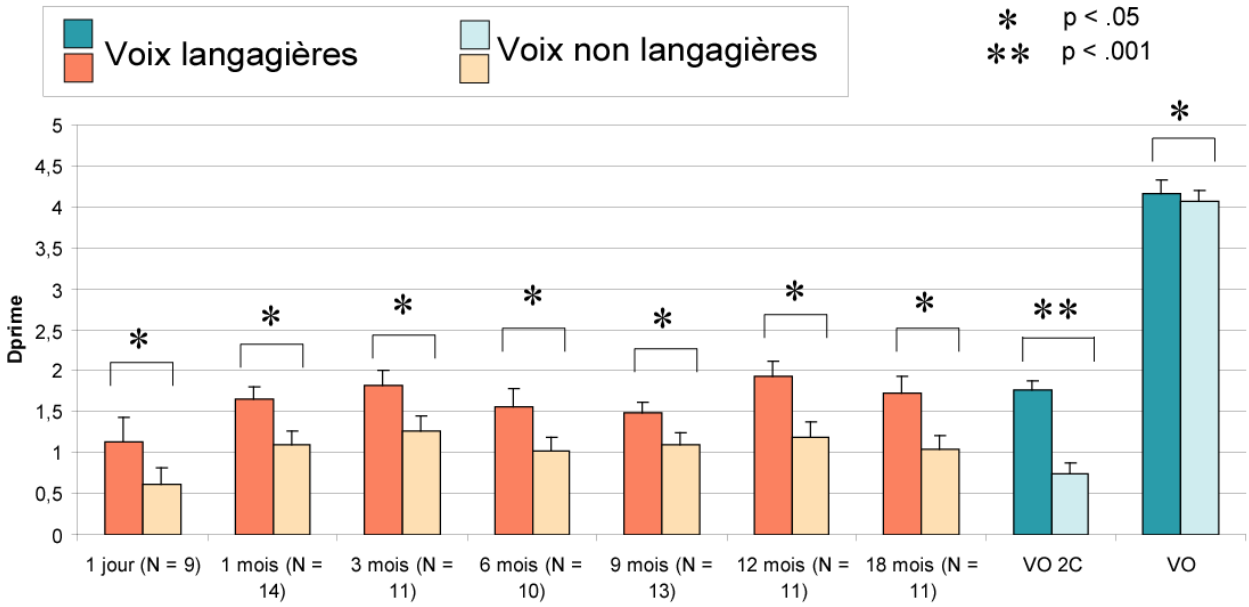


Figure 10.10 : Performance des patients suivis (en rouge) pour le test de détection de la voix en distinguant les stimuli langagiers et non langagiers en fonction du délai d'activation d'implant et comparaison avec des sujets normo-entendants (en bleu) en condition originale et en simulation d'implant par le vocoder. Chez les patients, la faible évolution des performances avec l'expérience de l'implant n'est pas significative (les mesures à 6 et 12 mois ont été exclues pour les statistiques en raison d'un nombre de mesures insuffisant), et leurs performances, même à 18 mois d'implantation étaient significativement très inférieures aux NHS en condition originale. Les valeurs de *d'* des stimuli vocaux langagiers étaient inférieures à celles des NHS en condition de simulation d'implant à 4 canaux, tandis que les valeurs pour les stimuli vocaux non langagiers étaient similaires à celles des NHS en condition de vocoder à 4 canaux.

La comparaison des performances de détection de la voix des patients en suivi pour les stimuli vocaux langagiers et non langagiers, avec celles des NHS nous a confirmé que les performances de ces patients semblaient un peu plus faibles que celles des patients du groupe transversal. En effet, pour toutes les mesures, les performances pour les stimuli vocaux langagiers étaient significativement inférieures à celles des normo-entendants en condition de vocoder à 4 canaux, alors que pour le groupe transversal les performances étaient similaires avec cette condition. En fait, le groupe en suivi avait des performances similaires à celles des NHS en condition de vocoder à 2 canaux pour les mesures à 3, 12 et 18 mois d'activation d'implant, et même inférieures pour les mesures à 1 jour, 1, 6 et 9 mois d'activation d'implant. Pour les stimuli vocaux non langagiers, on n'observait pas cette différence par rapport au groupe transversal. Les mesures des patients en suivis étaient toutes similaires à celles des NHS en condition de vocoder à 4 canaux.

On peut donc dire à partir de nos analyses que les patients étaient effectivement meilleurs pour détecter des stimuli vocaux quand ceux-ci étaient langagiers que lorsqu'ils étaient non langagiers, et ce à toutes les mesures du suivi. De plus ce groupe de patients en suivi avait des

performances pour les stimuli vocaux langagiers similaires à celles des patients du groupe transversal.

LES PERFORMANCES DE COMPREHENSION DU LANGAGE

Les mesures de compréhension du langage étaient évaluées par les performances de reconnaissances de mots dissyllabiques en auditif, avec l'implant seul. Les mesures étaient faites par l'orthophoniste (Marie-Laurence Laborde) de l'UIA de Purpan, lors de la prise en charge de suivi des patients, le même jour que le test de détection de la voix. Pour l'analyse, en raison d'un nombre insuffisant de valeur, nous n'avons pas pu faire de test statistique en incluant tous les temps du suivi. Le test statistique, comme pour la détection de la voix a été fait sans inclure les mesures à 6 et 12 mois. L'analyse des données nous montre que, tout comme pour le groupe transversal, l'évolution des performances étaient différentes de celles du test de détection de la voix : les performances de reconnaissance de mot progressaient de manière significative ($p = 0{,}0009$), jusqu'à atteindre une moyenne de 75,9 % de mots correctement reconnus pour le groupe de patients dont l'implant était activé depuis 12 mois (voir Figure 10.11). On observait une diminution à 18 mois (67,4 %), mais elle n'était pas significative. Le niveau de reconnaissance de mot restait significativement inférieur ($p < 0{,}0001$) à celle des normo-entendants en condition originale, qui reconnaissaient tous les mots sans erreurs. Lorsque nous comparons les patients à des normo-entendants en condition de vocoder à 4 canaux (données issus de Rouger, 2007, voir paragraphe « Les performances de compréhension du langage », p213) on peut remarquer que les patients, dès 1 mois d'activation d'implant étaient nettement meilleurs que des normo-entendants en condition de simulation d'implant à 4 canaux. Ceci confirme, la dissociation de récupération des fonctions du langage et de celles de détection de la voix avec l'expérience mis en évidence avec les patients du groupe transversal ; la compréhension du langage s'améliore significativement tandis que la détection de la voix ne s'améliore pas.

Les résultats chez les patients en suivis nous amène à proposer les mêmes explications en ce qui concerne les meilleures performances de reconnaissance du langage. Il s'agirait donc des modulations temporelles plus riches pour le langage, mais aussi de l'existence du contexte lexical qui permettrait aux patients de compléter l'information manquante lors d'une compréhension partielle ou encore l'idée que ces deux tâches impliquent des processus différents ayant des substrats neuronaux différents et ne bénéficieraient des mêmes récupérations fonctionnelles au cours du port de l'implant.

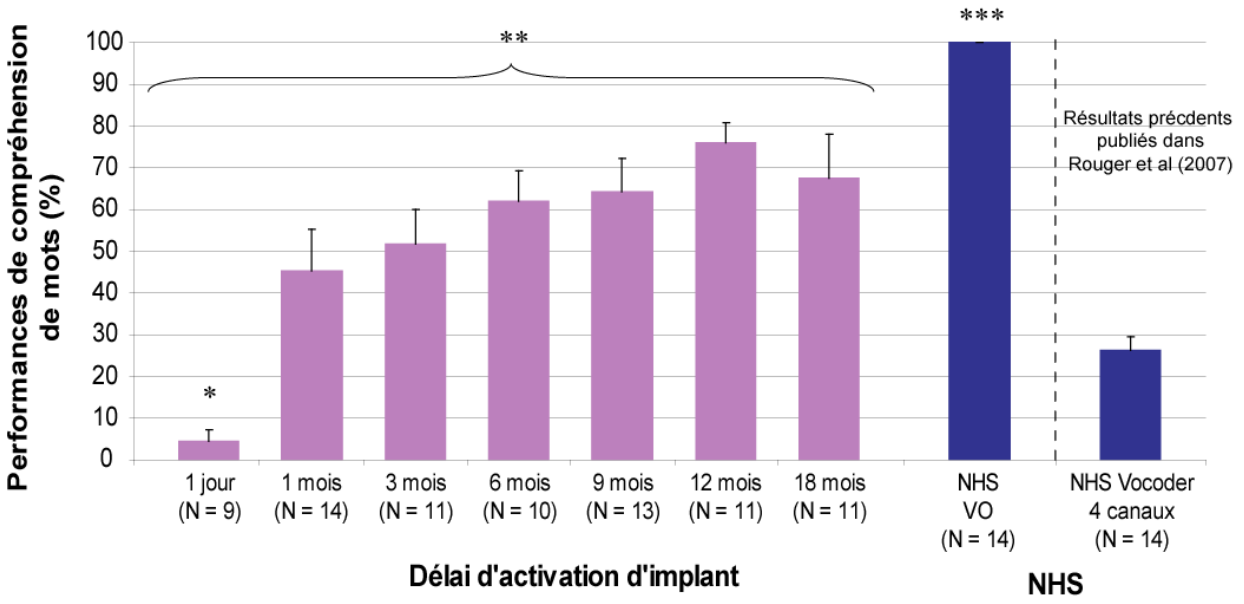


Figure 10.11 : Performances de reconnaissance de mots avec l'implant seul pour les patients implantés cochléaires du groupe transversal, et comparaison avec les sujets normo-entendants (NHS) de cette même étude, en condition originale, et à des sujets NHS d'une étude précédente (Rouger et al., 2007) en condition de vocoder à 4 canaux. Le niveau de reconnaissance de mots dissyllabiques augmente de manière significative au cours du temps pour les patients implantés cochléaires (IC) pour atteindre des performances très correctes (les mesures à 6 et 12 mois ont été exclues pour les statistiques en raison d'un nombre de mesures insuffisant). Toutefois, même pour les patients dont l'implant a été activé depuis 18 mois, les performances restent inférieures aux normo-entendants en condition originale.

De même que pour le groupe transversal, la corrélation entre les performances de détection de la voix et celles de compréhension de mots était positive (voir Figure 10.12). La corrélation était la même que pour le groupe transversal, avec un coefficient de corrélation R^2 de 0,1956, mais le seuil de significativité, p = 0,0096 était plus élevé.

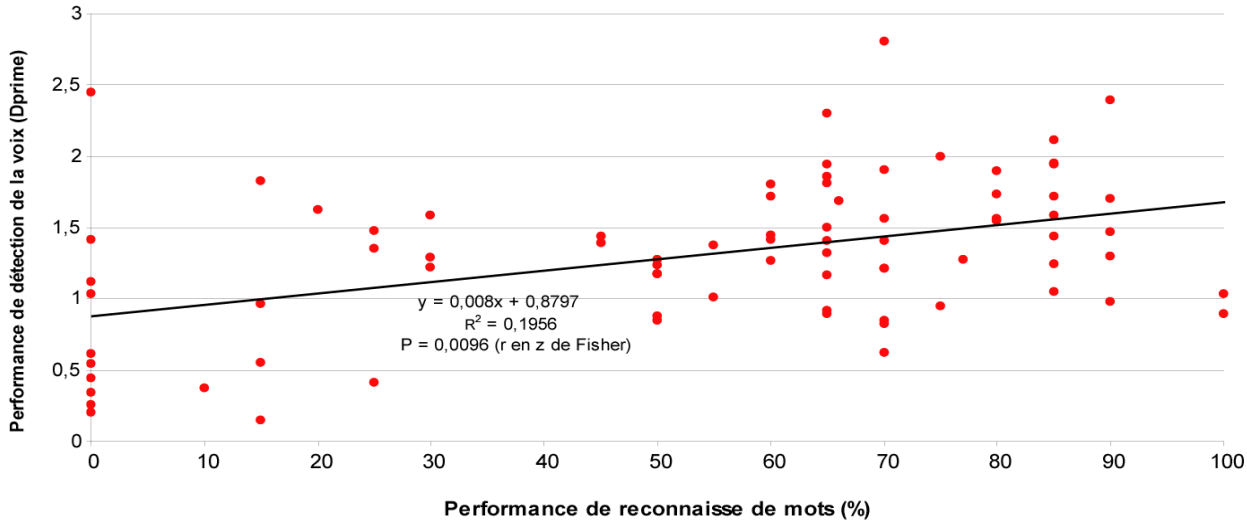


Figure 10.12 : Corrélation positive et significative pour les patients du groupe de suivi, entre les performances de détection de la voix en ordonnées (*d'*) et celles de compréhension de mots en abscisses (%). Les sujets ayants les meilleurs performances au test de compréhension de mots sont aussi les meilleurs pour la détection de la voix. Toutefois, les perfromances de détection de la voix restent très faibles pour tous les patients en comparaison aux sujets normo-entendant.

Le suivi nous a permis de regarder si ce résultat global était bien le reflet de ce qui se passe plus spécifiquement au niveau individuel. Sur les 14 patients, 8 montraient une corrélation positive, avec un R^2 compris entre 0,4 et 0,9. Donc d'un point de vue individuel, 8 patients avaient tendance à améliorer leurs performances en termes de détection de la voix à mesure qu'ils amélioraient leurs performances de compréhension de mots. La corrélation globale sur l'ensemble de données correspondait donc à la majorité des cas. En raison du petit nombre de données par patient la corrélation n'était clairement significative que pour 3 de ces 8 patients ayant une corrélation positive (voir Figure 10.13). Pour les 6 patients restants, la corrélation était soit nulle ou quasi-nulle (N = 4), avec un coefficient R^2 inférieur à 0,8, ou bien négative (N = 2), avec un coefficient R^2 égal a -0,4 pour les deux patients.

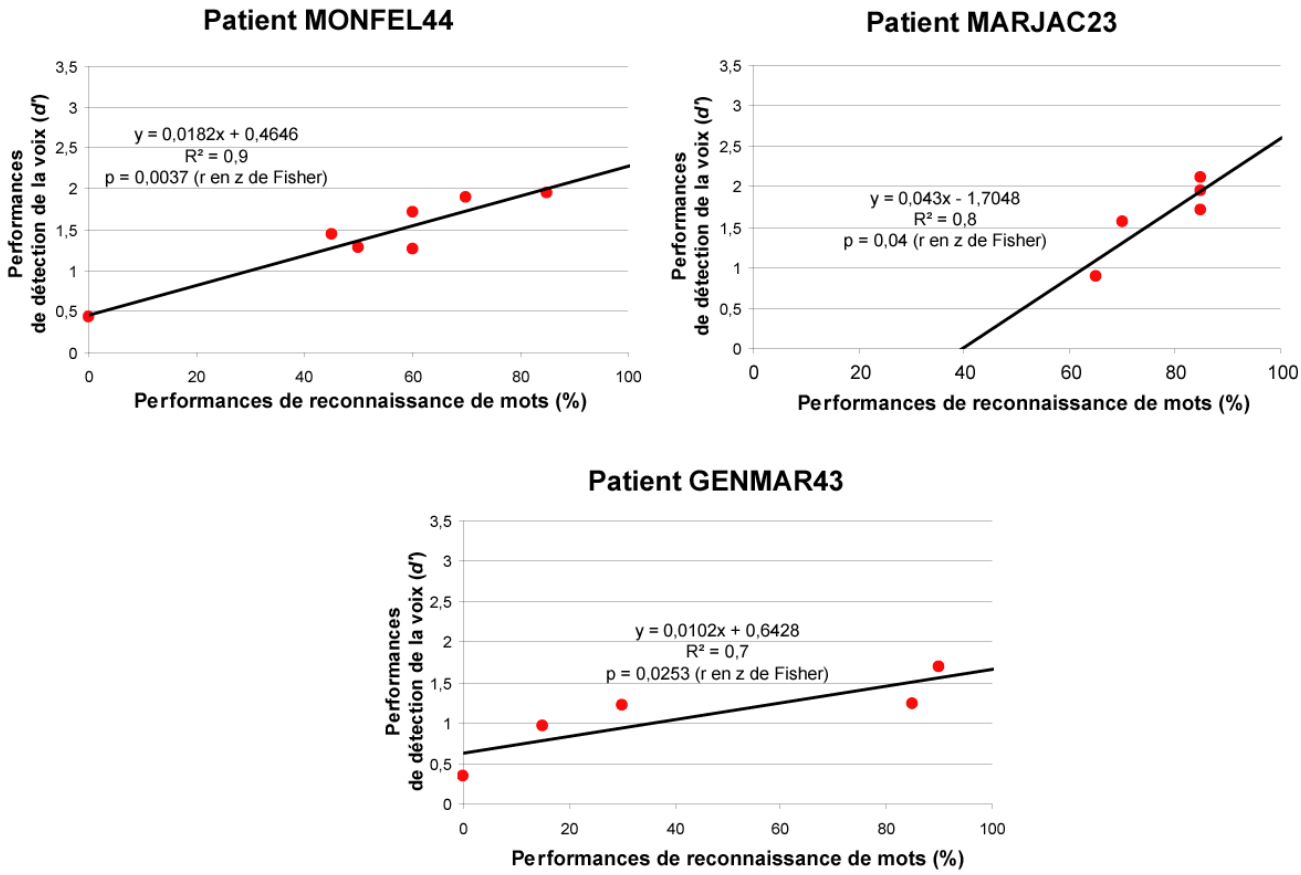


Figure 10.13 : Exemple de 3 patients pour qui la corrélation entre les performances de détection de la voix et les performances de compréhension de mots était positive. Au total, on retrouvait une corrélation positive pour 8 patients sur 14. Les 3 exemples présentés ici avaient une corrélation positive et significative.

Ceci confirme le fait que les patients qui étaient les meilleurs en compréhension de mots étaient également les meilleurs en détection de la voix. Toutefois cela ne réfute pas la dissociation d'évolution des performances, car même les meilleurs patients avaient un gros déficit pour la tâche de détection de la voix alors que le déficit était moindre, voire nul, pour certains patients en compréhension de mots.

LES PERFORMANCES DE LECTURE LABIALE

Comme pour le groupe transversal, nous nous sommes intéressés aux performances de lecture labiale afin de mesurer la variation de ces performances avec celles de la détection de la voix. Rappelons que le but de cette analyse était de savoir si une amélioration des performances en détection de la voix était en lien avec une diminution des performances de lecture labiale (voir paragraphe « Les performances de lecture labiale » p216). Comme cela a déjà été précisé, les performances de lecture labiale n'étaient pas réalisées systématiquement au cours du suivi à l'UIA, nos résultats sont donc à relativiser (voir paragraphe « Les performances de lecture labiale », p216). De plus, le faible nombre d'observations ne nous a pas permis de réaliser de test statistique. Toutefois on peut voir que les valeurs moyennes sont proches de celle retrouvées chez les patients du groupe transversal. Les performances variaient entre 19 et 37 % contre 22,5 et 34 % pour le groupe transversal (voir Figure 10.14).

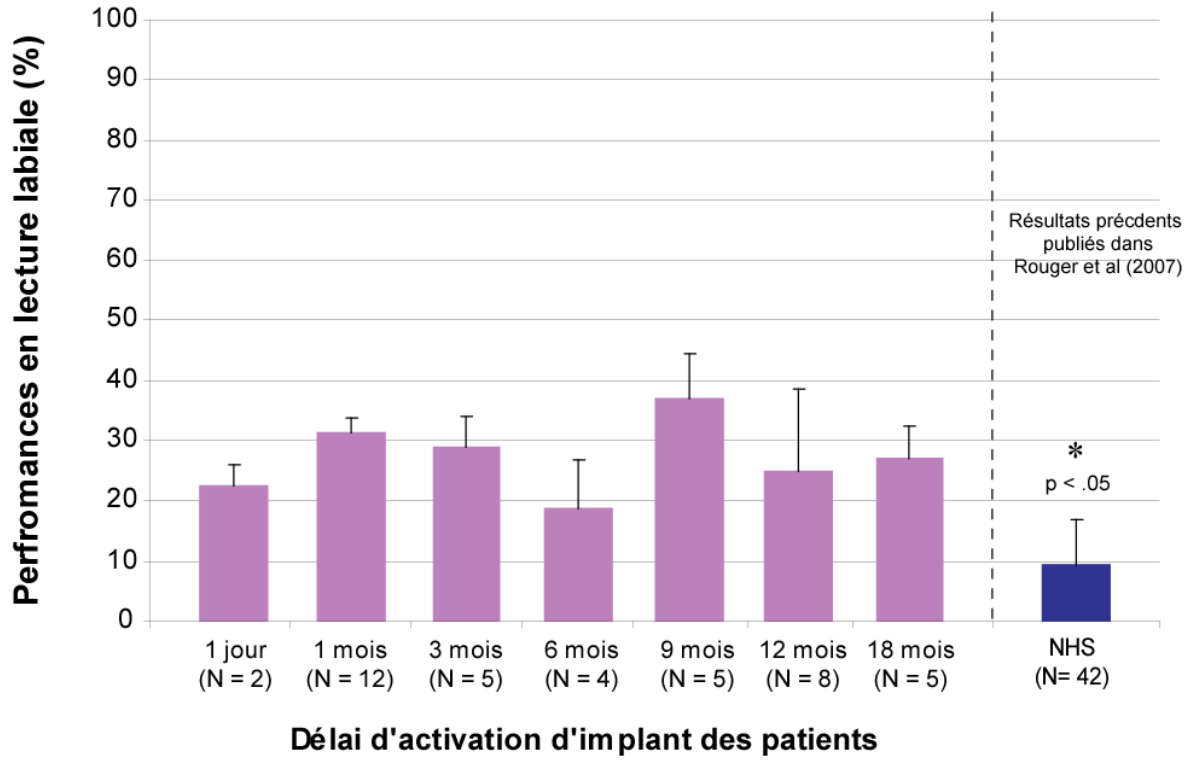


Figure 10.14 : Performance de lecture labiale pour les patients implantés cochléaires du groupe suivi, et comparaison avec des NHS d'une étude précédente (Rouger et al., 2007). Les variations du niveau de reconnaissance de mots dissyllabiques pour les patients implantés cochléaires (IC) n'étaient pas significatives, et les performances étaient toujours supérieures à celui des NHS.

Nos résultats confirment également ce que nous avions observé pour le groupe transversal, c'est-à-dire que les patients avaient de bien meilleures performances de lecture labiale que les normo-entendants (9,4%) et que ces performances, ne décroissaient pas au cours de l'expérience avec l'implant.

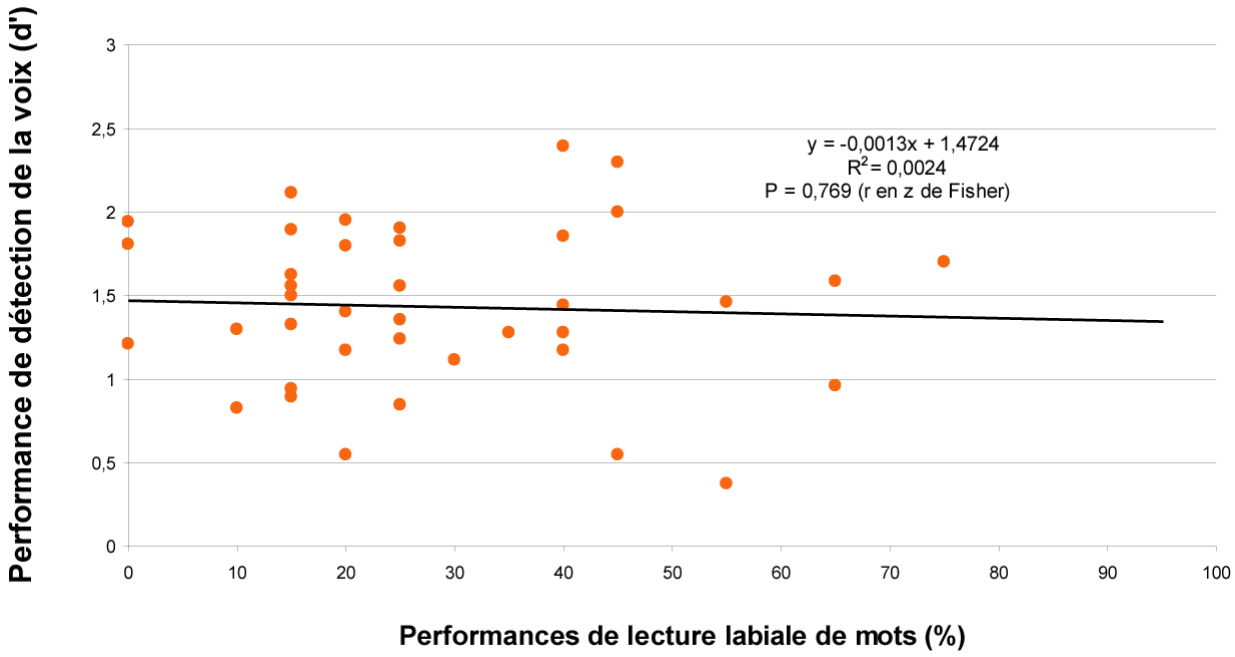


Figure 10.15 : Corrélation non significative entre les performances de détection de la voix (*d'*) et celles de lecture labiale (%).

En ce qui concerne la corrélation entre les performances de lecture labiale et celles de détection de la voix, nous n'avons pas retrouvé de corrélation significative (voir Figure 10.15). La corrélation ici est quasi-nulle ($R^2 = 0,0024$). Afin de savoir si les données globales de l'échantillon reflétaient bien ce qui se passait au niveau individuel, nous avons regardé les corrélations entre les performances de lectures labiales et les performances de détection de la voix pour chaque sujet, lorsque le nombre de données disponibles le permettait. La corrélation n'a été analysée au niveau individuel que lorsque les données de lecture labiale étaient disponibles pour au moins 4 mesures sur 7 pour un patient. Au total la corrélation individuelle a été établie pour 6 patients sur les 14 du groupe. Pour 3 patients la corrélation était négative avec des coefficients de corrélation R^2 compris entre 0,3 et 0,7. Pour 1 autre patient la corrélation était quasi-nulle (coefficient $R^2 = 0,04$) et pour les 2 derniers patients la corrélation était faiblement positive (coefficient $R^2 = 0,15$ et 0,1). Ainsi, pour les patients dont suffisamment de données étaient disponibles, lorsqu'elle était positive, la corrélation était plus forte que lorsqu'elle était négative. De façon générale on peut donc difficilement conclure à une corrélation, d'autant plus qu'il manque beaucoup de données de lecture labiale et que ces données manquantes pourraient changer le profil de cette corrélation.

Ces résultats ne corroborent pas ceux du groupe transversal où nous avions retrouvé une corrélation positive et significative. Toutefois ici, puisqu'il s'agit de mesures en suivi, et que nous

sommes au plus près de l'évolution individuelle des performances, nous pouvons penser que les valeurs sont plus proches de la réalité, et que les résultats de ce groupe sont plus fiables. Nous pouvons donc dire que les performances de lecture labiale ne sont pas du tout corrélées à celle de la détection de la voix.

10.1.1.4 ROLE DE LA PROTHESE AUDITIVE CONTROLATERALE SUR LES PERFORMANCES AVEC L'IMPLANT SEUL

Comme cela a déjà été évoqué, dans cette partie nous nous sommes attachés à rechercher un effet bénéfique du port de la prothèse auditive sur les performances de détection de la voix avec l'implant seul.

GROUPE TRANSVERSAL

En raison du profil comparable des patients implantés depuis 1 jour, avec et sans prothèses, aucune hypothèse n'était raisonnablement attendue. En conséquence, l'effet du port d'une prothèse auditive controlatérale sur les performances avec l'implant seul, a été analysé uniquement sur les patients implanté depuis plus d'un mois.

Sur les 28 patients du groupe transversal, 12 portaient une prothèse auditive controlatérale. La répartition du port de prothèse, au sein des groupes de délai d'activation étant inégale, nous avons choisi de regrouper tous les sujets dont l'implant était activé depuis 1 mois ou plus. Nous avons dons comparé le groupe de patients porteurs d'une prothèse controlatérale au groupe de patients qui n'en portaient pas, indépendamment du délai au-delà d'un mois post-activation. Parmi les patients porteurs d'une prothèse, les délais d'activation allaient de 1 mois à 2 ans. Chez ceux qui n'en portaient pas, les délais d'activation allaient de 1 mois à 10 ans. Pour ce dernier groupe, seul un patient était implanté depuis 10 ans, un autre depuis 6 ans, tandis que les autres étaient implantés depuis 3 ans ou moins.

Nos résultats montrent que chez les patients implantés depuis plus de 1 mois, les patients porteurs d'une prothèse controlatérale à l'implant avaient tendance à être meilleurs que ceux qui n'en portaient pas dans une tâche de reconnaissance de la voix, alors qu'ils étaient testés dans les mêmes conditions, c'est-à-dire avec l'implant seul. Mais cette tendance n'était pas significative. Lorsqu'on regarde plus en détail la détection de stimuli vocaux selon qu'ils sont langagiers ou non langagiers, on observe quand même certaines différences (voir Figure 10.16). En effet, à partir d'un mois, les patients porteurs de prothèse en plus de l'implant, avaient des performances significativement meilleures que ceux qui n'en portaient pas (Mann-Whitney ; $p = 0{,}0296$). Le port d'une prothèse controlatérale à l'implant semble donc améliorer les performances des patients, au moins en ce qui concerne les stimuli vocaux non langagiers. Toutefois cette différence ne se retrouvait pas sur l'ensemble des stimuli vocaux.

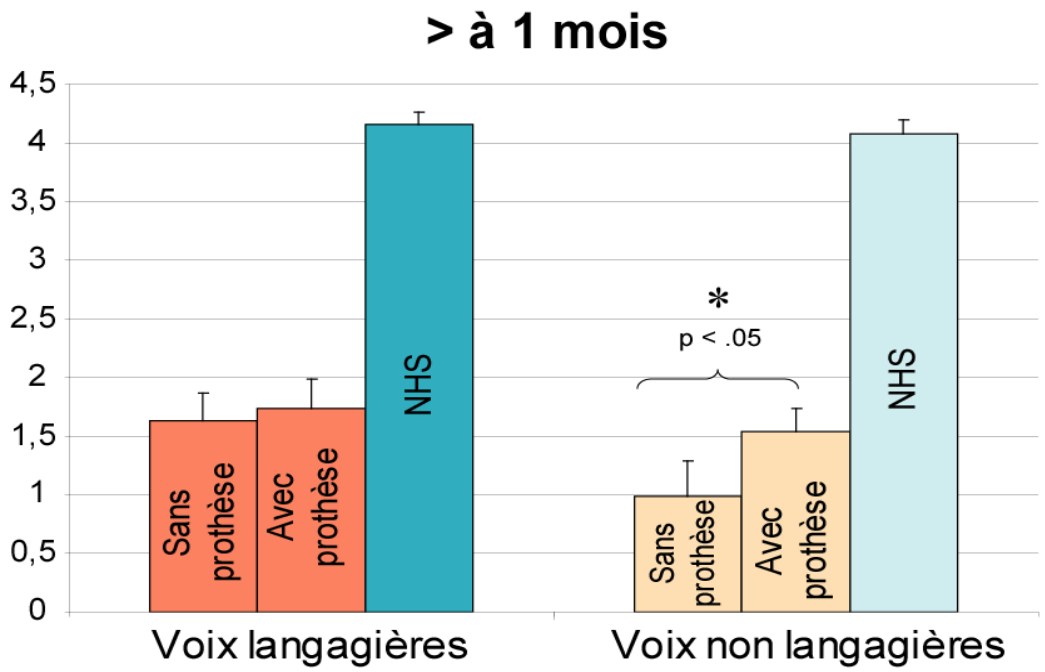


Figure 10.16 : Performances (±SE) des patients du groupe transversal, après 1 mois d'activation de l'implant, selon qu'ils portent une prothèse controlatérale au quotidien (N = 12) ou non (N = 10). Le test était toujours fait avec l'implant seul. Les patients implantés cochléaires porteurs d'une prothèse auditive controlatérale avaient de meilleures capacités de détection de la voix non langagière. Aucune différence n'était retrouvée pour la voix langagière. Dans les deux cas les patients restaient très déficitaires par rapport aux sujets normo-entendants (NHS ; N = 14).

En ce qui concerne les performances de compréhension de mots et les performances de lecture labiale, aucune différence, ni même tendance, n'a été retrouvée entre les patients implantés porteurs de prothèse et ceux qui n'en portaient pas.

GROUPE SUIVI

Pour le groupe du suivi, sur les 14 patients inclus, 5 portaient une prothèse controlatérale, mais rappelons que les patients étaient toujours testés avec l'implant seul.

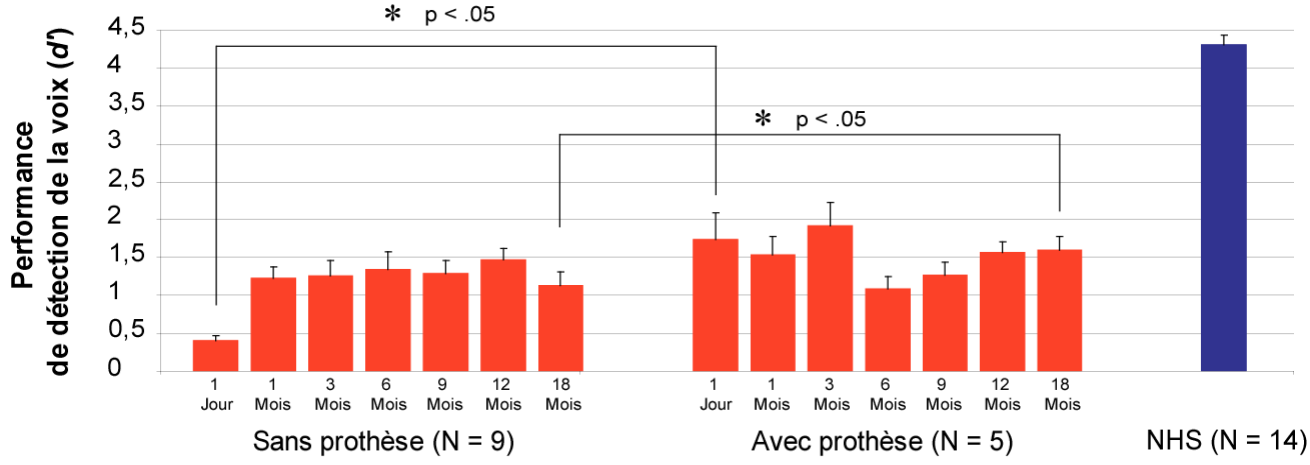


Figure 10.17 : Performances (±SE) des patients du groupe de suivi, selon qu'ils portent une prothèse controlatérale au quotidien ou non, pour chaque mesure du suivi. Le test était toujours fait avec l'implant seul. Les patients implantés cochléaires porteurs d'une prothèse auditive controlatérale avaient significativement de meilleurs capacités de détection de la voix, au premier jour de l'activation de l'implant et à plus de 18 mois.

Lors de l'activation de l'implant, on peut voir que les patients porteurs de prothèse avaient de meilleures performances que les patients qui n'en portaient pas (Mann-Whitney ; $p < 0,05$). Cette meilleure habileté des patients avec une prothèse est difficile à expliquer car au départ les sujets sont relativement égaux et naïfs face à la compréhension au travers de l'implant, et ont la même expérience de la prothèse. En regardant individuellement la clinique de ces patients, rien ne peut expliquer de façon évidente cette différence, que ce soit la marque ou le modèle de l'implant, la stratégie, ou encore l'étiologie de la surdité. L'explication qui semble la plus plausible est un biais d'échantillonnage. En effet, à 1 jour seul 3 patients sur 5 porteurs de prothèse ont passé le test, et seulement 6 sur 9 pour le groupe non porteurs de prothèse. En effet, lorsqu'on regarde à 1 mois les performances des sujets qui étaient manquants au premier jour de l'activation de l'implant, on remarque que c'est ceux là qui avaient de meilleures performances par rapport au groupe pour la mesure à 1 mois (voir Tableau 10.2). Inversement, pour le groupe avec prothèse, on remarque que les sujets manquants à 1 jours d'activation de l'implant étaient ceux qui avaient des mauvaises performances par rapport groupe à 1 mois d'activation de l'implant. Il est rare qu'un sujet diminue ses performances au moins entre 1 jour et 1 mois d'activation (seuls 2 patients ont des performances qui diminuent entre 1 jour et 1 mois), on peut donc penser que les patients qui sont absents à 1 jour d'activation, s'ils étaient présents, diminueraient les performances pour le groupe de patients porteurs de prothèse et les augmenteraient pour celui qui n'en porte pas. Ceci corrobore l'idée que les sujets manquants à 1 jour peuvent, au moins en partie, expliquer les différences de performances. Comme nous l'avons montré sur le Tableau 10.1 (voir p218), aux évaluations suivantes, les patients manquants sont moins nombreux.

	Sans prothèse	Avec Prothèse
Patients manquants au premier jour de l'activation	**1,16 (N = 2)**	**1,82 (N = 3)**
Patients présents au premier jour de l'activation	**1,35 (N = 3)**	**1,34 (N = 6)**

Tableau 10.2 : Performances moyennes (±SE) de détection de la voix pour les patients vus à 1 mois. Les patients étaient regroupés en fonction de leur absence ou non lors du 1er jour de l'activation et en fonction du port ou non d'un prothèse controlatérale à l'implant. La valeur indiquée correspond à la mesure de *d'* et le chiffre indiqué entre parenthèse correspond à l'effectif.

A partir d'un mois après l'activation de l'implant, on peut voir que les patients ayant une prothèse avaient tendance à avoir de meilleures performances que ceux qui n'en avaient pas, exceptées à 6 et 9 mois, mais la différence n'était significative qu'à 18 mois (voir Figure 10.17 ; Mann-Whitney ; $p < 0,05$). Le port d'une prothèse controlatérale à l'implant ne permet donc pas vraiment une amélioration des performances plus rapide, puisque la seule différence significative observée était assez tardive (18 mois post-activation de l'implant).

Afin de comparer au groupe transversal, nous avons moyenné les performances de chaque patient entre 1 et 18 mois (voir Figure 10.18). Ainsi on peut constater qu'on ne retrouve pas les résultats observés pour le groupe transversal au-delà de 1 mois d'implantation. Ici, après 1 mois d'activation de l'implant, les patients ayant une prothèse controlatérale avaient des résultats

similaires à ceux qui n'en portaient pas. Ce résultat était vrai que l'on considère les stimuli vocaux langagiers ou non langagiers. On peut voir sur la Figure 10.18 qu'aucune différence significative n'a été retrouvée lorsqu'on moyenne les données des sujets au cours du suivi après 1 mois d'activation d'implant.

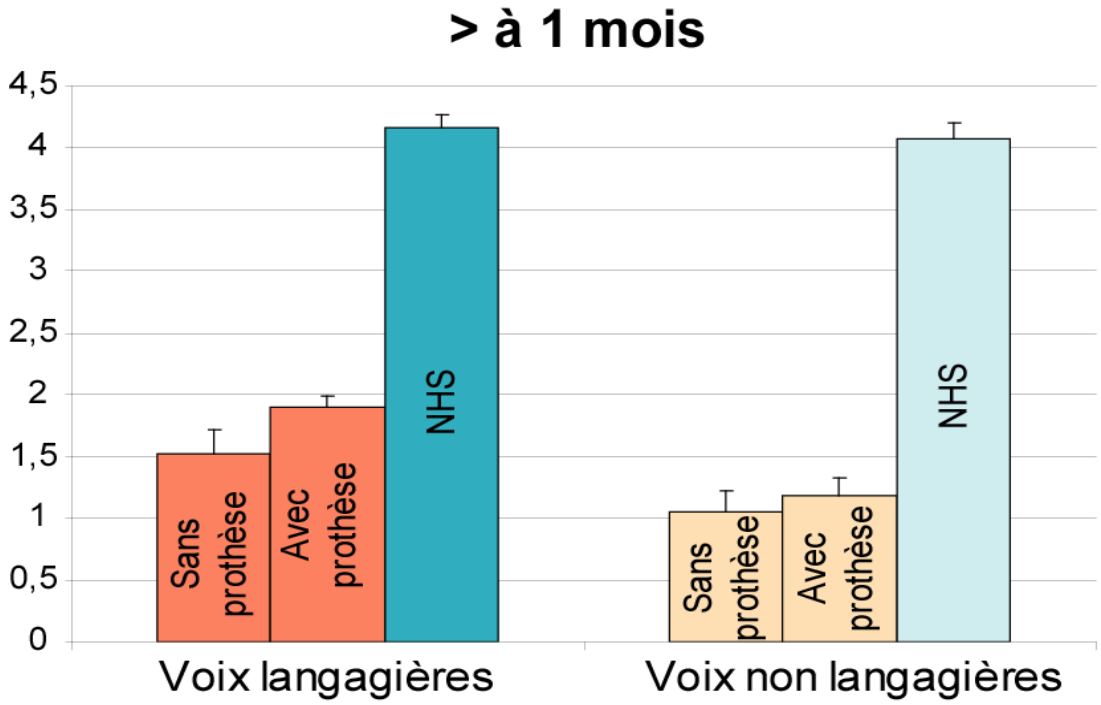


Figure 10.18 : Performances (±SE) moyennées des patients du groupe transversal, après 1 mois d'activation de l'implant, selon qu'ils portent une prothèse controlatérale (N = 5) au quotidien ou non (N = 9). Le test était toujours fait avec l'implant seul. Les patients implantés cochléaires porteurs d'une prothèse auditive controlatérale avaient des performances similaires à celles des patients qui n'en portaient pas.

En ce qui concerne les performances de compréhension de mots et les performances de lecture labiale, aucune différence, ni même tendance, n'a été retrouvée entre les patients implantés porteurs de prothèse et ceux qui n'en portaient pas.

10.1.2 RESUME ET DISCUSSION PRELIMINAIRE

Le premier élément important est que les résultats du groupe transversal et ceux du groupe de suivi ont donné des résultats très similaires. Les différences qui existaient entre les deux étaient rares. Les résultats du suivi ont essentiellement permis de préciser certaines variations, comme la dissociation entre les performances de reconnaissances de stimuli vocaux langagiers et non langagiers, qui n'apparaissaient pas pour les sous-groupes « 1 à 6 mois » et « 6 à 18 mois ». Mais il est probable qu'avec un plus gros échantillon de patients, ces différences seraient apparues. Il convient donc de dire qu'en ce qui concerne ce test de détection de la voix, l'approche transversale nous paraît plus appropriée, à condition qu'elle soit réalisée sur un échantillon conséquent,

puisqu'elle est moins lourde d'un point de vue de la durée de la période de test, et puisque les résultats étaient très semblables à ceux du suivi.

Les résultats pour le test de détection de voix montrent plusieurs choses. La première est que les patients implantés cochléaires, lors de l'activation de l'implant, avaient de gros déficits de détection de la voix par rapport aux sujets normo-entendants. Ce déficit semblait légèrement s'améliorer 1 mois après l'activation mais de manière non significative. Puis les performances stagnaient en restant très déficitaires. Lorsqu'on comparait les patients au sujets normo-entendants on a pu montrer que le déficit était équivalent à celui des normaux-entendant en condition de simulation d'implant à 4 canaux, voire 2 canaux. Ceci était vrai même pour des patients expérimentés (plus des 18 mois post-activation), alors que ces patients avaient un implant qui comprenait en moyenne 18 électrodes.

Ensuite, un autre résultat important est la dissociation des performances entre les stimuli vocaux de types langagiers et non langagiers. Les patients, de même que les NHS (voir paragraphe « Comparaison entre stimuli langagiers et non-langagiers", p156), arrivaient toujours à mieux détecter les stimuli langagiers, quelque soit la durée de l'activation de l'implant. Les analyses ont montré que mêmes si les pitches moyens de stimuli vocaux langagiers et non langagiers étaient significativement différents, ils ne différaient pas en revanche, du pitch moyen des stimuli non voix. Il semble donc que d'autres paramètres viennent influencer cette meilleure reconnaissance des stimuli langagiers. On peut notamment évoquer des modulations temporelles qui sont plus riches pour le langage. Or on sait que ces modulations temporelles sont bien conservées par le vocoder et les processeurs d'implants cochléaires. Il est donc probable que ces modulations temporelles soient des informations supplémentaires permettant aux sujets de mieux détecter ces types de stimuli vocaux. Pour les stimuli vocaux non langagiers, les sujets vont baser leur détection essentiellement sur les indices spectraux, qui eux, sont très dégradés par le processeur d'implant cochléaire et le vocoder. Cela peut donc expliquer la différence de performance entre stimuli langagiers et non langagiers. Mais pour les deux types de stimulus, les performances sont similaires à celles des NHS en condition de vocoder à 4 canaux. Or en moyenne les patients de ces deux groupes ont un implant comprenant 18 électrodes. Les performances sont donc plus faibles que ce à quoi il était raisonnable de s'attendre. De plus les performances en compréhension de mots sont très bonnes ce qui signifie que le processeur de l'implant seul n'est pas responsable du déficit en détection de la voix. Il y a donc bien une dissociation entre la récupération de la compréhension du langage et de la détection de la voix. Cette dissociation peut aussi être due à l'information temporelle qui est préservée au travers du processeur de l'implant et qui permet une compréhension du langage correcte (Q. J. Fu, Chinchilla et al., 2004; Shannon et al., 1995). Mais cette information temporelle ne peut pas expliquer toute la différence de récupération entre la tâche de compréhension de mots et la tâche de détection de la voix. En effet, si l'information temporelle permet la bonne compréhension de mots avec l'expérience de l'implant, les performances en détection de stimuli langagiers restaient déficitaires même si elles étaient meilleures que pour les stimuli non langagiers. L'idée du contexte

lexical qui aide la reconnaissance de mots partiellement compris est une hypothèse tout à fait probable. Toutefois, une autre hypothèse peut aussi être envisagée, il s'agit d'une bonne récupération des aires de la compréhension du langage et d'une mauvaise récupération de l'aire de la voix. Nous reviendrons sur cette dernière hypothèse ultérieurement, à la lumière de résultats nouveaux.

L'effet de la prothèse dans nos résultats, n'est pas évident. Toutefois, les données du groupe transversal nous permettent d'envisager que le port d'une prothèse controlatérale à l'implant serait bénéfique dans la tâche de détection de la voix. L'effet du port d'une prothèse en plus de l'implant est un élément intéressant qui mérite d'être approfondi par des tests spécifiques.

10.2 TESTS COMPORTEMENTAUX DE LA RECONNAISSANCE DE L'IDENTITE DE LA VOIX

Rappelons que l'objectif de cette partie des tests comportementaux était de s'intéressé à une autre partie de l'analyse de la voix, qui est la perception de l'identité.

10.2.1 TEST DE CATEGORISATION DU GENRE

10.2.1.1 MATERIEL ET METHODES

SUJETS

Trois groupes de sujets ont été inclus pour des approches différentes.

Un premier groupe de sujets, dit groupe transversal était constitué de 38 patients, dont 17 hommes et 21 femmes, dont la moyenne d'âge était de 54,5 ans (+/- 14,3). Les sujets avaient des délais d'activation variables allant de 1 jour à 131 mois et ont été répartis en 4 groupes : 1er jour d'activation (N = 9), 1 à 6 mois d'activation (N =4), 6 à 18 mois d'activation (N = 9) et plus de 18 mois d'activation (N = 10).

Le deuxième groupe était des patients suivi à partir du premier jour de leur activation jusqu'à 18 mois post-activation. Ce groupe était constitué de 14 patients, dont 5 hommes et 9 femmes, avec une moyenne d'âge à l'inclusion de 51,5 ans (+/- 14,5). Ce groupe de patients suivis était le même que pour le test de détection de la voix. Nous nous attacherons à comparer le profil d'évolution de ces patients dans les deux tests qu'ils ont passé.

Le troisième groupe de patient étaient recruté afin de tester les capacités perceptives des patients en ce qui concerne le pitch et le timbre. Autrement dit ce groupe serait testé avec le continuum faisant varier pitch et timbre, ainsi qu'avec les continua à pitch constant et timbre constant (voir « Test 2 » et « Test 3 » de la Figure 10.19). Pour cela nous avons donc recruté des

patients expérimentés avec l'implant. Le critère d'inclusion de ce groupe était de réussir la catégorisation du test du genre. Pour cela les patients devaient avoir une pente du point C50 à plus ou moins 2 écart-type de la moyenne du groupe ayant plus de 18 mois d'activation d'implant, ce groupe étant le plus expérimenté de notre échantillon. Autrement dit la pente des sujets de ce groupe, pour être inclus, devait être comprise entre 0,11 et 0,31 (voir les résultats concernant « La pente », p242). Nous avons donc réussi à inclure 10 patients qui correspondaient à ce critère d'inclusion. Le groupe était constitué de 5 hommes et 5 femmes ayant des délais d'activation de l'implant variable allant de 3 à 132 mois, avec une moyenne d'âge de 55 ans (+/-SD 12,4). Les patients de ce groupe n'étaient vus qu'une seule fois.

Certains patients portaient une prothèse auditive controlatérale au quotidien. Toutefois tous les patients passaient le test avec l'implant seul. L'effet du port de cette prothèse sera analysé à part. Un tableau descriptif pour l'ensemble des patients est donné en annexe (voir Tableau 14.2, p345).

STIMULI

Pour ce test, nous avons utilisé les mêmes stimuli que pour les NHS (voir paragraphe « Stimuli », p175). Il s'agissait donc de 3 continua réalisés par morphing entre une voix d'homme et une voix de femme avec le logiciel Matlab 6.5®. Le premier a consisté en un glissement progressif et linéaire des fréquences fondamentales (pitch) et des formants (timbre) de la voix d'homme vers ceux de la voix de femme (voir Figure 10.19, test 1). Pour le deuxième continuum, un pitch moyen (169 Hz) était maintenu constant et seuls les formants variaient de la voix d'homme vers la voix de femme (voir Figure 10.19, test 2). Pour le dernier continuum, des formants moyennés (timbre) entre la voix d'homme et de femme, étaient maintenus constants, et seule la fréquence fondamentale variait (voir Figure 10.19, test 3).

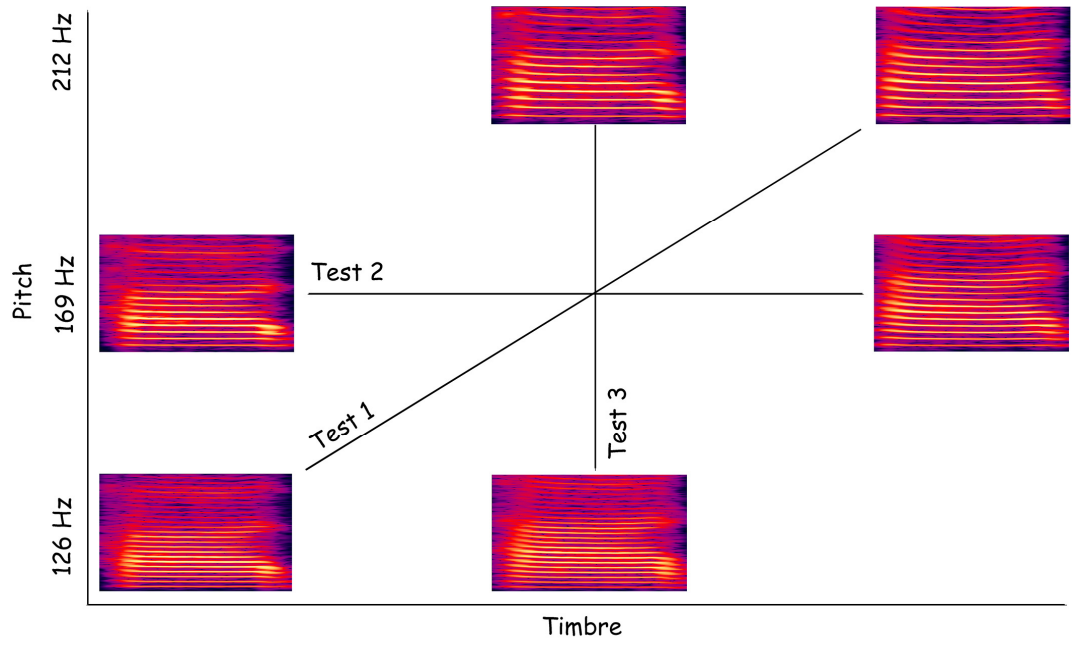


Figure 10.19 : Illustration des différents continua réalisés par morphing. Le test 1 fait varier à la fois le pitch et timbre de la voix original « homme » vers la voix originale « femme ». Le test 2 maintient un pitch intermédiaire en les 2 voix constant, et fait varier uniquement le pitch. Enfin, le test 3 maintient un timbre moyen constant en faisant varier le pitch.

Mille pas pour chaque continuum ont été générés avec ce morphing et nous disposions donc de 1001 voix pour chacun. Pour le test de la catégorisation du genre, nous avons utilisé 11 voix à égale distance physique pour chaque continuum.

PROCEDURE

Les sujets des 3 groupes étaient vus au cours de leur venue à l'hôpital dans le cadre de la prise en charge par l'unité d'implant auditif (UIA) de Purpan, et étaient testés dans une pièce insonorisée du service ORL. Les stimuli étaient présentés en champ libre via deux haut-parleurs (KINYO, modèle PS-240) et le volume était ajusté pour être environ à 72 dB SPL à 60 cm des enceintes. Les patients qui portaient une prothèse auditive controlatérale à l'implant, la retiraient et mettaient un bouchon d'oreille, afin qu'ils soient testés avec l'implant seul, et que les conditions de test puissent être comparables entre tous les patients. Les sujets étaient placés à cette même distance des haut-parleurs et face à un clavier, avec le quel ils avaient pour consigne de répondre si le stimulus présenté était une voix d'homme ou de femme, en appuyant sur une des deux touches « contrôle » correspondant à leur réponse. Pour des raisons de difficulté évidente chez les patients, aucune consigne de rapidité n'était donnée, et l'expérimentateur indiquait également qu'il leur était possible de ne pas répondre, mais qu'ils devaient essayer de répondre le plus possible. Après 10 secondes sans réponse du sujet, le programme de stimulation passait à l'essai suivant. Les stimuli étaient présentés aléatoirement et la durée entre deux essais était de 1,5 secondes. Le test durait entre 5 et 10 minutes, selon les temps de réponse des sujets et le nombre de non réponse. La présentation des stimuli et l'enregistrement des réponses des patients se faisaient à l'aide du logiciel Matlab 6.5®.

Les 38 sujets du groupe transversal n'étaient vus qu'une seule fois, ainsi que les patients du 3ème groupe qui passaient aussi les conditions pitch constant et timbre constant. Pour le groupe transversal, les patients vus au premier jour de l'activation, étaient des sujets qui furent inclus dans le groupe suivi. Pour ce deuxième groupe, les sujets en question, inclus au premier jour de l'activation et revus à des délais d'activation fixes qui correspondaient aux évaluations de la prise en charge par l'UIA : 1, 3, 6, 9, 12 et 18 mois post-activation. Le suivi impliquant des contraintes de temps particulières et notamment pour les sujets, tous les sujets n'ont pas pu être vus à tous les temps du suivi.

ANALYSES

Pour les analyses, seules les données des patients qui avaient fait moins de 25% de non réponses étaient analysés, et tous les essais où les patients n'avaient pas répondus étaient éliminés. Ensuite la même analyse que pour les sujets normo-entendants était appliquée (voir « Analyses », p177). On calculait ainsi pour chacune des 11 voix du continuum (10 présentations) le pourcentage de réponses « femme », ce qui nous permettait d'estimer une courbe sigmoïde (tangente hyperbolique, exemple présenté en Figure 9.16, p177). A partir du point d'inflexion de la courbe, le point C50, nous avons analysé la pente de la courbe en ce point et le seuil. En résumé, la pente va refléter la qualité de la catégorisation, et le seuil va refléter à partir de quel moment du continuum les

sujets vont catégoriser les voix comme étant plutôt des voix d'hommes ou plutôt des voix de femmes. Nous nous sommes également intéressés, comme pour les normo-entendants, au pourcentage de reconnaissance correcte de la voix d'homme et de la voix de femme.

Nous avons donc comparé les valeurs des pentes et des seuils au point C50, de l'erreur de l'estimation de la courbe et de la reconnaissance correcte de la voix d'homme et de la voix de femme, en fonction du délai d'activation d'implant pour le groupe transversal et le groupe de suivi. Pour le troisième groupe nous avons comparé ces mêmes valeurs entre les 3 continua : variation du pitch et timbre (voir Figure 10.19, test 1), pitch constant (voir Figure 10.19, test 2) et timbre constant (voir Figure 10.19, test 3).

10.2.1.2 RESULTATS

DES PATIENTS SUPRANORMAUX OU UNE LIMITE METHODOLOGIQUE ?

En analysant les données de la pente pour la catégorisation du genre, nous avons pu observer que certains patients avaient des performances extrêmement bonnes, que ce soit pour le groupe transversal ou pour le groupe en suivi. En fait, nous avons pu noter qu'il y avait deux types de patients, ceux dont les performances pour la pente étaient inférieures à celle des NHS et ceux dont les performances de la pente étaient bien au-delà de celles des NHS en condition originale, des patients « supranormaux ». Afin d'aborder les données de manière plus fine, nous avons considéré un critère pour distinguer les patients ayant des performances normales de ceux ayant des performances supranormales. Ce critère était de « plus ou moins 2 écart-types de la moyenne des NHS en condition originale », soit des valeurs comprises entre 0,039 et 0,805.

Groupe transversal

Dans le groupe transversal, les patients dit supranormaux étaient au nombre de 6 (2 hommes) et avaient des valeurs de pente de la courbe de catégorisation comprises entre 0,805 et 1,086, soit 94 à 160% de plus que la moyenne des NHS en condition originale, qui était de 0,417. Un exemple de réponses d'un patient supranormal est montré en Figure 10.20.

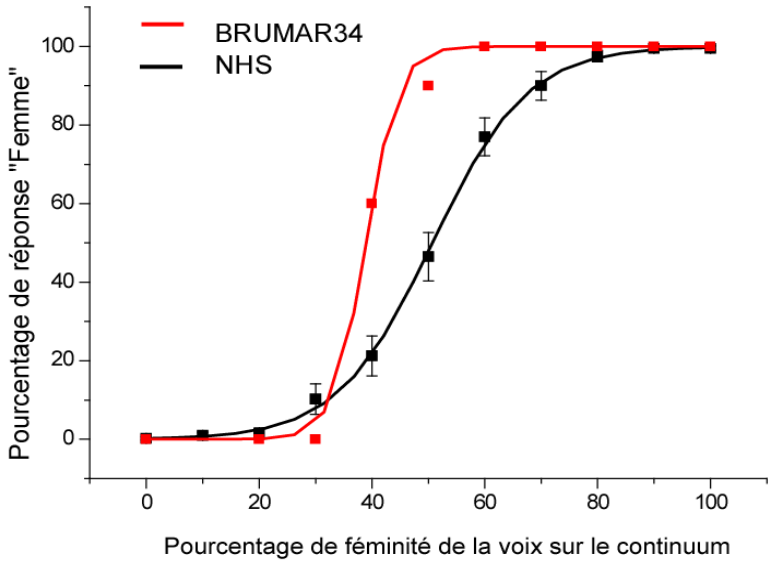


Figure 10.20: Exemple de réponse d'un patient supranormal, estimation de la courbe sigmoïde correspondante et comparaison avec les données moyennes des sujets normo-entendants.

On peut voir sur la Figure 10.21 la très bonne catégorisation de ce patient (en rouge) avec 100% de certitude pour les 4 voix les plus masculines et les 5 voix les plus féminines. Entre les deux, les valeurs variaient assez rapidement. Ce patients est très représentatif des 6 patients supranormaux. Pour ce groupe de patients supranormaux, les délais d'activation étaient variables, allant de 1 mois à 7 ans (voir Figure 10.21), et la moyenne d'âge était de 54,7 ans contre 54,5 ans pour l'ensemble du groupe. Aucun élément clinique évident ne pouvait expliquer ces meilleures performances de catégorisation, qu'il s'agisse du processeur de l'implant, de la stratégie de codage, la durée de surdité avant implant, etc.

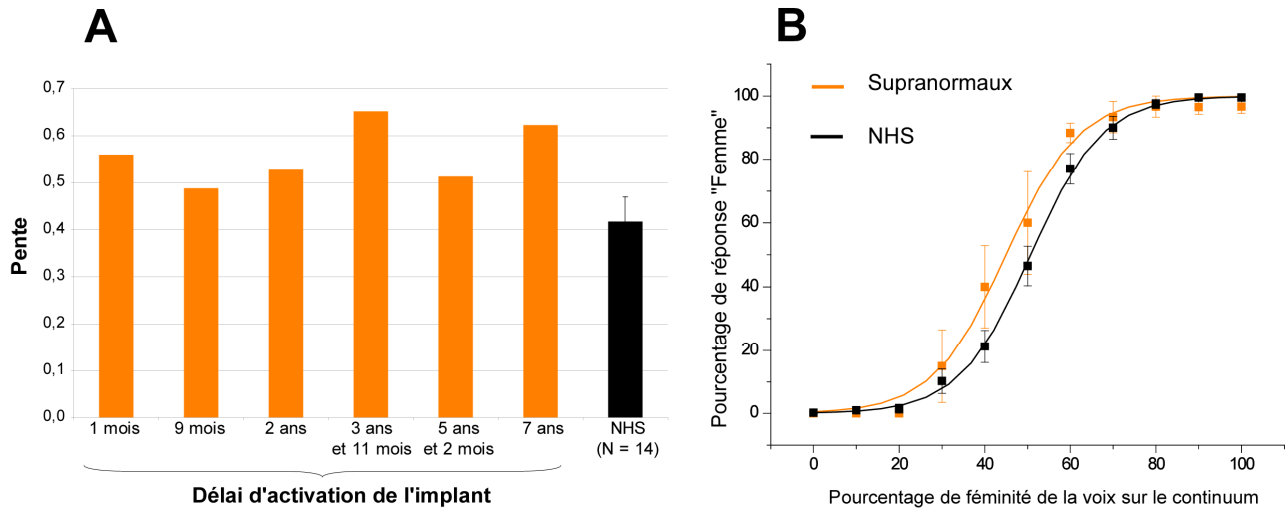


Figure 10.21 : **A.** Comparaison des pentes des patients supranormaux avec la pente moyenne des NHS. On peut voir que les délais d'activation de l'implant sont très variables et qu'on trouve même un patient supranormal dont l'implant est activé depuis 1 mois. **B.** Courbe psychophysique sur les données moyennées des patients supranormaux (orange) comparée à celle des NHS (noir). Le profil des courbes est proche sur les données moyennées, mais on peut voir qu'il y a plus de variabilité pour les supranormaux que pour les patients.

Groupe suivi

Pour le groupe du suivi, nous avons appliqué le même critère d'analyse que pour le groupe transversal, à savoir que nous avons analysé indépendamment les patients dont la pente, à au moins une mesure du suivi, dépassait 2 écart-types de la moyenne de NHS, c'est-à-dire avait une pente supérieure à 0,805, et les patients qui ne dépassaient pas ce seuil. Pour le suivi, 3 patients étaient supranormaux pour au moins une mesure du suivi, c'est-à-dire avaient une valeur de pente supérieure à 0,805. Les résultats sont présentés sur la Figure 10.22. On peut voir que, sur les 3 patients, 2 des patients dépassaient la valeur de 0,805 à 2 mesures du suivi pour l'un (GENMAR43), et à 3 mesures pour l'autre (ALAREN37). De plus lorsque cette valeur est dépassée, les valeurs sont parfois très importantes, allant jusqu'à 8,6 pour le patient GENMAR43 à 12 mois.

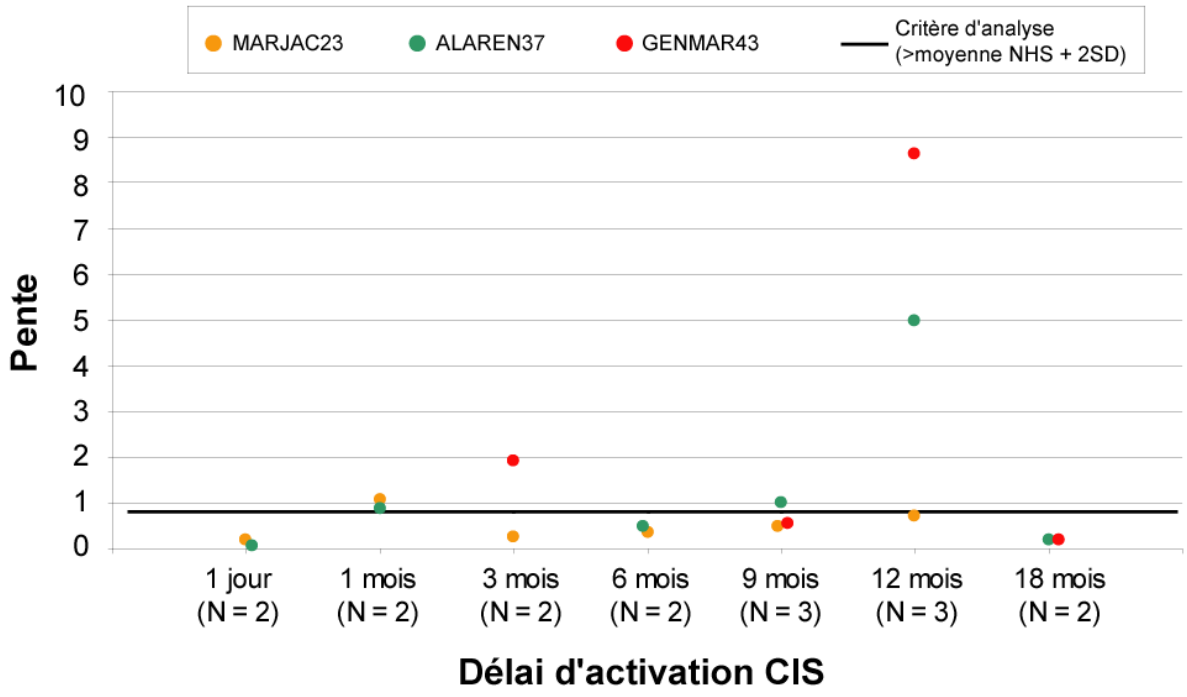


Figure 10.22 : Représentation des valeurs de la pente pour les patients ayant au moins une valeur supérieure au critère d'analyse au cours des mesures de suivis. Certains patients dépassent parfois plusieurs fois ce critère

La supériorité de ces patients à un moment donné, par rapport aux NHS reste difficile à expliquer. Les 3 patients supranormaux ont le même implant (Freedom Implant de Cochlear®), le même processeur (Freedom SP de Cochlear®) et la même stratégie de codage (ACE de Cochlear®). Mais ceci ne peut pas être la seule explication de leurs performances puisque d'autres patients du groupe de suivi n'ayant pas de performances supérieures aux NHS ont ces mêmes implant, processeur et stratégie de codage.

Une limite du modèle ?

On peut aussi remarquer sur la Figure 10.22, que le patient GENMAR43, alors qu'il avait une pente de catégorisation excessivement importante à 12 mois après l'activation de l'implant, retrouvait à 18 mois une valeur de pente tout à fait normale, c'est-à-dire semblable au reste des patients. Si on ne peut nier, en regardant les données brutes, que les performances de ces patients supranormaux sont très bonnes dans la tâche de catégorisation du genre (voir Figure 10.23), on peut en revanche se demander si les performances supranormales sont bien le reflet d'une compétence particulière de ces patients ou d'une limite du modèle en ce qui concerne les patients.

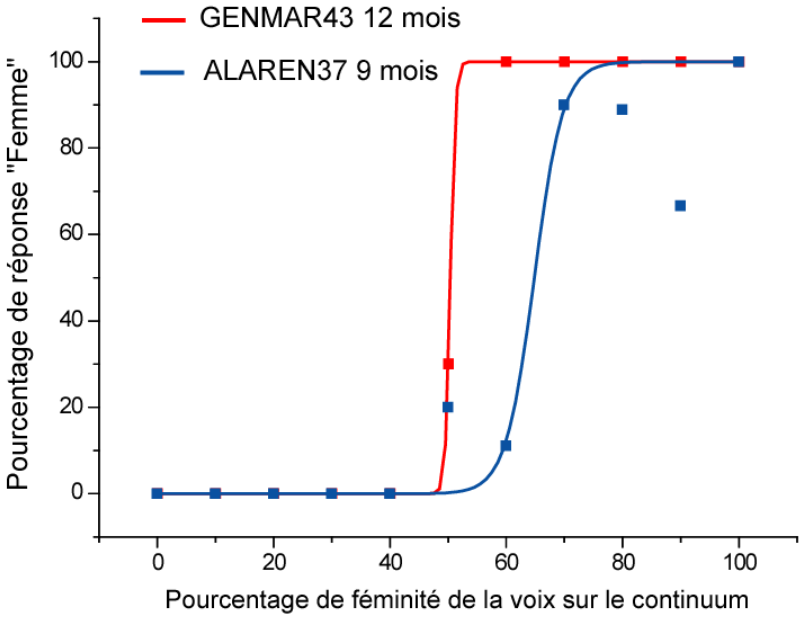


Figure 10.23 : Données du patient GENMAR43 à 12 mois (en rouge) et du patient ALAREN37 à 9 mois (en bleu). On peut voir que ces deux patients catégorisaient très bien.

Pour aborder la question des supranormaux, nous nous sommes également attardé sur les autres critères d'analyse des performances, comme le seuil, l'erreur de l'estimation de la courbe ainsi que les réponses aux extrêmes. Tout d'abord, en ce qui concerne le seuil, aucune différence significative n'a été retrouvée entre les valeurs observées chez les sujets supranormaux et les NHS, ni même entre les supranormaux et les autres patients. De même, en ce qui concerne les réponses aux extrêmes, aucune différence significative n'a été retrouvée entre les supranormaux et les NHS, excepté lors du suivi, ou les patients supranormaux avait des taux de reconnaissance correcte des voix d'homme inférieures à celles des NHS au premier jour de l'activation et 1 mois plus tard. Pour le reste, les performances des deux groupes étaient similaires, que ce soit pour le suivi ou pour le groupe transversal. En revanche, une analyse de l'erreur a montré que dans l'ensemble, l'erreur est significativement plus élevée dans le groupe de patients supranormaux que chez les NHS (Mann-Withney ; $p = 0{,}037$). Toutefois l'erreur était significativement plus faible chez les patients supranormaux en comparaison des autres patients. Si l'on recoupe cette information avec la pente, on peut donc dire que le modèle sigmoïde appliqué aux patients supranormaux montre que la

catégorisation est plus forte mais le modèle significativement moins fiable lorsqu'appliqué à ces données. Ceci est probablement lié à une variabilité plus importante des réponses chez les patients, même supranormaux. Il est possible que les 10 présentations de chaque voix ne soient pas suffisantes pour appréhender au plus juste les valeurs de catégorisation des patients. En effet, en tenant compte des non-réponses, pour certains patients, les pourcentages de réponses « femme » ne sont faits, que sur 9 ou 8 présentations. Une seule réponse du patient peut changer considérablement l'estimation de la courbe. Or on sait que, cette tâche peut s'avérer assez difficile et stressante pour les patients, plus que pour les NHS, et il se peut qu'il y ait parmi leurs réponses, des réponses au hasard. On peut donc dire que des présentations supplémentaires de chaque stimuli, peut-être jusqu'à 20 ou 30 présentations, permettraient de mesurer plus finement les réponses des patients pour chaque voix du continuum et ainsi d'avoir une estimation de la courbe plus cohérente avec les mesures du reste du groupe auquel les sujets appartiennent. Cependant, une telle augmentation des présentations amènerait au moins à doubler le temps de la tâche pour les patients. Ceci reviendrait pour eux à faire la tâche pendant 10 à 20 minutes, ce qui serait beaucoup plus lourd pour eux. Une autre alternative serait de pouvoir pondérer la valeur de la pente avec l'erreur d'estimation de la courbe.

L'ensemble de ces observations nous amènent à deux conclusions. La première, c'est la pertinence d'analyser ces patients supranormaux à part, au risque qu'ils noient les effets observés dans le reste du groupe, et la deuxième, c'est que les patients implantés cochléaires, au moins certains d'entre eux, sont capables de très bien catégoriser le genre. Les résultats pour les sujets supranormaux ayant été déjà présentés, ceux qui suivent porteront sur le reste des sujets du groupe transversal dans une première partie puis ceux du groupe de suivi dans une deuxième partie.

RESULTATS DU GROUPE TRANSVERSAL

Pour certains patients du groupe transversal, l'estimation d'une courbe sigmoïde n'a pas pu être réalisée, car les réponses étaient faites au hasard ce qui donnait des valeurs de seuil et de pente complètement aberrantes. La Figure 10.24 montre un exemple de patients à 1 jour d'activation de l'implant avec des réponses au hasard. Nous avons donc exclu des analyses de seuil, de pente et d'erreur 3 patients sur 9 du groupe à 1 jour post activation et 1 patient sur 4 du groupe dont l'implant était activés depuis 1 à 6 mois. En revanche, les données de ces patients ont été conservées pour le calcul des reconnaissances correctes de la voix d'homme et de la voix de femme aux extrêmes du continuum. On peut tout de même noter que le fait que le nombre de patients répondant au hasard diminue au cours de l'expérience de l'implant est un indice de meilleure perception de la voix par les patients.

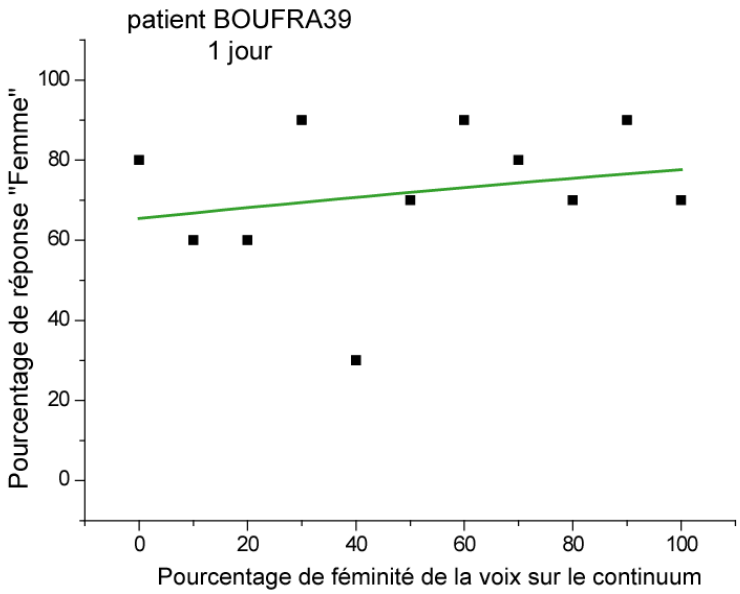


Figure 10.24 : Estimation de courbe sigmoïde (vert) pour un patient dont l'implant est activé depuis 1 jour. Les données brutes sont représentées par les carrés noirs. On peut voir ici que le modèlme sigmoîde ne semble pas très approprié.

Le seuil

L'analyse du seuil du C50 de la courbe nous a montré qu'il n'existait pas de différence entre les valeurs des patients des différents groupes de délais d'activation d'implants. En comparaison des NHS, le seuil à 1 jour d'activation de l'implant était différent de celui des NHS en condition originale (Mann-Whitney ; $p = 0,017$), et était décalé du côté de la voix de l'homme (voir Figure 10.25). On observait la même différence avec la condition 32 canaux de vocoder (Mann-Whitney ; $p = 0,026$). Toutefois ces différences étaient assez peu importantes, et dans tous les cas les valeurs du seuil se trouvaient au milieu du continuum, entre 45 et 55%. Pour les autres comparaisons, aucune différence n'a été retrouvée, excepté la condition 4 canaux de vocoder qui était différente de tous les groupes de patients mais ceci s'explique par le fait que cette mesure était déviante chez les NHS.

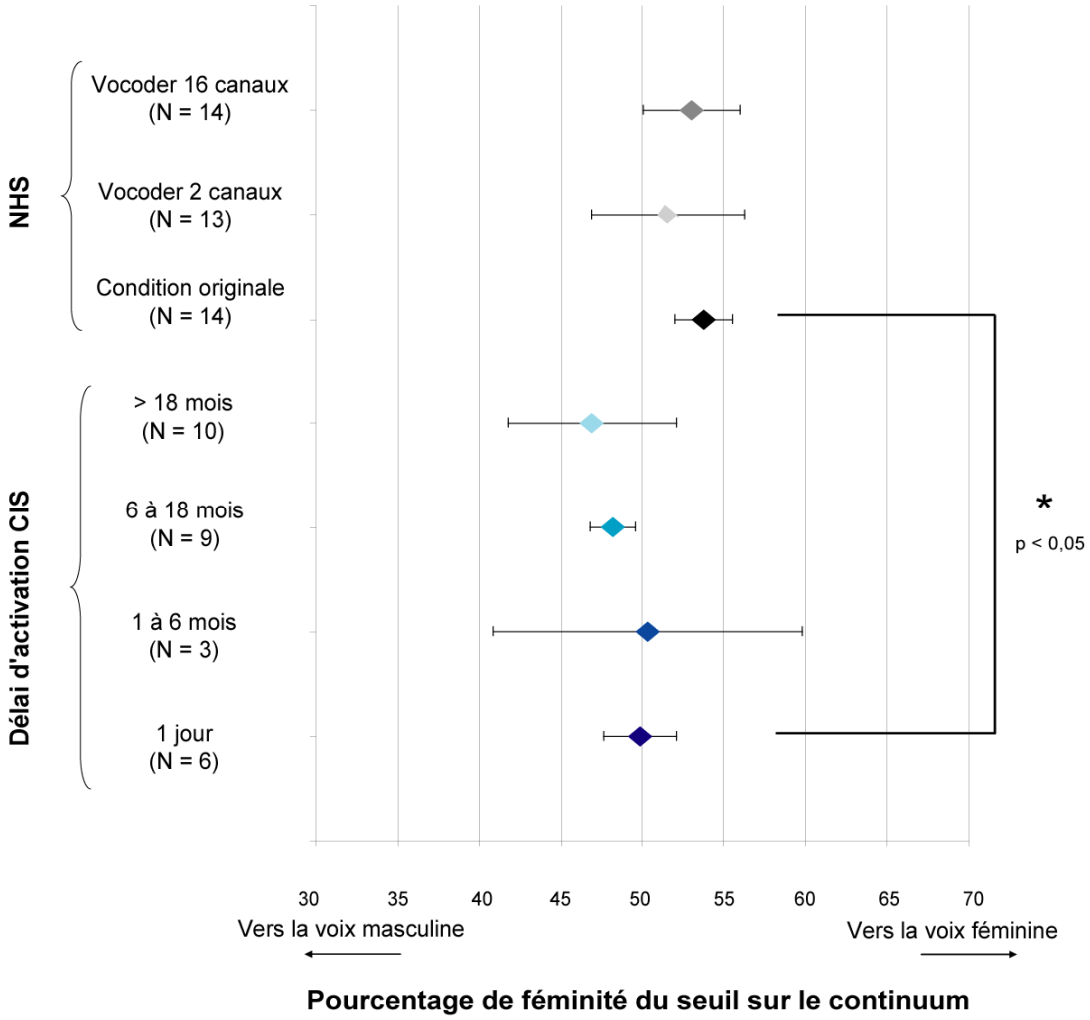


Figure 10.25 : Valeurs moyennes (± SE) du seuil du point C50 (point d'inflexion de la courbe) pour les patients du groupe transversal, en fonction des groupes de délai d'activation de l'implant. Aucune différence n'a été retrouvée entre les différents groupes de patients. Le groupe dont l'implant était activé depuis 1 jour avait un seuil significativement plus du côté de la voix masculine par rapport aux NHS en condition originale.

En résumé, l'évolution du seuil en fonction des délais d'activation de l'implant n'était pas significative, ce qui signifie que, pour les patients qui arrivaient à faire le test, le milieu perceptif du continuum correspond bien à ce que l'on retrouve chez les NHS. Les patients n'avaient donc pas tendance à entendre les voix du continuum comme plus masculines ou plus féminines.

La pente

Dans ce groupe, comme cela à été dit, sur les 38 patients, 6 avaient des valeurs de pente au dessus du critère fixé de 0,805, et ne seront pas inclus dans les données qui suivent. Les résultats porteront donc sur les 32 patients restant.

Pour le groupe transversal, on peut observer sur la Figure 10.26 que les patients étaient très déficitaires lors de l'activation de l'implant, avec une pente de catégorisation très inférieure à celle des NHS en condition originale (Mann-Whitney ; $p = 0,0065$). Les performances des patients avaient

une nette tendance à s'améliorer pour les patients dont l'implant était activé depuis au moins 6 mois, avec une pente de catégorisation qui augmentait, mais cette évolution n'était pas significative (Kruskal-Wallis ; $p > 0,05$). On peut également voir que pour les patients dont l'implant était activé depuis plus de 18 mois, les performances stagnaient et restaient inférieures à celles des NHS en condition originale (Mann-Whitney ; $p = 0,0019$).

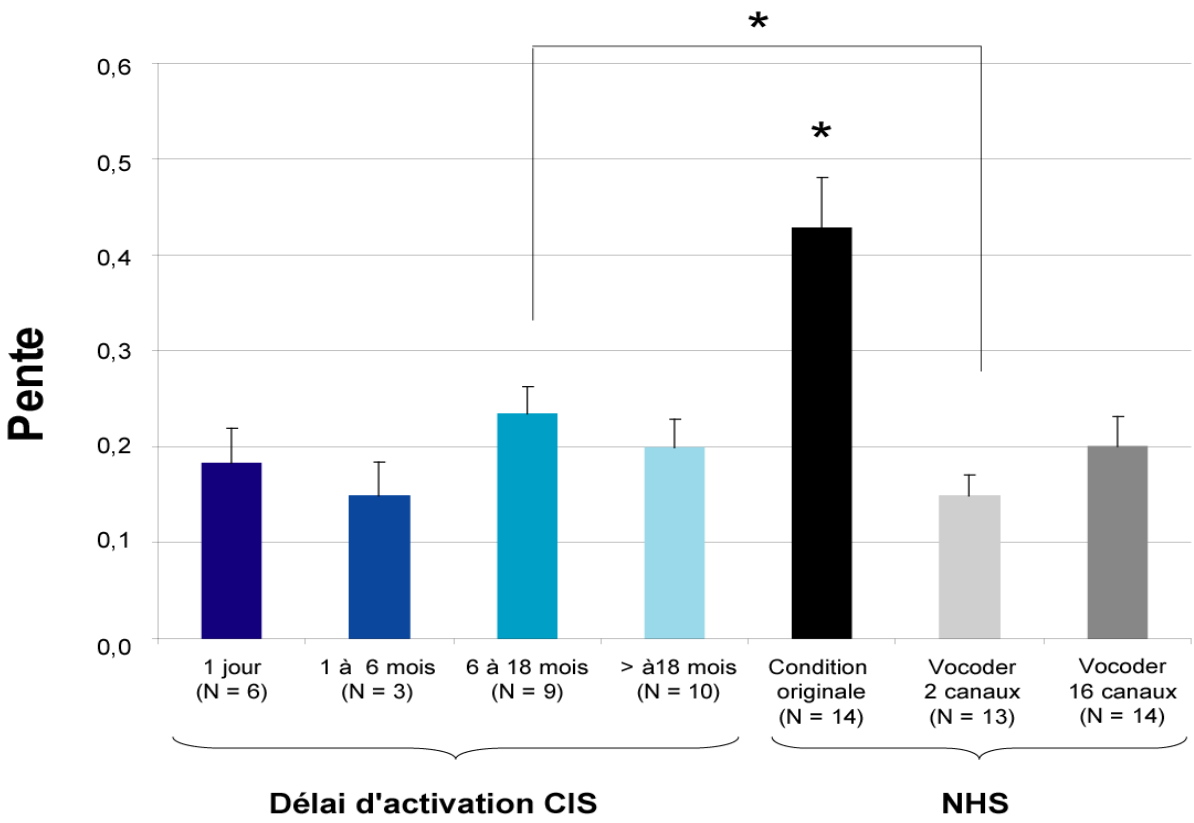


Figure 10.26 : Valeurs moyennes (± SE) de la pente de la courbe psychophysique calaculées à partir des réponses des patients du groupe transversal. Les valeurs sont présentées pour les sous-groupes de délai d'activation et comparés aux NHS. L'évolution de la pente avec le délai d'activation d'implant n'était pas significative. Les patients sont toujours significativement inférieur aux NHS en condition originale. Pour le groupe 6 à 18 mois, on retrouvait une pente significativement plus élevée que pour les NHS en condition de vocoder à 2 canaux. A plus de 18 mois, les patients ont des valeurs de pentes similaires à celles des NHS en conditions de vododer à 16 et 32 canaux.

Les analyses ont montré que tous les patients implantés avaient des valeurs de pentes significativement inférieures à celles des NHS en condition originale (Mann-Whitney ; $p < 0,02$). Pour la condition de vocoder à 64 canaux, la situation était mitigée ; le groupe des patients dont l'implant était activé depuis 6 à 18 mois avait des pentes similaires tandis que pour les 3 autres groupes la pente était inférieure (Mann-Whitney ; $p < 0,04$). Pour les autres conditions de vocoder, 32, 16, 8, 4 et 2 canaux, on ne retrouvait aucune différence significative au niveau de la pente, excepté à 2 canaux, où les valeurs de pentes étaient significativement inférieures à celles des patients dont l'implant était activé depuis 6 à 18 mois. Rappelons que pour les NHS, aucune différence n'avait été retrouvée entre les conditions de vocoder à 16, 8, 4 et 2 canaux. Ceci explique donc que les valeurs de la pente des patients soient similaires à ces 4 conditions. L'absence de différence entre

les 4 groupes de patients et les NHS en condition 32 canaux de vocoder s'explique très probablement par une variabilité importante des valeurs pour les patients. Nous pouvons donc dire que les performances de catégorisation des patients étaient similaires à celles NHS en condition de vocoder à 4, 8, 16 ou 32 canaux, et ce quelque soit le délai d'activation de l'implant. Toutefois on peut aussi noter qu'à 1 jour ainsi qu'à 1 à 6 mois d'activation de l'implant, les valeurs moyennes étaient plus proches de la condition de vocoder à 2 canaux que les 2 autres groupes de patients.

Il semble donc que le processeur de l'implant cochléaire, de même que le vocoder, dégrade moins la capacité perceptive du genre que la capacité de détection de la voix, puisque les patients sont similaires aux NHS en condition de simulation d'implant à 16 canaux, ce qui correspond davantage à leur nombre moyen d'électrodes (18 électrodes en moyenne). On peut voir sur la Figure 10.27 pour le groupe dont l'implant était activé depuis plus de 6 mois que, même si la pente était significativement différente de celle des NHS en condition originale, la courbe de catégorisation était très bonne et semblable à celles des NHS.

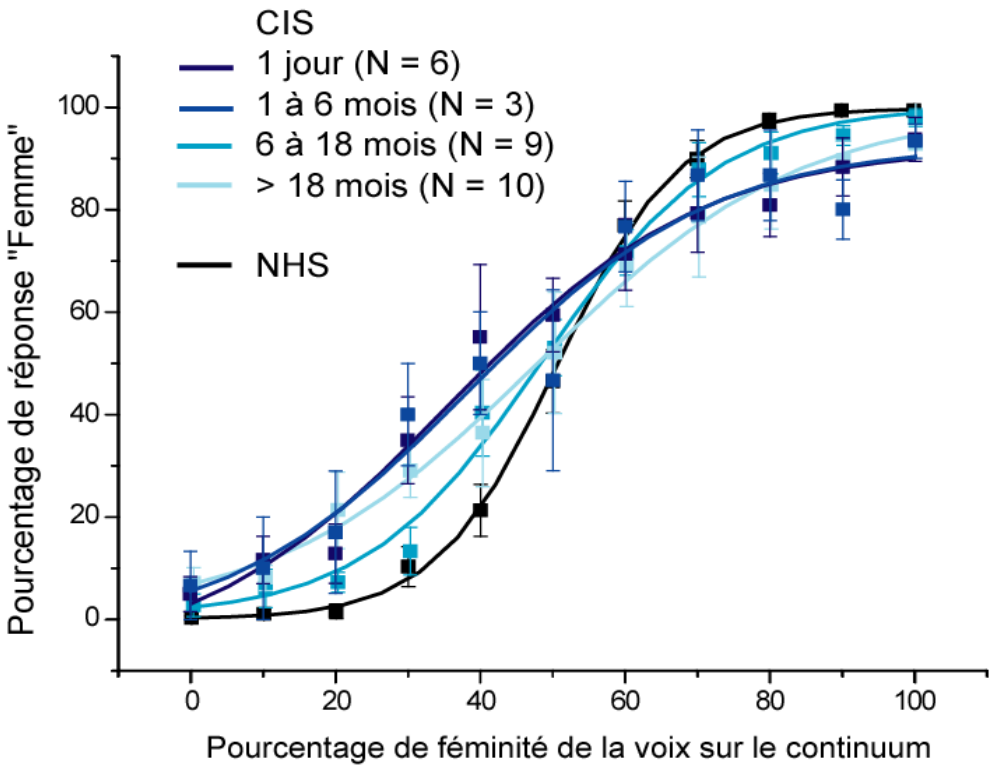


Figure 10.27 : Les courbes représentent l'estimation de la sigmoïde et les carrés représentent les valeurs moyennes (+/- SE) des sujets. Les données sont présentées pour les NHS (en noir) et les patients en fonction des différents délais d'activation (couleurs bleues). Avant 6 mois d'activation de l'implant, il n'y a pas vraiment de catégorisation, la courbe ressemble davantage à une droite. En revanche après 6 mois d'activation, il existe une bonne catégorisation du genre de la voix chez les patients.

Lorsqu'on observe les données individuelles on peut s'apercevoir que certains patients avaient des performances au moins aussi bonnes que les normo-entendant, tandis que d'autres étaient assez mauvais pour cette tache, même à des délais d'activation d'implant comparables (voir Figure 10.28).

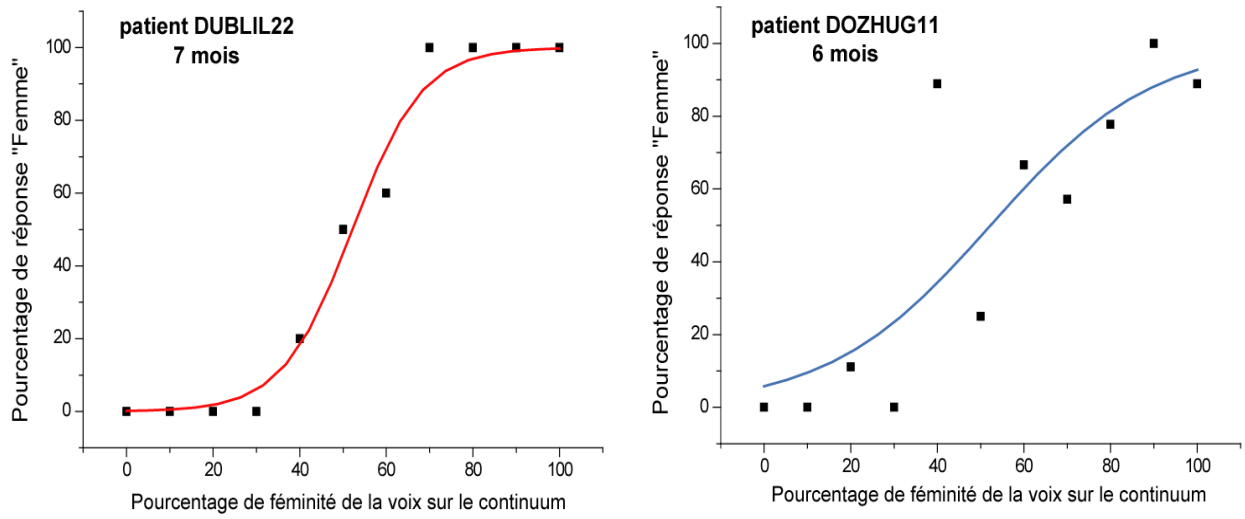


Figure 10.28 : Exemple de deux patients du groupe d'activation d'implant « 6 à 18 mois ». Le premier patient DUBLIL22 (rouge) a un implant activé depuis 7 mois et arrive très bien à catégoriser avec une courbe très semblable à celle des NHS. Le deuxième patient DOZHUG11 (bleu) a un implant activé depuis 6 mois n'arrive pas bien à catégoriser, avec une courbe assez plate.

On peut donc dire que, au moins pour certains patients, la catégorisation du genre était bonne jusqu'à être équivalente à celle des normo-entendants. De plus, rappelons que certains patients, les supranormaux, catégorisaient de façon beaucoup plus franche, avec une pente beaucoup plus importante que celle des NHS.

L'erreur

L'analyse de l'erreur nous montre que l'erreur de la courbe, c'est-à-dire l'adéquation du modèle aux données comportementales, au premier jour de l'activation de l'implant était significativement plus importante que celle des NHS en condition originale (voir Figure 10.29). Par la suite la valeur moyenne avait tendance à diminuer avec l'expérience de l'implant, mais l'effet n'était pas significatif. L'erreur restait significativement plus élevée que celle des NHS en condition originale pour tous les groupes de patients. La comparaison des patients avec les NHS en condition de vocoder n'a révélé aucune différence significative avec les conditions 64 à 4 canaux. L'adéquation du modèle pour les patients est donc similaire à celle des NHS pour toutes ces conditions. En revanche à 2 canaux de vocoder chez les NHS, on observe une différence significative avec les groupes patients implantés depuis plus de 6 mois (Mann-Whitney : $p < 0{,}04$).

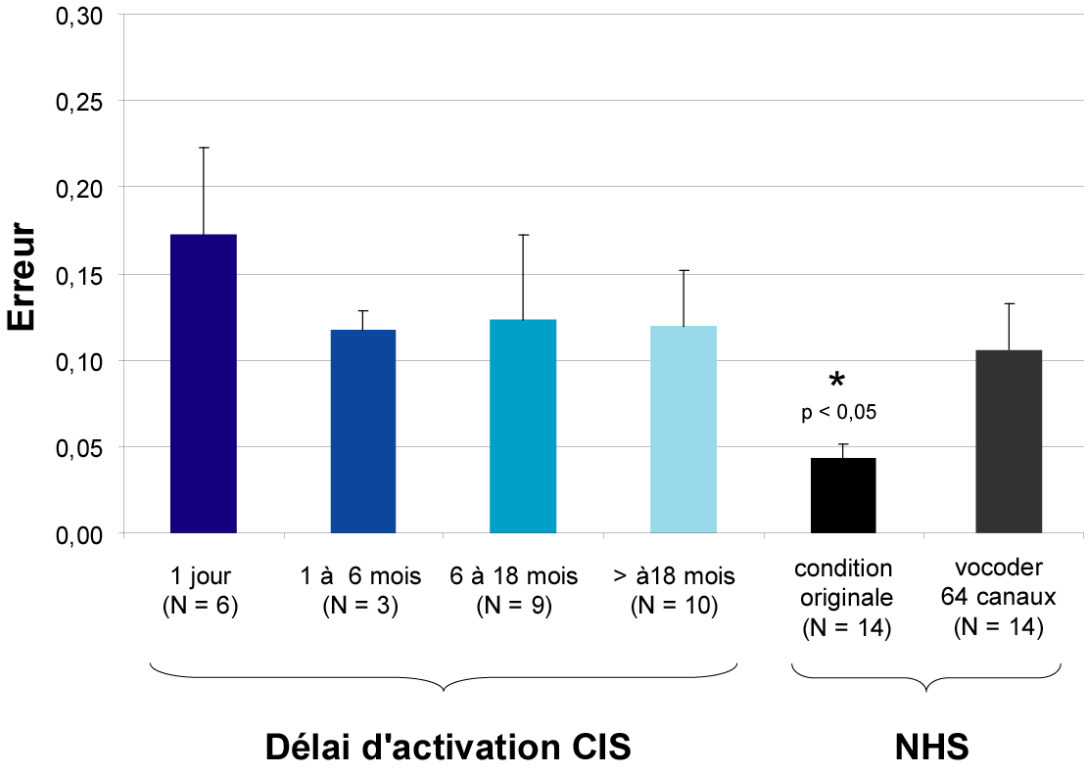


Figure 10.29 : Valeur moyenne (±SE) de la l'erreur de la courbe sigmoïde estimée à partir des réponses brutes des patients du groupe transversal. Aucune différence significative n'a été retrouvée entre les différents groupes de patients. Les 4 groupes de patients avaient une erreur moyenne de la courbe significativement plus grande que les NHS en condition originale.

En résumé, les valeurs de l'erreur de la courbe variaient donc en sens inverse de la pente, mais cette variation n'était pas significative.

L'analyse des réponses aux extrêmes

Les résultats de l'analyse de la reconnaissance correcte de la voix d'homme et de la voix de femme sont présentés sur la Figure 10.30. Au premier jour de l'activation d'implant, les performances, autant pour la voix d'homme que pour la voix de femme, étaient significativement plus faibles que celles des NHS en condition originale (Mann-Whitney ; $p = 0{,}0071$). Avec l'expérience de l'implant les performances augmentaient mais l'amélioration n'était pas significative. Toutefois, les valeurs de reconnaissance de la voix d'homme et de la voix de femme des patients « 1 à 6 mois » restaient significativement inférieures à celles des NHS en condition originale (Mann-Whitney ; $p = 0{,}0043$). Pour la voix de femme, les comparaisons avec les NHS en condition de vocoder montraient que les performances pour les groupes de patient à 6 mois d'activation ou moins étaient significativement plus faibles que les conditions de vocoder 64 et 16 canaux chez les patients. Après 6 mois d'activation d'implant, les performances étaient significativement plus importantes qu'à 4 canaux de vocoder chez les NHS (Mann-Whitney ; $p < 0{,}02$). Ceci va dans le sens d'une amélioration des performances, pour la voix de femme, chez les patients au cours du temps. Pour la voix d'homme, les comparaisons avec les NHS en condition

de vocoder montraient que les performances des patients étaient significativement plus faibles que celles des NHS en conditions de vocoder à 64 et 32 canaux (Mann-Whitney ; $p < 0,02$). Ceci était vrai pour tous les groupes de patients excepté pour le groupe « 6 à 18 mois ». Par la suite les autres comparaisons entre les patients et les autres conditions de vocoder chez les NHS, n'ont révélé aucune autre différence significative, pour la voix d'homme.

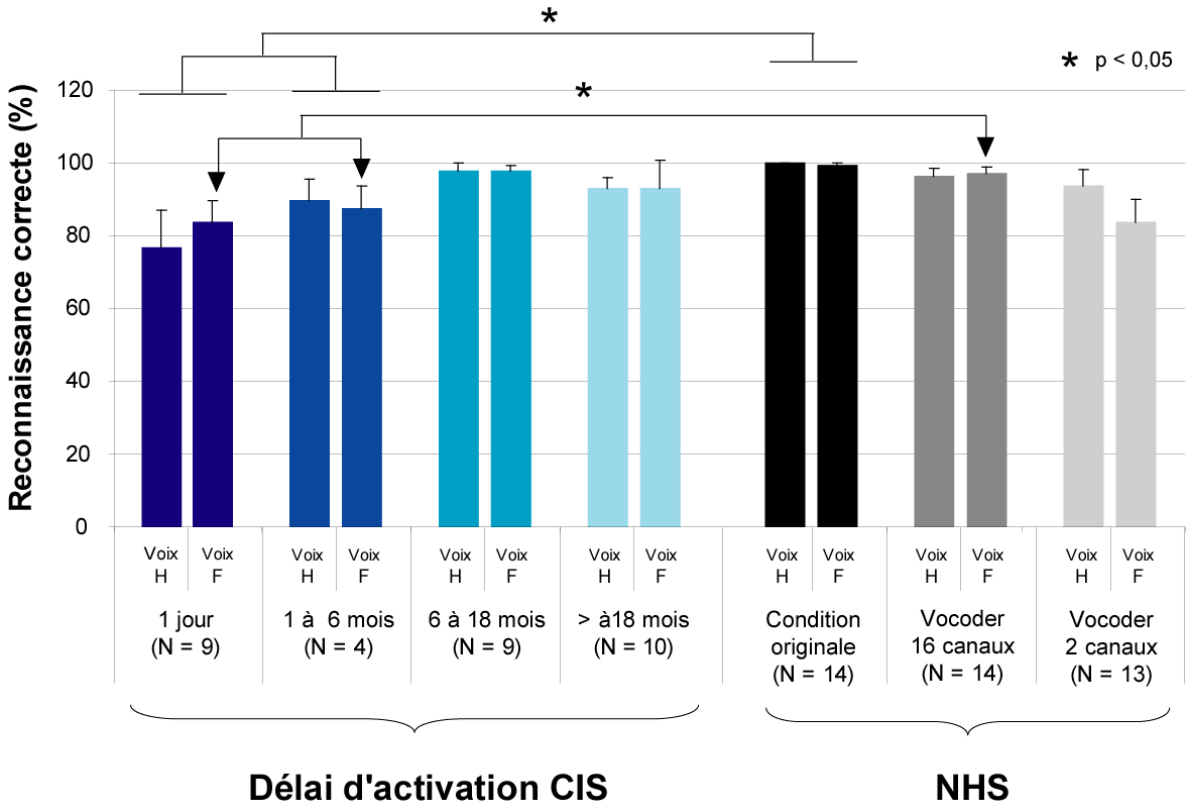


Figure 10.30 : Valeur moyenne (±SE) de la reconnaissance correcte de la voix d'homme (Voix H) et de la voix de femme (Voix F) pour les patients du groupe transversal et comparaison avec les NHS. L'évolution des performances chez les patients n'était pas significative. En comparaison des NHS en condition originale, les premiers implantés depuis 6 mois ou moins avaient des performances significativement plus faible. Les performances de reconnaissance de la voix de femme étaient également significativement plus faibles que celle des NHS en condition de vocoder à 16 canaux.

En résumé, de même que pour la pente et l'erreur, les valeurs allaient dans le sens d'une amélioration, mais celle-ci n'était pas significative. La comparaison avec les NHS en condition de vocoder montrait aussi une amélioration ; les patients à moins de 6 mois d'activation d'implant avaient des performances similaires à celles des NHS en condition de vocoder les plus dégradées, tandis que les patients à plus de 6 mois d'activation d'implant avaient des performances similaires à celle des NHS en condition de vocoder les moins dégradées.

Les performances de compréhension du langage

Tout comme pour le test de détection de la voix, nous nous sommes intéressés aux performances de reconnaissance de mots dissyllabiques des patients. Ces mesures étaient faites en condition auditive, avec l'implant seul. Les valeurs présentées ici, sont celles mesurées par l'orthophoniste (Marie-Laurence Laborde) de l'UIA de Purpan, lors de la prise en charge de suivi des patients (voir Figure 10.31). Les valeurs indiquées ici ont été mesurées le même jour que le test de détection de la voix. En analysant les données, on peut remarquer que, les performances de reconnaissance de mots progressaient de manière significative, jusqu'à atteindre une moyenne de 74,8 % de mots correctement reconnus pour le groupe de patients dont l'implant était activé depuis plus de 18 mois. Toutefois le niveau de reconnaissance des patients restait significativement inférieur à celui des normo-entendants en condition originale, qui eux reconnaissaient tous les mots sans erreurs. Nous présentons également sur la Figure 10.31 des données issus de travaux antérieurs au sein du laboratoire, réalisés et publiés par Julien Rouger (Rouger et al., 2007). Ces données correspondent aux performances de reconnaissance correcte de mots dissyllabiques par des normo-entendants en condition de vocoder à 4 canaux (voir « Les performances de compréhension du langage » pour le test « Voix-Non voix », p213). Ainsi, on peut remarquer que les patients, au premier jour d'activation de l'implant avaient des performances très faibles. Mais les performances augmentaient rapidement et significativement (Kruskal-Wallis ; $p = 0,002$). Ainsi, dès 1 à 6 mois d'activation ils étaient nettement meilleurs que des normo-entendants en condition de simulation d'implant à 4 canaux. Ces résultats vont dans le sens de ceux obtenus dans le test de détection de la voix : l'évolution des performances de compréhension du langage sont bien meilleures que celle de la perception de la voix. Toutefois, la dissociation que l'on observait pour le test de détection de la voix n'est pas aussi nette pour le test de catégorisation du genre. En effet nous avons vu que les patients étaient capables de s'améliorer pour la catégorisation du genre jusqu'au niveau des normo-entendants à 16 canaux de vocoder, voire 32, ce qui correspondait à leur nombre moyen d'électrode.

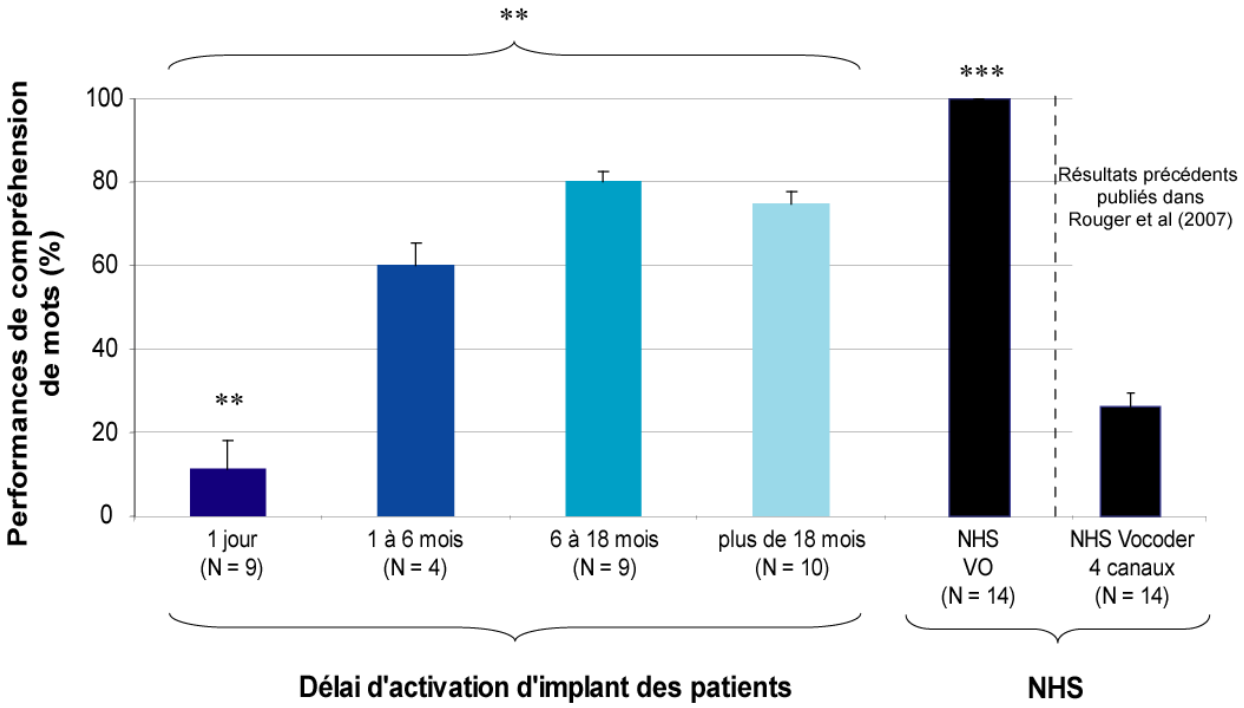


Figure 10.31 : Performances (±SE) de compréhension de mots pour les patients du groupe transversal et comparaison aux NHS. L'amélioration des performances avec l'expérience de l'implant étaient significative. On n'observait pas de différence significative entre les groupes de patients dont l'implant était activé depuis plus d'un mois. Les patients restaient toujours inférieurs au NHS en condition originale.

Afin d'affiner la mesure de la relation entre performance de compréhension du langage et catégorisation du genre, nous avons corrélé les performances des patients pour ces deux tâches (voir Figure 10.32). Les corrélations ont donc été réalisées entre les performances de compréhension de mots et les différents critères d'analyse de la catégorisation du genre, à savoir, l'erreur de la courbe par rapport aux données, la pente et le seuil au point d'inflexion et la reconnaissance correcte de la voix d'homme et de la voix de femme.

Sur l'ensemble du groupe (voir Figure 10.32, données en rouge), les différentes corrélations étaient toutes nulles avec un coefficient de corrélation R^2 maximum de 0,07 pour la pente. Aucune corrélation ne s'est avérée significative. Lorsqu'on distinguait les différents groupes de patients, on pouvait remarquer que pour le groupe de patients au premier jour d'activation, les corrélations entre la compréhension de mots et la catégorisation du genre était toujours nulle et non significative et ressemblait à la corrélation globale (voir Figure 10.32, données en bleu). Ceci était également vrai pour les groupes de patients implantés depuis moins de 18 mois. En revanche au-delà de 18 mois, on voyait apparaitre des corrélations beaucoup plus fortes pour tous les critères d'analyse de la catégorisation (voir Figure 10.32, données en vert). Une corrélation négative significative à même été retrouvée pour l'erreur (Fisher ; p = 04).

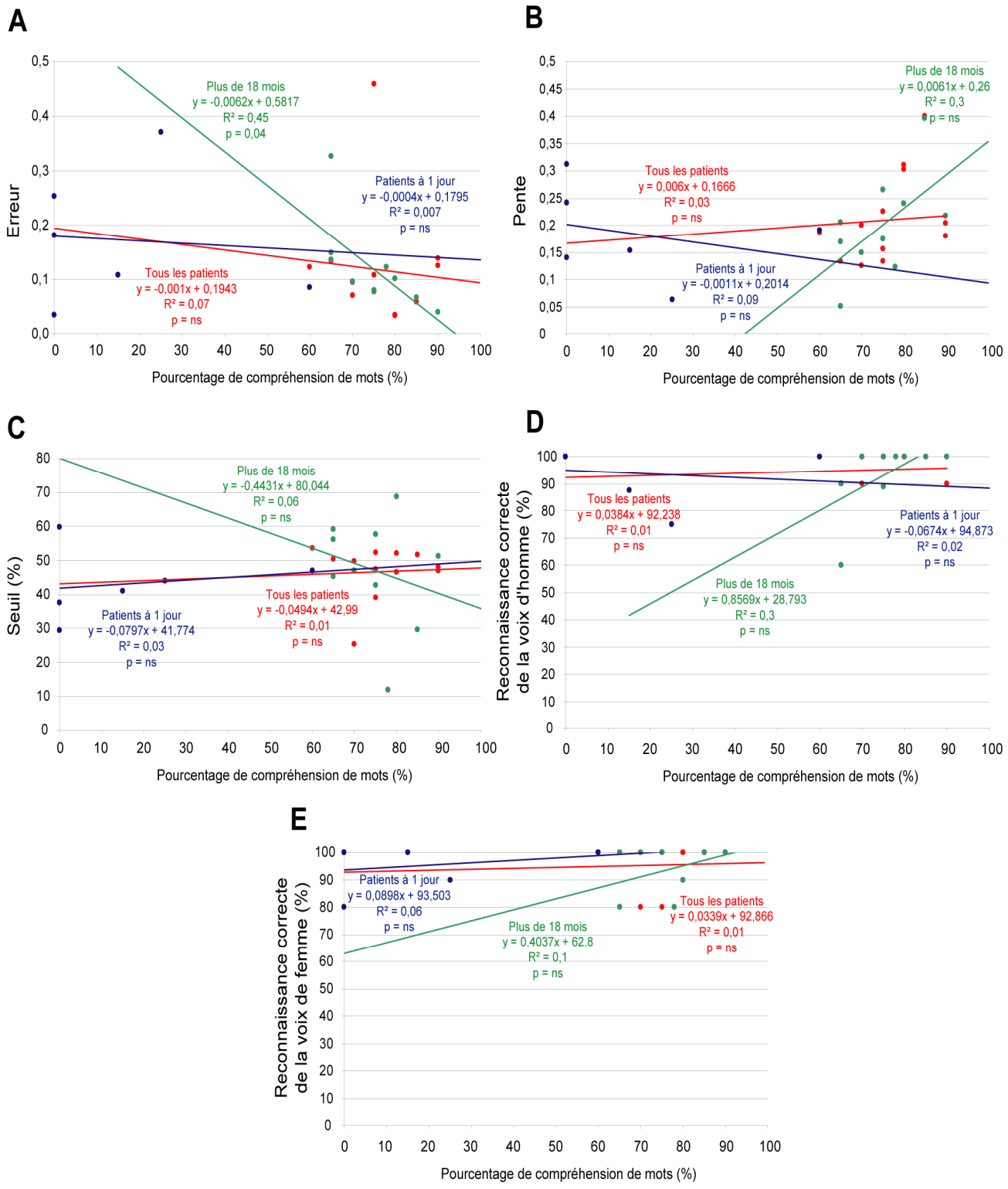


Figure 10.32 : Corrélations entre les performances de compréhension de mots et l'erreur (A), la pente (B), le seuil (C) et la reconnaissance correcte de la voix d'homme (D) et de la voix de femme (E), pour les patients du groupe transversal. La corrélation était réalisée pour l'ensemble du groupe (rouge), et pour chacun des groupes de patients. Sur la figure était représentée, la corrélation pour le groupe à 1 jour (bleu) et pour le groupe le plus expérimenté, au-delà de 18 mois (vert). Pour l'ensemble du groupe, aucune corrélation significative n'était retrouvée.

L'absence de corrélation sur l'ensemble du groupe est surprenante étant donné qu'une amélioration semblait avoir lieu également pour la catégorisation du genre. Ceci indique donc que ces deux tâches ne sont pas corrélées et que l'aptitude des patients pour chacune d'elle évoluait différemment. Ceci va donc dans le sens de la dissociation de récupération entre les performances de compréhension du langage et celles de la perception de la voix déjà évoquée précédemment. Toutefois, les résultats pour le groupe à plus de 18 mois d'activation d'implant venaient nuancer cette idée en montrant une corrélation beaucoup plus forte pour ce groupe, indiquant pour l'ensemble des critères, que plus les patients étaient bons en compréhension du langage, meilleures étaient leurs performances en catégorisation du genre. Il semble donc que chez les patients les plus expérimentés avec l'implant, la récupération de perception de la voix et celle de la compréhension du langage aille de pair. Toutefois, rappelons que pour ces patients, une seule corrélation significative a été retrouvée. Ce résultat reflète probablement l'effet de l'expérience de l'implant.

Les performances de lecture labiale

Toujours dans le but de voir si de bonnes performances en lecture labiale pouvaient être corrélées à de moins bonnes performances de perception de la voix, nous avons comparé les performances de catégorisation du genre avec celles de lecture labiale. Rappelons que le test de lecture était réalisé le même jour que le test de détection de la voix par l'orthophoniste (Marie-Laurence Laborde) de l'UIA de Purpan, lors de la prise en charge de suivi des patients (voir Figure 10.33). Le test n'étant pas réalisé systématiquement, cela explique le plus grand nombre de données manquantes.

Tout d'abord, on peut voir que les performances des patients étaient relativement stables au cours du temps (entre 22,5 et 34 %), avec une évolution non significative, et toujours meilleures que celles des normo-entendants (9,4%). Ceci est cohérent avec les résultats de lecture labiale mesurés dans le test de détection de la voix et correspond donc à ce qui a déjà été publié par Rouger *et al.* (Rouger et al., 2007) ; les patients, bien qu'ils aient retrouvé un très bon niveau de compréhension de la parole, conservent des performances de lecture labiale très supérieures à celles des sujets normo-entendants.

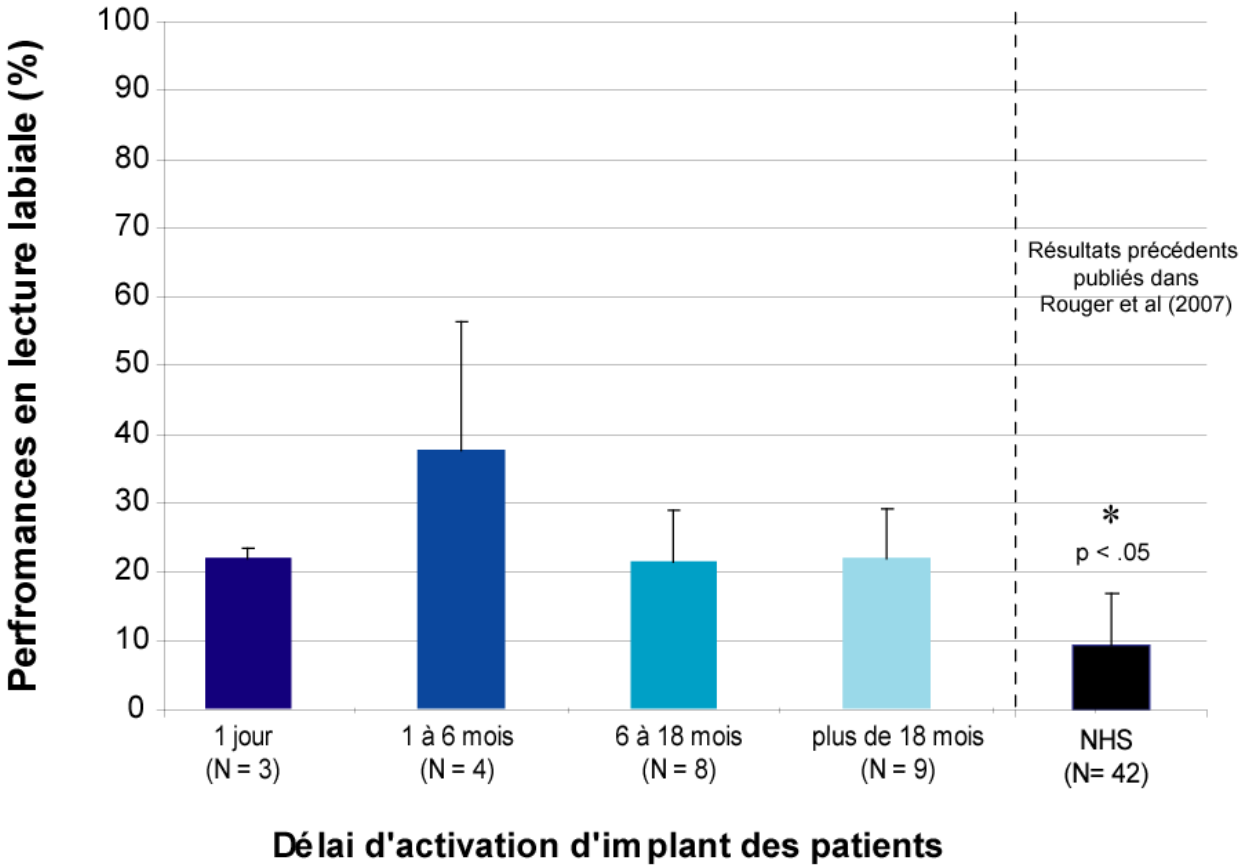


Figure 10.33 : Performances (±SE) de compréhension de mots pour les patients du groupe transversal et comparaison aux NHS. L'amélioration des performances avec l'expérience de l'implant n'étaient pas significative. Les patients restaient toujours meilleurs que les NHS.

La Figure 10.34 montre les corrélations de ces performances de lecture labiale avec les différents critères d'analyse que nous avons étudiés pour la catégorisation du genre ; l'erreur (A), la pente (B), le seuil, (C) et la reconnaissance correcte de la voix d'homme (D) et de la voix de femme (E). La corrélation a été réalisée pour l'ensemble des patients (données en rouge) et en fonction des groupes de délai d'activation de l'implant (aucune corrélation n'a pu être réalisée avant 6 mois d'activation en raison du petit nombre de données). Les résultats montrent que, pour l'ensemble du groupe, aucune corrélation n'a été retrouvée, avec un coefficient de corrélation R^2 compris entre 0,001 et 0,1. Ces résultats étaient valables pour les groupes de patients à plus de 6 mois d'activation d'implant (données du groupe à plus de 18 mois en vert sur la Figure 10.34).

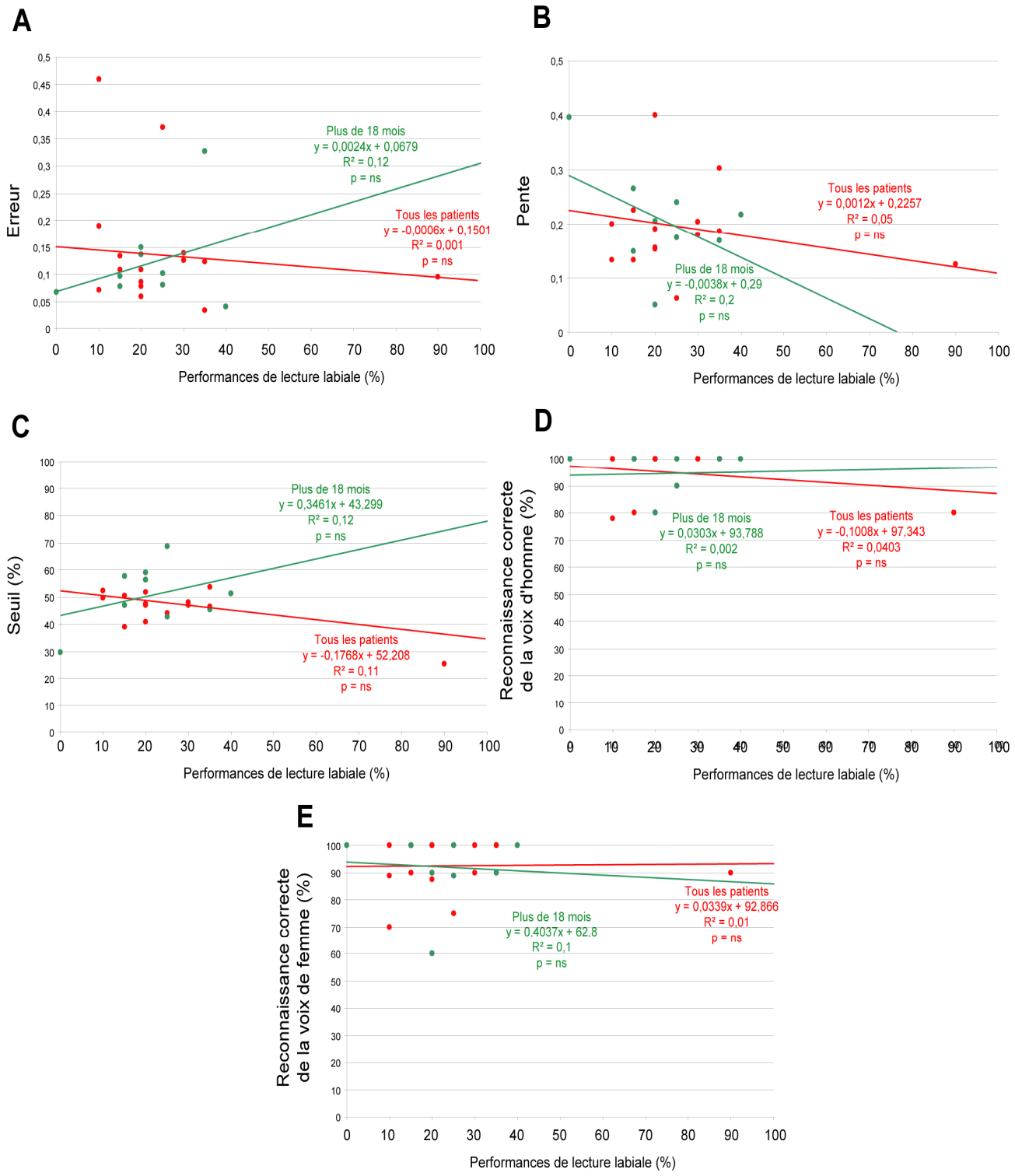


Figure 10.34 : Corrélations entre les performances de lecture labiale et l'erreur (A), la pente (B), le seuil (C) et la reconnaissance correcte de la voix d'homme (D) et de la voix de femme (E), pour les patients du groupe transversal. La corrélation était réalisée pour l'ensemble du groupe (rouge), et pour chacun des groupes de patients. Sur la figure était représentée, la corrélation pour le groupe à 1 jour (bleu) et pour le groupe le plus expérimenté, au-delà de 18 mois (vert). Pour l'ensemble du groupe, aucune corrélation significative n'était retrouvée.

Une fois de plus, nous ne pouvons pas conclure que de meilleures performances en lecture labiale soient associées à de moins bonnes performances en détection de voix. Bien sûr ce résultat doit être confirmé en raison du nombre de données manquantes importantes au premier jour de l'activation d'implant.

RESULTATS DU GROUPE EN SUIVI

Pour l'analyse de la pente, nous avons appliqué le même critère que pour le groupe transversal, à savoir que nous n'avons pas inclus dans les analyses les patients dont la pente, à au moins une mesure du suivi, dépassait 2 écart-types de la moyenne de NHS, c'est-à-dire avait une pente supérieure à 0,805. Sur les 14 patients inclus dans le groupe de suivi, 3 dépassaient le critère fixé. Les données pour ces patients ont déjà été présentées et discutées dans le paragraphe « Des patients supranormaux ou une limite méthodologique ? », p236. Les résultats qui suivent porteront donc uniquement sur les 11 patients dont les valeurs de pente ne dépassaient pas 0,805.

Pour le groupe de sujets qui a été suivi du premier jour de l'activation de l'implant à 18 mois post-activation, tous les sujets n'ont pas pu être vus à tous les temps. De plus, il est arrivé que certains patients n'arrivent pas à faire le test du tout. Le Tableau 10.3 résume le nombre de patients qui ont été vus, le nombre de patients qui n'ont pas été vus et le nombre de ceux qui n'ont pas réussi à faire le test. Certains patients (3 au premier jour d'activation et 1 à 1 mois post-activation) faisaient le test mais avaient des réponses complètement hasardeuses. L'estimation d'une courbe sigmoïde n'a donc pas pu être réalisée pour ces patients, car les valeurs de seuil et de pente étaient complètement aberrantes. Ces 4 patients n'ont donc pas été pris en compte pour les analyses de l'erreur de la courbe sigmoïde, de la pente et du seuil. En revanche, leurs réponses ont été prises en compte pour l'analyse de la reconnaissance correcte de la voix de l'homme et de la voix de femme. Ceci explique que pour toutes ces analyses, le nombre de patients pour les deux premières mesures du suivi ne soit pas toujours le même.

Délai d'activation	1 jour	1 mois	3 mois	6 mois	9 mois	12 mois	18 mois
Nombre de patients manquants	2	0	2	3	1	2	2
Nombre de patients qui échouent	2	3	2	1	1	0	0
Nombre de patients restant	7	8	7	7	9	9	9
Nombre de patients répondant au hasard	3	1	0	0	0	0	0

Tableau 10.3 : Nombre de patients qui n'arrivent pas à faire le test, qui ont manqué la date de suivi et nombre patients qui ont pu être vu et ont réussi à faire le test.

Sur ce tableau ont peut constater que le nombre de patients qui n'arrivaient pas à faire la tâche ou la faisait en répondant au hasard diminue au cours du suivi, ce qui va dans le sens d'une familiarisation avec la tâche, voire d'une amélioration des performances au cours du temps.

Le seuil

En raison d'un nombre insuffisant de valeurs les statistiques ont dues être réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois. L'analyse du seuil de la courbe au point d'inflexion (C50) n'a révélé aucune différence significative au cours du suivi. Les valeurs moyennes du seuil restaient concentrées entre 38 et 56 % sur le continuum, et la variabilité des valeurs était assez importante.

Lorsqu'on a comparé les valeurs du seuil de catégorisation des patients avec celles des normo-entendants en condition originale, on a pu constater qu'aucune des mesures au cours du suivi n'était significativement différente de celle des NHS en condition originale. En comparaison avec les conditions de vocoder, la seule différence qui a été retrouvée était entre avec la condition à 4 canaux et toutes les mesures du suivi (Mann-Whitney ; $p < 0,05$). Cette différence paraît compréhensible étant donné que la condition de vocoder à 4 canaux était la seule qui différait des autres conditions passées par les NHS. L'écart de cette condition paraît mineur et a déjà été discuté dans le paragraphe « L'analyse du seuil », p180.

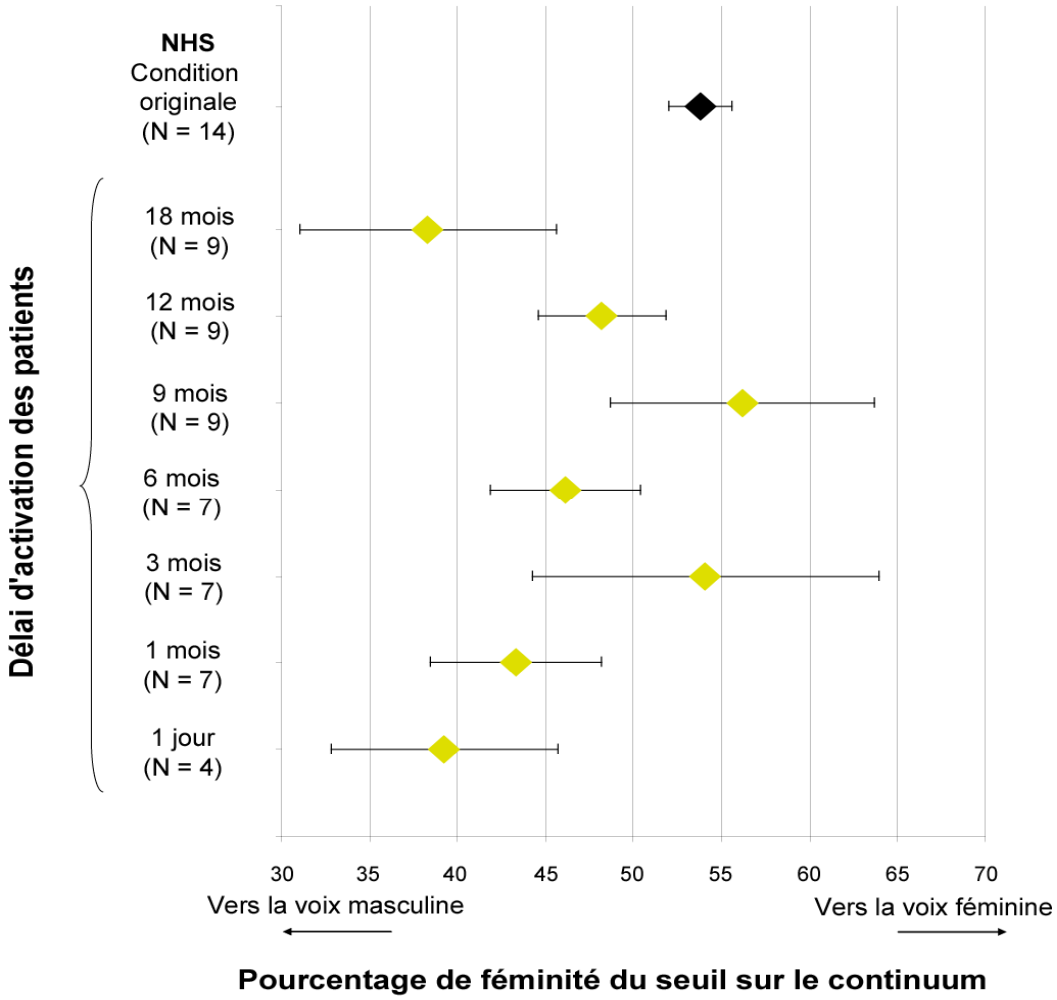


Figure 10.35 : Valeur moyenne (±SE) du seuil pour les patients implantés du groupe de suivi (en jaune) et comparaison avec les NHS en condition originale (en noir). Aucune différence n'a été retrouvée au cours du suivi, probablement en raison de la grande variabilité des valeurs des patients, ni en comparaison avec les NHS en condition originale. En raison d'un nombre de valeur insuffisant de valeur lors des mesures du suivi, les analyses statistiques ont été réalisées sans les mesures à 1 jour, 3 et 6 mois.

Nous conclurons en disant qu'il n'y a pas d'effet de l'expérience du port de l'implant sur le seuil des patients sourds implantés cochléaires. De plus, le seuil des patients était similaire à celui des NHS en condition originale. L'implant, n'introduit donc pas de biais au niveau du seuil dans le test de catégorisation du genre.

La pente

Pour le groupe du suivi, en raison d'un nombre insuffisant de valeurs les analyses statistiques ont dues être réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois.

On peut voir sur Figure 10.36 que, un jour après l'activation de l'implant, les performances étaient déficitaires en comparaison des normo-entendants en condition originale (Mann-Whitney ; $p < 0{,}034$). Puis les valeurs de la pente, qui reflète la capacité à catégoriser, augmentaient. A 9 et 12 mois les valeurs avaient presque doublé. Les analyses ont révélé que, chez les patients en suivis, l'augmentation des performances au cours du temps n'était pas significative. On peut penser que cette absence de significativité est due à une grande variabilité des valeurs dans le groupe de patients. Cependant cette tendance à l'amélioration confirme celle que nous avions trouvée dans le groupe transversal. Il semble donc probable qu'avec un plus grand nombre de patients cette augmentation devienne significative. De plus lorsqu'on compare les valeurs de la pente chez les patients à celles des NHS en condition de vocoder, cela confirme la tendance à l'amélioration.

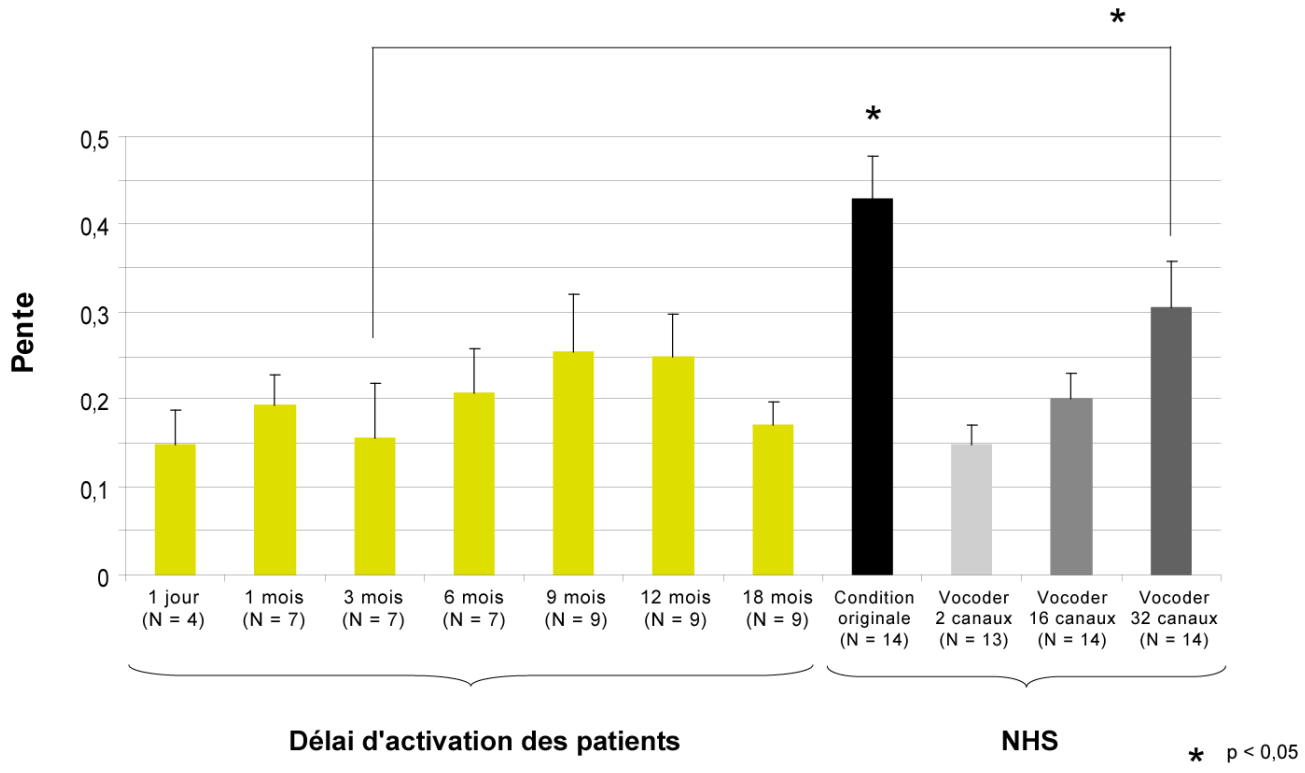


Figure 10.36 : Valeurs moyennes (± SE) de la pente de la courbe psychophysique pour le groupe de suivi. On peut voir que les patients, qui étaient significativement déficitaires par rapport aux NHS en condition originale, semblaient s'améliorer au cours du temps. Cette augmentation des valeurs de la pente n'était pas significative (en raison d'un nombre de valeurs insuffisant, les analyses statistiques ont été réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois).

En effet, toutes les mesures de la pente étaient significativement similaires aux valeurs des NHS en condition de vocoder à 2, 4, 8 et 16 canaux. Cela signifie qu'à tous les temps du suivi, les patients avaient des performances de catégorisation au moins égales aux NHS en condition de vocoder à 16 canaux. Rappelons tout de même que chez les NHS, la diminution du nombre de canaux de 16 à 2, n'avait pas d'effet sur les performances. Il semble donc logique que si les patients avaient des performances similaires à la condition de vocoder 16 canaux, elles soient aussi similaires à 8, 4 et 2 canaux de vocoder. Les analyses ont aussi montré qu'à 3 mois, les patients avaient une pente moyenne significativement inférieure à celles des NHS en condition de vocoder à 32 et 64 canaux (Mann-Whitney : $p = 0{,}0072$). La différence n'était pas retrouvée pour les deux premières mesures (1 jour et 1 mois), alors que les valeurs de pente étaient assez proches. On peut raisonnablement penser que, au moins pour la première mesure, cette absence de différence est liée au petit nombre de valeurs. De plus, rappelons que les patients qui étaient absents de ces deux premières mesures étaient en majorité des patients qui n'arrivaient pas à faire la tâche ou répondaient au hasard. Or, si on pouvait estimer les valeurs de pente de ces patients, elles tireraient probablement la moyenne vers le bas. On peut donc dire que les patients de 1 jour à 3 mois post-activation de l'implant étaient similaires à 16 canaux, ce qui correspond au nombre moyen d'électrodes de ce groupe (18 électrodes en moyenne). Puis, avec l'expérience de l'implant, les performances augmentaient pour devenir similaires à celles des normo-entendants à 32 canaux et 64 canaux de vocoder. Chez les NHS, les performances à 32 canaux étaient significativement supérieures à celles en condition 16 canaux de vocoder.

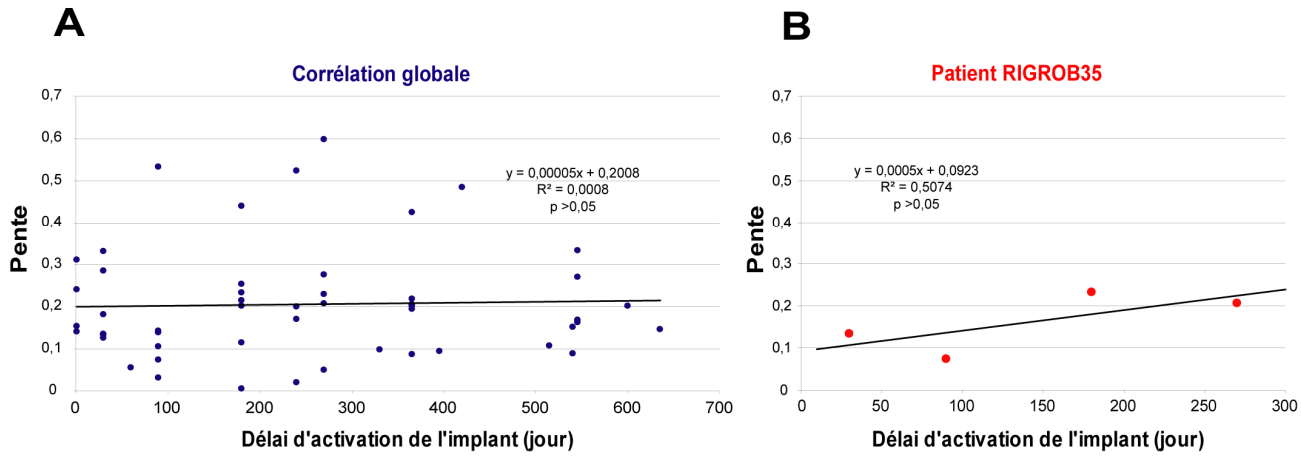


Figure 10.37 : **A.** Absence de corrélation pour les patients du groupe de suivi, entre les la pente de la courbe sigmoïde, c'est-à-dire la capacité de catégorisation, et le délai d'activation de l'implant. **B.** Exemple d'un patient pour qui la corrélation était positive mais non significative.

La Figure 10.37.A montre les données de l'ensemble du groupe suivi. Aucune corrélation n'a été retrouvée ($R^2 = 0{,}001$). D'un point de vue individuel, seul un patient semblait avoir une corrélation positive entre la pente et le délai d'activation d'implant ($R^2 = 0{,}5$), mais elle n'était pas significative et cette corrélation n'était réalisée que sur 4 points (voir Figure 10.37.B). Pour les autres patients, on ne retrouve aucune corrélation.

Nos résultats confirment donc ceux du groupe transversal, et nous amènent à conclure deux choses. Premièrement, nous pouvons affirmer qu'en moyenne, même avec l'expérience de l'implant, les patients restaient déficitaires par rapports aux NHS en condition originale. Et deuxièmement, les patients semblaient s'améliorer au cours du temps, même si la tendance n'était pas significative. On peut parler d'amélioration en raison de la comparaison avec les NHS qui confirme que les performances des patients se rapprochaient de plus en plus de celles des NHS à 64 canaux. De plus, la présence de sujets supranormaux dans ce groupe avec des capacités de catégorisation supérieures à celles des normo-entendants, montre bien que les patients sont capables d'une très bonne catégorisation.

L'erreur

L'analyse de l'erreur montrait qu'à toutes les mesures du suivi, les patients avaient une erreur de la courbe significativement plus importante que les NHS en condition originale (Mann-Whitney ; $p < 0{,}05$), excepté à 12 mois d'activation d'implant, probablement en raison d'une variation très importante (voir Figure 10.38). Au cours du suivi, on pouvait voir que par rapport aux deux premières mesures, les mesures à 3, 6 et 9 mois montraient une forte augmentation de l'erreur, alors que pour les mêmes temps on notait une augmentation de la pente. Puis à 12 et 18 mois, on pouvait voir que le modèle sigmoïde était beaucoup plus adapté aux données puisque l'erreur de l'estimation diminuait. Toutefois ces variations de l'erreur au cours du temps n'étaient pas significatives (en raison d'un nombre de valeurs insuffisant, les analyses statistiques ont été réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois).

Sur les données de l'ensemble du groupe suivi, aucune corrélation n'a été retrouvée ($R^2 = 0{,}01$), ni même au niveau individuel.

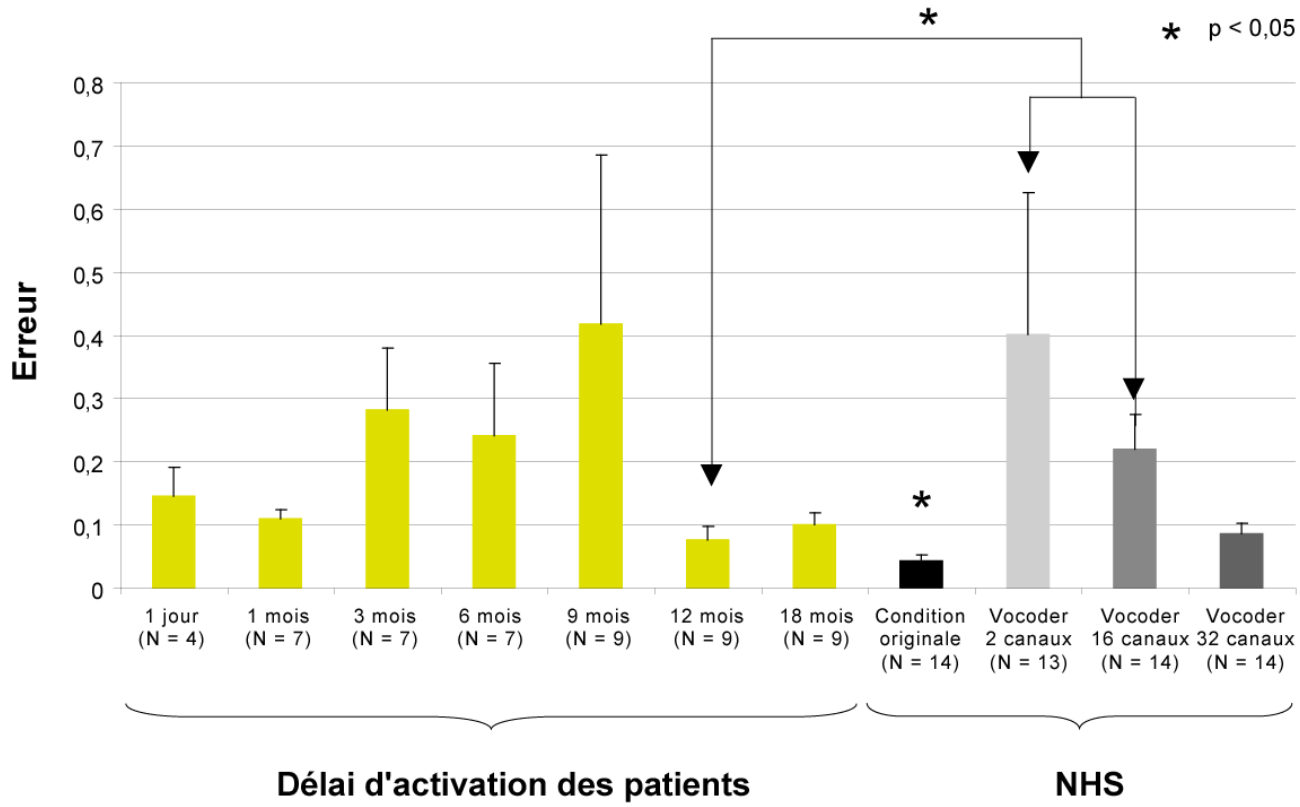


Figure 10.38 : Valeurs moyennes (± SE) de l'erreur de l'estimation de la courbe au cours du suivi et comparaison avec celle des NHS. L'erreur reflétait l'adéquation du modèle sigmoïde aux données. L'évolution de l'erreur au cours du temps n'était pas significative (en raison d'un nombre de valeurs insuffisant, les analyses statistiques ont été réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois). Les NHS en condition originale avaient une pente significativement plus faible que les patients à tous les temps du suivi sauf à 12 mois. La mesure à 12 mois était significativement plus faible que celle des NHS en condition de vocoder à 16 canaux et en-deçà.

Analyse des réponses aux extrêmes

Dans l'ensemble, l'analyse des reconnaissances de la voix d'homme et de la voix de femme allait dans le sens d'une amélioration des performances avec l'expérience de l'implant. Toutefois l'évolution moyenne des performances au cours du suivi n'a pas révélé d'amélioration significative, que ce soit pour la voix d'homme ou pour la voix de femme (voir Figure 10.39). Mais lorsqu'on comparait les données des patients à celles des normo-entendants, on pouvait voir que les performances au premier jour de l'activation d'implant, pour la voix d'homme et pour la voix de femme, étaient significativement inférieures à celles des NHS en condition de vocoder à 16 canaux et au-delà (32 et 64) ainsi qu'en condition originale (Mann-Whitney : $p < 0,05$). Pour la voix d'homme, les performances de 1 mois à 6 mois s'amélioraient légèrement et devenaient similaires à celles des NHS à 16 canaux de vocoder (mais pas à 64 et 32). Puis à 9 et 12 mois les performances étaient similaires à celle des NHS en condition originale. A 18 mois, une légère diminution des performances les rendaient de nouveau inférieurs au NHS en condition originale. En résumé, pour la voix d'homme, on pouvait voir une amélioration des performances jusqu'à devenir semblables à celles des NHS en condition originale, au moins à 9 et 12 mois. Pour la voix de femme, le profil de

progression était très semblable, à ceci près qu'il y avait une diminution inexpliquée à 3 mois post-activation d'implant. Les performances à ce délai d'activation étaient plus faibles que celles des NHS à 8 canaux (Mann-Whitney : $p < 0{,}01$). En revanche on observait aussi des performances similaires à celles des NHS en condition originale, pour les mesures à 6 et 9 mois. L'analyse des performances de reconnaissance correcte de la voix de femme montrait donc aussi une amélioration.

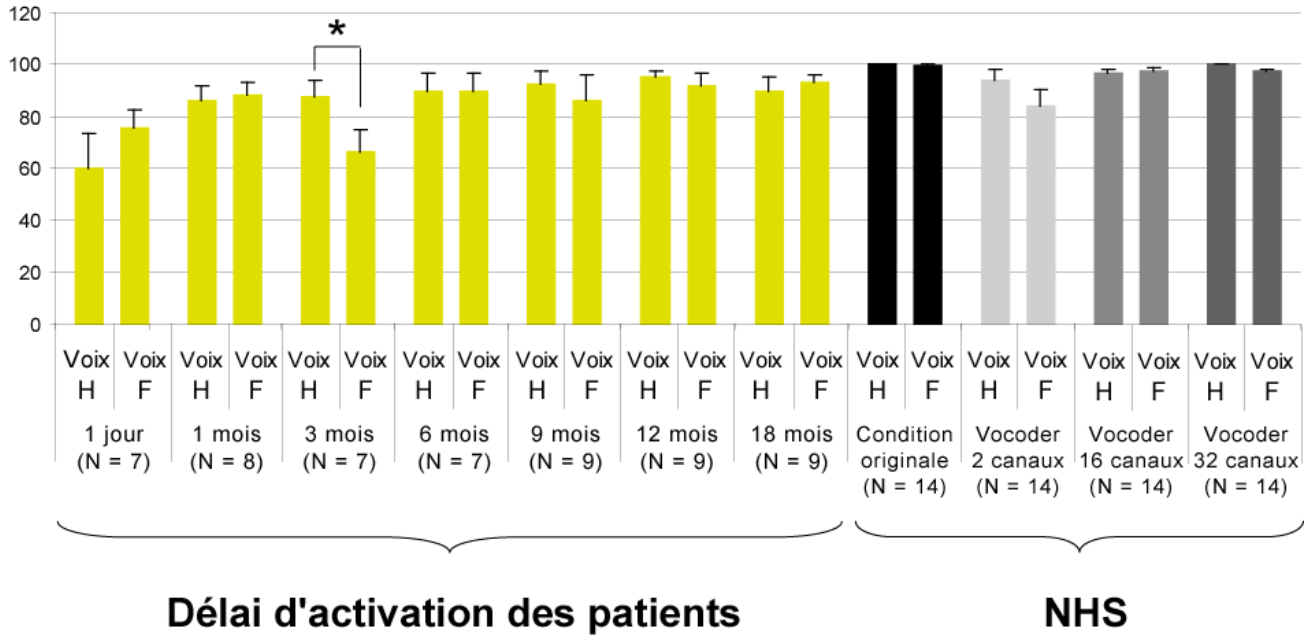


Figure 10.39 : Valeurs moyennes (± SE) de la reconnaissance correcte de la voix d'homme (voix H) et de la reconnaissance correcte de la voix de femme (Voix F) des patients du groupe suivi et comparaison avec celles des normo-entendants. L'évolution au cours du temps pour les patients n'était pas significative (en raison d'un nombre de valeurs insuffisant, les analyses statistiques ont été réalisées sans prendre en compte les mesures à 1 jour, 3 et 6 mois). Mais la comparaison avec les NHS allait dans le sens d'une amélioration au cours du suivi.

On peut noter que les mesures à 6, 9 et 12 mois, où nous observions ici des performances similaires à celles des NHS en condition originale, correspondait aussi aux mesures où les valeurs de pentes étaient les meilleures (voir paragraphe « La pente », p256).

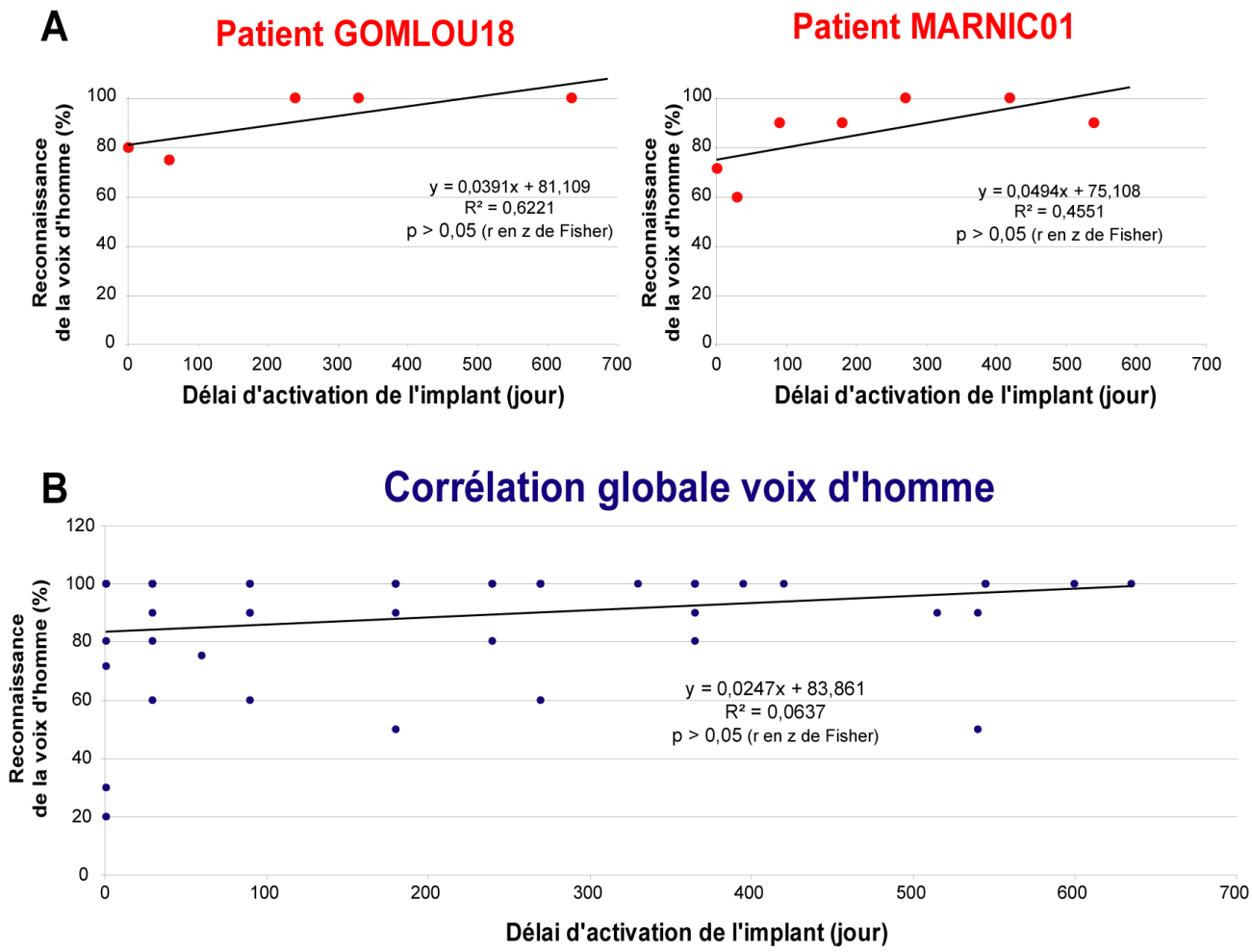


Figure 10.40 : **A.** Exemple de deux patients pour qui la corrélation était positive mais non significative. **B.** Faible corrélation non significative pour les patients du groupe de suivi, entre la reconnaissance correcte de la vois de l'homme et le délai d'activation de l'implant.

Une autre manière d'évaluer l'évolution des performances était de faire une corrélation entre les performances de reconnaissance correcte et le délai d'activation d'implant (voir Figure 10.40.B. pour la voix d'homme, et Figure 10.41.B. pour la voix de femme). Les corrélations individuelles n'ont été réalisées que lorsque les données étaient disponibles pour au moins 4 mesures du suivi, ce qui concernait 9 des 11 patients non supranormaux du suivi. Pour la voix d'homme on a pu constater au niveau individuel que 4 des 11 patients avaient une corrélation positive (voir Figure 10.40.A), mais elle n'était significative pour aucun d'entre eux. La corrélation était négative pour un seul patient ($R^2 = 0,5895$) et nulle pour les autres patients. Au niveau du groupe (voir Figure 10.40.B.), la corrélation était positive mais faible et non significative.

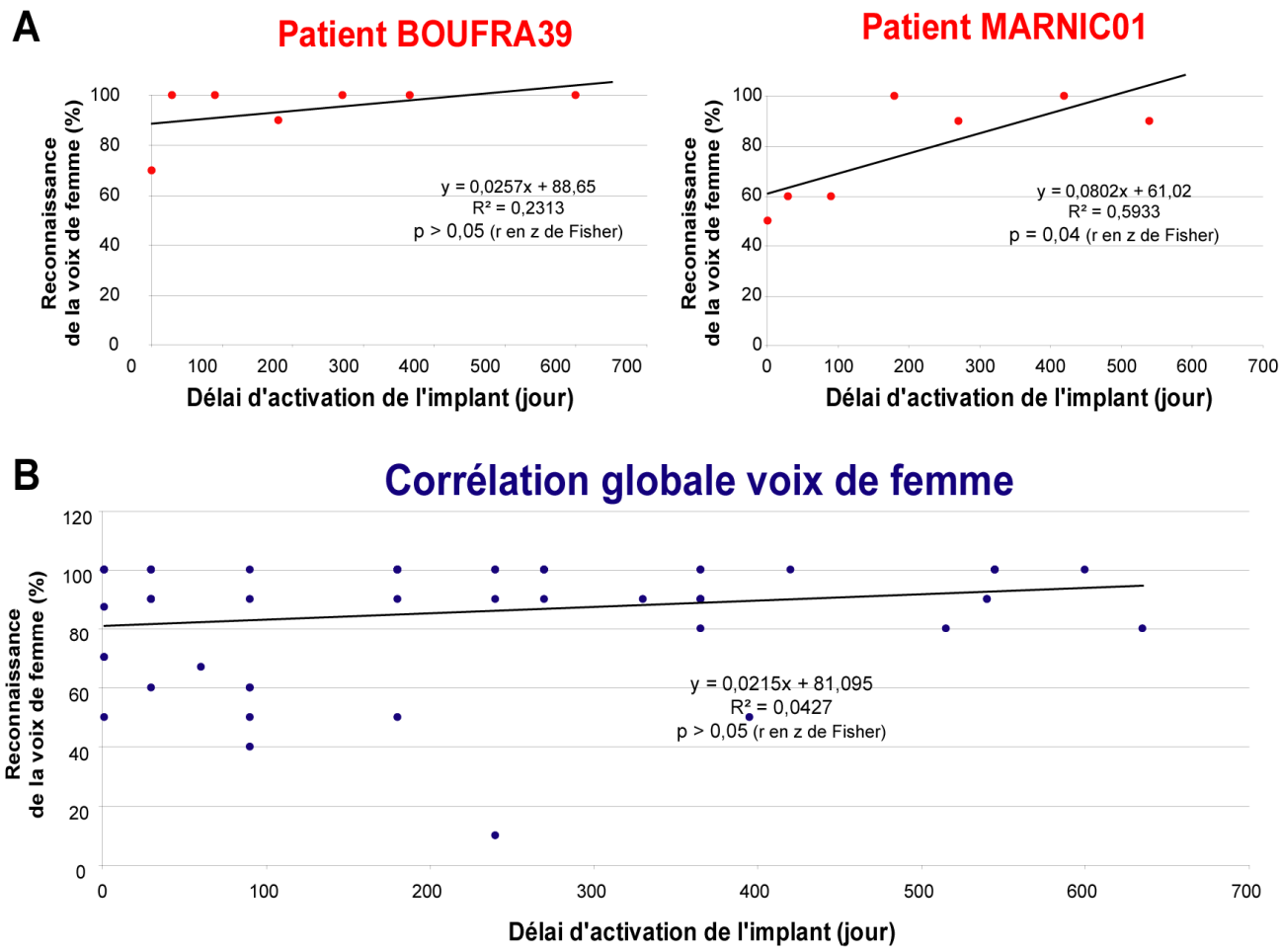


Figure 10.41 : **A.** Exemple de deux patients pour qui la corrélation était positive. La corrélation était significative pour le deuxième patient (MARNIC01) mais pas pour le premier. **B.** Très faible corrélation non significative pour les patients du groupe de suivi, entre la reconnaissance correcte de la vois de la femme et le délai d'activation de l'implant.

En ce qui concerne la voix de femme, la situation est un peu semblable mais moindre. Au niveau individuel, 5 patients avaient une corrélation positive (R^2 compris entre 0,6 et 0,1) mais elle n'était significative que pour l'un d'entre eux (voir Figure 10.41.A). Un patient avait une corrélation négative ($R^2 = 0,1533$) tandis qu'elle était nulle pour les autres. Toutefois aucune autre corrélation ne s'était révélée significative au niveau individuel. Au niveau du groupe (Figure 10.41.B.) la corrélation était positive mais encore plus faible que pour la voix d'homme et toujours non significative.

En résumé, l'analyse des voix extrêmes du continuum ne montrait aucun effet de groupe significatif. Mais on peut noter que les performances avaient nettement tendance à s'améliorer et que la comparaison avec les NHS confortait cette idée.

Les performances de compréhension du langage

Le groupe de patients suivis pour ce test de catégorisation du genre était le même que pour celui de la détection de la voix. Nous avons donc déjà exposé leurs résultats en termes de compréhension de mots, avec une nette amélioration des performances au cours de l'activation de l'implant (voir paragraphe « Les performances de compréhension du langage », p223). La corrélation de ces performances avec les performances obtenues pour les différents critères d'analyses (pente, seuil, erreur et reconnaissance correcte des extrêmes) sur le groupe de suivi à montré que ces deux tâches évoluaient relativement conjointement (voir Figure 10.42). En effet, les meilleures performances de compréhension de mots étaient associées à de meilleures performances de catégorisation du genre, c'est-à-dire à une erreur de la courbe plus faible et des reconnaissances correctes de la voix d'homme et de la voix de femme plus importantes. Les corrélations étaient significatives (Fisher ; $p < 0,05$).

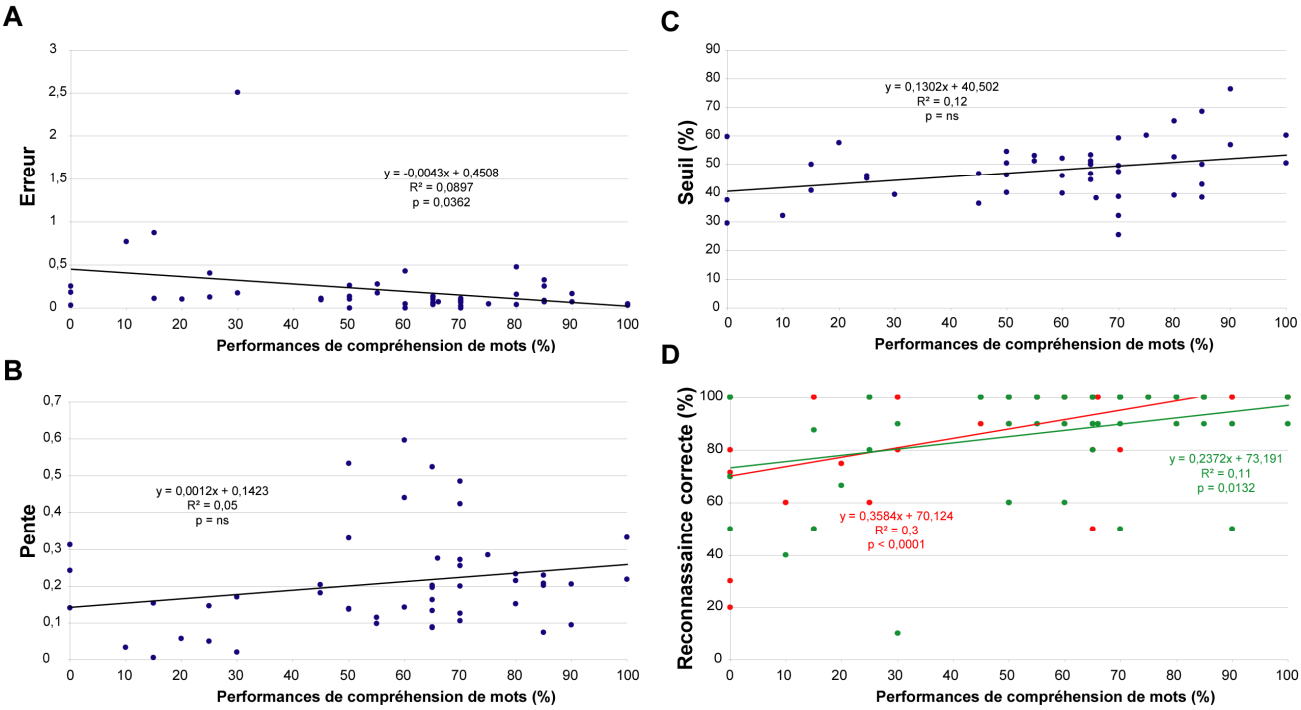


Figure 10.42 : Corrélations entre les performances de compréhension de mots et l'erreur (A), la pente (B), le seuil (C) et la reconnaissance correcte de la voix d'homme et de la voix de femme (D), pour les patients du groupe suivi. La corrélation était significative pour l'erreur et la reconnaissance correcte.

Nous avons également regardé les corrélations individuelles afin de s'assurer que la corrélation globale reflétait bien ces données. En raison d'un nombre de données insuffisantes, les corrélations n'ont pu être réalisées que pour 8 sur 11 patients non supranormaux. Pour la pente et le seuil, 6 corrélations sur 8 étaient nulles ($R^2 < 0,01$), 1 était positive ($R^2 > 0,22$) et 1 autre négative ($R^2 > 0,2$). Pour les deux critères il ne s'agissait pas des mêmes patients. Ceci indique donc que la corrélation pour la pente et le seuil sur l'ensemble des données du groupe reflète la majorité des

patients. Pour l'erreur et les reconnaissances correctes, à chaque fois 3 individus (pas les mêmes pour chaque critère) avaient une corrélation qui allait dans le sens de la corrélation globale ($R^2 > 0,3$) contre 4 (5 pour la reconnaissance de la voix d'homme) pour qui aucune corrélation n'était retrouvée et 1 patient (0 pour la reconnaissance de la voix d'homme) qui avait une corrélation inverse à la corrélation globale ($R = 0,2$ dans les deux cas). Pour ces 3 critères, le résultat du groupe était plus mitigé. On peut voir que pour la majorité la corrélation est nulle, mais ce qui la rend significative sur l'ensemble de l'échantillon sont les données de 3 individus qui ont une forte corrélation.

Dans leur ensemble, ces résultats rejoignent donc ceux du groupe transversal et nous amènent donc à dire que les performances de compréhension du langage s'amélioraient conjointement à celles de la catégorisation du genre. Toutefois, ces résultats, sont portés par une minorité d'individus, pour qui la corrélation est très fortement significative. Les données individuelles nuançaient donc légèrement les résultats du groupe.

Les performances de lecture labiale

De même que pour les performances de compréhension du langage, les résultats pour le test de lecture labiale des patients en suivi on déjà été exposés (voir paragraphe « Les performances de lecture labiale », p226). Les performances des patients restaient relativement stables au cours du temps et meilleures que celles des NHS.

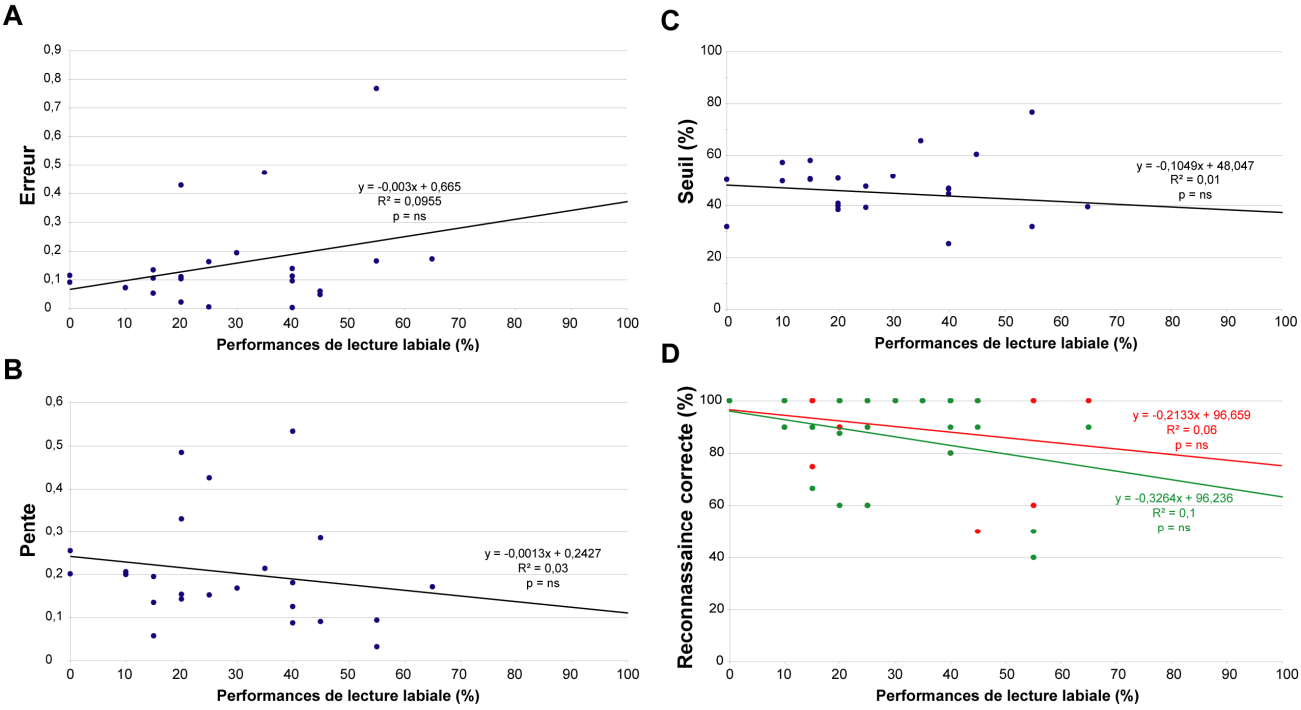


Figure 10.43 : Corrélations entre les performances de lecture labiale et l'erreur (A), la pente (B), le seuil (C) et la reconnaissance correcte de la voix d'homme et de la voix de femme (D), pour les patients du groupe suivi. Aucune corrélation significative n'a été retrouvée.

En ce qui concerne la corrélation entre les performances de lecture labiale et celles des différents critères de catégorisation du genre de la voix, nous n'avons pas retrouvé de corrélation significative (voir Figure 10.43). Toutefois, on peut remarquer que la tendance allait bien dans le sens de notre hypothèse de départ ; des meilleures performances de lecture labiale associées avec de moins bonnes performances de reconnaissance de la voix. En effet, les critères de catégorisation du genre montraient que lorsque les patients avaient de meilleures performances dans ce domaine (*i.e.* pente élevée, erreur de la courbe faible, bonne reconnaissance des voix extrêmes) ils en avaient de moins bonnes en lecture labiale. Même si les corrélations étaient toutes très faibles et non significatives, les corrélations des différents critères d'analyse de la catégorisation variaient de manière cohérente. Malheureusement le nombre important de données manquantes ne nous a pas permis de réaliser de corrélations individuelles et nous amènent donc à relativiser cette tendance.

Ces résultats ne corroborent pas ceux du groupe transversal où nous avions retrouvé une corrélation positive et significative. En revanche la tendance rejoint celle retrouvée pour le test de détection de la voix. Il est toutefois assez difficile de trancher sur les relations entre lecture labiale et perception de la voix, tant les tendances sont faibles et contradictoires entre le groupe de suivi et le groupe transversal.

LA CAPACITE PERCEPTIVE DU PITCH ET DU TIMBRE (3EME GROUPE)

L'objectif de ce test était de mesurer les capacités des patients dans la tâche de catégorisation du genre, avec les stimuli originaux, ainsi que sur la seule base des variations de pitch (hauteur de la voix) ou celle des variations de timbre.

La pente

Sur les 10 patients qui ont été inclus dans ce groupe, on peut voir que la pente moyenne de catégorisation en condition originale, était similaire à celle des patients du suivi à 18 mois et à celle des patients du groupe transversal à plus de 18 mois post-activation de l'implant (voir Figure 10.44). Ceci est cohérent avec notre critère d'inclusion qui était d'avoir une pente de catégorisation à plus ou moins 2 écart-types autour de la moyenne de la pente des patients au-delà de 18 mois.

Lorsqu'on regarde les capacités de catégorisation sur la seule information du timbre (condition « pitch constant » sur la Figure 10.44), on constate que les performances étaient très déficitaires, avec une pente moyenne de 0,066. Cette valeur était inférieure à toutes celles retrouvées en condition de vocoder chez les NHS et à celles retrouvées chez les patients. On peut donc dire que les patients ne perçoivent pas, ou très mal, l'information du timbre au travers de l'implant. En revanche, lorsqu'il n'y avait que les variations de pitch (condition « timbre constant »), les patients arrivaient à catégoriser aussi bien qu'en condition originale, et aussi bien que les NHS dans la même condition. Les analyses montrent que les performances des patients étaient similaires pour les conditions originale et timbre constant. Les pentes moyennes pour ces deux dernières conditions étaient significativement supérieures à celle de la condition pitch constant. Les analyses statistiques

ont aussi montré que les pentes moyennes des patients étaient significativement inférieures à celles des NHS, pour les conditions originale et pitch constant (Mann-Whitney : $p < 0{,}001$), tandis qu'aucune différence n'était retrouvée pour la condition timbre constant. Chez les patients, aucune différence n'a été trouvée entre les sujets masculins et féminins.

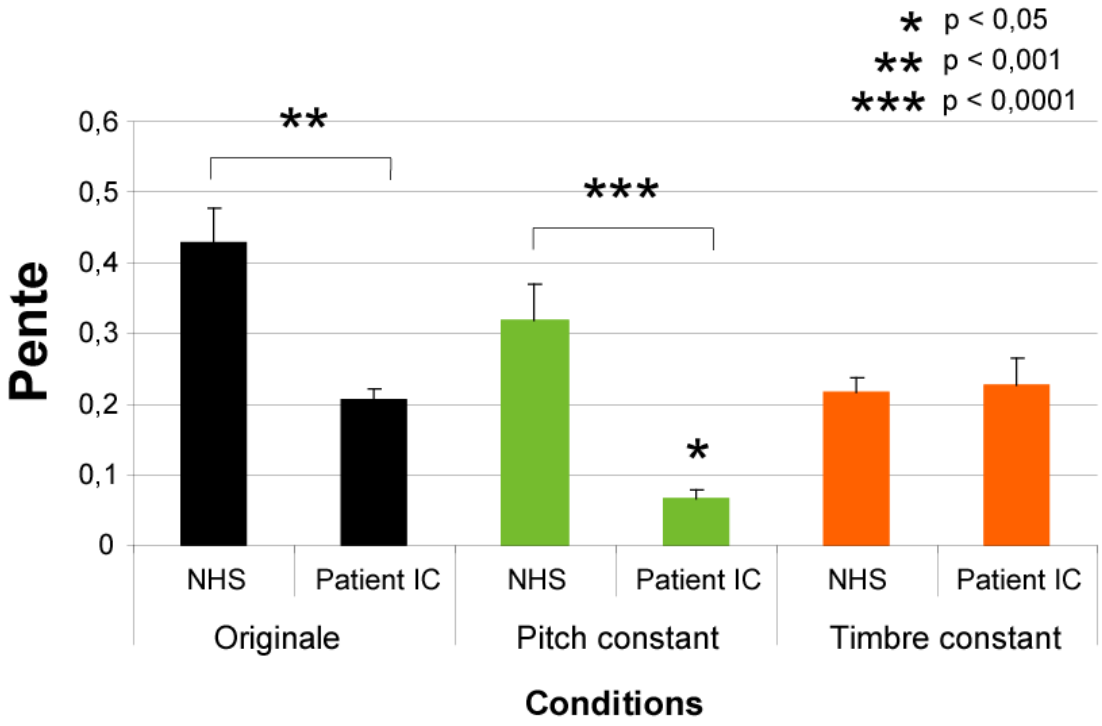


Figure 10.44 : Capacités de catégorisation représentées par les valeurs (±SE) de la pente chez les patients (N = 10) en condition originale, pitch constant et timbre constant, et comparaison avec les valeurs de pente des NHS (N = 14) dans les mêmes conditions. Les patients en timbre constant catégorisaient aussi bien qu'en condition originale. En condition pitch constant il y avait un déficit significatif des performances. Les patients avaient des performances inférieures aux NHS en condition originale et pitch constant, et similaires aux NHS en condition timbre constant.

Ces résultats très intéressants nous indiquent plusieurs choses. D'une part, sur la base des résultats des NHS nous avons pu montrer que vraisemblablement, les variations de timbre étaient des indices auditifs très utiles pour la catégorisation du genre chez les sujets normo-entendants, et particulièrement chez les femmes. Cette information acoustique seule semblait permettre une meilleure catégorisation que le pitch seul. Chez les patients le profil était différent. On remarque que, même si l'information de pitch est dégradée par le processeur de l'implant cochléaire, elle était suffisante et suffisamment bien perçue par les patients implantés pour permettre une catégorisation du genre, aussi bonne que les NHS. En revanche, les patients percevaient très mal les informations de timbre, ne leur permettant pas de se baser sur cet indice pour catégoriser le genre. On peut en conclure que si l'information de timbre est un indice important permettant une bonne catégorisation, cette information est très mal perçue chez les patients au travers de l'implant cochléaire. Or nous avons vu que l'information de pitch était elle bien perçue, il est donc évident que la catégorisation du genre chez le patient ne se fait que sur l'indice de pitch. On peut d'ailleurs remarquer qu'en

condition originale, les performances sont les mêmes qu'en condition timbre constant, où seules les variations de pitch sont présentes. Les indices acoustiques de timbre, présents en condition originale, n'amélioraient pas les performances de catégorisation.

Le seuil

En ce qui concerne le seuil de catégorisation, c'est-à-dire le point d'inflexion de la courbe, on peut voir sur la Figure 10.45 qu'il ne variait pas beaucoup chez les patients selon les conditions pitch constant ou timbre constant. Il y avait un léger décalage du seuil vers la voix masculine pour la condition timbre constant, c'est-à-dire que les patients avaient tendance à trouver davantage que les voix étaient féminines, mais cette tendance n'était pas significative. Mais d'un point de vue statistique, les variations de pitch seules (timbre constant) ou les variations de timbre seules (pitch constant) ne modifiaient pas significativement le seuil, bien que la variabilité de cette valeur au sein du groupe de patients augmente. La comparaison avec les NHS n'a révélé aucune différence significative. Autrement dit, pour chaque condition, les seuils n'étaient pas significativement différents entre les patients implantés et les sujets normo-entendants.

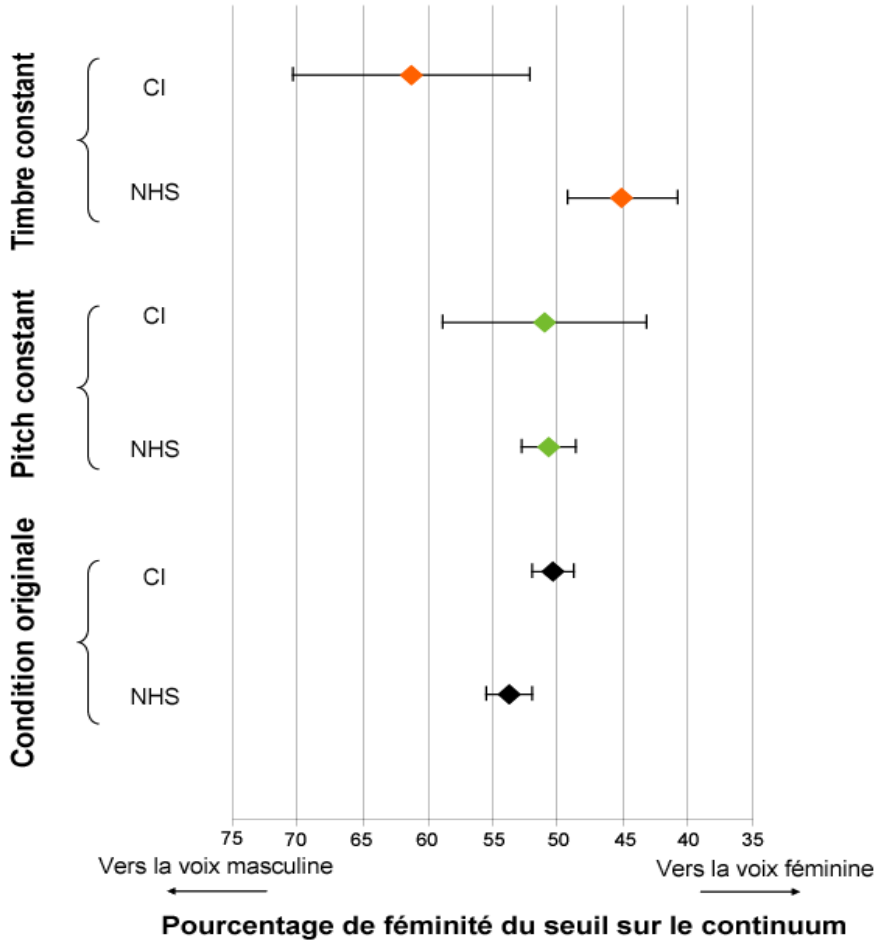


Figure 10.45 : Valeurs moyennes (±SE) du seuil de catégorisation chez les patients (N = 10) en conditions originale, pitch constant et timbre constant, et comparaison avec les valeurs de pente des NHS (N = 14) dans les mêmes conditions. Aucune différence significative n'a été retrouvée au sein des patients ni en comparaison avec les sujets normo-entendants.

L'erreur

L'analyse de l'adéquation du modèle (erreur de la courbe) a révélé que pour les patients, il y avait une moins bonne adéquation pour les conditions de pitch constant et timbre constant que pour la condition originale, ce qui se traduisait par une légère augmentation de l'erreur (voir Figure 10.46). La réponse de catégorisation du genre pour les patients était donc meilleure pour les voix originales que pour les deux autres conditions. Toutefois, cette variation chez les patients ne s'est pas révélée significative.

En comparaison des sujets normo-entendants, les réponses des patients correspondaient significativement moins bien à une catégorisation pour les conditions originale (Mann-Whitney : $p = 0{,}0005$) et pitch constant (Mann-Whitney : $p = 0{,}0006$). La différence allait dans le même sens pour la condition timbre constant, mais en raison d'une forte variabilité la différence n'était pas significative.

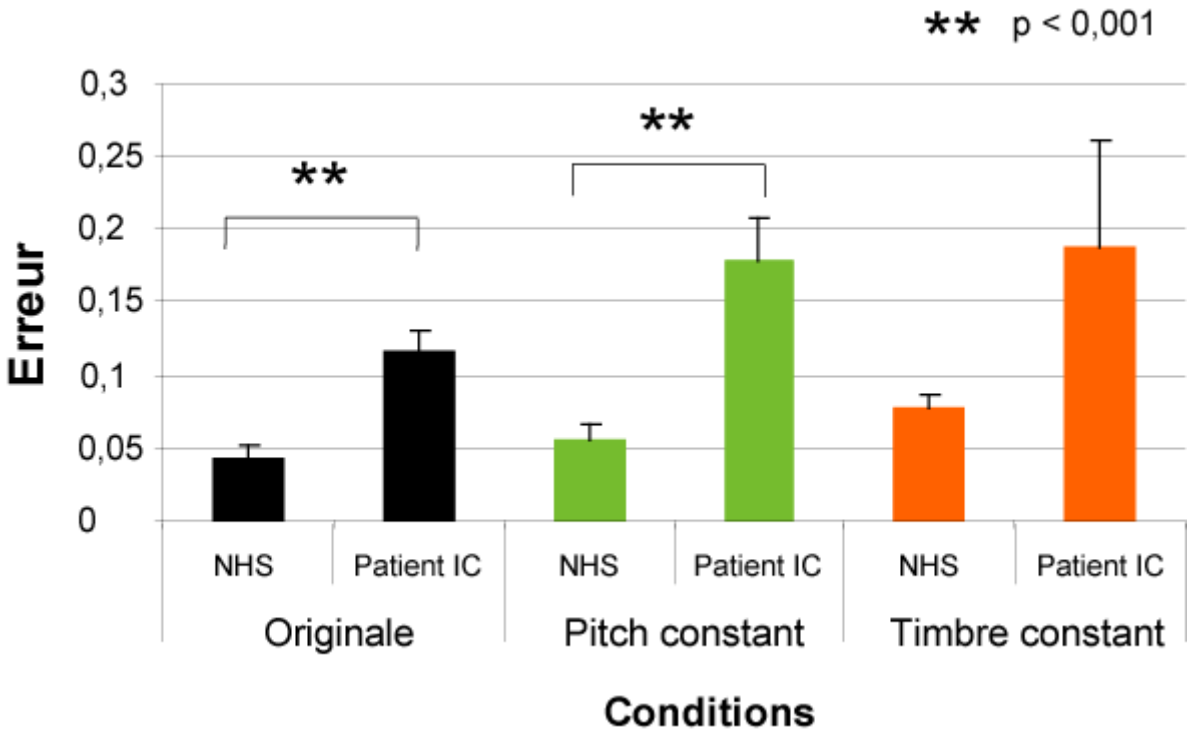


Figure 10.46 : Valeurs moyennes (±SE) de l'erreur de l'adéquation du modèle sigmoïde aux données des patients, dans la tâche de catégorisation du genre, dans les conditions originale, pitch constant et timbre constant, et comparaison aux NHS.

En comparaison de l'analyse de la pente, ces résultats confirment que les patients ont une moins bonne catégorisation du genre que les NHS, en condition originale et en condition pitch constant. Si la variation de l'erreur chez les patients n'était pas différente d'une condition à l'autre, on peut tout de même remarquer une certaine cohérence dans la variabilité de ces deux critères (pente et erreur). On peut conclure que la condition originale était celle où les patients étaient les meilleurs avec une pente maximum et une erreur minimum. Inversement, la condition pitch constant

était celle où la catégorisation du genre était la plus difficile avec une pente minimum et une erreur maximum.

Analyse des réponses aux extrêmes

Les reconnaissances correctes de la voix d'homme et de la voix de femme ont également confirmé l'analyse de la pente et de l'erreur (voir Figure 10.47). En effet, pour les deux voix, on observait des performances maximum, proches de 100% pour les voix originales. Les moins bonnes performances étaient observées pour la condition pitch constant. Les variations entres les différentes conditions étaient significatives pour la voix d'homme et la voix de femme (Friedman ; $p < 0,01$). Pour les deux voix, les performances en condition de pitch constant étaient inférieures à celles de la condition originale (Wilcoxon ; $p < 0,01$). Puis, pour la voix d'homme en condition timbre constant, les performances étaient les mêmes qu'en condition originale et significativement supérieures à celles de la condition pitch constant (Wilcoxon ; $p < 0,01$).

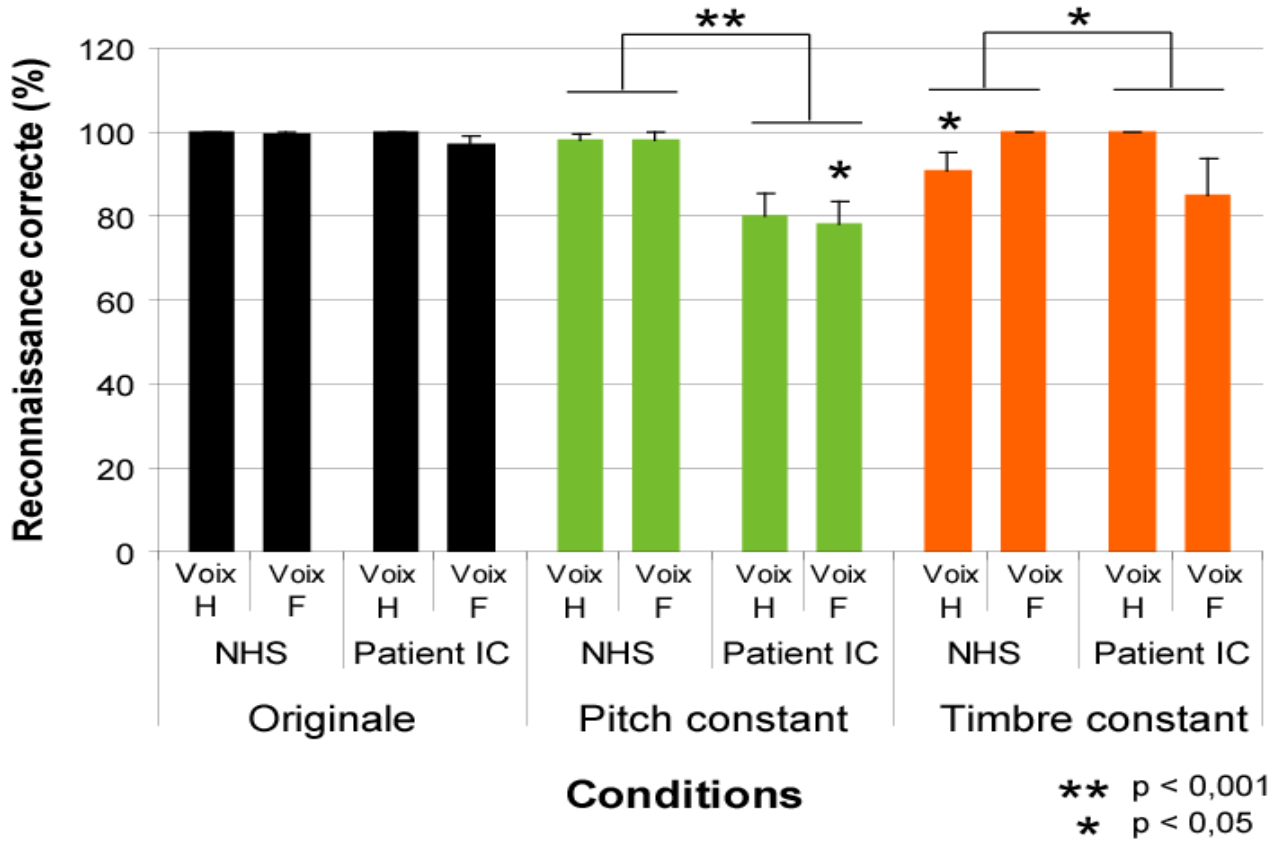


Figure 10.47 : Valeurs moyennes (±SE) de la reconnaissance correcte de la voix d'homme (Voix H) et de la voix de femme (Voix F) dans la tâche de catégorisation du genre, dans les conditions originale, pitch constant et timbre constant, et comparaison aux NHS.

En comparaison des NHS, aucune différence n'a été retrouvée en condition originale, où les performances étaient très semblables. En condition de pitch constant, les performances pour les deux types de voix étaient significativement inférieures aux NHS. Enfin en condition timbre constant, le profil inverse des NHS était observé ; une reconnaissance quasi-parfaite de la voix d'homme (Mann-Whitney ; $p < 0,01$) et une nette diminution de la reconnaissance de la voix de femme (Mann-Whitney ; $p < 0,01$). Cette dernière différence reste difficile à expliquer, mais au sein de chaque groupe, la différence de reconnaissance entre les deux voix n'était pas significative.

Dans tous les cas cette analyse confirme que pour les patients, la catégorisation du genre est déficitaire par rapport aux NHS en condition pitch constant. De plus elle confirme également la plus grande difficulté des patients à reconnaitre, et donc catégoriser, le genre en condition pitch constant. La condition originale restaient celle où les performances des patients étaient les meilleures tandis que la condition timbre constant était intermédiaire.

10.2.1.3 ROLE DE LA PROTHESE CONTROLATERALE SUR LES PERFORMENCES AVEC L'IMPLANT SEUL

Rappelons que ce que nous comparons ici est l'effet du port d'une prothèse au quotidien sur la compréhension au travers de l'implant seul. Pour étudier cela, nous ne disposions des données nécessaires que pour les groupes transversal et de suivi.

Pour ces deux groupes, étant donné qu'il y avait des sujets supranormaux qui ont été analysés à part, les résultats qui suivent ne porteront que sur le reste des sujets des groupes, non supranormaux. Toutefois, il est intéressant de noter que parmi les 6 sujets supranormaux du groupe transversal, c'est-à-dire ceux qui arrivaient mieux à catégoriser que les NHS, 5 d'entre eux ne portaient pas de prothèse auditive controlatérale. Parmi les 3 sujets supranormaux du groupe de suivi, un seul portait une prothèse. Ces observations ne vont pas dans le sens d'une tendance à ce que le port de la prothèse permette une meilleure perception de la voix et notamment du genre.

GROUPE TRANSVERSAL

En raison du profil comparable des patients implantés depuis 1 jour, avec et sans prothèse, aucune hypothèse n'était raisonnablement attendue. En conséquence, l'effet du port d'une prothèse auditive controlatérale sur les performances avec l'implant seul, a été analysé uniquement sur les patients implantés depuis plus d'un mois. Etant donné la répartition non équilibrée du port d'une prothèse auditive selon les différents groupes de patients, nous avons choisi, à l'instar du test de détection de la voix, de regrouper les patients dont l'implant était activé depuis plus d'un mois.

Les résultats sur les sujets (non supranormaux) sont présentés sur la Figure 10.48. Sur l'ensemble des patients dont l'implant était activé depuis plus d'un mois, 11 ne portaient pas de

prothèse auditive et 12 en portaient une (mais un patient ayant répondu au hasard, n'a pas de valeur de pente sur la Figure 10.48).

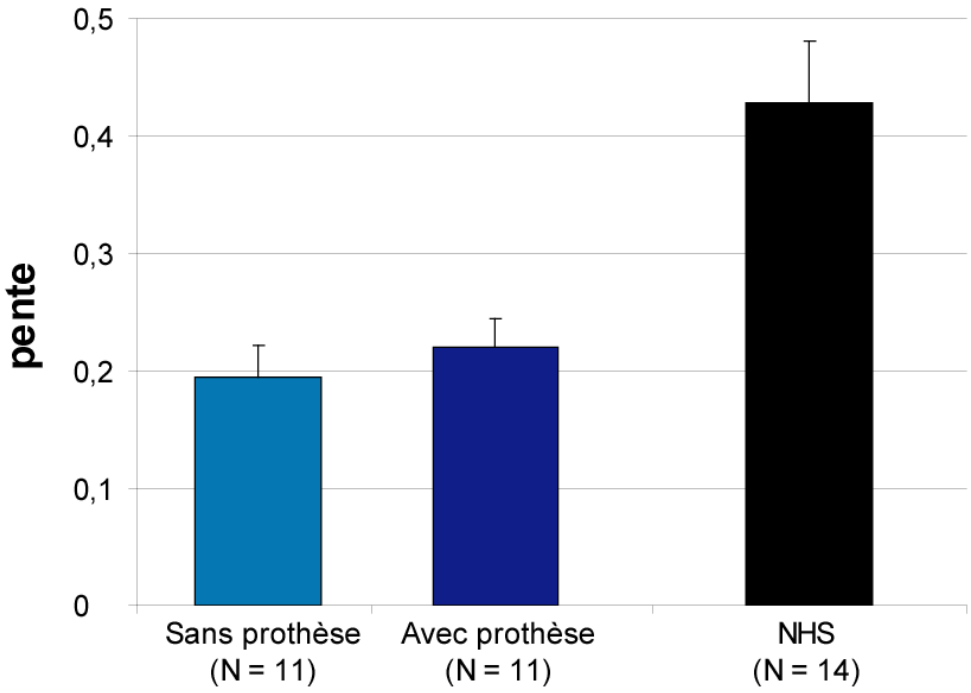


Figure 10.48 : Valeur moyenne (± SE) de la pente chez les patients implantés du groupe transversal, selon qu'ils portaient une prothèse auditive ou non et comparaison avec les NHS. Aucune différence n'a été retrouvée entre les patients porteurs de prothèse et ceux qui n'en portaient pas. Tous les patients ont été testés avec l'implant seul.

On peut voir qu'à partir d'un mois d'activation de l'implant, aucune différence entre les patients n'était retrouvée. On peut donc dire que le port d'une prothèse au quotidien n'améliorait pas la catégorisation du genre au travers de l'implant. Ceci peut être opposé à l'effet du port de la prothèse dans le test de détection de la voix où, pour le groupe transversal, on trouve une légère facilitation par le port de la prothèse. Cette différence peut s'expliquer par le fait que dans la tâche du genre, nous avons montré que les patients, même s'ils restaient déficitaires par rapport aux NHS, s'amélioraient au même niveau que les NHS en condition de vocoder à 16 et 32 canaux, ce qui correspondait au nombre moyen d'électrode de leur implant (18 électrode en moyenne). Il leur serait donc difficile de s'améliorer d'avantage en fonction du signal transmis par l'implant. Cette absence de différence entre les patients porteurs d'une prothèse et ceux qui n'en portaient pas était également vraie pour les autres critères d'analyse, comme le seuil, l'erreur de la courbe sigmoïde ou encore les reconnaissances correctes des voix extrêmes de l'homme et de la femme.

Pour résumer, les patients, qu'ils soient porteur de prothèse ou non, ont probablement atteint un plafond de performances correspondant au signal délivré par l'implant, ce qui ne permet pas de voir de facilitation des performances avec le port d'une prothèse controlatérale. Le port de la prothèse pouvait peut-être accélérer la récupération des performances chez les patients. C'est ce que nous avons étudié pour le groupe suivi, dans le paragraphe suivant.

GROUPE SUIVI

Pour le groupe du suivi, sur 11 patients qui n'étaient pas supranormaux, 4 portaient une prothèse et 7 n'en portaient pas. Dans un premier temps nous avons analysé les données des patients à chaque temps du suivi en comparant les performances entre les 2 groupes de patients. Les résultats sont présentés sur la Figure 10.49. On peut voir qu'à toutes les mesures du suivi, les patients qui ne portaient pas de prothèse avaient tendance à avoir de meilleures performances que ceux qui en portaient une. Cette différence en faveur des non porteurs de prothèse n'était pas significative, probablement en raison d'une variabilité assez importante au sein de chaque groupe. Ceci répond donc à la question du paragraphe précédent : le port de la prothèse ne permet pas de récupération plus rapide des performances. La tendance allait dans le sens inverse de que celle du test de détection de la voix où on observait plutôt une facilitation des performances avec le port de la prothèse. Toutefois ici, la variabilité était trop grande pour révéler une significativité, nous pouvons donc conclure que les patients implantés qui portaient une prothèse avaient des performances similaires à celles des patients qui n'en portaient pas.

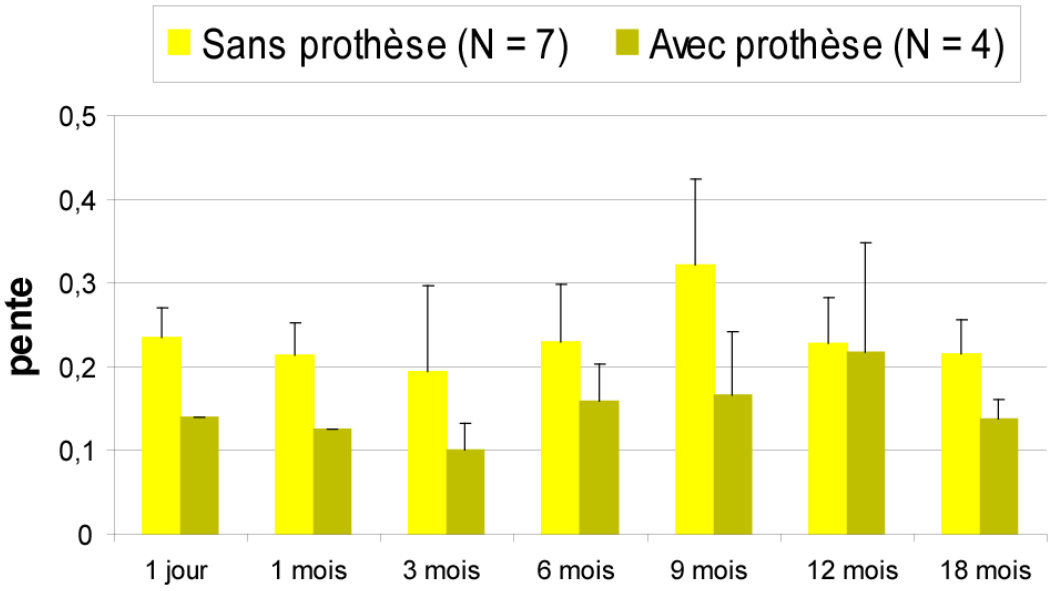


Figure 10.49 : Valeur moyenne (± SE) de la pente chez les patients implantés du groupe suivi, selon qu'ils portaient une prothèse auditive ou non. Les patients ont été différenciés pour chaque mesure du suivi. Tous les patients ont été testés avec l'implant seul. Aucun effet significatif n'a été retrouvé.

Dans un second temps, nous avons moyenné les performances de chaque sujet pour les mesures de 1 à 18 mois du suivi. Ceci afin de pouvoir plus facilement comparer aux résultats du groupe transversal.

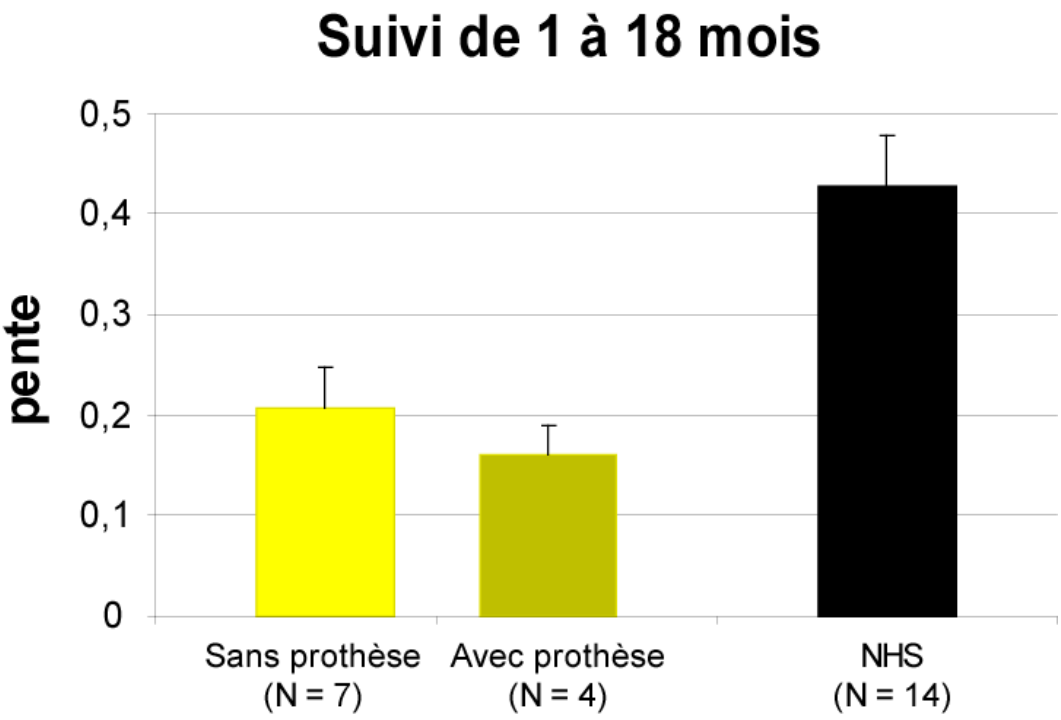


Figure 10.50 ; Valeur moyenne (± SE) de la pente chez les patients implantés du groupe suivi, selon qu'ils portaient une prothèse auditive ou non et comparaison avec les NHS. Aucune différence n'a été retrouvée entre les patients porteurs de prothèse et ceux qui n'en portaient pas. Tous les patients ont été testés avec l'implant seul.

Sur la Figure 10.50, on peut voir que lorsqu'on moyenne les performances pour chaque patient entre les mesures de 1 à 18 mois, on retrouve des valeurs de pente très similaires à celle du groupe transversal et toujours aucune différence significative entre les porteurs de prothèse et ceux qui n'en portaient pas.

10.2.2 TEST DE DISCRIMINATION ENTRE DEUX VOIX

SUJETS

Dans ce groupe le critère d'inclusion était le même que pour le test de catégorisation du genre en conditions de pitch constant et timbre constant, à savoir une pente au point d'inflexion comprise entre 0,11 et 0,31. Actuellement pour ce test, 4 patients ont déjà été recrutés, et nous espérons poursuivre les inclusions. Il s'agit donc ici de résultats préliminaires et les observations qui en seront faites seront à prendre avec précaution. Au moment des analyses, le groupe était constitué de 1 homme et 3 femmes, avec une moyenne d'âge de 41,5 ans (+/- 15,8). Deux d'entre eux avait un implant activé depuis 9 mois et les deux autres depuis 1 an. Parmi ces 4 patients, aucun ne portaient de prothèse auditive controlatérale à l'implant.

STIMULI

Pour ce test, nous avons utilisé les mêmes stimuli que pour les NHS (voir paragraphe « Stimuli », p189). Il s'agissait donc du continuum réalisé par morphing entre une voix d'homme et une voix de femme avec le logiciel Matlab 6.5®, qui consistait en un glissement progressif et linéaire des fréquences fondamentales (pitch) et des formants (timbre) de la voix d'homme vers ceux de la voix de femme (voir Figure 10.19, test 1). Les 1000 pas du continuum générés avec ce morphing, soient 1001 voix, étaient utilisés pour le test.

PROCEDURE

Les sujets étaient vus au cours de la prise en charge par l'Unité d'Implant Auditif (UIA) de Purpan, et étaient testés dans une pièce insonorisée du service ORL. Les stimuli étaient présentés en champ libre via deux haut-parleurs (KINYO, modèle PS-240) et le volume était ajusté pour être environ à 72 dB SPL à 60 cm des enceintes. Les sujets étaient placés à cette même distance des haut-parleurs et face à un clavier, avec lequel ils avaient pour consigne de répondre si les deux voix présentées étaient identiques ou différentes, en appuyant sur une des deux touches « contrôle » correspondant à leur réponse. Tout comme pour les sujets normo-entendants, nous avons considéré le continuum en 3 parties, voix d'hommes, voix intermédiaires et voix de femmes (pour les détails du test voir paragraphe « Procédure » chez les sujets normo-entendants, p190). Au travers de ce test nous avons déterminé la distance minimum entre deux voix pour laquelle les patients entendaient une différence pour chacune des parties du continuum. Pour des raisons de difficulté évidente chez les patients, aucune consigne de rapidité n'était donnée. Les stimuli étaient présentés aléatoirement et la durée entre deux essais était de 1,5 secondes. Le test durait entre 7 et 12 minutes, selon les temps de réponse des sujets. La présentation des stimuli et l'enregistrement des réponses des patients se faisaient à l'aide du logiciel Matlab 6.5®.

ANALYSES

Pour les analyses, nous avons relevé la distance minimum entre deux voix pour que le patient entende une différence, pour chaque partie du continuum (homme, intermédiaire et femme). Nous avons comparé les différentes distances entre elles pour voir s'il y avait des zones du continuum où les patients avaient une meilleur perception de la différence entre deux voix. De plus, pour chaque zone du continuum, nous avons comparé la distance obtenue pour les patients avec celle obtenue chez les sujets normo-entendant (voir paragraphe « La distance nécessaire pour percevoir une différence entre 2 voix », p191).

10.2.2.1 RESULTATS

Les résultats préliminaires montrent que, pour les patients la distance sur le continuum pour entendre une différence entre deux voix était similaire pour chaque zone du continuum, homme, intermédiaire et femme. Même si on doit tout à fait relativiser les analyses statistiques sur 4 patients, les résultats montraient qu'il n'y avait pas de différence significative.

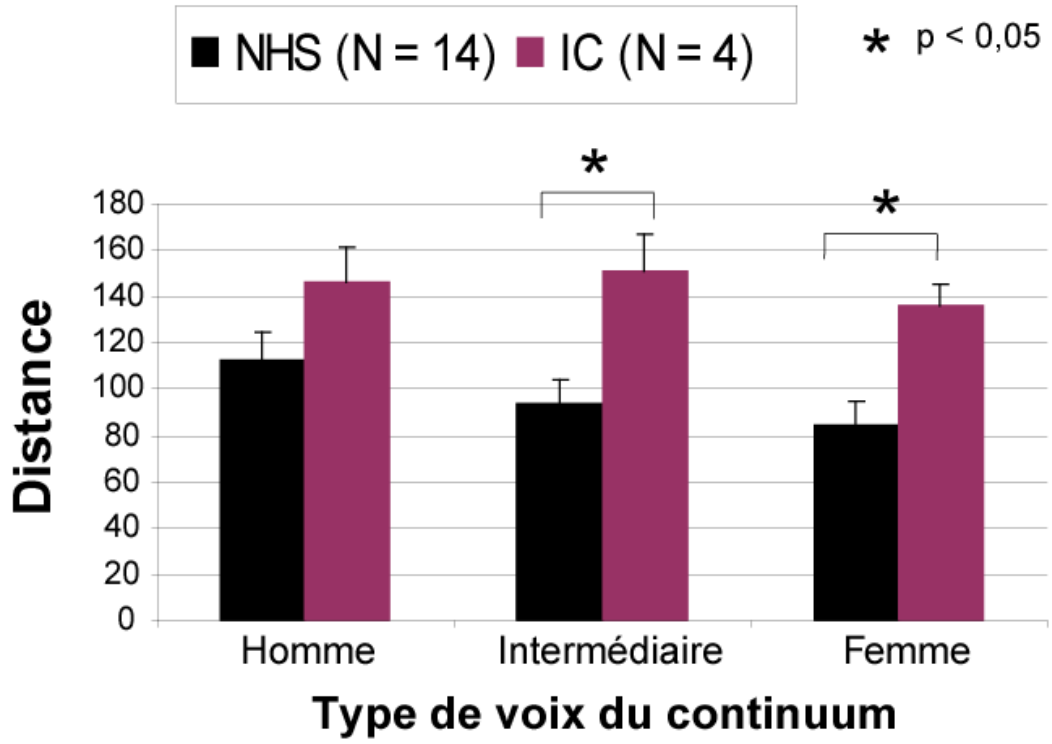


Figure 10.51 : distance moyenne perçue entre deux voix par les patients à plusieurs endroits du continuum ; sur le premier tiers de voix côté homme (homme), sur le deuxième tiers de voix intermédiaires (intermédiaire) et sur les dernier tiers de voix côté femme (femme). Les performances des patients sont comparées à celle des NHS.

Lorsqu'on a comparé les performances des patients à celles des NHS, on a pu voir que les patients avaient de moins bonnes performances pour toutes les zones du continuum, avec une distance nécessaire pour entendre une différence plus grande. La différence entre les patients et les NHS n'était significative que pour les zones voix intermédiaire et voix de femme (Mann-Whitney : $p < 0,039$). La différence n'était pas significative pour les voix homme car les NHS étaient un peu mois bons que pour les deux autres types de voix, mais ils restaient meilleurs que les patients. La question que l'on peut se poser maintenant, est que se passera-t-il avec des patients supplémentaires ? Va-t-on retrouver les mêmes différences entre les 3 parties du continuum que chez les NHS, c'est-à-dire une distance qui diminuait des voix hommes vers les voix de femme ? Ou va-t-on observer la même chose que sur nos 4 patients, à savoir une distance relativement similaire entre les différents types de voix ? Pour cela nous continuons à inclure des patients et espérons que l'échantillon soit bientôt suffisamment important pour finaliser les analyses. Toutefois il paraît évident que les patients sont déficitaires en comparaison aux NHS. Lorsqu'on compare les patients aux NHS en condition de simulation d'implant avec le vocoder, on remarque que les résultats pour les voix d'hommes et intermédiaires étaient similaires à ceux des NHS en condition de vocoder à 16 canaux et en-deçà. Ce résultat semble cohérent, étant donné que ce nombre de canaux correspond au

nombre moyen d'électrodes des patients (18 électrodes). En revanche pour les voix de femmes, les performances des patients étaient similaires à celles des NHS à 8 canaux de vocoder et en-deçà.

On peut donc conclure de ces résultats préliminaires que les patients semblaient déficitaires en comparaison des NHS, et que ce déficit semblait encore plus important du côté des voix de femme. Toutefois, l'inclusion d'autre patients est indispensable pour confirmer ces résultats.

10.3 DISCUSSION

En premier lieu, rappelons que cette approche comportementale chez les patients sourds implantés cochléaires était réalisée en parallèle à celle faite chez le normo-entendant. L'objectif en comparant les deux groupes était de pouvoir faire la part des choses dans la perception de la voix, entre ce qui relève de la nature du signal et ce qui relève de la surdité et les éventuelles réorganisations neurofonctionnelles qu'elle implique.

Dans l'ensemble on peut dire que les mesures perceptives que nous avons réalisées étaient aux limites de la perception des patients (et des sujets), en utilisant des durées de stimuli très courtes (400 ou 500 ms). L'idée en plaçant les sujets dans des conditions de perception difficiles était d'éviter tout effet plafond des performances afin de pouvoir évaluer l'évolution de ces performances.

10.3.1 LA DETECTION DE LA VOIX

Les résultats du test de détection de la voix donne un résultat intéressant essentiellement par son caractère de nouveauté. En effet elle n'avait jamais été étudiée dans la littérature scientifique. Or son évaluation chez les patients apporte des renseignements précieux sur un des premiers niveaux d'analyse auditive de la voix du modèle de perception de la voix (Belin et al., 2004). Nos résultats montrent plusieurs choses.

Tout d'abord, les groupes de suivi et transversal ont des résultats très similaires, ce qui nous permet d'aborder les résultats de manière commune aux deux groupes. En effet, pour les deux groupes, on observait un déficit important par rapport aux sujets normo-entendants en condition originale.

Un point important est que ce déficit ne portait pas sur les performances de compréhension du langage, qui avec l'expérience de l'implant atteignaient un niveau de compréhension très satisfaisant. Le déficit observé chez les patients portait donc spécifiquement sur la détection de la voix. Cette dissociation de récupération s'explique notamment par deux phénomènes déjà évoqués (voir « Les performances de compréhension du langage », p213). Le premier est le fait que les patients savaient déjà que les sons qui seraient présentés pour le test de compréhension du langage étaient de la voix, et le second est l'effet du contexte lexical, qui permettait aux patients en cas de compréhension partielle de compléter l'information avec leurs connaissances lexicales. Ces deux

effets rendent donc délicat la comparaison directe entre ces deux tests. On remarque tout de même que lorsque la voix comportait des signaux vocaux langagiers le déficit était moins important. Or dans ce cas il n'y avait aucun effet du contexte lexical. Ce résultat indique donc que la détection de la voix langagière est mieux perçue, ce qui s'explique notamment par des stratégies de codage axées sur la reconnaissance du langage.

Cependant, le déficit observé chez les patients implanté par rapport aux normo-entendants est bien plus important que celui que l'on observait chez ces derniers dans des conditions de simulation d'implant similaires (16 canaux). En effet, on retrouvait pour les patients des performances équivalentes à celles des normo-entendants naïfs par rapport à ce type de signal, en condition de simulation d'implant à 4 voire 2 canaux. De plus ce déficit n'évoluait quasiment pas au cours du temps ; les performances stagnaient avec le délai d'activation de l'implant.

Il semble donc évident au travers de ces résultats que le déficit des patients ne s'expliquait pas uniquement par la nature du signal transmis par le processeur de l'implant, puisqu'on a pu voir que le déficit des patients était très supérieur à ce à quoi on aurait pu s'attendre en fonction des résultats obtenus chez les normo-entendants, pourtant non expérimentés, tandis qu'ils étaient capables de s'améliorer au niveau de la compréhension de la parole.

10.3.2 LA CATEGORISATION DU GENRE

Pour ce test nous avons pu montrer que les patients implantés cochléaires étaient déficitaires par rapport à des sujets normo-entendants en condition originale. Ce résultat n'est pas nouveau, il a déjà été étudié dans la littérature, et les études montraient toutes un déficit de la perception du genre chez les patients (Q. J. Fu, S. Chinchilla et al., 2005; Kovacic & Balaban, 2009; Meister et al., 2009). Dans la plupart de ces études, les résultats montraient des performances correctes proches de 80 à 90 %, bien qu'inférieures à celles des normo-entendants (généralement égales à 100%), lorsque les pitches des voix d'hommes et de femmes étaient suffisamment distants. Les performances des patients chutaient drastiquement lorsque les pitches des voix d'hommes et de femmes se retrouvaient dans la même zone fréquentielle. Les auteurs suggéraient ainsi que la discrimination du genre chez les patients implantés cochléaires se faisait essentiellement sur la base du pitch. Les études chez les normo-entendants ont également montré que la distinction du genre du locuteur se faisait essentiellement sur la base du pitch, laissant le rôle du timbre et des autres indices acoustiques secondaires, et davantage pertinents pour les cas d'ambiguïté du pitch. Ceci confirme nos résultats obtenus dans le test de catégorisation. Les patients sont aussi bons pour catégoriser le genre en condition originale que lorsqu'il y a seulement l'information du pitch, bien que toujours inférieurs au normo-entendants. Les patients perçoivent donc suffisamment bien le pitch pour catégoriser le genre et l'information qui fait défaut aux patients pour avoir un niveau de perception du genre similaire à celui des normo-entendants est plutôt relative au timbre.

Toutefois, la comparaison avec la simulation d'implant chez les sujets sains naïfs a également montré que ce déficit était moins important que pour la détection de la voix. Le déficit correspondait à celui des normo-entendants dans les mêmes conditions de simulations d'implant, 16 canaux. Rappelons cependant que les normo-entendants expérimentaient le vocoder pour la première fois, constat qui a pour conséquence de confirmer l'existence d'un déficit chez les patients, qui ne serait pas seulement dû au processeur de l'implant.

Ce déficit moins important dans la tâche de catégorisation du genre s'explique de plusieurs manières. Tout d'abord comme pour le test de la perception de la parole les patients savaient qu'il s'agissait de stimuli vocaux, ce qui leur permettait de se concentrer davantage sur la perception du genre, et sur la perception d'indices qui leur seront spécifiques à cela. Par exemple si on envisage une tâche où les sujets auraient d'abord à reconnaître une voix avant de la catégoriser comme féminine ou masculine, on peut aisément imaginer que les performances chuteraient de façon importante. Ensuite, nous avons vu que le pitch, d'une part était un bon indicateur du genre, et d'autre part passait suffisamment bien pour réaliser la tâche. Certains auteurs ont même rapporté de très bons résultats de perception du genre mais avec des stimuli plus longs (Kovacic & Balaban, 2009). Or les patients que nous avons testés l'étaient avec des stimuli très courts de 400 ms. Il semble donc que les patients, en ayant accès à davantage d'information acoustique à mesure que la durée du stimulus augmente, soient capables de développer des stratégies compensatrices. Plusieurs travaux vont dans ce sens. On peut notamment citer Chatterjee et Peng qui ont montré que, dans une étude de perception du pitch au travers de l'intonation, les patients implantés pouvaient se montrer plus sensibles au pattern temporal afin de mieux percevoir le pitch (Chatterjee & Peng, 2008). D'autres travaux, non publiés, ont été réalisés au sein du laboratoire du CerCo, par Mathieu Marx et Jessica Foxton. Des patients et des sujets normo-entendants devaient décider si la phrase qu'il entendait était une question ou une affirmation, alors que le pitch était constant sur toute la phrase. Les résultats ont montré que, chez les patients implantés depuis plus d'un an, les performances étaient meilleures que celles des normo-entendants (voir Figure 10.52). Ces résultats laissaient entendre que les patients étaient capables de compenser la perception du pitch par des habiletés particulières de perception des modulations d'amplitude et de rythme (Marx, 2009).

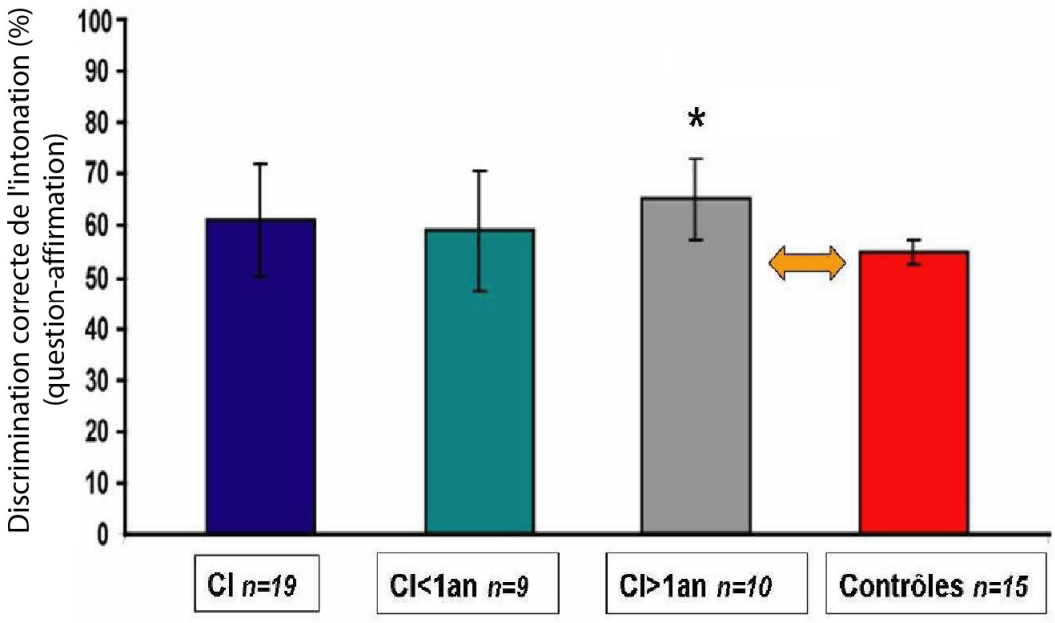


Figure 10.52 : Résultats des patients implantés cochléaires (CI) et des sujets normo-entendants (Contrôles) au test de discrimination de l'intonation entre question et affirmation, en condition de pitch constant. Les diagrammes représentent les pourcentages moyens (±SD) de discrimination. Extrait de Marx, 2009.

En conclusion, on peut dire que le déficit des patients en matière de perception du genre est essentiellement lié à une mauvaise perception de la résolution spectrale. Mais ce déficit est probablement partiellement compensé, comme le suggèrent d'autres travaux, par une meilleure perception d'autres indices comme les variations d'amplitude et de rythme.

10.3.3 LA DISCRIMINATION ENTRE DEUX VOIX

Pour ce test nous devons rappeler qu'il s'agit là de résultats préliminaires, et qu'ils nécessitent donc d'être complété. Toutefois il semble, qu'avec ce nombre restreint de patient, on observe un grand déficit de discrimination de voix. En comparaison des sujets normo-entendants, il ne semble pas y avoir de différences entre les différents types de voix.

Toutefois l'observation d'un déficit dans le domaine de la discrimination de voix chez les patients implantés n'est pas étonnant. De telles observations ont déjà été rapportées au travers de la littérature (Meister et al., 2009; Muhler, Ziese, & Rostalski, 2009; Wilson, Lawson, Muller, Tyler, & Kiefer, 2003; Zeng, 2004). On s'attend donc en incluant davantage de sujet à confirmer ce déficit. Dans des travaux récents, non publiés, réalisés par Mathieu Marx (doctorant au CerCo), ont permis d'évaluer les performances de discrimination de voix en lien avec l'information du pitch (Marx, 2009). Pour cela, des patients et sujets normo-entendants étaient soumis à plusieurs essais, où ils devaient écouter une phrase prononcées deux fois consécutives. La tâche était de dire si les deux prononciations correspondaient au même locuteur ou à deux locuteurs différents. Notons que même

lorsqu'il s'agissait du même locuteur, il s'agissait bien de la même phrase mais pas du même échantillon de voix. Les analyses ont tenu compte de la distance du pitch moyen entre les deux locuteurs, lorsque les voix étaient différentes, et ont calculé l'indice de sensibilité *d'* pour 3 niveaux de difficulté différents : difficile (différence de pitches < 10 Hz), intermédiaire (différences de pitches entre 15 et 25 Hz) et facile (différence de pitches > 35 Hz). Les résultats montrent que lorsque le niveau de difficulté était plus élevé (difficile), les performances de discrimination de voix diminuaient (voir Figure 10.53.A). Mais la différence n'était pas significative. En revanche, en comparaison avec les NHS, les patients implantés étaient toujours significativement déficitaires, quelque soit le niveau de difficulté de l'indice de pitch.

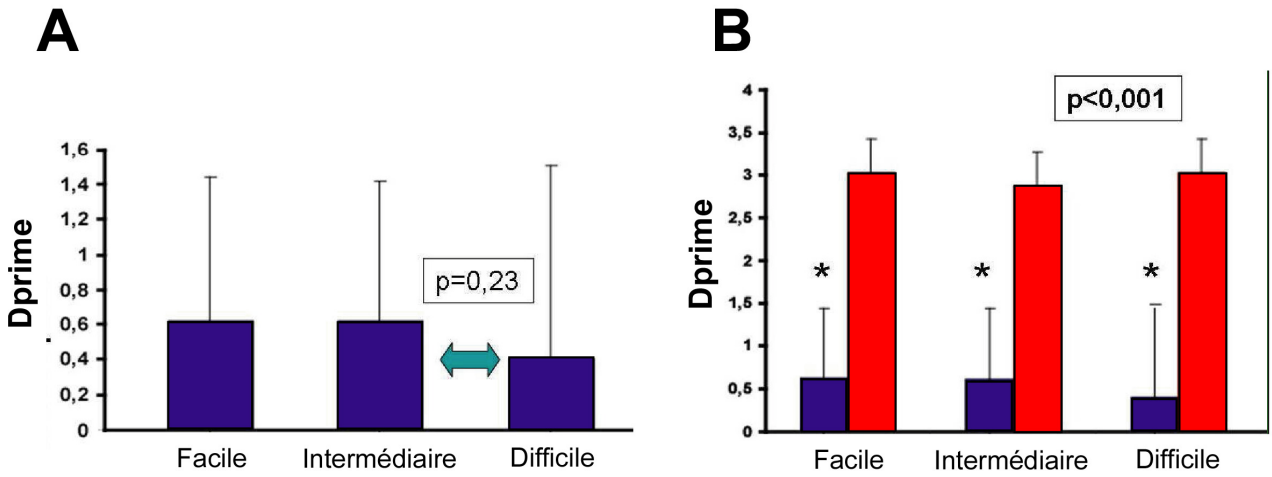


Figure 10.53 : **A.** Valeur des *d'* (±SD) en fonction du niveau de difficulté pour le groupe de patients implantés (n=11). Bien qu'on constate une diminution des performances en condition difficile par rapport aux deux autres conditions, la différence n'était pas significative. **B.** D des *d'* (±SD) en fonction du niveau de difficulté pour le groupe de patients implantés et celui du groupe de normo-entendants. La différence entre les deux groupes était significative pour chaque niveau de difficulté. Extrait de Marx, 2009.

Nous envisageons à l'avenir, d'une part de continuer à inclure des patients pour ce test, et d'autre part d'étudier l'effet de la familiarité de la voix sur la discrimination de voix. On peut en effet imaginer que lorsqu'une voix est familière à un patient, il est plus performant pour la reconnaitre et la différencier des autres voix. De plus nous mesurons en parallèle les capacités de discrimination de pitch des patients afin de voir s'il existe une corrélation entre leurs performances à ces deux tests.

10.3.4 DISSOCIATION DE RECUPERATION ENTRE VOIX ET LANGAGE

Tout comme chez les sujets normo-entendants, nous avons montré le statut particulier du langage en termes de perception, dans la tâche de détection de la voix. Ce résultat confirme donc que le vocodeur affecte moins les stimuli langagiers, de par la richesse de leurs modulations temporelles. Toutefois les résultats chez les patients implantés cochléaires vont plus loin que ce constat et

montrent une dissociation de récupération entre les performances de compréhension du langage et les performances de perception de la voix. En effet, même si nous avons vu que le déficit était plus important dans la tâche de détection de la voix que dans le test de catégorisation, il reste que les performances des patients expérimentés n'augmentaient pas significativement par rapport au premier jour d'activation de l'implant. Or même si dans le cas du test du genre on notait des performances similaires à celles des normo-entendants dans des conditions similaires de vocoder (16 canaux contre 18 électrodes en moyenne), il reste que les normo-entendants étaient tous naïfs. On peut donc comprendre que des patients nouvellement implantés soient au même niveau, mais on s'attend, comme c'est le cas pour le langage à ce que les performances soient meilleures chez des patients expérimentés avec l'implant. Ceci est donc le signe d'un déficit de perception de la voix chez les patients en comparaison des normo-entendants.

En résumé, ce déficit se constate donc d'une part en comparaison avec les normo-entendants en condition de simulation d'implant mais également au regard de l'absence d'évolution de leurs performances avec l'expérience de l'implant. Or comme nous l'avons vu, et comme l'on montré de nombreuses études (voir « Conséquences de l'implantation cochléaire sur les performances auditives des patients » et plus spécifiquement « Le langage », p128) la compréhension du langage évolue fortement au cours du temps chez les patients implantés, allant même pour certains d'entre eux, jusqu'à des niveaux similaires à celui des normo-entendants.

En réponse à l'objectif de départ de ce travail de thèse nous pouvons donc affirmer que les patients implantés cochléaires présentent, de façon générale un déficit de perception de la voix, qui ne semble pas s'améliorer au cours du temps, contrairement aux capacités de compréhension du langage.

10.3.5 INTERET DU SUIVI

Dans notre étude nous avons choisi d'aborder deux approches : la mesure transversale et le suivi. D'un point de vue général, le suivi est plus contraignant que l'approche transversale, car il implique de suivre les mêmes patients, depuis leur activation d'implant. Le problème est que, sur la durée, il est très difficile de voir les patients à tous les temps du suivi. Or tout l'intérêt du suivi est de pouvoir avoir toutes ces mesures afin de contrôler les variabilités interindividuelles. De ce point de vue là, il parait préférable de se contenter de l'approche transversale, qui permet d'inclure beaucoup de patients sur une durée beaucoup plus courte, car il n'y a pas de données manquantes. D'un autre côté, nos résultats ont pu montrer que l'approche en suivi, en diminuant la variabilité individuelle permettait de mettre en évidence des résultats plus cohérents au cours du temps. Par exemple pour le test de détection de la voix, la différence entre les voix langagières et non langagières apparaissait à 1 jour post-activation et à plus de 18 mois post-activation, mais n'apparaissait pas pour les autres patients intermédiaires. En revanche cette différence était bien présente pour toutes les mesures du suivi. Toutefois, même si les mesures du suivi paraissaient plus cohérentes, dans l'ensemble elles

révélaient les mêmes variations et les mêmes résultats que le groupe transversal. Il nous parait donc plus simple dans les tests psychophysiques à venir d'utiliser une approche transversale.

10.3.6 EFFET DE L'AUDITION RESIDUELLE SUR LA PERCEPTION DES INFORMATIONS PARALINGUISTIQUES

Dans notre étude, nous n'avons pas pu mettre en évidence de bénéfice direct d'une audition résiduelle et du port d'une prothèse controlatérale sur la perception au travers de l'implant. Seule une légère tendance se dégageait pour le test de détection de la voix. Notons que cette tendance était observée pour le test où les patients dans leur ensemble présentaient le plus gros déficit par rapport aux sujets normo-entendants. Il n'est pas exclu que des tests approfondis puissent mettre en évidence un réel bénéfice.

De nombreuses études ont montré l'effet bénéfique d'information d'autres modalités, notamment celle de la vision sur les performances auditives. Les résultats sont même allés jusqu'à montrer qu'un entraînement audiovisuel améliorait la perception auditive seule. Ce résultat a été montré aussi bien chez les normo-entendants que chez les patients implantés (Q.J. Fu, Galvin, Wang, & Nogaki, 2005; Q. J. Fu, G. Nogaki et al., 2005; Gfeller, Witt, Adamek et al., 2002; Li & Fu, 2007; Loebach & Pisoni, 2008; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Nogaki, Fu, & Galvin, 2007). L'idée étant que l'information supplémentaire à l'implant viendrait agir comme feed-back et ainsi corriger et améliorer la perception au travers de l'implant seul. En ce qui concerne l'audition résiduelle chez les patients, des auteurs ont déjà montré que l'utilisation d'un implant cochléaire et de l'audition résiduelle quand il y en a, ont un effet bénéfique sur l'audition (M. F. Dorman, Gifford, Spahr, & McKarns, 2008). On peut donc logiquement s'attendre à ce qu'en jouant ce rôle de feed-back, elle améliorerait la perception au travers de l'implant. Des analyses complémentaires sont donc nécessaires pour confirmer cette hypothèse.

10.3.7 IMPLICATIONS TECHNOLOGIQUE ET CLINIQUE

Les données que nous avons obtenues, peuvent amener les stratégies de rééducation et les techniques des implants cochléaires à se développer dans au moins deux directions. Tout d'abord, il ressort de ces travaux, que l'information du pitch est relativement bien perçue au travers de l'implant cochléaire. En revanche beaucoup des déficits observés chez les patients semblent expliqués par une mauvaise perception du timbre. Or, d'après les résultats obtenus chez les sujets normo-entendants en condition de simulation d'implant, il semble que l'augmentation du nombre de canaux apporte un réel bénéfice sur la perception de la voix. En effet l'augmentation du nombre de canaux de 16 à 32 permet parfois d'améliorer les performances des normo-entendants non expérimentés avec le signal du processeur, à des niveaux proches de la condition originale. Toutefois, comme nous l'avons

évoqué précédemment, il semble que les patients, en raison de la résolution spectrale limitée, soient capables de développer des stratégies compensatrices, basées sur la perception des variations d'amplitude ou de rythme. Il semble difficile de dire actuellement si leur meilleure aptitude provient d'une plus grande sensibilité qui se serait développée ou bien d'une meilleure utilisation des indices naturellement perçus. Toutefois, le traitement de ces indices d'intensité et de variation temporelle, peuvent également être une piste à explorer pour les stratégies de codage, par exemple en mettant plus d'emphase sur les variations d'intensité propre à l'intonation.

Par ailleurs, nous avons pu constater que, le déficit observé chez les patients ne provenait probablement pas uniquement de la nature du signal transmis par le processeur, mais aussi surement de la suite des réorganisations corticales conséquentes à la surdité. Or, après l'implantation, la rééducation orthophonique est essentiellement axée sur la perception et la compréhension du langage, qui fait que les patients voient leurs capacités de communication augmenter de façon extrêmement satisfaisante. Cette rééducation participe très certainement à la réactivation des réseaux impliqués dans le langage, même si comme on a pu l'exposer dans la partie « Réorganisation des voix auditives », ces réseaux sont probablement différents (voir p138). Il est donc possible qu'une rééducation spécifique axée sur la détection et la perception des aspects paralinguistiques de la voix puisse améliorer les performances des patients. Nous envisageons donc de pouvoir comparer des patients qui auraient un entrainement supplémentaire sur la perception de la voix, avec d'autres patients qui auraient une rééducation plus classique. Les patients seraient évalués sur ces mêmes tests, ainsi que sur d'autres aspects paralinguistiques de la voix tels que la familiarité ou l'information affective de la voix. Nous espérons pouvoir observer un bénéfice d'une rééducation spécifique à la voix. De plus, l'hypothèse d'une plus grande réactivation de la TVA chez des patients entrainés, nous semblent également une hypothèse probable, qui mériterait d'âtre explorée en imagerie.

11 ÉTUDE D'IMAGERIE FONCTIONNELLE CHEZ LE SUJET SAIN EN CONDITION DE SIMULATION D'IMPLANT

11.1 METHODES

Cette étude IRMf s'inscrit dans une étude plus large, réalisée à Glasgow, en collaboration avec Kusma Strelnikov, Julien Rouger et Pascal Belin. Nous ne présenterons ici qu'une partie des résultats en rapport avec le sujet qui nous préoccupe, la reconnaissance de la voix. L'objectif de notre analyse était d'étudier l'effet du vocoder sur l'activation de l'aire de la voix. Pour cela nous avons dans un premier temps définit les aires qui répondaient spécifiquement aux stimuli vocaux. Nous avons utilisé le *voice localizer* proposé par Pascal Belin, gratuit et disponible sur le site de l'auteur (stimuli disponibles sur http://vnl.psy.gla.ac.uk/ressources_main.php). Dans un second temps, après avoir défini les aires répondant spécifiquement à la voix, nous avons mesuré l'activation cérébrale en réponse à des sons vocaux vocodés.

De plus, une partie dite, comportementale, a été réalisée afin d'évaluer les performances des sujets sur la détection de stimuli utilisés dans l'IRMf.

11.1.1 SUJETS

Pour cette étude nous avions recruté 15 sujets normo-entendants, de langue maternelle anglaise, dont 9 hommes et 6 femmes avec une moyenne d'âge de 26 ans. Les sujets n'avaient rapporté aucune pathologie neurologique ou auditive. Tous les participants avaient donné leur consentement éclairé, selon la déclaration d'Helsinki (1968). L'étude avait été approuvée par le comité d'éthique local de l'Université de Glasgow.

11.1.2 STIMULI

Les stimuli utilisés pour le *voice localizer* étaient ceux proposé par Pascal Belin sur son site (stimuli disponibles sur http://vnl.psy.gla.ac.uk/ressources_main.php), et utilisé dans Belin *et al*., 2000. Il s'agissait de 40 blocks de sons (16 bit, mono, 22050Hz) de 8 secondes. Parmi eux, 20 étaient des blocks de sons vocaux composés d'un tiers de stimuli langagiers (mots anglophones et

étrangers) et de deux tiers des stimuli non langagiers (rires, de toux, bâillements, onomatopées). Tous les stimuli étaient des sons qui faisaient vibrer les cordes vocales. Les 20 autres blocks étaient des blocks de stimuli non vocaux composés de bruit environnementaux (eau, vent, etc.) de sons industriels (voiture, téléphone, sonnerie, instruments de musique) ainsi que de quelques vocalisations animales. Il y avait en tout 99 stimuli vocaux et 99 stimuli non vocaux.

Pour l'analyse des régions d'intérêt (i.e. les aires mises en évidence avec le *voice localizer*) avec le vocoder, ce qui correspond à la partie du protocole *event-related*, nous avons utilisé d'autres types de stimuli. Ceux-ci étaient divisés en 3 catégories : Mots, Voix et Non-voix. La première (« Mots ») était constituée de 26 (30 pour la partie « mesures comportementales ») mots anglais fréquents avec une durée moyenne de 0,9 sec (+/-SD : 0,05). La deuxième (« Voix ») était constituée de 26 (30 pour la partie « mesures comportementales ») stimuli vocaux « voix » c'est-à-dire de mots provenant d'une langue étrangère aux participants (i.e., finlandais et japonais). La durée moyenne des sons de cette catégorie était de 0,8 sec (+/-SD : 0,1). Enfin la troisième catégorie (« Non-voix ») était constituée de 26 (30 pour la partie « mesures comportementales ») sons issus de l'environnement. Ces sons de l'environnement étaient extraits du set de 55 stimuli « non-voix » du test de détection de la voix (voir le paragraphe « Stimuli », p152) et avaient une durée moyenne de 0,7 secondes (+/-SD : 0,04). Pour les 3 catégories de stimuli, la moitié d'entre eux (i.e. 13) étaient prononcés par une femme et l'autre moitié par un homme. Dans le cas des non-mots, les mots en japonais étaient tous prononcés par un homme et les mots en finlandais étaient tous prononcés par une femme. L'intensité des 3 catégories avaient été égalisée avec Matlab 6.5®. Ces 3 catégories de sons étaient présentées en condition originale ainsi que dans 4 conditions de vocoder, 2, 4, 8 et 16 canaux. Dans les analyses IRMf qui suivent nous avons regroupé les stimuli « Mots » et « Voix », et avons nommé ces groupes les stimuli « Mots et Voix ».

11.1.3 PROTOCOLE EXPERIMENTAL

11.1.3.1 MESURES COMPORTEMENTALES

A partir des stimuli décrits précédemment, 3 tâches à 2 choix forcés (2AFC) ont été proposées aux sujets, mais pour des raisons de pertinence avec notre étude, nous ne présenterons que deux d'entre elles. Pour la première (« voix versus non-voix »), les sons présentés au sujet étaient 30 stimuli « voix » et 30 stimuli « non-voix » dans les 5 conditions de dégradation (originale, 16, 8, 4 et 2 canaux de vocoder), soient 300 stimuli, dans un ordre aléatoire, et la tâche du sujet était de dire s'il s'agissait d'une voix où non. Pour la deuxième tâche (« mot versus non-mot »), on présentait au sujet les 30 stimuli « mots » et les 30 stimuli « voix », toujours dans les 5 conditions de dégradation (300 stimuli) et aléatoirement, et on lui demandait de dire si le son qu'il avait entendu était un mot ou non. La présentation des stimuli et l'enregistrement des réponses et temps de réaction, étaient faits avec le logiciel Matlab 2007b®. La durée totale de la partie comportementale était d'environ 40

minutes. Notons que pour éviter des effets de fatigue, les sujets étaient vus pour cette partie dans une durée de 1 à 15 jours après leur passage dans l'IRMf.

A partir des résultats obtenus, nous avons appliqué la théorie de détection du signal et nous avons pu calculer l'indice de sensibilité *d'* pour chaque tâche, chaque condition de dégradation et chaque sujet. Nous avons utilisé des tests non-paramétriques en raison du nombre restreint de mesures par condition.

11.1.3.2 VOICE LOCALIZER

Les sujets étaient placés dans l'IRM et les stimuli étaient présentés binauralement à une intensité de 85 à 90 dB SPL, au travers d'un casque qui procurait une atténuation sonore de 30 dB des sons extérieurs. Les sujets étaient placés dans le scan de l'IRM avec le casque sur les oreilles. Ils avaient reçu pour instructions d'écouter attentivement les sons présentés. Il s'agissait donc d'une écoute passive de blocks de 8 secondes, de sons vocaux, non vocaux ou de silence. Les blocks étaient suivis d'un délai de 2 secondes. L'ordre de la présentation, en seconde, était fixé par le Tableau 11.1. Les temps de stimulation non stipulés dans le tableau correspondaient aux blocks de silence (N = 20).

Voix	22	62	82	112	132	162	202	222	242	262	312	352	372	402	432	462	482	512	542	572
Non Voix	12	32	52	102	122	142	182	232	282	302	322	342	382	422	442	472	502	522	552	592

Tableau 11.1 : Ordre en seconde des stimuli Voix et Non voix du protocole *Voice Localizer*. Les temps manquants correspondaient au blocks de silence.

Il y avait donc au total 3 x 20 blocks (voix, non-voix et silence) de 8 secondes entrecoupés de 2 secondes de silence. Le début de chaque block était synchronisé avec le début du scan.

11.1.3.3 ACTIVATION DE L'AIRE DE LA VOIX AVEC LE VOCODER

Pour cette partie du protocole, nous avons utilisé une analyse factorielle complète, avec les 3 catégories de stimulus (mots, voix et non-voix) passés dans 5 niveaux d'intelligibilité : en condition originale ainsi qu'en vocoder à 2, 4, 8 et 16 canaux. Il y avait 26 stimuli par catégorie passés dans les 5 conditions de stimulation, ce qui faisait 390 sons par catégorie. Les stimuli de toutes catégories et tous niveaux d'intelligibilité était présentés dans un ordre pseudo-aléatoire dans deux sessions IRMf. La durée qui séparait chaque début de stimuli (SOA) était de 4 secondes. Pour chaque stimulus, un scan IRMf spécifique avait lieu pendant les 4 secondes à partir du début de la présentation du stimulus selon un protocole *event-related* classique. A ce paradigme était ajouté 20% d'événements nuls, c'est-à-dire où aucun stimulus n'était présenté. Ceci correspondait à la condition repos. Afin de maintenir l'attention des sujets dans cette écoute passive, nous avions introduit une tâche : pour chaque catégorie de stimuli, 2 stimuli sur 26 étaient répétés aléatoirement, et la consigne donnée au sujet était de détecter ces répétitions en appuyant sur le bouton d'un boitier.

Tout comme pour le *voice localizer*, les sujets étaient placés dans l'IRM avec le casque sur les oreilles qui procurait une atténuation sonore de 30 dB des sons extérieurs et les stimuli étaient présentés binauralement à une intensité de 85 à 90 dB SPL.

11.1.3.4 ACQUISITION DES IMAGES

Le scan des images IRM a eu lieu au centre de Neuroimagerie de l'Université de Glasgow et ont été réalisées avec un system IRM à 3T de Siemens. Les scans fonctionnels étaient acquis pour la totalité du plan (single-shot) avec une méthode d'écho planar imaging (EPI) avec séquence d'impulsion de radiofréquences ou RF (TR = 2s, TP = 30ms, flip angle = 77°, FOV = 215 mm, matrice = 128 x 128). Les 32 coupes axiales (résolution 3,75 x 3,75 mm dans le plan et 5 mm d'épaisseur) de chaque volume étaient alignées avec la ligne AC-PC (cingulaire antérieur-cingulaire postérieur). Un total de 310 volumes par sujet étaient acquis pour la partie *voice localizer* et 498 volumes pour la partie *event-related*. Apres le scanner fonctionnel, des images anatomiques pondérées en T1 étaient obtenues pour chaque participant (résolution de 1 x 1 x 1 mm).

11.1.3.5 PRETRAITEMENT DES IMAGES

Les traitements des images et les analyses statistiques ont été réalisés avec SPM5. Les mêmes prétraitements des images ont été réalisés pour les analyses *voice localizer* et *event-related*.

La première correction était une correction temporelle (*slice timing*) qui a consisté en une interpolation temporelle entre toutes les coupes d'acquisition d'un même volume, pour chaque sujet. Ensuite on procédait aux corrections spatiales. Pour chaque sujet, toutes les images étaient recalées (*realigned*) sur le premier volume acquis afin de corriger les mouvements de têtes inter scans. Ensuite, les images IRM anatomiques obtenues pour chaque sujet étaient mises en correspondance (*coregistration*) avec l'image fonctionnelle recalée et normalisée en utilisant les paramètres déterminés pour les images fonctionnelles. Puis, les images fonctionnelles étaient normalisées spatialement avec l'espace commun MNI (*normalisation*) afin de pouvoir procéder aux analyses de groupes. Enfin un lissage (*smoothing*) spatial était réalisé avec un filtre gaussien de 6 mm de la largeur à mi-hauteur (FWHM).

11.1.3.6 ANALYSES IRMF

Pour le *voice localizer*, le contraste « voix versus non-voix » était réalisé pour chaque sujet. Dans un deuxième temps, les contrastes étaient comparés pour l'ensemble du groupe par une corrélation. Les aires cérébrales issues de cette analyse étaient utilisées comme région d'intérêt (ROI). Une analyse de pic d'activation a ensuite été utilisée.

Pour la partie *event-related*, 15 contrastes par sujets ont été réalisés : pour chaque catégorie les 5 niveaux de dégradation (stimuli originaux, et 4 niveaux de vocoder) ont été comparés au repos. Ces contrastes ont été ensuite utilisés dans une analyse factorielle complète où un premier facteur

était la catégorie, avec 3 niveaux (mot, voix, non-voix), et l'autre facteur était la dégradation avec 5 niveaux (original, 16, 8, 4 et 2 canaux).

Pour chaque ROI révélée avec l'analyse du *voice localizer*, nous avons extrait les valeurs d'activation pour chaque pic de chaque cluster correspondant aux 15 contrastes de l'analyse *event-related* (3 catégories x 5 conditions de dégradation). Ensuite nous avons réalisé une ANOVA pour mesurer l'effet des catégories de sons et des niveaux de dégradation sur l'activité de chaque pic.

11.2 RESULTATS

Dans un premier temps nous allons présenter les résultats comportementaux. Puis, nous présenterons les résultats du *voice localizer* et l'effet de la simulation d'implant sur l'activité de l'aire de la voix. Enfin, bien que pour les mesures comportementales et IRMf les tâches demandées aux sujets soient quelque peu différentes (i.e., active versus passive), nous essayerons de corréler les performances comportementales avec l'activité de la voix en fonction du niveau de dégradation des stimuli.

11.2.1 RESULTATS COMPORTEMENTAUX

Pour les résultats comportementaux, 2 sujets sur 15 n'ont pu être revus à la suite de leur passage dans l'IRMf. Les résultats présentés ci-après portent donc sur les 13 autres sujets.

11.2.1.1 TACHE VOIX VERSUS NON-VOIX

Pour cette tâche on peut voir qu'en condition originale les performances des sujets étaient bonnes. Puis avec l'effet du vocoder les performances diminuaient de manière significative (Friedman : $p = 0{,}0054$). Toutefois les analyses post-hoc ont montré que cet effet était dû à la plus grande diminution des performances en condition de vocoder à 2 canaux. En effet seule la condition 2 canaux était significativement différente de toutes les autre (Wilcoxon, $p < 0{,}0343$). Notons que les performances en condition de vocoder à 8 canaux étaient significativement inférieures à celles de la condition originale. Mais elle n'était pas différente des conditions 16 et 4 canaux. Les analyses statistiques nous indiquent donc que les sujets étaient aussi bons en condition originale qu'en condition à 16 et 4 canaux.

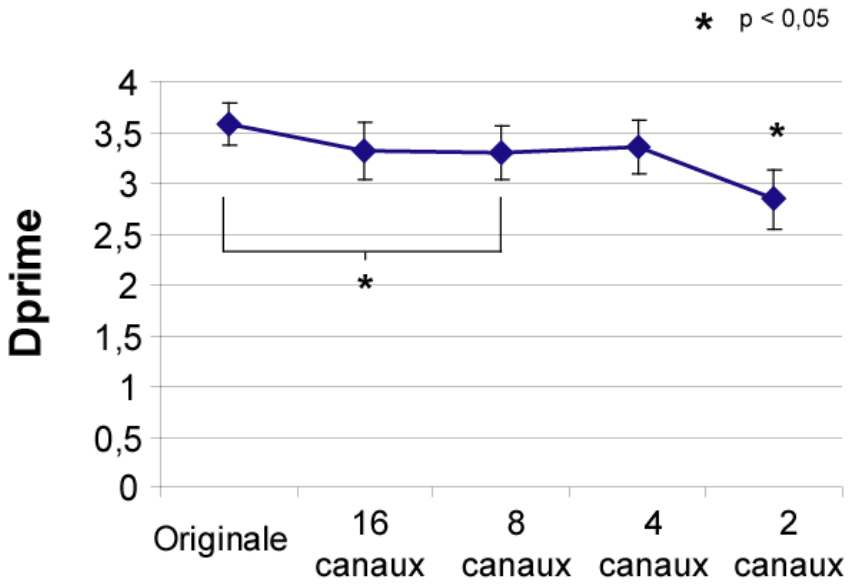


Figure 11.1 : Effet des conditions de dégradation sur les performances (*d'*) de la tâche « voix versus non voix ». L'effet global était significatif mais essentiellement dû à la condition de vocoder 2 canaux.

En comparaison des performances que nous avions obtenues dans la partie « Voix humaine versus sons environnementaux, stimulation binaurale » (p152, voir les Résultats, p163), on peut voir que les résultats des sujets sont particulièrement bons ici dans le même type de tâche. Alors qu'ils étaient un petit peu moins bons en condition originale ($\Delta d'$ = 0,72), les sujets étaient nettement meilleurs dans toutes les autres conditions avec des performances supérieures de 42, 66, 123 et 131%, respectivement pour les conditions 16, 8, 4 et 2 canaux de vocoder. Cette différence en faveur des sujets de l'IRMf peut être expliquée par la nature des stimuli. Les 30 stimuli « non voix » étaient issus du set de sons « non-voix » de la tâche de détection de la voix, donc à priori des stimuli très similaires, qui différaient seulement par leur durée. Ceux-ci étaient plus longs de 200ms par rapport à ceux de la tâche de détection de la voix. En ce qui concerne les stimuli « voix », dans la présente tâche il s'agissait de mot étranger (finlandais et japonais). Ces stimuli sont plus apparentés à de la parole (non sémantique) pour les sujets. Or nous avons déjà montré dans la partie « Test de détection de la voix humaine chez les patients (p151) que les performances de détection de la voix (versus des sons de l'environnement) sont bien meilleures pour les stimuli vocaux langagiers que pour des sons vocaux non langagiers. Ceci s'explique principalement en raison d'une information temporelle plus riche pour les sons vocaux langagiers. D'ailleurs lorsqu'on compare aux performances « voix langagières » des sujets de la tâche de détection de la voix en stimulation binaurale (voir paragraphe « Comparaison entre stimuli langagiers et non-langagiers », p156) on constate que la supériorité des patients ici présents n'était plus que de 9, 26, 64 et 61% (respectivement pour les conditions de vocoder à 16, 8, 4 et 2 canaux). Cette différence persistante peut s'expliquer par la plus grande durée des stimuli. Les sujets bénéficiaient de davantage d'information par le fait que les stimuli vocaux

étaient langagiers, mais ils en bénéficiaient également sur une durée plus longue. Ces deux arguments peuvent expliquer pourquoi ici les performances sont très peu sensibles au vocoder.

11.2.1.2 TACHE MOT VERSUS NON-MOT

Les résultats pour la tâche « mots versus non-mots », montrent que la dégradation des stimuli par le vocoder affectait significativement les performances (Friedman : $p < 0{,}0001$). Toutes les comparaisons deux à deux montraient une différence significative entre les performances (Wilcoxon : $p < 0{,}04$), excepté lors de la comparaison des conditions des vocoder à 16 et 8 canaux.

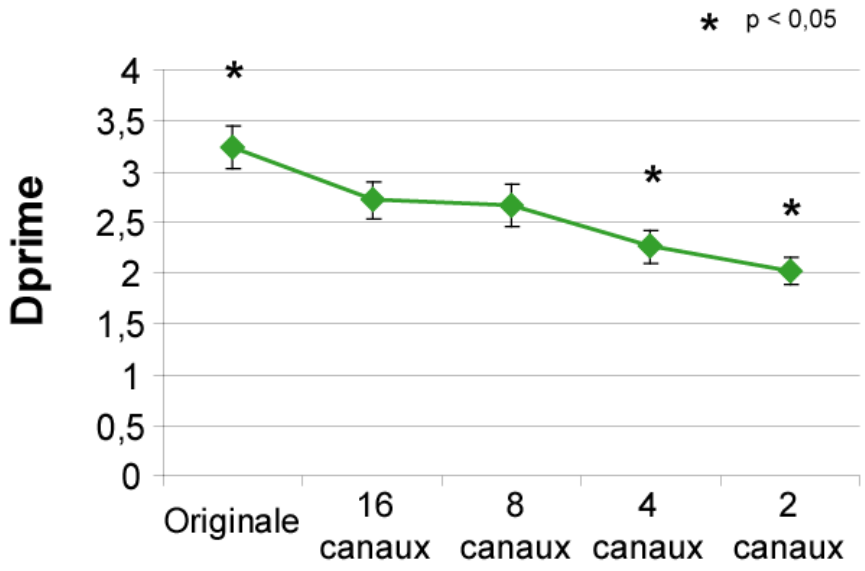


Figure 11.2 : Effet des conditions de dégradation sur les performances (*d'*) de la tâche « mot versus non-mots ». L'effet global était significatif et les analyses post-hoc ont montré que toutes les conditions étaient significativement différentes en comparaison deux à deux, sauf entre les conditions 16 et 8 canaux de vocoder.

Cette diminution des performances était probablement en lien avec la diminution de l'information spectrale puisque l'information temporelle était bien préservée par le vocoder. De plus, la distinction entre mot et non-mot doit être aidée par la compréhension du mot puisque les deux types de stimuli, « mot » et « non-mot » (les stimuli « non-mot » sont les stimuli « voix ») contiennent des informations langagières, donc des modulations temporelles probablement assez semblables. Ces résultats indiquent donc que la dégradation du signal par le vocoder à bien une incidence sur les performances.

11.2.2 VOICE LOCALIZER

Les résultats du *voice localizer* sont présentés sur la Figure 11.3. Sur cette figure on peut voir que 6 clusters ressortaient significativement plus activés par la voix par rapport aux sons de l'environnement. On peut également constater que les activations étaient très symétriques entre les deux hémisphères.

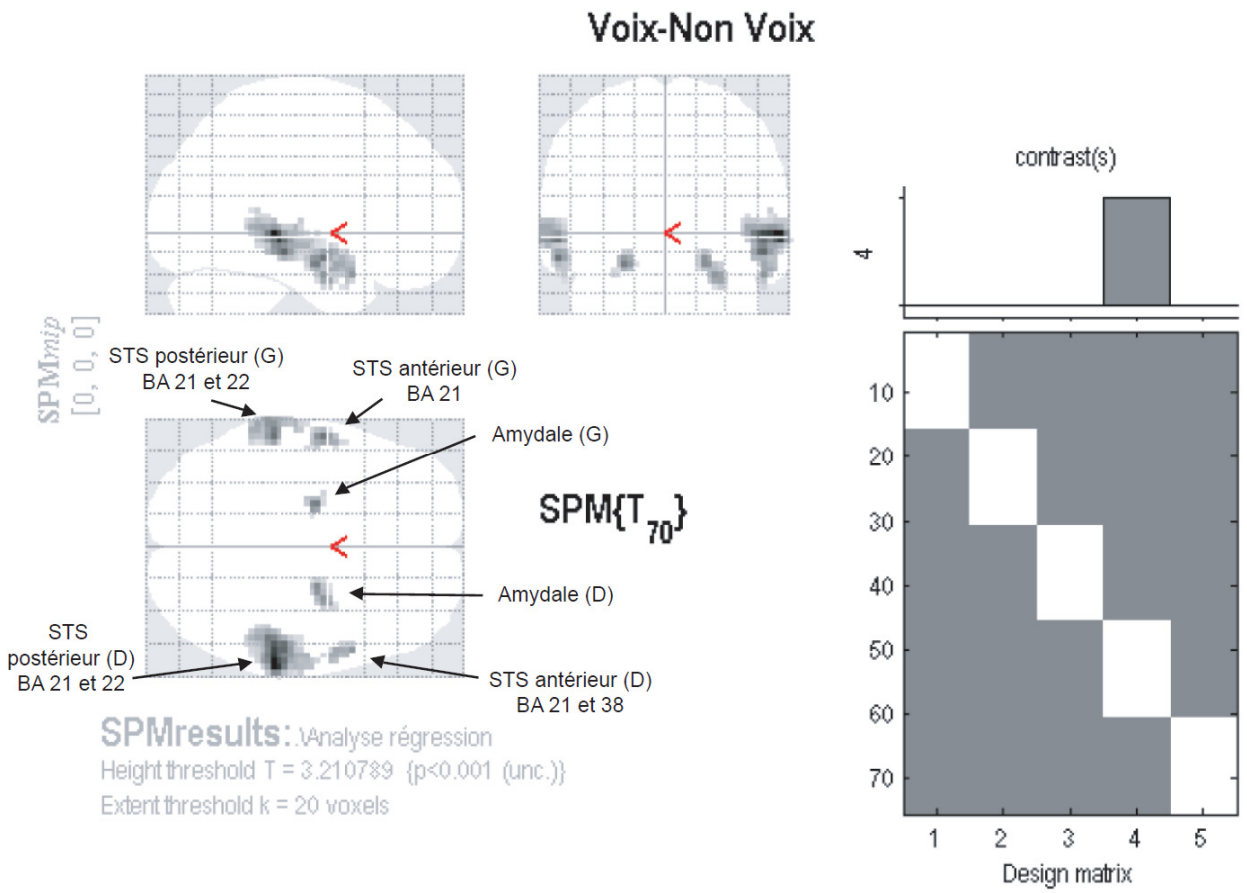


Figure 11.3 : Résultats du *voice localizer*, contraste voix versus non-voix. 6 clusters sont significativement activés par la voix (seuil non corrigés)

Tout d'abord on peut voir que, pour les deux hémisphères, l'amygdale était activée par le contraste « voix » versus « non-voix ». Pour l'hémisphère gauche, le pic et l'ensemble du cluster se trouvait dans l'amygdale. Pour l'hémisphère droit, le cluster était légèrement étendu à l'hippocampe (6,5 %) et le pic du cluster se trouvait dans cette zone. Toutefois, étant donné que la majorité du cluster se trouvait dans l'amygdale, nous parlerons du cluster de l'amygdale droite. Cette activation bilatérale de l'amygdale en réponse à des stimuli vocaux paraît cohérente avec des études antérieures, qui montre que l'amygdale est activée par des stimuli vocaux non linguistiques, tels que

des rires, des cris de peur, etc. (Fecteau et al., 2007; Scott et al., 1997). Or de tels stimuli avec un contenu émotionnel se trouvaient dans notre set de stimuli vocaux et ceci peut expliquer l'activation de l'amygdale en réponse aux stimuli vocaux.

Enfin dans chaque hémisphère deux autres clusters répondaient spécifiquement à la voix. Ces deux clusters étaient situés dans le lobe temporal, et à chaque fois dans le sillon temporal supérieur (STS). Dans l'hémisphère droit nous avons retrouvé une activation du gyrus temporal médian se situant sur les berges du STS postérieur, à cheval sur les aires de Brodmann (BA) 21 et 22. Un autre cluster a été identifié, toujours dans le STS mais un peu plus antérieur, dans les gyri temporaux médian et supérieur, à cheval entre les aires de Brodmann 21 et 38, se situant de part et d'autre du STS. Dans l'hémisphère gauche, on trouve un profil symétrique. Outre l'amygdale qui a déjà été évoquée plus haut, on trouvait un cluster dans le STS postérieur, plus précisément dans le gyrus temporal supérieur, à cheval entre les BA21 et BA22. Le deuxième cluster se situait dans le gyrus temporal médian, sur les berges du STS antérieur, essentiellement dans l'aire 21 de Brodmann.

Ces clusters d'activation identifiés dans le STS paraissaient cohérents avec ceux d'autres études. Par exemple en ce qui concerne l'activation du STS postérieur droit, le pic d'activation est quasiment le même (distance = 7 mm, *smoothing* = 6 mm) que celui identifié par Belin *et al.* en 2000 et le même que celui identifié par Fecteau *et al.* en 2004 et Coez *et al.* en 2004. Le pic du STS antérieur droit est également assez proche de celui identifié dans Belin *et al.* en 2002 (distance = 14 mm). Ces deux pics sont aussi proches de ceux identifiés dans d'autres études faisant le contraste « voix versus non-voix », tel que les études de Belin *et al.* en 2002 et Beliziare *et al.* (2007), avec des distances comprises entre 12 et 14 mm entre les pics. Dans l'hémisphère gauche on retrouve aussi des pics similaires à ceux d'autres études. Le pic du STS antérieur était le même que celui identifié dans Belin *et al.* (2000) et Coez *et al.* (2008). Le pic du STS postérieur était proche de celui de Belin *et al.* 2000 et 2002), de Fecteau *et al.* (2004), Beliziare *et al.* (2007) et Coez *et al.* (2008) avec des distances comprises entre 9 et 15 mm, toujours supérieures à la valeur du *smoothing*. Nous avons donc montré que les régions d'intérêt (ROI) que nous avons pu mettre en évidence sont très cohérentes avec celles révélées par d'autres auteurs (Belin, Zatorre et al., 2002; Belin et al., 2000; Coez et al., 2008; Fecteau, Armony, Joanette, & Belin, 2004), ce qui nous permet d'affirmer la pertinence de nos résultats. Les différences de distance observées entre nos résultats et ceux d'études antérieures sur le même sujet étaient de l'ordre de 1 à 1,5 cm. Ceci correspond typiquement à la variabilité observée entre les études de neuroimagerie, et peut refléter les différences individuelles entre les groupes ou peuvent simplement être dues à des différences dans le protocole expérimental, ou encore à différents paramètres d'analyses.

Région cérébrale	BA	Taille	p	Coordonnées du pic (Talairach)		
				x	*y*	*z*
Hémisphère droit						
STS postérieur	21 et 22	259	0,000	60	-30	0
STS antérieur	21 et 38	49	0,003	54	12	-21
Amygdale		55	0,001	21	-6	-15
Hémisphère gauche						
STS postérieur	21 et 22	113	0,000	-57	-30	-3
STS antérieur	21	65	0,000	-57	-3	-12
Amygdale		31	0,031	-21	-6	-15

Tableau 11.2 : Résumé des clusters activés par le contraste "voix versus non voix", dans les hémisphères droit et gauche. Sont également indiqués les aires de Brodmann correspondante (BA), la taille du cluster en nombre de voxel (Taille) et le seuil de significativité corrigé pour le cluster (p). Les 3 dernières colonnes indiquent les coordonnées du pic du cluster selon l'atlas de Talairach.

Nos résultats montrent donc que les pics d'activité spécifiques aux stimuli vocaux, se situaient dans les STS antérieur et postérieur bilatéraux (voir Figure 11.4) et dans l'amygdale bilatérale. La symétrie de ces pics entre les deux hémisphères était très forte pour l'amygdale et le STS postérieur, et un peu moins importante pour le STS antérieur, avec des distances respectives de 0 mm, 3 mm et 17 mm. La distance était calculée en prenant la valeur absolue pour l'axe des *x*, et en conservant les valeurs exactes pour les axes *y* et *z*. Etant donné que la distance pour le cluster était plus importante et supérieure au *smoothing* nous avons mesuré le recouvrement des clusters des STS antérieurs droit et gauche lorsque l'on projetait un hémisphère sur l'autre. Les analyses ont montré, que le cluster du STS antérieur droit, qui était plus petit que le gauche (75% du cluster du STS antérieur gauche), recouvrait 50 % du cluster du STS gauche. Ceci indique donc que même si la distance entre les pics était plus importante que pour les autres clusters, la correspondance spatiale était tout de même assez bonne.

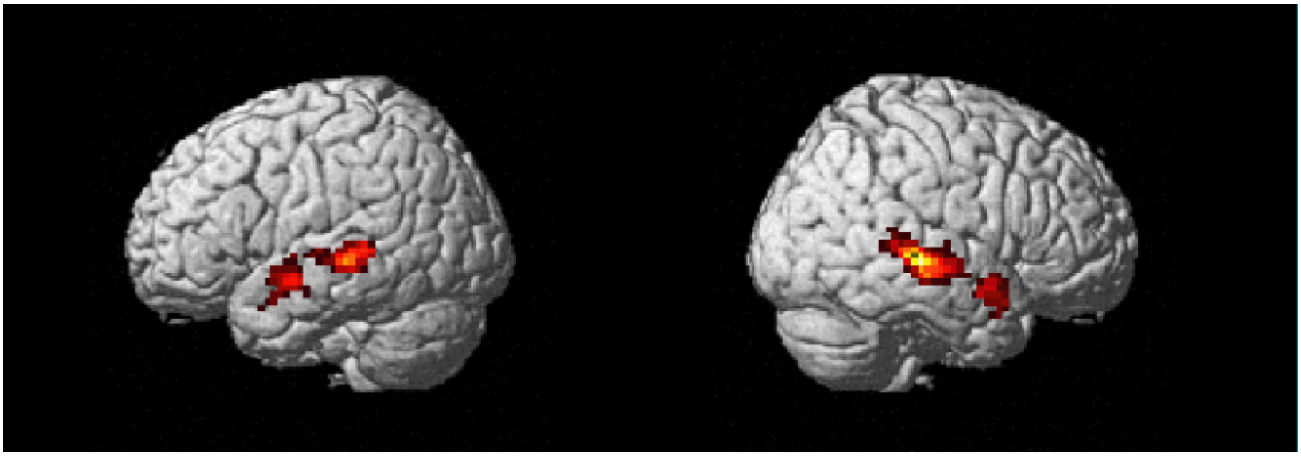


Figure 11.4 : Vue latérale des activations spécifiques à la voix (*voice localizer*) pour l'hémisphère droit (à droite) et l'hémisphère gauche (à gauche).

L'analyse suivante sera donc une analyse de variation du signal BOLD dans ces ROI, en fonction des différents niveaux de vocoder pour des stimuli vocaux.

11.2.3 EFFET DE LA SIMULATION D'IMPLANT PAR LE VOCODER

L'activation de l'aire de la voix par les stimuli « voix » et par les stimuli « mot » était statistiquement similaire, ce qui paraît évident étant donné qu'il s'agissait de deux stimuli vocaux. D'ailleurs une telle distinction entre les stimuli était pour les besoins de l'étude générale dont l'objectif était aussi d'évaluer l'effet du vocoder selon que les stimuli étaient des mots ou simplement de la voix. Nous avons donc choisi de regrouper les activations dues à ces deux types de stimuli dans nos analyses.

11.2.3.1 REPONSE AUX STIMULI ORIGINAUX

Sur la Figure 11.5 on peut voir l'effet des stimuli originaux « mot et voix » et « non-voix » sur la variation du signal BOLD, pour chacun des pics des clusters. Le résultat de l'ANOVA a montré qu'il n'y avait pas d'effet de l'interaction entre les différents pics et la catégorie de stimuli (« mot et voix » versus « non-voix »). En revanche il y avait un effet global de la catégorie de stimuli ($p = 0{,}0091$) et des différences de variation du signal entre les différents pics ($p = 0{,}0008$). Les analyses post-hoc ont montré notamment que les différences entre les pics se situaient entre la variation de signal BOLD du gyrus temporal médian et supérieur droit et les deux amygdales (droite et gauche), le gyrus temporal médian et supérieur gauche et le gyrus temporal médian droit. En ce qui concerne l'effet de la catégorie des stimuli on remarque que globalement les stimuli « mot et voix » provoquaient une plus grande variation de signal BOLD que les stimuli « non voix » (excepté pour l'amygdale Gauche). Cette observation paraît logique étant donné que ces ROI ont été définies parce qu'elles répondaient spécifiquement à des stimuli vocaux versus des sons non vocaux.

Toutefois la différence de variation du signal BOLD entre stimuli « mot et voix » et stimuli « non voix » n'était significative que pour le pic se situant dans le gyrus temporal médian et supérieur gauche. Toutefois, pour les autres pics, excepté pour l'amygdale gauche, la tendance même non significative allait dans le même sens.

On peut tout de même s'étonner de constater que, notamment dans le STS antérieur droit et les amygdales droite et gauche, la différence de variation de signal entre stimuli « mot et voix » et « non-voix » soit très faible. D'ailleurs pour ces 3 clusters, la variation de signal BOLD en réponse à des stimuli vocaux (« mot et voix ») était assez faible, ce qui est surprenant étant donné que ces régions sont celles qui répondent spécifiquement à la voix chez les mêmes sujets. Cette variation peut provenir des stimuli utilisés, qui différaient dans le *voice localizer* et la partie *event-related* de l'étude. Les stimuli « non voix » étaient très semblables entre les deux parties et ne peuvent donc pas vraiment expliquer cette différence. En revanche, les stimuli vocaux étaient différents. En effet, dans le *voice localizer* les stimuli vocaux étaient beaucoup plus diversifiés et donc plus riches en termes de modulations spectro-temporelles, puisqu'ils étaient composés de stimuli langagiers (mot, mots étrangers, syllabes, etc.) et non langagiers (toux, bâillement, rire, etc.). Le set de stimuli « mot et voix » utilisé dans la partie *event-related* pour tester la simulation d'implant par le vocoder, n'était composé que de mots et de mots étrangers. Le fait que les stimuli « mot et voix » de cette partie IRMf-vocoder soient plus pauvres et terme de contenu spectral et temporal pourrait expliquer cette moindre activation des clusters en réponse à ces stimuli.

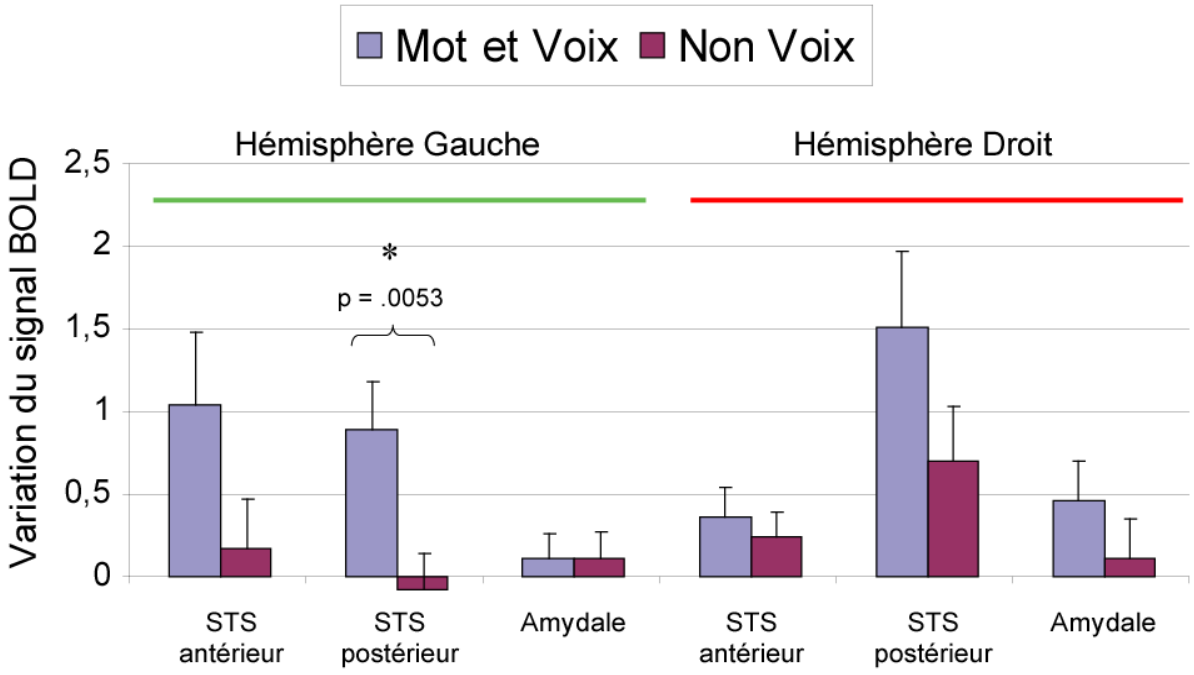


Figure 11.5 : Effet des stimuli « mot et voix » et des stimuli « non voix » sur la variation du signal BOLD (+/-SE) pour les pics des différents clusters. L'ANOVA à mesures répétées montre un effet significatif de la catégorie (« mot et voix » et « non voix ») et une différence significative de variation du signal entre les pics mais pas d'effet d'interaction entre ces deux paramètres. Les analyses post-hoc montraient la différence entre les variations du signal BOLD des stimuli « mots et voix » et « non voix » n'était significative que pour le pic du cluster se trouvant dans le STS postérieur gauche.

11.2.3.2 CONTRASTE VOIX-NON VOIX ET VOCODER

Une seconde analyse a été de mesurer la différence de variation de signal BOLD entre la condition « mot et voix » et « non-voix », et ce dans les différentes conditions de dégradation (voir Figure 11.6). L'objectif de cette analyse était de vérifier si l'aire de la voix identifiée par contraste entre des stimuli vocaux et non vocaux dans le *voice localizer*, révélait un contraste équivalent entre ces deux types de stimuli, mais dans des conditions de vocoder. Pour cela, nous avons moyenné le signal obtenu (variation de signal BOLD) lors de l'écoute des stimuli « mot » et lors de l'écoute des stimuli « voix », et cela pour chaque sujet à chaque pic des 6 clusters identifiés. Ensuite, a été soustrait à cette moyenne de variation de signal « mot et voix » la variation d'activité du signal BOLD lors de l'écoute des stimuli « non voix ». Ainsi, nous avions un contraste « mot et voix » versus « non voix » pour chaque patient et chaque pic de cluster.

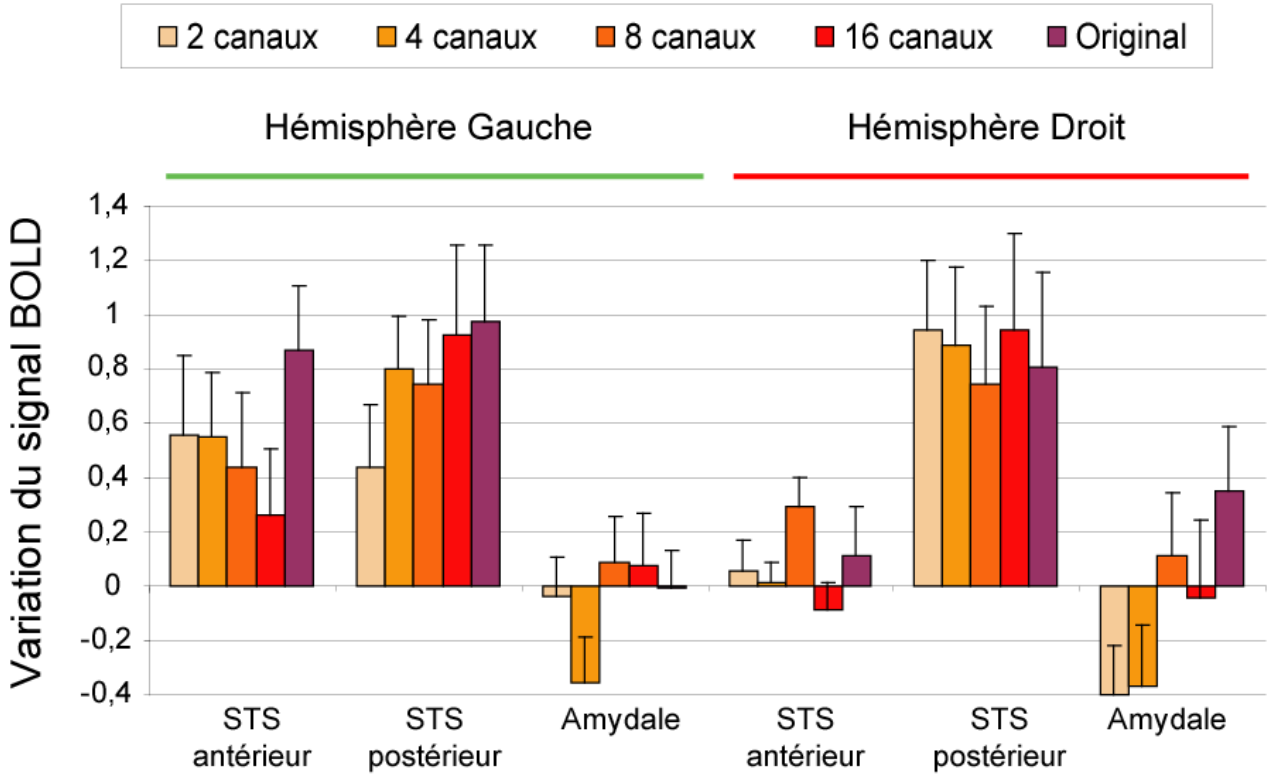


Figure 11.6 : Contraste de variation du signal BOLD (+/- SE) entre les conditions de stimuli « mot et voix » et « non voix » pour chaque pic de cluster. L'ANOVA à mesures répétées montre une différence significative de contraste entre les pics, mais aucun un effet significatif du niveau de dégradation.

Les résultats de cette analyse montrent que pour les STS postérieurs gauche et droit, ainsi que pour le STS antérieur droit, le contraste entre « mot et voix » et « non voix » est toujours positif, pour toutes les conditions de vocoder et la condition originale. Ceci qui signifie que, au moins pour ces différents clusters, la variation de signal pour les stimuli « mot et voix » était toujours plus importante que pour les stimuli « non voix ». Pour les amygdales droite et gauche ce contraste était

soit négatif soit proche de 0, signifiant une plus grande activation ou une activation similaire pour les stimuli « non voix » que pour les stimuli « mot et voix ». Le fait que certains pics aient une réponse aux stimuli vocaux assez faible par rapport aux stimuli non vocaux semble étonnant. En effet, ces pics ont été définis par le même contraste, voix versus non voix, mais avec des stimuli différents du *voice localizer*. Comme cela a déjà été discuté dans le paragraphe précédant, cette différence provient probablement des différents stimuli utilisés.

Les analyses statistiques révèlent donc une différence de contraste entre les différents pics. Mais pour chaque pic, le contraste entre « mot et voix » et « non voix » n'était significativement pas affecté par la dégradation du signal par le vocoder. Ceci nous indique également que le vocoder n'a pas d'effet sur la différence d'activation entre les stimuli vocaux et non vocaux. Nous pouvons donc affirmer qu'en condition de vocoder, ces aires étaient capables, d'un point vu statistique, de révéler un contraste similaire à celui qu'on trouve en condition originale. En condition de vocoder, l'aire de la voix, est donc toujours sensible à la différence entre les stimuli vocaux et non vocaux qu'en condition originale. Cette seconde analyse nous amène donc à dire que la dégradation du signal par le vocoder n'affecte pas les propriétés fonctionnelles de l'aire de la voix.

11.2.3.3 REPONSE AUX STIMULI DEGRADES PAR LE VOCODER

Dans un troisième temps nous avons regardé l'effet de la manipulation du vocoder sur la variation du signal BOLD dans les 6 clusters identifiés par le *voice localizer* (voir Figure 11.7). Lorsqu'on stimule les ROI, on peut constater que globalement, excepté pour l'amygdale gauche, la variation du signal BOLD était plus grande pour les stimuli orignaux que pour toutes les conditions de vocoder. De plus on peut remarquer que pour les STS postérieurs droit et gauche ainsi que pour l'amygdale droite, plus le signal était dégradé par le vocoder, plus la variation de signal était faible. Pour les 3 autres pics, un tel effet n'était pas observé ; la variation de signal entre les différentes conditions de vocoder était similaire. L'ANOVA à mesures répétées à montré que l'effet des niveaux de dégradation n'était pas significatif. En revanche il y avait des différences de variation de signal entre les pics des différents clusters. Ceci s'illustre par une variation de signal moins grande en réponse aux stimuli « mot et voix » pour certains clusters que pour d'autres.

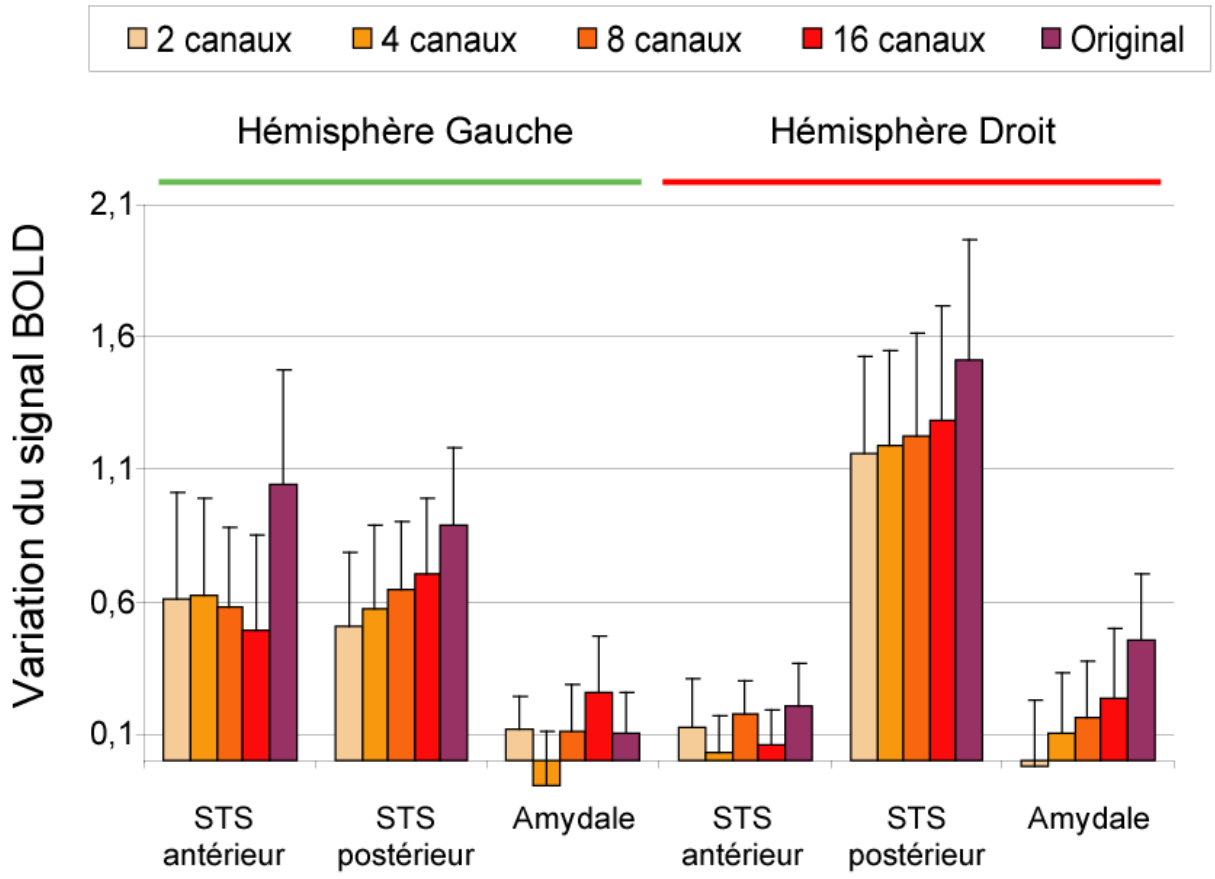


Figure 11.7 : Effet des différents niveaux de dégradation (original, 2, 4, 8 et 16 canaux de vocoder) sur la variation du signal BOLD (+/-SE) des pics des différents clusters lors de la stimulation avec les stimuli « mot et voix ». L'ANOVA à mesures répétées montre une différence significative de variation du signal entre les pics, mais aucun un effet significatif du niveau de dégradation.

Il ressort toutefois de cette analyse un point intéressant ; il semble que lors de la stimulation des ROI (révélées par le *voice localizer*) par les stimuli vocaux (« mot et voix »), la variation de signal BOLD des STS antérieurs et postérieurs soit asymétrique entre les deux hémisphères. En effet, on peut remarquer que pour le STS antérieur, la variation de signal BOLD était plus importante pour l'hémisphère gauche que pour l'hémisphère droit. Pour le STS postérieur, on remarquait le tableau inverse. Nous nous sommes donc posé la question de savoir si cette asymétrie entre STS postérieur et antérieur, droit et gauche était significative. Pour cela nous avons regroupé les valeurs de vocoder dans une seule et même condition, puisqu'il n'existait pas de différence significative entre les différentes conditions. Nous avons conservé la condition originale à part. Un test-t a montré que la différence entre STS postérieurs droit et gauche était significative pour le vocoder ($p < 0,0001$) et qu'elle était à la limite de la significativité pour la condition originale ($p = 0,0516$). Cette différence de significativité entre les deux tests était probablement due à un nombre d'observations supérieur pour la condition vocoder (N = 4 conditions de vocoder x 15 sujets, contre N = 15 sujets x 1 condition originale). Si on moyennait pour chaque sujet les valeurs de variation de signal BOLD obtenues pour toutes les conditions de vocoder des stimuli « mot et voix », la différence ne devenait plus significative ($p = 0,0992$). De plus la plus grande activité pour le STS postérieur droit par rapport au gauche, a déjà été montrée dans des études antérieures. Par exemple, en 2002, Belin *et al.* montraient que, autant pour les sons originaux vocaux linguistiques que pour les non linguistiques, l'activation était plus importante dans le STS postérieur (et antérieur) droit (Belin, Zatorre et al., 2002). Nous trouvons le même résultat pour le STS postérieur. Ceci confirme donc l'idée que si le nombre d'observations pour la condition originale était plus important la différence serait nettement significative. La plus grande variation de signal dans le STS antérieur gauche par rapport au droit en réponse à des stimuli vocaux était significative en condition originale ($p = 0,0353$) et en condition de vocoder ($p = 0,0001$). Cette asymétrie de l'activation du STS antérieur est plus contradictoire dans la littérature. Cette différence en faveur du STS antérieur gauche n'était pas retrouvée par Belin *et al.* (Belin, Zatorre et al., 2002; Belin et al., 2000). En revanche dans une publication plus récente en TEP, de Coez *et al.* (Coez et al., 2008), on pouvait retrouver une activité du STS antérieur gauche correspondante à celle que nous avons trouvé. Chez Coez *et al.*, cette activation très antérieure du STS existait dans l'hémisphère gauche mais pas dans l'hémisphère droit, où alors la partie antérieure du STS droit était bien plus postérieure qu'à gauche. Ces résultats sont en accord avec nos données. De plus nos résultats confirment que l'activation du STS antérieur était plus importante à gauche qu'à droite non seulement pour les stimuli originaux mais également pour les stimuli vocodés.

Cette asymétrie de variation du signal BOLD peut être due à la nature des stimuli que nous avons utilisés. En effet, pour le *voice localizer*, les stimuli vocaux utilisés contenaient aussi des stimuli vocaux non langagiers, tel que des rires, de la toux, des bâillements, etc. Or, Belin *et al.* ont montré que le STS antérieur droit répondait aussi fortement aux sons vocaux non langagiers. Notre set de stimuli utilisés pour étudier l'effet de la dégradation par le vocoder ne contenait pas de sons

vocaux non langagiers. Nos résultats vont donc dans le sens d'une asymétrie en ce qui concerne le traitement des stimuli vocaux langagiers.

11.2.3.4 CORRELATION AVEC LES DONNEES COMPORTEMENTALES

Nous avons également analysé la relation entre les performances pour les tâches comportementales (« mot versus non-mot » et « voix versus non-voix ») et les variations du signal BOLD.

Tout d'abord, aucune corrélation positive n'a été retrouvé en ce qui concerne la variation du signal BOLD et les performances lors de la tâche « mot versus non-mot », et ce, que l'on considère l'ensemble des mesures, ou chaque condition de dégradation indépendamment. En revanche, pour la tâche « voix versus non-voix », nous avons trouvé une corrélation positive et très fortement significative avec la variation de signal BOLD dans les STS postérieurs droit et gauche (voir Figure 11.8). La corrélation était positive mais n'était pas significative lorsqu'on considérait les conditions de dégradation indépendamment, et ce, en raison du plus petit nombre de valeurs. Mais étant donné qu'aucune différence d'activation, entre ces conditions, n'a été retrouvée, il était justifié de les prendre ensemble pour chercher une corrélation.

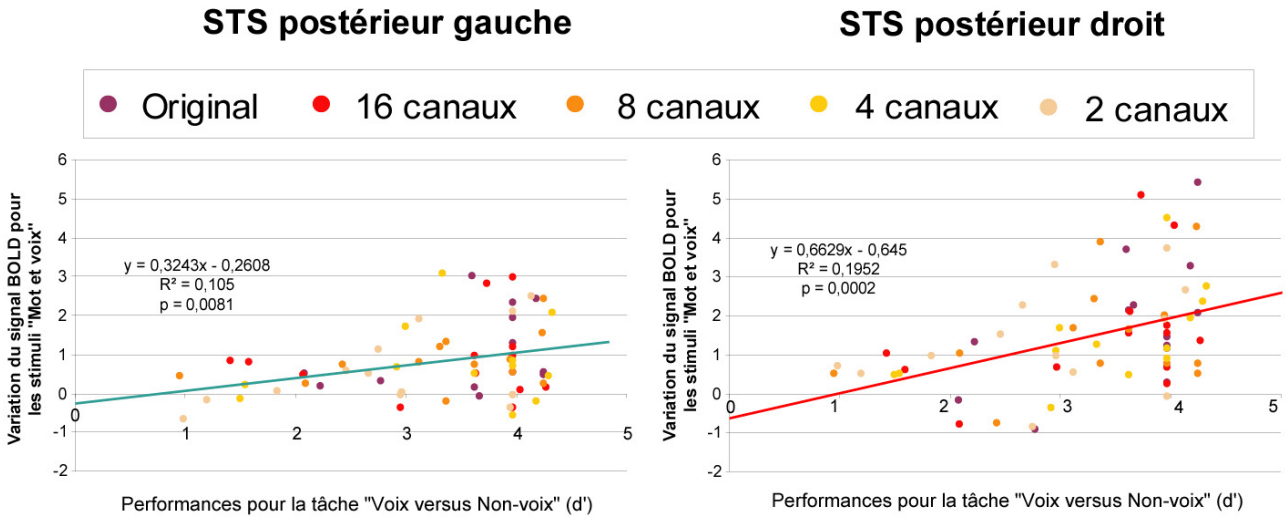


Figure 11.8 : Corrélation positive et significative pour les pics d'activité des STS postérieur gauche (à gauche) et droit (à droite), entre les performances comportementales lors de la tâche « voix versus non voix » (*d'*) et la variation du signal BOLD.

Ces résultats nous permettent de dire que probablement les deux STS postérieurs, droit et gauche, seraient impliqués dans la tâche de détection de la voix parmi des sons de l'environnement. Ce résultat est intéressant car la tâche pendant le scan de la caméra de l'IRM était une tâche passive d'écoute des stimuli. Or, la tâche comportementale était une tâche active où le sujet avait pour consigne de différencier les deux types de stimuli en répondant à l'aide d'une touche correspondant à sa réponse. On peut donc dire que les STS postérieurs bilatéraux étaient impliqués dans ces deux

types de tâche concernant le traitement de la voix versus des sons de l'environnement. Toutefois, ceci n'exclu pas l'implication d'autres aires dans le traitement de la tâche comportementale. Ces résultats démontrent la pertinence des aires définies par le *voice localizer*, et leur probable implication dans la détection de la voix, au moins en ce qui concerne le STS postérieur bilatéral.

11.3 DISCUSSION

En premier lieu de discussion il paraît important de souligner la correspondance entre l'aire de la voix mise en évidence dans notre étude avec celle d'autres études (Belin, Zatorre et al., 2002; Belin et al., 2000; Coez et al., 2008; Fecteau et al., 2004). Ceci corrobore donc la définition et la localisation de cette aire et valide la pertinence de nos données IRMf. En ce qui concerne la latéralité de cette aire, nos résultats ne vont pas dans le sens d'une préférence hémisphérique droite comme cela à parfois été retrouvé par les auteurs. Mais rappelons que cette latéralité n'était pas retrouvée systématiquement et, lorsqu'elle l'était, n'était pas toujours franche. Ceci corrobore les résultats comportementaux que nous avons obtenus au test de détection de la voix en stimulation monaurale (voir partie « Voix humaine versus sons de l'environnement, stimulation monaurale », p162). Les résultats ne montraient aucune différence selon que la détection de la voix était faite par l'oreille gauche ou par l'oreille droite, suggérant aucune facilitation du traitement par l'hémisphère droit.

Nos résultats montraient également une activation bilatérale de l'amygdale, qui n'avait pas été montré jusqu'ici par le *voice localizer*. Ce résultat est assez nouveau mais s'explique facilement par le contenu émotionnel (rire, cri de surprise, etc.) de nos stimuli vocaux. Cependant le même protocole de *voice localizer* a déjà été utilisé dans d'autre étude, avec les mêmes stimuli sans que cela ne révèle une activation de l'amygdale chez les sujets, du moins du point de vue de l'analyse de groupe. Il est possible que cette activation apparaisse d'un point de vue individuel chez certain patients mais chez un nombre insuffisant de sujets pour que cela ressorte sur l'analyse de groupe. On peut donc penser que dans notre étude, cette activation de l'amygdale bilatérale est apparue chez un suffisamment grand nombre de patients pour que cette activation ressorte significative pour le groupe. On remarque qu'en ce qui concerne les activations du STS, elles apparaissent systématiquement dans l'analyse de groupe pour les différentes études l'ayant utilisé.

Un autre point intéressant dans ces résultats est l'asymétrie dans la latéralité des réponses des STS antérieurs et postérieurs. Le STS antérieur était plus activé dans l'hémisphère gauche que dans l'hémisphère droit, et l'inverse était observé pour le STS postérieur qui était plus activé en réponse à des stimuli vocaux dans l'hémisphère droit. Ceci était vrai pour les stimuli originaux et pour les stimuli vocodés. L'asymétrie pour les stimuli originaux est assez cohérente avec les données de la littérature. En revanche l'effet des stimuli vocaux dégradés par le vocoder sur l'aire de la voix n'ayant encore jamais été étudié, les résultats sont très nouveaux. Il est intéressant de constater que l'asymétrie en faveur des STS antérieur gauche (par rapport au droit) et STS postérieur droit (par

rapport au gauche) est aussi vérifiée pour les stimuli dégradés par le vocoder. Ceci va donc dans le sens d'un effet similaire des stimuli vocodés et originaux sur l'aire de la voix.

Le résultat essentiel de cette analyse est bien sur l'absence d'effet significatif de la dégradation des stimuli par le vocoder sur l'activité de l'aire de la voix. Cette absence d'effet du vocoder a été démontrée par deux analyses différentes. Tout d'abord, le contraste de variation de signal entre les stimuli vocaux et non vocaux était similaire entre la condition originale et toutes les conditions de vocoder, suggérant un maintien de la sensibilité de l'aire de la voix en condition de vocoder comparable à celle observée en condition originale. Ensuite, nous avons montré que la variation de signal BOLD pour les stimuli vocaux n'était pas significativement affectée par le vocoder. Toutefois, pour cette dernière analyse, on a pu constater une tendance à la diminution de la variation de signal BOLD à mesure que l'on dégradait les stimuli par le vocoder pour le STS postérieur bilatéral et l'amygdale droite. Mais, probablement en raison d'une variation trop importante des valeurs, la différence n'était pas significative, même en comparant directement la condition de vocoder à 2 canaux à la condition originale.

De cette absence d'effet significatif, on peut conclure que le traitement du processeur de l'implant cochléaire est capable d'activer l'aire de la voix de façon similaire à ce que ferait un système auditif préservé. Ce résultat est très intéressant car, même si on ne peut pas complètement exclure l'idée que le processeur active de façon moindre l'aire de la voix, il est évident que les stimuli vocaux au travers du processeur sont capables d'activer cette aire. Si l'aire de la voix peut être sensible à la voix au travers du processeur de l'implant cochléaire, alors la question se pose de savoir pourquoi nous avons observé un déficit des patients dans le test de détection de la voix (voir paragraphe « Test de détection de la voix humaine », p209). La réponse à cette question est probablement reliée à des processus de réorganisation corticale. Des auteurs ont déjà montré que chez certains patients l'activation de l'aire de la voix pouvait être altérée (Coez et al., 2008). Les auteurs montraient que l'absence de sensitivité de l'aire de la voix était reliée à de faibles performances de reconnaissance de mots. L'étude avait consisté à mesurer l'activité de l'aire de la voix chez 3 groupes de sujets ; un groupe de patients normo-entendants, un groupe de patients ayant un bon niveau de reconnaissance de mots (>80%) et un autre groupe de patients ayant de mauvaises performances de compréhension de mots (<20%). L'activation du STS durant le *voice localizer* était significativement et bilatéralement diminuée chez le dernier groupe de sujet (mauvaises performances) comparé au groupe contrôle, tandis que le groupe de patients ayant de bonnes performances de reconnaissance de mots n'avait pas de différence significative avec le groupe contrôle. Notons toutefois qu'une corrélation de cette activité avec des performances de détection de la voix n'a pas été étudiée. Or nous avons montré au cours de ce travail qu'il y avait une dissociation de récupération entre cette aptitude et la compréhension du langage. En ce qui concerne la récupération de la compréhension du langage des études ont montré qu'elle était liée à des réorganisations impliquant l'aire de la voix. Il a notamment été montré que les patients implantés cochléaires s'appuyaient beaucoup sur les informations visuelles disponibles, telles que la lecture

labiale, pour compenser l'insuffisance de l'information auditive de l'implant (Rouger et al., 2008; Strelnikov et al., 2009). On sait également que ces habiletés visuelles sont préservés après l'implantation, permettant ainsi aux patients d'obtenir des performances presque parfaites lors de la compréhension audiovisuelle de mots (Rouger et al., 2007). Rouger *et al.* en 2007 ont montré dans une étude TEP, qu'une tâche de lecture labiale chez des patients récemment implantés (0 à 1 mois d'activation) induisait une activation de l'aire de la voix. Un an plus tard, alors que les patients avaient récupéré la compréhension du langage, cette aire de la voix n'était plus impliquée dans la tâche de lecture labiale. On peut donc penser que dans l'étude de Coez *et al.* les patients du groupe ayant de mauvaises performances s'appuyaient sans doute fortement sur la lecture labiale, ce qui impliquait une réorganisation crossmodale de l'aire de la voix, celle-ci répondant donc moins aux stimuli vocaux. A l'inverse, les patients du groupe qui avait de bonnes performances de reconnaissance de mots, sont sans doute moins dépendants de la lecture labiale. Celle-ci serait alors moins supportée par le STS, ce qui permettrait une activation quasi-normale de ce dernier en réponse à des stimuli vocaux. Ceci permettrait donc d'expliquer les différences d'activation de l'aire de la voix, observées chez les patients. Le signal transmis au travers du processeur de l'implant cochléaire serait donc capable d'activer l'aire de la voix, comme l'étude IRMf chez le sujet normal nous l'a montré, mais cette activation ne serait possible qu'à condition que l'aire de la voix ne soit pas « colonisée » par d'autres fonctions, comme la lecture labiale.

Cette réflexion pose toutefois une autre question. Comme l'étude de Coez *et al.* l'a montré, au moins certains patients sont capables d'activer l'aire de la voix de la même manière que les sujets normaux. Cette capacité était liée à une bonne récupération de la compréhension de la parole. Or, nous avons pu montrer dans ce travail que les patients avait un déficit important dans le test de détection de la voix, que ce déficit n'évoluait pas, et qu'il était présent même chez les patients ayant de bonnes performances en compréhension de langage (voir paragraphe « Test de détection de la voix humaine », p209). Bien que cela n'ait pas été testé dans l'étude de Coez *et a*l., on peut penser que probablement, mêmes les patients ayant de bonnes performances de reconnaissances de mots avaient également de mauvaises performances dans la détection de la voix. Comment expliquer ainsi que, au moins certains patients puissent avoir une activation normale de l'aire de la voix, alors que dans l'ensemble, les performances des patients dans un test de détection de la voix soient mauvaises ? Si ces résultats montrent que l'activation de l'aire de la voix peut être déficitaire, ou quasi normale pour d'autres patients, comment expliquer que leurs performances de détection de la voix soient pour tous, déficitaires, comme nous l'avons montré au cours de ce travail ? Dans le même sens de cette question, nous pouvons nous demander comment, les sujets normaux passés dans l'IRMf avaient une activation similaire de l'aire de la voix, dans une tâche d'écoute passive, alors qu'ils étaient déficitaires dans la tâche psychophysique (hors IRMf) de détection de la voix en condition 2 canaux de vocoder ? La réponse à cette question pourrait probablement venir d'une différence entre l'aspect perceptif et l'aspect décisionnel de la détection de la voix. En effet, ce que nous avons pu montrer, c'est que l'aire de la voix s'activait de façon similaire pour les stimuli originaux et pour les stimuli dégradés par le vocoder, lors d'une écoute passive. D'autre part, nos

résultats ont montré que dans une tâche active de détection de la voix, les performances se dégradaient avec la dégradation du signal par le vocoder. Il y a donc l'aspect décisionnel qui s'ajoute lors de la tâche de détection. Nous avons également montré une corrélation entre l'activation de l'aire de la voix les performances au test comportemental (STS postérieur bilatéral, voir précédemment « Corrélation avec les données comportementales », p301). Cette corrélation, comme cela a déjà été dit, indique que l'aire de la voix participe effectivement à la tâche de détection de la voix, mais cela n'exclut aucunement l'implication d'autres aires ou d'autres réseaux dans cette tâche. Ce pourrait donc être ces mêmes réseaux, qui seraient dysfonctionnels lors de conditions de signal dégradé par le processeur/vocoder et/ou en raison de réorganisations subséquentes à la surdité, dans la tâche de détection de la voix. Ce dysfonctionnement expliquerait ainsi, d'une part la diminution des performances avec la dégradation du signal par le vocoder ou l'implant cochléaire, et d'autre part que même si l'aire de la voix peut être « fonctionnelle » chez le patient et le normo-entendant en condition de vocoder, elle ne suffirait pas à permettre une bonne détection (décision) de la voix.

Cette idée de réseau impliqué dans la tâche de détection de la voix est corroborée par les résultats qui ont été obtenus dans l'étude générale par Strelnikov *et al.* (Strelnikov, Massida, Rouger, Barone, & Belin). Rappelons que notre étude IRMf s'inscrivait dans une étude plus large, réalisée en collaboration avec Kusma Strelnikov, Julien Rouger, Pascal Belin et Pascal Barone. Les objectifs généraux de l'étude étaient ; (1) de savoir si les aires affectées par la dégradation du vocoder étaient les mêmes selon que les stimuli étaient vocaux langagiers, vocaux non langagiers ou des sons de l'environnement, (2) et de savoir si les réseaux impliqués dans les traitements des sons originaux étaient les mêmes que ceux impliqués dans le traitement des sons dégradés par le vocoder. Les résultats de l'étude montrent que l'activation du gyrus temporal supérieur (STG) par les sons vocaux langagiers (« mots ») et non langagiers (« voix »), tous deux dégradés par le vocoder, étaient statistiquement les mêmes (voir Figure 11.9). Cette stimulation du STG était significativement plus importante pour les stimuli vocaux (« mots et voix ») que pour les sons de l'environnement (« non voix »). Cela signifiait donc que l'activité du STG était plus importante pour traiter la voix humaine dégradée que les sons de l'environnement dégradés et que la sélectivité des ces aires à la voix humaine survivait même à des hauts niveaux de dégradation par le vocoder. Ceci va donc dans le même sens que ce que nous avons trouvé au niveau de l'aire de la voix.

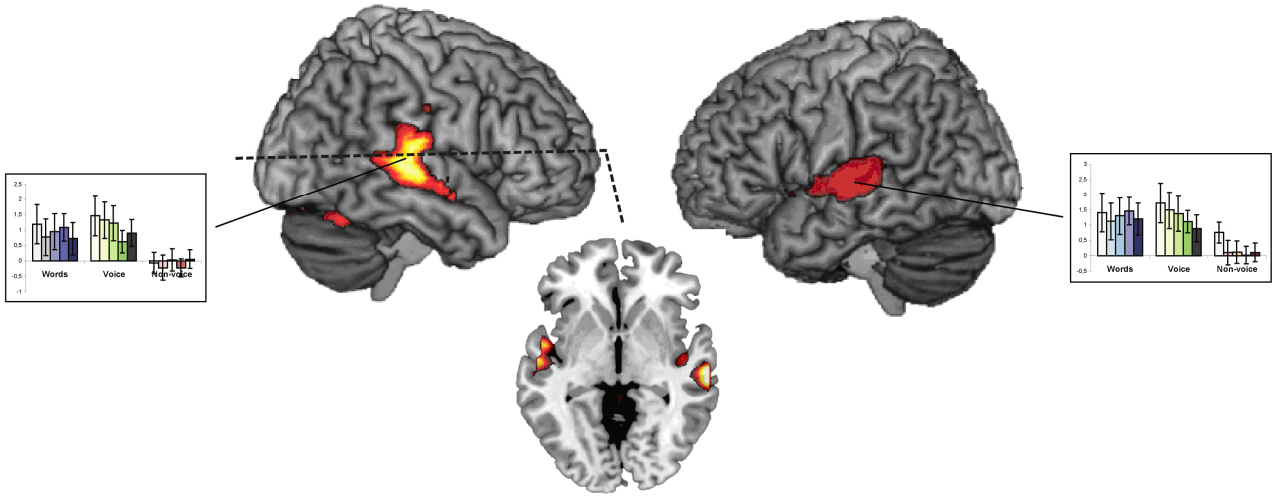


Figure 11.9 : Vues latérales (en haut) des activations spécifiques aux stimuli vocaux dégradés par le vocoder (« mots et voix ») versus les stimuli non vocaux dégradés par le vocoder (« non voix »). La coupe horizontale (en bas) correspond à la coordonnée z = 3. L'ordre des groupes de diagrammes correspondait de gauche à droite aux stimuli « mot », « voix » et « non voix ». Pour chaque groupe de stimuli, la couleur la plus foncée correspondait au niveau de dégradation le plus élevé. Extrait de Strelnikov *et al.* (*in preparation*).

En revanche les stimuli dégradés par le vocoder et les stimuli originaux n'entrainaient pas la même activation cérébrale quand on les comparait : les stimuli originaux « mot et voix » entrainaient une plus grande activation du cortex frontal médian que les stimuli vocodés, tandis que ces derniers entrainaient une plus grande activation du STG bilatéral et du cervelet droit (voir Figure 11.10).

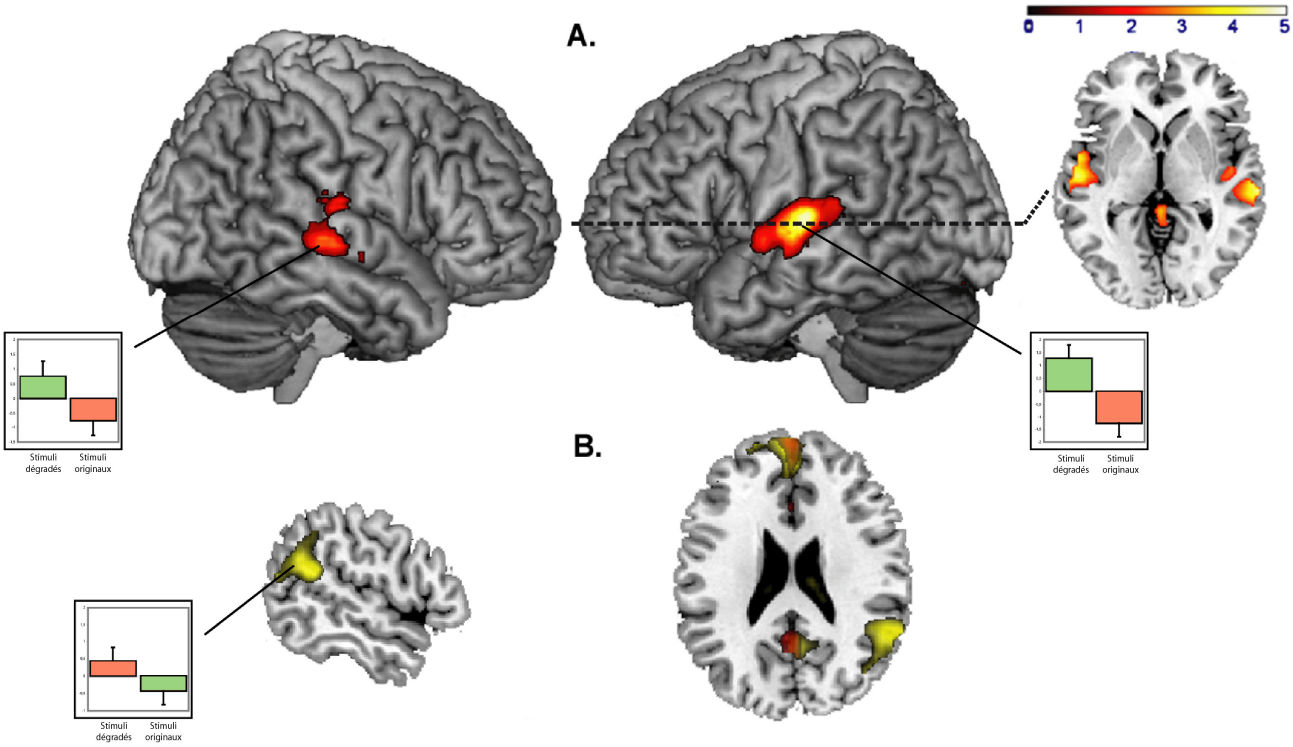


Figure 11.10 : **A.** Vues latérales du contraste pour les stimuli vocaux (« mot et voix ») dégradés versus originaux. **B.** Vue latérale droite et coupe horizontale du contraste pour les stimuli vocaux (« mot et voix ») originaux versus dégradés. Les activations colorés en rouge correspondent à ce que l'on retrouve pour les stimuli « mot » et celles colorées en jaunes à ce que l'on retrouve pour les stimuli « voix ». Extrait de Strelnikov *et al.* (*in preparation*)

Il semble donc au travers des résultats de Strelnikov *et al.* que les stimuli vocaux, qu'ils soient langagiers ou non, activent une ensemble d'aires cérébrales qui correspondrait à un réseau. Ce réseau serait donc différemment activé selon le niveau de dégradation. L'aire de la voix mise en évidence par Belin *et al*. en 2000 appartient très probablement à ce réseau. On peut donc penser que les aspects décisionnels présents lors de la tâche comportementale de détection de la voix correspondent à une partie de ce réseau, mais bien distincte de l'aire de la voix. Ceci expliquerait pourquoi l'aire de la voix peut-être activée par des stimuli vocaux alors que les performances des sujets restent faibles, comme nous avons pu le montrer dans notre étude. D'ailleurs on peut remarquer que les études des lésions sur la phonagnosie concernaient des lésions très diverses et souvent assez variées (Garrido et al., 2009; D. R. Van Lancker et al., 1988; D. R. Van Lancker et al., 1989). Il y a là aussi de quoi corroborer l'hypothèse d'un réseau d'aires cérébrales impliqué dans la perception de la voix.

Un tel rôle de la réorganisation de l'aire de la voix (impliquant la lecture labiale ou d'autres habiletés) et d'autres réseaux impliqués, sur les performances des patients à détecter la voix, doit être mesuré finement, ainsi que l'évolution de ces deux paramètres au cours de l'expérience avec l'implant, afin de mieux comprendre ce phénomène. Nous espérons pouvoir poursuivre cette étude avec un protocole qui nous permettrait de répondre à plusieurs questions. La première serait, est-ce que des aires ou réseaux supplémentaires sont impliqués dans une tâche active de détection de la voix, par rapport à une tâche d'écoute passive. La seconde serait, est-ce que ces différents réseaux, perceptif et décisionnel, sont affectés par le signal du processeur de l'implant cochléaire. Nous avons déjà montré, que l'aire de la voix (tâche d'écoute passive) ne semblait pas affectée, l'idée serait donc de comparé ce résultat avec celui concernant d'autres aires potentiellement impliquées dans la tâche décisionnelle. Enfin, un dernier point intéressant, qui mériterait d'être approfondi, serait la question de l'effet de la déprivation auditive sur ces réseaux, et de l'évolution de leur fonctionnalité lors de la récupération auditive au travers de l'implant, en corrélant ces aspects avec des données comportementales. En matière de perception du langage, des auteurs ont déjà montré qu'à performances égales, les sujets normo-entendants et les patients implantés cochléaires n'activaient pas les même réseaux pour traiter l'information (voir « Effet de l'implantation cochléaire sur l'organisation neurofonctionnelle », p137). Ces différences de réseaux on plusieurs explications. Premièrement, on peut invoquer les réorganisations fonctionnelles, plus ou moins importantes, conséquentes à la surdité. Deuxièmement, il est aisé d'imaginer que de la dégradation du signal auditif par le vocoder, amènerait les patients à développer des stratégies compensatrices pour traiter l'information auditive, et ainsi activer, suractiver et/ou sousactiver, certains réseaux plus que d'autres.

12 DISCUSSION GENERALE ET CONCLUSION

12.1 SYNTHESE DES RESULTATS

Rappelons que l'objectif de ce travail était de faire un état des lieux de la perception de la voix chez les patients sourds implantés cochléaires ainsi que son évolution au cours du temps.

Tout d'abord, les tests comportementaux chez les sujets normo-entendants nous ont montré que le processeur était responsable d'un certain déficit de la perception de la voix. Il semble que ce déficit provienne essentiellement d'une mauvaise perception du timbre au travers de l'implant, liée à une résolution spectrale particulièrement limitée.

Toutefois les résultats chez les patients, nous ont amenés à des conclusions supplémentaires. Le déficit chez les patients semble être dû à d'autres facteurs que le signal transmis par le processeur, puisqu'à nombre de canaux similaires, les patients, même expérimentés, étaient moins bons que les normo-entendants naïfs vis-à-vis du signal. De plus le déficit était propre à la perception voix et ses aspects paralinguistiques, puisque les performances des patients en matière de compréhension du langage atteignait un niveau très satisfaisant avec l'expérience de l'implant, tandis que celles qui concernaient la perception de la voix stagnaient. Si le processeur de l'implant ne suffit pas à expliquer le déficit, on peut penser à l'influence de facteurs neurofonctionnels liés à la déprivation auditive et à la rééducation post-implantation.

L'étude IRMf ne nous a pas permis de corroborer l'idée selon laquelle ce déficit en matière de perception de la voix serait lié à une mauvaise ou sous-activation de l'aire de la voix par le signal du processeur de l'implant cochléaire. Ceci dit nous n'excluons pas qu'une mauvaise fonctionnalité de l'aire de la voix, qui serait alors plutôt liée aux réorganisations neurofonctionnelles subséquentes à la surdité. Les résultats de cette expérience nous amènent plutôt à penser que l'aire de la voix jouerait davantage un rôle de traitement ou d'extraction des informations vocales et serait impliquée dans un réseau, dont l'aspect décisionnel relèverait d'une autre région cérébrale. De plus comme nous l'avons évoqué, les patients sont susceptibles de développer des stratégies compensatrices pour palier à la dégradation du signal auditif par le processeur de l'implant, qui pourrait entrainer l'implication de réseaux différents pour les mêmes tâches.

En conclusion, on peut dire que le déficit des patients en matière de perception de la voix est lié au signal limité transmis par le processeur de l'implant, mais également aux stratégies de codage

utilisées par les processeurs, à l'absence d'une rééducation spécifique à la voix ainsi qu'à la surdité et les réorganisations neurofonctionnelle subséquentes.

12.2 DISCUSSION ET PERSPECTIVES

Au cours de ce travail, nous nous sommes intéressés à la perception de la voix au travers de l'implant cochléaire. Cependant on peut noter que les résultats présentés ici ne considèrent que certains aspects de la perception de la voix. En effet, nous avons actuellement testé la détection de la voix, une analyse assez bas-niveau selon le modèle de perception de la voix de Pascal Belin. Nous avons également testé la catégorisation du genre et de l'âge ainsi que la discrimination de voix, qui correspond à la perception de l'identité de la voix dans le dit modèle.

Nous espérons à l'avenir, pouvoir finir l'acquisition des données chez les patients pour ces tests, mais également, nous intéresser à d'autres aspects de la perception de la voix. Toujours dans la perception de l'identité, nous souhaitons particulièrement nous intéresser à la perception de la familiarité et à l'impact de la familiarité d'une voix sur la compréhension du langage, ou la capacité discriminative de voix. Plus précisément, nous souhaitons étudier l'effet de la familiarité, donc de la mémoire d'une voix, sur les capacités perceptives relatives aux informations contenues dans cette voix. Nous nous attendons à observer des performances meilleures pour les voix familières aux patients.

Un autre aspect intéressant est la perception de l'information émotionnelle dans la voix. Dans ce but, l'acquisition de données a déjà été entamée dans le cadre de la thèse de doctorat de Mathieu Marx. Ce test, basé sur le modèle du test de catégorisation du genre, utilise un continuum réalisé par morphing entre deux émotions, la joie et la tristesse. La tâche des sujets est de catégoriser l'émotion perçue. Les données préliminaires montrent que les patients ont un déficit important dans ce type de tâche (voir Figure 12.1). De plus, ce déficit était plus marqué pour ce test que pour celui que nous avons obtenu pour la catégorisation du genre ou la détection de la voix, et cela même au-delà d'un an après l'activation de l'implant (Marx, 2009).

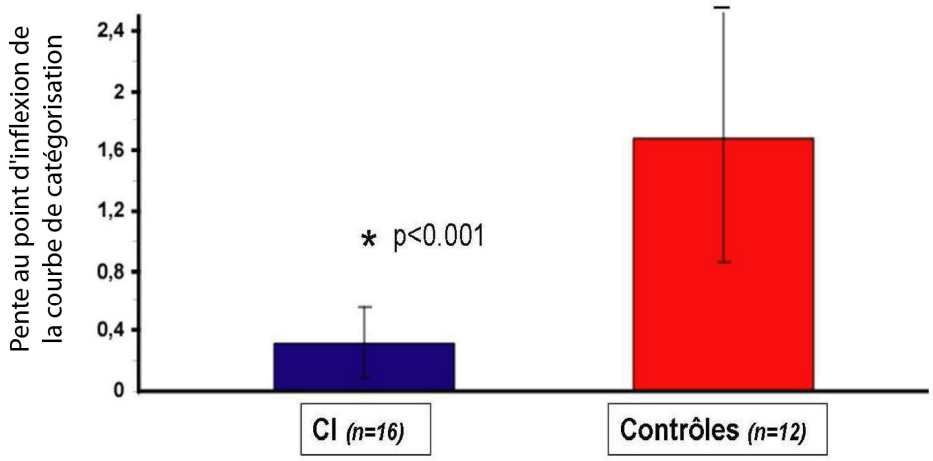


Figure 12.1 : Résultats préliminaires des patients implantés cochléaires (CI) et des sujets normo-entendants (contrôles) dans le test de catégorisation de l'émotion (joie-tristesse). Les valeurs correspondent aux moyennes (± SD) de la pente au point d'inflexion de la courbe de catégorisation. La différence entre les deux groupes était significative. Adapté d'après Marx, 2009.

La dimension dynamique du pitch va également être étudiée au travers de l'étude de la perception de la prosodie. Cela constitue l'essentiel du travail de Jessica Foxton déjà citée et post-doctorante au laboratoire. A terme nous espérons pouvoir déterminer dans quel domaine de perception de la voix, le déficit est le plus important chez les patients. Nous espérons également pouvoir corréler ce ou ces déficits à une mesure spécifique de perception des différents indices acoustiques indépendants de la voix, tels le pitch et le timbre.

L'effet de l'entrainement est également un point que nous souhaitons étudier, chez les patients implantés et chez les normo-entendants en condition de simulation d'implant. Nous espérons ainsi confirmer que les stratégies adaptatives dépendent bien de l'expérience de ce type de signal. De plus nous souhaitons étudier chez les patients si une rééducation spécifique de la perception de la voix peut améliorer les performances par rapport aux rééducations classiques. Si oui cela nous permettrait de savoir si l'amélioration est liée à une meilleure perception des indices acoustiques ou, comme on s'y attend, au développement de stratégies adaptatives basées sur la perception d'autres indices mieux perçus tels que le rythme, ou l'amplitude.

Dans ce travail, mais aussi au travers de la littérature, nous avons également vu la part importante que tiens le pitch pour la perception de la voix. Les études rapportent qu'il est le premier élément acoustique permettant de discriminer des voix et particulièrement de genre ou d'âges différents. Nos résultats mettent également en avant le fait que, au moins en ce qui concerne les capacités de catégorisation du genre, le pitch est suffisamment bien perçu pour réaliser la tâche. Toutefois, ce résultat n'indique pas quelles sont les capacités exactes et spécifiques à la perception du pitch, lorsque la tâche porte uniquement sur la perception de celui-ci. Mathieu Marx dans son

travail d'acquisition des données, évalue donc les performances des patients en matière de perception du pitch afin de savoir s'il existe une corrélation entre ces performances et celles concernant la perception d'autres aspects de la voix.

Egalement, on se doit de souligner que dans ce travail, la perception des patients n'est évaluée qu'en condition auditive seule. Or, les patients et les normo-entendants sont exposés en continu à des stimuli multisensoriels et plusieurs études ont déjà montré l'influence considérable qu'a la modalité visuelle sur la modalité auditive chez les patients sourds. En cela, le travail de doctorat de Julien Rouger sur la perception audiovisuelle de la parole chez les patients implantés cochléaires (Rouger, 2007), a conduit à plusieurs publications sur le sujet (Rouger et al., 2008; Rouger et al., 2007; Strelnikov et al., 2009). Il nous parait donc indispensable, en plus de la description détaillée des performances auditives seules des patients, de s'intéresser à la perception multimodale et notamment audio-visuelle. L'objectif de ces mesures est de voir à quel point la vision influence la perception auditive de la voix. Dans le cadre de son Master recherche de Neuropsychologie, Laure Chambaudie, actuellement doctorante au CerCo, a étudié l'influence d'un visage sur dans la tâche de catégorisation du genre. Ses résultats ont montré que les performances des patients étaient biaisées par la présentation d'un visage masculin ou féminin simultanément à la présentation du stimulus auditif. Le biais perceptif était bien entendu du côté du genre du visage qui était présenté. Ce biais était présent même lorsque la consigne était donnée aux patients de catégoriser la voix sans faire attention aux visages (Chambaudie, 2010). Nous espérons donc ainsi quantifier l'impact de la modalité visuelle en ce qui concerne la perception de la voix et comparer ces données à celles concernant la perception de la parole.

Enfin, la conclusion de ce travail reste toute de même un point d'interrogation en ce qui concerne la fonctionnalité de l'aire de la voix chez les patients implantés cochléaires. En effet, la meilleures façon de répondre à grand nombre de questions restées sans réponses jusque là est de pouvoir explorer ce domaine. Le projet de départ de ce travail incluait cet aspect expérimental, et pour des raisons techniques nous avons dû le mettre de côté momentanément. Nous avons bon espoir de pouvoir réaliser ce projet en TEP chez les patients implantés cochléaires. L'objectif serait d'évaluer l'activité de l'aire de la voix lors d'un protocole TEP de *voice localizer* similaire à celui proposé par Pascal Belin en IRMf, mais également lors d'une tâche de détection active de la voix pendant le scan. Cette activité serait mesurée au moment de l'activation de l'implant et au-delà d'un certain temps correspondant généralement au moment où les patients atteignent un niveau plafond de performances dans les tests comportementaux. Ces données TEP pourraient être comparées à l'évolution hors scan des performances des sujets au cours de la rééducation. On pourrait également inclure deux groupes de patients ; l'un recevant une rééducation classique et l'autre recevant en plus un entrainement spécifique à la perception de la voix. Nous espérons ainsi montrer que le déficit de perception de la voix observé au niveau comportemental dans ce travail de thèse est corrélé à un déficit fonctionnel de l'aire de la voix et de l'éventuel réseau auquel elle appartient.

12.3 CONCLUSION

Pour conclure nous dirons que ce travail constitue une première avancée dans le domaine de la perception de la voix chez les patients implantés cochléaires, avec des résultats nouveaux. Mais surtout, ce travail de thèse, apporte une formidable quantité de pistes à explorer dans un travail ultérieur et particulièrement en ce qui concerne l'imagerie TEP chez les patients implantés.

13 REFERENCES

AFNOR. (1977). *Recueil des normes françaises de l'acoustique* (Vol. 1 (Vocabulaire), NF S 30-107).

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn Sci, 4*(7), 267-278.

Andrews, M. L., & Schmidt, C. P. (1997). Gender presentation: Perceptual and acoustical analysesof voice. *Jounal of voice, 11*(3), 307-313.

Aronovitch, C. D. (1976). The voice of personality: stereotyped judgments and their relation to voice quality and sex of speaker. *J Soc Psychol, 99*(2), 207-220.

Assal, G., Aubert, C., & Buttet, J. (1981). [Cerebral asymmetry and voice recognition (author's transl)]. *Rev Neurol (Paris), 137*(4), 255-268.

Assal, G., Zander, E., Kremin, H., & Buttet, J. (1976). [Voice discrimination in patients with cerebral cortical lesions]. *Schweiz Arch Neurol Neurochir Psychiatr, 119*(2), 307-315.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: a new approach to psychophysical scaling. *Am J Psychol, 84*(2), 147-166.

Bachorowski, J. A., & Owren, M. J. (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J Acoust Soc Am, 106*(2), 1054-1063.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: an approach for identification of genetic and nongenetic components. *Am J Hum Genet, 62*(2), 224-231.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *Am J Hum Genet, 67*(3), 755-758.

Baken, R. J., & Olikoff, R. F. (2000). *Clinical Measurement of Speech and voice*. San Diego: CA: Singular Publishing Group Thomson Learning.

Baskent, D., & Shannon, R. V. (2003). Speech recognition under conditions of frequency-place compression and expansion. *J Acoust Soc Am, 113*(4 Pt 1), 2064-2076.

Baskent, D., & Shannon, R. V. (2004). Frequency-place compression and expansion in cochlear implant listeners. *J Acoust Soc Am, 116*(5), 3130-3140.

Baumann, O., & Belin, P. (2008). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychol Res*.

Bavelier, D., Brozinsky, C., Tomann, A., Mitchell, T., Neville, H., & Liu, G. (2001). Impact of early deafness and early exposure to sign language on the cerebral organization for motion processing. *J Neurosci, 21*(22), 8931-8942.

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends Cogn Sci, 10*(11), 512-518.

Bavelier, D., & Neville, H. J. (2002). Cross-modal plasticity: where and how? *Nat Rev Neurosci, 3*(6), 443-452.

Bear, M. F., Connors, B. W., & Paradisio, M. A. (2002). *Neurosciences : A la découverte du cerveau* (2de ed.). Baltimore: Lippincot Williams & Wilkins.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., et al. (2006). Electrophysiological markers of voice familiarity. *Eur J Neurosci, 23*(11), 3081-3086.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci, 8*(3), 129-135.

Belin, P., McAdams, S., Thivard, L., Smith, B., Savel, S., Zilbovicius, M., et al. (2002). The neuroanatomical substrate of sound duration discrimination. *Neuropsychologia, 40*(12), 1956-1964.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Curr Opin Neurobiol, 16*(4), 391-399.

Benjamin, B. J. (1981). Frequency variability in the aged voice. *J Gerontol, 36*(6), 722-726.

Bennett, S. (1981). Vowel formant frequency characteristics of preadolescent males and females. *J Acoust Soc Am, 69*, 231-238.

Bergeson, T. R., Pisoni, D. B., & Davis, R. A. (2003). A longitudinal study of audiovisual speech perception by hearing-impaired children with cochlear implants. *Volta Rev., 103*, 347$^{\perp}$ 370.

Biever, D., & Bless, D. (1989). Frequency variability in the aged voice. *Journal of Gerontology, 36*, 722-726.

Bilecen, D., Seifritz, E., Scheffler, K., Henning, J., & Schulte, A. C. (2002). Amplitopicity of the human auditory cortex: an fMRI study. *Neuroimage, 17*(2), 710-718.

Birkett, P. B., Hunter, M. D., Parks, R. W., Farrow, T. F., Lowe, H., Wilkinson, I. D., et al. (2007). Voice familiarity engages auditory cortex. *Neuroreport, 18*(13), 1375-1378.

Bladon, A. (1983). Acoustic phonetics, auditory phonetics, speaker sex and speech recognition: a thread. In F. Fallside & A. Woods (Eds.), *Computer speech processing*. Englewood Cliffs: NJ: Prentice-Hall.

Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., et al. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiol Neurootol, 1*(5), 293-306.

Boddaert, N., Chabane, N., Gervais, H., Good, C. D., Bourgeois, M., Plumet, M. H., et al. (2004). Superior temporal sulcus anatomical abnormalities in childhood autism: a voxel-based morphometry MRI study. *Neuroimage, 23*(1), 364-369.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci, 8*(3), 389-395.

Bosworth, R. G., & Dobkins, K. R. (1999). Left hemisphere dominance for motion processing in deaf signers. *Psychol. Sci., 10*, 256.

Bosworth, R. G., & Dobkins, K. R. (2002a). The effects of spatial attention on motion processing in deaf signers, hearing signers, and hearing nonsigners. *Brain Cogn, 49*(1), 152-169.

Bosworth, R. G., & Dobkins, K. R. (2002b). Visual field asymmetries for motion processing in deaf and hearing signers. *Brain Cogn, 49*(1), 170-181.

Bregman, A. S. (1990). *Auditory scene analysis : The perceptual organization of sound*. Cambridge: Mass.

Brosch, M., Selezneva, E., & Scheich, H. (2005). Nonauditory events of a behavioral procedure activate auditory cortex of highly trained monkeys. *J Neurosci, 25*(29), 6797-6806.

Bross, M. (1979). Residual sensory capacities of the deaf: a signal detection analysis of a visual discrimination task. *Percept Mot Skills, 48*(1), 187-194.

Bross, M., & Sauerwein, H. (1980). Signal detection analysis of visual flicker in deaf and hearing individuals. *Percept Mot Skills, 51*(3 Pt 1), 839-843.

Brown, C. J., Abbas, P. J., Borland, J., & Bertschy, M. R. (1996). Electrically evoked whole nerve action potentials in Ineraid cochlear implant users: responses to different stimulating electrode configurations and comparison to psychophysical responses. *J Speech Hear Res, 39*(3), 453-467.

Brown, C. J., Abbas, P. J., & Gantz, B. (1990). Electrically evoked whole-nerve action potentials: data from human cochlear implant users. *J Acoust Soc Am, 88*(3), 1385-1391.

Brown, M., Perry, A., Cheesman, A. D., & Pring, T. (2000). Pitch change in male-to-female transsexuals: has phonosurgery a role to play? *Int J Lang Commun Disord, 35*(1), 129-136.

Brozinsky, C. J., & Bavelier, D. (2004). Motion velocity thresholds in deaf signers: changes in lateralization but not in overall sensitivity. *Brain Res Cogn Brain Res, 21*(1), 1-10.

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res Cogn Brain Res, 9*(3), 227-238.

Burns, E. M., & Campbell, S. L. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: examples of categorical perception. *J Acoust Soc Am, 96*(5 Pt 1), 2704-2719.

Busby, P. A., & Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *J Acoust Soc Am, 97*(4), 2603-2606.

Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci, 2*(8), 759-766.

Calmels, M. N., Saliba, I., Wanna, G., Cochard, N., Fillaux, J., Deguine, O., et al. (2004). Speech perception and speech intelligibility in children after cochlear implantation. *Int J Pediatr Otorhinolaryngol, 68*(3), 347-351.

Carper, R. A., Moses, P., Tigue, Z. D., & Courchesne, E. (2002). Cerebral lobes in autism: early hyperplasia and abnormal age effects. *Neuroimage, 16*(4), 1038-1051.

Castellengo, m. (1987). Les sources acoustiques. In D. Mercier (Ed.), *Le livre des techniques du son* (pp. 45-70). Paris: Fréquences.

Chambaudie, L. (2010). *Perception multisensorielle des indices paralinguistiques : étude comparative psychophysique chez le sujet sourd postlingual implanté cochléaire et le sujet normo-entendant*.Unpublished manuscript, Université de Toulouse.

Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neurosci, 10*, 127.

Chatterjee, M., & Peng, S. C. (2008). Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. *Hear Res, 235*(1-2), 143-156.

Chatterjee, M., Peredo, F., Nelson, D., & Baskent, D. (2010). Recognition of interrupted sentences under conditions of spectral degradation. *J Acoust Soc Am, 127*(2), EL37-41.

Childers, D. G., & Wu, K. (1991). Gender recognition from speech. Part II: Fine analysis. *J Acoust Soc Am, 90*(4 Pt 1), 1841-1856.

Chouard, C. H., Fugain, C., Meyer, B., & Lacombe, H. (1983). Long-term results of the multichannel cochlear implant. *Ann N Y Acad Sci, 405*, 387-411.

Chouard, C. H., & MacLeod, P. (1976). Implantation of multiple intracochlear electrodes for rehabilitation of total deafness: preliminary report. *Laryngoscope, 86*(11), 1743-1751.

Ciocca, V. (1999). Evidence against an effect of grouping by spectral regularity on the perception of virtual pitch. *J Acoust Soc Am, 106*(5), 2746-2751.

Ciocca, V. (2008). The auditory organization of complex sounds. *Front Biosci, 13*, 148-169.

Clark, G. (2003). Cochlear implants in children: safety as well as speech and language. *Int J Pediatr Otorhinolaryngol, 67 Suppl 1*, S7-20.

Clark, G. M., Black, R., Forster, I. C., Patrick, J. F., & Tong, Y. C. (1978). Design criteria of a multiple-electrode cochlear implant hearing prosthesis [43.66.Ts, 43.66.Sr]. *J Acoust Soc Am, 63*(2), 631-633.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2005). Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. *J Speech Lang Hear Res, 48*(1), 204-223.

Coez, A., Zilbovicius, M., Ferrary, E., Bouccara, D., Mosnier, I., Ambert-Dahan, E., et al. (2008). Cochlear implant benefits in deafness rehabilitation: PET study of temporal voice activations. *J Nucl Med, 49*(1), 60-67.

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *J Speech Hear Res, 14*(3), 565-577.

Coleman, R. O. (1976a). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *J Speech Hear Res, 19*(1), 168-180.

Coleman, R. O. (1976b). Male and female voice quality and its relationship to vowel formant frequencies. *J Speech Hear Res, 14*, 565-577.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal behaviour, 65*(5), 997-1004.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

Deguine, O., Fraysse, B., Uziel, A., Cochard, N., Reuillard-Artieres, F., Piron, J. P., et al. (1993). Predictive factors in cochlear implant surgery. *Adv Otorhinolaryngol, 48*, 142-145.

Deguine, O., Garcia de Quevedo, S., Fraysse, B., Cormary, X., Uziel, A., & Demonet, J. F. (1995). Criteria for selecting the side for cochlear implantation. *Ann Otol Rhinol Laryngol Suppl, 166*, 403-406.

Delattre, P. (1965). La nasalit´e vocalique en fran¸cais et en anglais. *The French Review, 39*(1), 92-109.

Demonet, J. F., Thierry, G., & Cardebat, D. (2005). Renewal of the neurophysiology of language: functional neuroimaging. *Physiol Rev, 85*(1), 49-95.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Speech patterns heard early in life influence later perception of the tritone paradox. *Music Perception, 21*, 357-372.

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for a speech-related critical period. *J Acoust Soc Am, 119*(2), 719-722.

Deutsch, D., Le, J., Shen, J., & Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *J Acoust Soc Am, 125*(5), EL208-213.

Djourno, A., Eyries, C., & Vallancien, B. (1957). [Electric excitation of the cochlear nerve in man by induction at a distance with the aid of micro-coil included in the fixture.]. *C R Seances Soc Biol Fil, 151*(3), 423-425.

Dorman, M. F., Basham, K., McCandless, G., & Dove, H. (1991). Speech understanding and music appreciation with the Ineraid cochlear implant. *The Hearing Journal, 44*, 34-37.

Dorman, M. F., Gifford, R. H., Spahr, A. J., & McKarns, S. A. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol Neurootol, 13*(2), 105-112.

Dorman, M. F., & Ketten, D. (2003). Adaptation by a cochlear-implant patient to upward shifts in the frequency representation of speech. *Ear Hear, 24*(5), 457-460.

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am, 102*(5 Pt 1), 2993-2996.

Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., & Parkin, J. (1990). Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *J Acoust Soc Am, 88*(5), 2074-2079.

Dorman, M. F., Spahr, T., Gifford, R., Loiselle, L., McKarns, S., Holden, T., et al. (2007). An electric frequency-to-place map for a cochlear implant patient with hearing in the nonimplanted ear. *J Assoc Res Otolaryngol, 8*(2), 234-240.

Dowell, R. C., Martin, L. F., Tong, Y. C., Clark, G. M., Seligman, P. M., & Patrick, J. F. (1982). A 12-consonant confusion study on a multiple-channel cochlear implant patient. *J Speech Hear Res, 25*(4), 509-516.

Doyle, K. J., Mills, D., Larky, J., Kessler, D., Luxford, W. M., & Schindler, R. A. (1995). Consonant perception by users of Nucleus and Clarion multichannel cochlear implants. *Am J Otol, 16*(5), 676-681.

Dromey, C., Stathopoulos, E. T., & Sapienza, C. M. (1992). Glottal airflowand electroglottographic measures of vocal function at multiple intensities. *J Voice, 6*, 44-54.

Dudley, H. (1939a). The Automatic Synthesis of Speech. *Proc Natl Acad Sci U S A, 25*(7), 377-383.

Dudley, H. (1939b). The vocoder. *Bell labs record, 17*, 122-126.

Duvernoy, H. M. (1999). *The human brain: surface, three-dimensional sectional anatomy with MRI, and blood supply.* (2nd ed.). Vienna: Springer.

Eckel, H. E., & Sittel, C. (1995). Morphometry of the larynx in horizontal sections. *Am J Otolaryngol, 16*(1), 40-48.

Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput Biol, 5*(3), e1000302.

Ellis, H. D., Jones, D. M., & Mosdell, N. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol, 88 (Pt 1)*, 143-156.

Endres, W., Bambach, W., & Flosser, G. (1971). Voice spectrograms as a function of age, voice disguise, and voice imitation. *J Acoust Soc Am, 49*(6), 1842-1848.

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage, 47*(4), 1778-1791.

Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Curr Biol, 19*(12), 1028-1033.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage, 23*(3), 840-848.

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *Neuroimage, 36*(2), 480-487.

Fifer, W. P., & Moon, C. M. (1994). The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatr Suppl, 397*, 86-93.

Fine, I., Finney, E. M., Boynton, G. M., & Dobkins, K. R. (2005). Comparing the effects of auditory deprivation and sign language within the auditory and visual cortex. *J Cogn Neurosci, 17*(10), 1621-1637.

Finney, E. M., Clementz, B. A., Hickok, G., & Dobkins, K. R. (2003). Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. *Neuroreport, 14*(11), 1425-1427.

Finney, E. M., Fine, I., & Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nat Neurosci, 4*(12), 1171-1173.

Fishman, K. E., Shannon, R. V., & Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J Speech Lang Hear Res, 40*(5), 1201-1215.

Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: a study using magnetic resonance imaging. *J Acoust Soc Am, 106*(3 Pt 1), 1511-1522.

Fletcher, H. F. (1940). Auditory patterns. *Rev. Mod. Phys., 12*, 47-65.

Fletcher, H. F., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J Acoust Soc Am, 5*, 82-108.

Formby, C. (1985). Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners. *J Acoust Soc Am, 78*(1 Pt 1), 70-77.

Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron, 40*(4), 859-869.

Fritzell, B., Hammarberg, B., Gauffin, J., Karlson, I., & Sundberg, J. (1986). Breathiness and insufficient vocal fold closure. *J Phon, 14*, 549-553.

Fu, Q. J., Chinchilla, S., & Galvin, J. J. (2004). The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol, 5*(3), 253-260.

Fu, Q. J., Chinchilla, S., Nogaki, G., & Galvin, J. J., 3rd. (2005). Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. *J Acoust Soc Am, 118*(3 Pt 1), 1711-1718.

Fu, Q. J., Galvin, J. J., Wang, X., & Nogaki, G. (2005). Moderate auditory training can improve speech performance of adult cochlear implant users. *Acoustics Research Letters Online, 6*(3), 106-111.

Fu, Q. J., Hsu, C. J., & Horng, M. J. (2004). Effects of speech processing strategy on Chinese tone recognition by nucleus-24 cochlear implant users. *Ear Hear, 25*(5), 501-508.

Fu, Q. J., Nogaki, G., & Galvin, J. J., 3rd. (2005). Auditory training with spectrally shifted speech: implications for cochlear implant patient auditory rehabilitation. *J Assoc Res Otolaryngol, 6*(2), 180-189.

Fu, Q. J., & Shannon, R. V. (1999a). Effect of acoustic dynamic range on phoneme recognition in quiet and noise by cochlear implant users. *J Acoust Soc Am, 106*(6), L65-70.

Fu, Q. J., & Shannon, R. V. (1999b). Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing. *J Acoust Soc Am, 105*(3), 1889-1900.

Fu, Q. J., & Shannon, R. V. (2000). Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners. *J Acoust Soc Am, 107*(1), 589-597.

Fujiki, N., Naito, Y., Hirano, S., Kojima, H., Shiomi, Y., Nishizawa, S., et al. (1999). Correlation between rCBF and speech perception in cochlear implant users. *Auris Nasus Larynx, 26*(3), 229-236.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2005). Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogn Neurosci, 17*(10), 1578-1592.

Fujisaki, W., & Kashino, M. (2002). The basic hearing abilities of absolute pitch possessors. *Acoust. Sci. & Tech, 23*(2), 77-83.

Galifret, Y. (1991). Acoustique physiologique. In E. universalis (Ed.), *Encyclopaedia universalis* (Vol. 1, pp. 198-210).

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., et al. (2009). Developmental phonagnosia: a selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123-131.

Gelfer, M. P. (1988). Perceptual attributes of voices: Development and use of rating scales. *J Voice, 2*(4), 320-326.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J Voice, 19*(4), 544-554.

Gelfer, M. P., & Young, S. R. (1997). Comparisons of intensity measures and their stability in male and female speakers. *J Voice, 11*(2), 178-186.

Gendry Meresse, I., Zilbovicius, M., Boddaert, N., Robel, L., Philippe, A., Sfaello, I., et al. (2005). Autism severity and temporal lobe functional abnormalities. *Ann Neurol, 58*(3), 466-469.

George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., et al. (1996). Understanding emotional prosody activates right hemisphere regions. *Arch Neurol, 53*(7), 665-670.

Gervais, H., Belin, P., Boddaert, N., Leboyer, M., Coez, A., Sfaello, I., et al. (2004). Abnormal cortical voice processing in autism. *Nat Neurosci, 7*(8), 801-802.

Gfeller, K., Christ, A., Knutson, J., Witt, S., & Mehr, M. (2003). The effects of familiarity and complexity on appraisal of complex songs by cochlear implant recipients and normal hearing adults. *J Music Ther, 40*(2), 78-112.

Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S., & DeBus, B. (1998). Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults. *J Am Acad Audiol, 9*(1), 1-19.

Gfeller, K., & Lansing, C. R. (1991). Melodic, rhythmic, and timbral perception of adult cochlear implant users. *J Speech Hear Res, 34*(4), 916-920.

Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., et al. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Cochlear Implants Int, 3*(1), 29-53.

Gfeller, K., Witt, S., Adamek, M., Mehr, M., Rogers, J., Stordahl, J., et al. (2002). Effects of training on timbre recognition and appraisal by postlingually deafened cochlear implant recipients. *J Am Acad Audiol, 13*(3), 132-145.

Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., & Knutson, J. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann Otol Rhinol Laryngol, 111*(4), 349-356.

Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., & Knutson, J. F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear Hear, 18*(3), 252-260.

Giraud, A. L., Price, C. J., Graham, J. M., & Frackowiak, R. S. (2001). Functional plasticity of language-related brain areas after cochlear implantation. *Brain, 124*(Pt 7), 1307-1316.

Giraud, A. L., Price, C. J., Graham, J. M., Truy, E., & Frackowiak, R. S. (2001). Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron, 30*(3), 657-663.

Giraud, A. L., & Truy, E. (2002). The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia, 40*(9), 1562-1569.

Giraud, A. L., Truy, E., Frackowiak, R. S., Gregoire, M. C., Pujol, J. F., & Collet, L. (2000). Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. *Brain, 123 (Pt 7)*, 1391-1402.

Goldberger, Z. D. (2001). Music of the left hemisphere: exploring the neurobiology of absolute pitch. *Yale J Biol Med, 74*(5), 323-327.

Gonzalez, J. (2004). Formant frequencies and body size of speaker: A weak relationship in adult humans. *J. Phonetics, 32*, 277-287.

Gonzalez, J., & Carpi, A. (2004). Early effects of smoking on the voice: a multidimensional study. *Med Sci Monit, 10*(12), CR649-656.

Gorham-Rowan, M. M., & Laures-Gore, J. (2006). Acoustic-perceptual correlates of voices quality in elderly men and women. *Journal of Communication Disorders, 39*, 171-184.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., et al. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci, 8*(2), 145-146.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. *J Acoust Soc Am, 103*(5 Pt 1), 2677-2690.

Green, D., & Swets, J. (1966). *Signal Detection Theory and*

Psychophysics. New York: John Wiley and Sons.
Greenwood, D. D. (1990). A cochlear frequency-position function for several species--29 years later. *J Acoust Soc Am, 87*(6), 2592-2605.
Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluation of synthesized musical instrument tones. *J Acoust Soc Am, 62*, 454-462.
Griffiths, T. D., Green, G. G., Rees, A., & Rees, G. (2000). Human brain areas involved in the analysis of auditory movement. *Hum Brain Mapp, 9*(2), 72-80.
Griffiths, T. D., Rees, A., Witton, C., Shakir, R. A., Henning, G. B., & Green, G. G. (1996). Evidence for a sound movement area in the human cerebral cortex. *Nature, 383*(6599), 425-427.
Griffiths, T. D., Rees, G., Rees, A., Green, G. G., Witton, C., Rowe, D., et al. (1998). Right parietal cortex is involved in the perception of sound movement in humans. *Nat Neurosci, 1*(1), 74-79.
Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron, 65*(6), 852-858.
Guimaraes, I., & Abberton, E. (2005). Fundamental frequency in speakers of Portuguese for different voice samples. *J Voice, 19*(4), 592-606.
Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., et al. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci Lett, 348*(1), 13-16.
Hackett, T. A., Stepniewska, I., & Kaas, J. H. (1998). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol, 394*(4), 475-495.
Hadjikhani, N., Joseph, R. M., Snyder, J., & Tager-Flusberg, H. (2006). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cereb Cortex, 16*(9), 1276-1282.
Hajda, J. M. (1999). *The effect of time-variant acoustical properties on orchestral instrument timbres.* University of California, Los Angeles.
Hall, D. A. (2003). Auditory pathways: are 'what' and 'where' appropriate? *Curr Biol, 13*(10), R406-408.
Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cereb Cortex, 19*(3), 576-585.
Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42*(9), 1281-1292.
Hamzavi, J., Baumgartner, W. D., Pok, S. M., Franz, P., & Gstoettner, W. (2003). Variables affecting speech perception in postlingually deaf adults following cochlear implantation. *Acta Otolaryngol, 123*(4), 493-498.
Handel, S. (1995). Timbre perception and auditory object identication. In B. C. J. Moore (Ed.), *Hearing* (pp. 425-461). San Diego: Academic Press.
Hanson, H. M. (1997). Glottal characteristics of female speakers: acoustic correlates. *J Acoust Soc Am, 101*(1), 466-481.
Helmholtz, H. L. F. (1930). *Sensations of Tone* (A. J. Ellis, Trans.). New York: Longmans.
Henton, C., & Bladon, R. (1985). Breathiness in normal female speech: inefficiency versus desirability. *Lang Commun, 5*, 221-227.
Herzog, H., Lamprecht, A., Kuhn, A., Roden, W., Vosteen, K. H., & Feinendegen, L. E. (1991). Cortical activation in profoundly deaf patients during cochlear implant stimulation demonstrated by H2(15)O PET. *J Comput Assist Tomogr, 15*(3), 369-375.
Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci, 4*(4), 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99.
Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J Acoust Soc Am, 97*(5 Pt 1), 3099-3111.
Hirose, H., Kubota, M., Kimura, I., Ohsawa, M., Yumoto, M., & Sakakihara, Y. (2002). People with absolute pitch process tones with producing P300. *Neurosci Lett, 330*(3), 247-250.
Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in males. *J Acoust Soc Am, 96*(5 Pt 1), 2646-2654.
Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *J Acoust Soc Am, 84*(2), 511-529.
Huber, J. E., Stathopoulos, E. T., Curione, G. M., Ash, T. A., & Johnson, K. (1999). Formants of children, women, and men: the effects of vocal intensity variation. *J Acoust Soc Am, 106*(3 Pt 1), 1532-1542.
Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia, 46*(2), 632-639.
Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport, 8*(12), 2809-2812.
Inverso, Y., & Limb, C. J. (2010). Cochlear implant-mediated perception of nonlinguistic sounds. *Ear Hear, 31*(4), 505-514.
Iseli, M., Shue, Y. L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *J Acoust Soc Am, 121*(4), 2283-2295.
Ito, J., Sakakibara, J., Iwasaki, Y., & Yonekura, Y. (1993). Positron emission tomography of auditory sensation in deaf patients and patients with cochlear implants. *Ann Otol Rhinol Laryngol, 102*(10), 797-801.
Iverson, P., & Krumhansl, C. L. (1993). Isolating the dynamic attributes of musical timbre. *J Acoust Soc Am, 94*(5), 2595-2603.
Ives, D. T., Smith, D. R., & Patterson, R. D. (2005). Discrimination of speaker size from syllable phrases. *J Acoust Soc Am, 118*(6), 3816-3822.
Jaramillo, M., Paavilainen, P., & Naatanen, R. (2000). Mismatch negativity and behavioural discrimination in humans as a function of the magnitude of change in sound duration. *Neurosci Lett, 290*(2), 101-104.
Jeannon, J. P., Carding, P. N., & Wilson, J. A. (1998). Vocim analysis of laryngeal images: is breathiness related to the glottic area? *Clin Otolaryngol Allied Sci, 23*(4), 351-353.
Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123 (Pt 1)*, 155-163.
Jyung, R. W., Miller, J. M., & Cannon, S. C. (1989). Evaluation of eighth nerve integrity by the electrically evoked middle latency response. *Otolaryngol Head Neck Surg, 101*(6), 670-682.
Kaas, J. H., & Hackett, T. A. (1998). Subdivisions of auditory cortex and levels of processing in primates. *Audiol Neurootol, 3*(2-3), 73-85.
Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A, 97*(22), 11793-11799.
Kaiser, A. R., Kirk, K. I., Lachs, L., & Pisoni, D. B. (2003). Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J Speech Lang Hear Res, 46*(2), 390-404.
Kansaku, K., & Kitazawa, S. (2001). Imaging studies on sex differences in the lateralization of language. *Neurosci Res, 41*(4), 333-337.
Kansaku, K., Yamaura, A., & Kitazawa, S. (2000). Sex differences in lateralization revealed in the posterior language areas. *Cereb Cortex, 10*(9), 866-872.

Kaukoranta, E., Sams, M., Hari, R., Hamalainen, M., & Naatanen, R. (1989). Reactions of human auditory cortex to a change in tone duration. *Hear Res, 41*(1), 15-21.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage, 14*(6), 1402-1408.

Kent, R. D., & Forner, L. L. (1979). Developmental study of vowel formant frequencies in an imitation task. *J Acoust Soc Am, 65*(1), 208-217.

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., & Gstoettner, W. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology, 40*(1), 32-42.

Kiefer, J., von Ilberg, C., Reimer, B., Knecht, R., Gall, V., Diller, G., et al. (1998). Results of cochlear implantation in patients with severe to profound hearing loss--implications for patient selection. *Audiology, 37*(6), 382-395.

Kiefer, J., von Ilberg, C., Rupprecht, V., Hubner-Egner, J., & Knecht, R. (2000). Optimized speech understanding with the continuous interleaved sampling speech coding strategy in patients with cochlear implants: effect of variations in stimulation rate and number of channels. *Ann Otol Rhinol Laryngol, 109*(11), 1009-1020.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychol Sci, 14*(3), 220-224.

Klajman, S., Koldej, E., & Kowalska, A. (1982). Investigation of musical abilities in hearing-impaired and normal-hearing children. *Folia Phoniatr (Basel), 34*(5), 229-233.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am, 87*(2), 820-857.

Kocak, M., Ulmer, J. L., Biswal, B. B., Aralasmak, A., Daniels, D. L., & Mark, L. P. (2005). The influence of gender on auditory and language cortical activation patterns: preliminary data. *AJNR Am J Neuroradiol, 26*(9), 2248-2255.

Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal cues in acoustic and electric hearing. *Ear Hear, 25*(2), 173-185.

Kovacic, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *J Acoust Soc Am, 126*(2), 762-775.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., et al. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hear Res, 65*(1-2), 118-124.

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. analyses acoustiques et quantification psychophysique. *Journal de Physique, 4*, C5-625-628.

Krumhansl, C. L. (1989). Why is musical timbre so hard to understand ? In S. Nielsen & O. Olsson (Eds.), *Structure and perception of electroacoustic sound and music* (pp. 43-53). Amsterdam: Elsevier.

Kunzel, H. J. (1989). How well does average fundamental frequency correlate with speaker height and weight? *Phonetica, 46*(1-3), 117-125.

Lambertz, N., Gizewski, E. R., de Greiff, A., & Forsting, M. (2005). Cross-modal plasticity in deaf subjects dependent on the extent of hearing loss. *Brain Res Cogn Brain Res, 25*(3), 884-890.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp, 24*(1), 11-20.

Lauter, J. L., Herscovitch, P., Formby, C., & Raichle, M. E. (1985). Tonotopic organization in human auditory cortex revealed by positron emission tomography. *Hear Res, 20*(3), 199-205.

Leal, M. C., Shin, Y. J., Laborde, M. L., Calmels, M. N., Verges, S., Lugardon, S., et al. (2003). Music perception in adult cochlear implant recipients. *Acta Otolaryngol, 123*(7), 826-835.

Lechevalier, B. (2003). *Le Cerveau de Mozart*: Odile Jacob.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: developmental changes of temporal and spectral parameters. *J Acoust Soc Am, 105*(3), 1455-1468.
Levanen, S., & Hamdorf, D. (2001). Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. *Neurosci Lett, 301*(1), 75-77.
Levanen, S., Jousmaki, V., & Hari, R. (1998). Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Curr Biol, 8*(15), 869-872.
Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *Neuroimage, 20*(4), 2142-2152.
Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends Cogn Sci, 9*(1), 26-33.
Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport, 12*(12), 2653-2657.
Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40*(2), 291-305.
Li, T., & Fu, Q. J. (2007). Perceptual adaptation to spectrally shifted vowels: training with nonlexical labels. *J Assoc Res Otolaryngol, 8*(1), 32-41.
Linville, S. E. (1996). The sound of senescence. *J Voice, 10*(2), 190-200.
Linville, S. E., & Fisher, H. B. (1985). Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *J Acoust Soc Am, 78*(1 Pt 1), 40-48.
Lockhead, G. R., & Byrd, R. (1981). Practically perfect pitch. *J Acoust Soc Am, 70*, 387-389.
Lockwood, A. H., Salvi, R. J., Coad, M. L., Arnold, S. A., Wack, D. S., Murphy, B. W., et al. (1999). The functional anatomy of the normal human auditory system: responses to 0.5 and 4.0 kHz tones at varied intensities. *Cereb Cortex, 9*(1), 65-76.
Loebach, J. L., & Pisoni, D. B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *J Acoust Soc Am, 123*(2), 1126-1139.
Loizou, P. C., Poroy, O., & Dorman, M. (2000). The effect of parametric variations of cochlear implant processors on speech understanding. *J Acoust Soc Am, 108*(2), 790-802.
Loke, W. H., & Song, S. (1991). Central and peripheral visual processing in hearing and nonhearing individuals. *Bull. Psychon. Soc., 29*, 437.
Looi, V., & Arnephy, J. (2009). Environmental sound perception of cochlear implant users. *Cochlear Implants Int*.
Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear, 29*(3), 421-434.
Marx, M. (2009). *Perception de la voix et de la prosodie par le patient implanté cochléaire*.Unpublished manuscript, Université de Toulouse.
Matsumoto, H., Hiki, S., Sone, T., & Nimura, T. (1973). Multidimensional representation of personal quality of vowels and its acoustical correlates. *IEEE Transactions on Audio and Electroacoustics, 21*, 428-436.
Mazziotta, J. C., Phelps, M. E., Carson, R. E., & Kuhl, D. E. (1982). Tomographic mapping of human cerebral metabolism: auditory stimulation. *Neurology, 32*(9), 921-937.
McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound : the cognitive psychology of human audition* (pp. 146-198). Oxford: Oxford University Press.
McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res, 58*(3), 177-192.
McKay, C. M., & Henshall, K. R. (2002). Frequency-to-electrode allocation and speech perception with cochlear implants. *J Acoust Soc Am, 111*(2), 1036-1044.

Meister, H., Landwehr, M., Pyschny, V., Walger, M., & von Wedel, H. (2009). The perception of prosody and speaker gender in normal-hearing listeners and cochlear implant recipients. *Int J Audiol, 48*(1), 38-48.

Mendoza, E., Valencia, N., Munoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *J Voice, 10*(1), 59-66.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17*(4), 1742-1754.

Meyer, M., Baumann, S., & Jancke, L. (2006). Electrical brain imaging reveals spatio-temporal dynamics of timbre perception in humans. *Neuroimage, 32*(4), 1510-1523.

Micco, A. G., Kraus, N., Koch, D. B., McGee, T. J., Carrell, T. D., Sharma, A., et al. (1995). Speech-evoked cognitive P300 potentials in cochlear implant recipients. *Am J Otol, 16*(4), 514-520.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res, 219*(1-2), 36-47.

Mills, C. (1985). Perception of visual temporal patterns by deaf and hearing adults. *Bull. Psychon. Soc., 23*, 483.

Milner, B. (1962). Laterality effects in audition. In V. Mountcastel (Ed.), *Interhemispheric Relations and Cerebral Dominance* (pp. 173-201). Baltimore: John Hopkins University Press.

Mitchell, R. L., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41*(10), 1410-1421.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Percept Psychophys, 44*(6), 501-512.

Miyazaki, K. (1989). Absolute pitch identification: effects of timbre and pitch region. *Music Perception, 7*, 1-14.

Miyazaki, K. (2004). Recognition of transposed melodies by absolute-pitch possessors. *Jpn. Psychol. Res., 46*, 266-282.

Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Percept Psychophys, 64*(8), 1337-1345.

Monrad-Krohn, G. H. (1963). The third element of speech: prosody and its disorders. In L. Halpern (Ed.), *Problems of Dynamic Neurology* (pp. 101–117): Hebrew University Press.

Moore, B. C., & Glasberg, B. R. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *J Acoust Soc Am, 87*(5), 2163-2177.

Moore, B. C., Glasberg, B. R., & Peters, R. W. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J Acoust Soc Am, 80*(2), 479-483.

Moore, B. C. J. (1989). Frequency selectivity, masking and the critical band. In *An introduction to the psychology of hearing* (3rd ed., pp. 54-136). London: Academic Press Limited.

Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1982). *An introduction to the psychlogy of hearing* (2d ed.). London: Academic Press.

Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *J Acoust Soc Am, 127*(2), 1047-1058.

Morel, A., Garraghty, P. E., & Kaas, J. H. (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol, 335*(3), 437-459.

Morel, A., & Kaas, J. H. (1992). Subdivisions and connections of auditory cortex in owl monkeys. *J Comp Neurol, 318*(1), 27-63.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage, 13*(4), 684-701.

Morris, J. S., Scott, S. K., & Dolan, R. J. (1999). Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia, 37*(10), 1155-1163.

Morris, R. J., & Brown, W. S., Jr. (1994). Age-related differences in speech variability among women. *J Commun Disord, 27*(1), 49-64.
Mortensen, M. V., Mirz, F., & Gjedde, A. (2006). Restored speech comprehension linked to activity in left inferior prefrontal and right temporal cortices in postlingual deafness. *Neuroimage, 31*(2), 842-852.
Muhler, R., Ziese, M., & Rostalski, D. (2009). Development of a speaker discrimination test for cochlear implant users based on the Oldenburg Logatome corpus. *ORL J Otorhinolaryngol Relat Spec, 71*(1), 14-20.
Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *J Acoust Soc Am, 98*(6), 3080-3095.
Murphy, C. H., & Doyle, P. C. (1987). The effects of cigarette smoking on voice-fundamental frequency. *Otolaryngol Head Neck Surg, 97*(4), 376-380.
Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *J Neurosci, 26*(4), 1293-1302.
Murry, T., & Singh, S. (1980). Multidimensional analysis of male and female voices. *J Acoust Soc Am, 68*(5), 1294-1300.
Naito, Y., Okazawa, H., Honjo, I., Hirano, S., Takahashi, H., Shiomi, Y., et al. (1995). Cortical activation with sound stimulation in cochlear implant users demonstrated by positron emission tomography. *Brain Res Cogn Brain Res, 2*(3), 207-214.
Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., et al. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39*(10), 1047-1054.
Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn, 44*(3), 342-366.
Neville, H. J., Bavelier, D., Corina, D., Rauschecker, J., Karni, A., Lalwani, A., et al. (1998). Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proc Natl Acad Sci U S A, 95*(3), 922-929.
Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Res, 405*(2), 268-283.
Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am, 95*(2), 1085-1099.
Nishimura, H., Hashikawa, K., Doi, K., Iwaki, T., Watanabe, Y., Kusuoka, H., et al. (1999). Sign language 'heard' in the auditory cortex. *Nature, 397*(6715), 116.
Nogaki, G., Fu, Q. J., & Galvin, J. J., 3rd. (2007). Effect of training rate on recognition of spectrally shifted speech. *Ear Hear, 28*(2), 132-140.
Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Hum Dev, 18*(1), 27-36.
Ohnishi, T., Matsuda, H., Hashimoto, T., Kunihiro, T., Nishikawa, M., Uema, T., et al. (2000). Abnormal regional cerebral blood flow in childhood autism. *Brain, 123 (Pt 9)*, 1838-1844.
Okazawa, H., Naito, Y., Yonekura, Y., Sadato, N., Hirano, S., Nishizawa, S., et al. (1996). Cochlear implant efficiency in pre- and postlingually deaf subjects. A study with H2(15)O and PET. *Brain, 119 (Pt 4)*, 1297-1306.
Orlikoff, R. F. (1990). The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. *J Speech Hear Res, 33*(3), 450-457.
Oviatt, D. L., & Kileny, P. R. (1991). Auditory event-related potentials elicited from cochlear implant recipients and hearing subjects. *Am. J. Audiol., 1*, 48-55.
Owren, M. J., Berkowitz, M., & Bachorowski, J. A. (2007). Listeners judge talker sex more efficiently from male than from female vowels. *Percept Psychophys, 69*(6), 930-941.

Pandya, D. N., & Sanides, F. (1973). Architectonic parcellation of the temporal operculum in rhesus monkey and its projection pattern. *Z Anat Entwicklungsgesch, 139*(2), 127-161.

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., et al. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr Clin Neurophysiol, 94*(1), 26-40.

Pantev, C., Hoke, M., Lehnertz, K., & Lutkenhoner, B. (1989). Neuromagnetic evidence of an amplitopic organization of the human auditory cortex. *Electroencephalogr Clin Neurophysiol, 72*(3), 225-231.

Parasnis, I. (1992). Allocation of attention, reading skills, and deafness. *Brain Lang, 43*(4), 583-596.

Parasnis, I., & Samar, V. J. (1985). Parafoveal attention in congenitally deaf and hearing young adults. *Brain Cogn, 4*(3), 313-327.

Parasnis, I., Samar, V. J., & Berent, G. P. (2003). Deaf adults without attention deficit hyperactivity disorder display reduced perceptual sensitivity and elevated impulsivity on the Test of Variables of Attention (T.O.V.A.). *J Speech Lang Hear Res, 46*(5), 1166-1183.

Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., & Koch, U. (1996). Neural delays shape selectivity to interaural intensity differences in the lateral superior olive. *J Neurosci, 16*(20), 6554-6566.

Parkinson, A. J., Parkinson, W. S., Tyler, R. S., Lowder, M. W., & Gantz, B. J. (1998). Speech perception performance in experienced cochlear-implant patients receiving the SPEAK processing strategy in the Nucleus Spectra-22 cochlear implant. *J Speech Lang Hear Res, 41*(5), 1073-1087.

Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *J Acoust Soc Am, 59*(3), 640-654.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*(4), 767-776.

Pegaroo Krook, M. I. (1988). Speaking fundamental frequency characteristics of normal Swedish subjects obtained by glottal frequency analysis. *Folia Phoniatr (Basel), 40*, 82-90.

Pelizzone, M., Cosendai, G., & Tinembart, J. (1999). Within-patient longitudinal speech reception measures with continuous interleaved sampling processors for ineraid implanted subjects. *Ear Hear, 20*(3), 228-237.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci, 24*(30), 6810-6815.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117 (Pt 6)*, 1283-1301.

Petitto, L. A., Zatorre, R. J., Gauna, K., Nikelski, E. J., Dostie, D., & Evans, A. C. (2000). Speech-like cerebral activity in profoundly deaf people processing signed languages: implications for the neural basis of human language. *Proc Natl Acad Sci U S A, 97*(25), 13961-13966.

Phelps, M. E., & Mazziotta, J. C. (1985). Positron emission tomography: human brain function and biochemistry. *Science, 228*(4701), 799-809.

Phillips, M. L., Young, A. W., Scott, S. K., Calder, A. J., Andrew, C., Giampietro, V., et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc Biol Sci, 265*(1408), 1809-1817.

Pialoux, P., Chouard, C. H., & MacLeod, P. (1976). Physiological and clinical aspects of the rehabilitation of total deafness by implantation of multiple intracochlear electrodes. *Acta Otolaryngol, 81*(5-6), 436-441.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *J Exp Psychol Hum Percept Perform, 20*(5), 976-986.

Platel, H., Price, C., Baron, J. C., Wise, R., Lambert, J., Frackowiak, R. S., et al. (1997). The structural components of music perception. A functional anatomical study. *Brain, 120 (Pt 2)*, 229-243.

Poizner, H., & Tallal, P. (1987). Temporal processing in deaf signers. *Brain Lang, 30*(1), 52-62.

Profita, J., & Bidder, T. G. (1988). Perfect pitch. *Am J Med Genet, 29*(4), 763-771.

Proksch, J., & Bavelier, D. (2002). Changes in the spatial distribution of visual attention after early deafness. *J Cogn Neurosci, 14*(5), 687-701.

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage, 49*(2), 1641-1649.

Rao, S. M., Mayer, A. R., & Harrington, D. L. (2001). The evolution of brain activation during temporal processing. *Nat Neurosci, 4*(3), 317-323.

Rastatter, M. P., & Jacques, R. D. (1990). Formant frequency structure of the aging male and female vocal tract. *Folia Phoniatr (Basel), 42*(6), 312-319.

Rauschecker, J. P. (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl, 532*, 34-38.

Rauschecker, J. P., Tian, B., & Hauser, M. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science, 268*(5207), 111-114.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci Biobehav Rev, 32*(1), 123-142.

Reed, C. M., & Delhorne, L. A. (2005). Reception of environmental sounds through cochlear implants. *Ear Hear, 26*(1), 48-61.

Rendall, D., Kollias, S., Ney, C., & Lloyd, P. (2005). Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. *J Acoust Soc Am, 117*(2), 944-955.

Rettenbach, R., Diller, G., & Sireteanu, R. (1999). Do deaf people see better? Texture segmentation and visual search compensate in adult but not in juvenile subjects. *J Cogn Neurosci, 11*(5), 560-583.

Ritter, S., Dosch, H. G., Specht, H. J., Schneider, P., & Rupp, A. (2007). Latency effect of the pitch response due to variations of frequency and spectral envelope. *Clin Neurophysiol, 118*(10), 2276-2281.

Rodero, E. (2010). Intonation and Emotion: Influence of Pitch Levels and Contour Type on Creating Emotions. *J Voice*.

Roers, F., Murbe, D., & Sundberg, J. (2009). Predicted singers' vocal fold lengths and voice classification-a study of x-ray morphological measures. *J Voice, 23*(4), 408-413.

Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. *Int J Psychophysiol, 75*(1), 44-47.

Romanski, L. M., Averbeck, B. B., & Diltz, M. (2005). Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. *J Neurophysiol, 93*(2), 734-747.

Romanski, L. M., & Goldman-Rakic, P. S. (2002). An auditory domain in primate prefrontal cortex. *Nat Neurosci, 5*(1), 15-16.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 367-373.

Rosen, S., Faulkner, A., & Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J Acoust Soc Am, 106*(6), 3629-3636.

Ross, D. A., Gore, J. C., & Marks, L. E. (2005). Absolute pitch: music and beyond. *Epilepsy Behav, 7*(4), 578-601.

Rouger, J. (2007). *Perception audiovisuelle de la parole chez le sourd postlingual implanté cochléaire et le sujet normo-entendant : étude longitudinale psychophysique et neurofonctionnelle.*, Université de Toulouse 3 - Paul Sabatier, Toulouse.

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Res, 1188*, 87-99.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci U S A, 104*(17), 7295-7300.

Rubinstein, J. T., Parkinson, W. S., Tyler, R. S., & Gantz, B. J. (1999). Residual speech recognition and cochlear implant performance: effects of implantation criteria. *Am J Otol, 20*(4), 445-452.

Rupp, A., Gutschalk, A., Uppenkamp, S., & Scherg, M. (2004). Middle latency auditory-evoked fields reflect psychoacoustic gap detection thresholds in human listeners. *J Neurophysiol, 92*(4), 2239-2247.

Russell, A., Penny, L., & Pemberton, C. (1995). Speaking fundamental frequency changes over time in women: a longitudinal study. *J Speech Hear Res, 38*(1), 101-109.

Ruytjens, L., Georgiadis, J. R., Holstege, G., Wit, H. P., Albers, F. W., & Willemsen, A. T. (2007). Functional sex differences in human primary auditory cortex. *Eur J Nucl Med Mol Imaging, 34*(12), 2073-2081.

Sadato, N., Yamada, H., Okada, T., Yoshida, M., Hasegawa, T., Matsuki, K., et al. (2004). Age-dependent plasticity in the superior temporal sulcus in deaf humans: a functional MRI study. *BMC Neurosci, 5*, 56.

Samson, S. (2003). Neuropsychological studies of musical timbre. *Ann N Y Acad Sci, 999*, 144-151.

Samson, S., & Zatorre, R. J. (1994). Contribution of the right temporal lobe to musical timbre discrimination. *Neuropsychologia, 32*(2), 231-240.

Samson, S., Zatorre, R. J., & Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain, 125*(Pt 3), 511-523.

Samson, Y., Belin, P., Thivard, L., Boddaert, N., Crozier, S., & Zilbovicius, M. (2001). [Auditory perception and language: functional imaging of speech sensitive auditory cortex]. *Rev Neurol (Paris), 157*(8-9 Pt 1), 837-846.

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res Cogn Brain Res, 12*(2), 181-198.

Sapienza, C. M., Stathopoulos, E. T., & Dromey, C. (1998). Approximations of open quotient and speed quotient from glottal airflow and EGG waveforms: effects of measurement criteria and sound pressure level. *J Voice, 12*(1), 31-43.

Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., et al. (2008). Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A, 105*(46), 18035-18040.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*(5198), 699-701.

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat Neurosci, 8*(9), 1241-1247.

Schonwiesner, M., von Cramon, D. Y., & Rubsamen, R. (2002). Is it tonotopy after all? *Neuroimage, 17*(3), 1144-1161.

Schroeder, C. E., Smiley, J., Fu, K. G., McGinnis, T., O'Connell, M. N., & Hackett, T. A. (2003). Anatomical mechanisms and functional implications of multisensory convergence in early cortical processing. *Int J Psychophysiol, 50*(1-2), 5-17.

Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci, 26*(2), 100-107.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature, 385*(6613), 254-257.

Shafiro, V. (2008). Identification of environmental sounds with varying spectral resolution. *Ear Hear, 29*(3), 401-420.

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci, 23*(13), 5545-5552.

Shallop, J. K., Arndt, P. L., & Turnacliff, K. A. (1992). Expanded indications for cochlear implantation: Perceptual results in seven adults with residual hearing. *J. Speech Lang. Path., 16*, 141-148.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*(5234), 303-304.

Shaw, A. (1974). The external ear. In W. Keidel & W. Neff (Eds.), *Handbook of Sensory Physiology* (pp. 455-490). Berlin: Springer.

Shepherd, R. K., Clark, G. M., & Black, R. C. (1983). Chronic electrical stimulation of the auditory nerve in cats. Physiological and histopathological results. *Acta Otolaryngol Suppl, 399*, 19-31.

Shipp, D. B., & Nedzelski, J. M. (1994). Prognostic value of round-window psychophysical testing with cochlear-implant candidates. *J Otolaryngol, 23*(3), 172-176.

Singh, P. G., & Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *J Acoust Soc Am, 92*(5), 2650-2661.

Skinner, M. W., Clark, G. M., & Whitford, L. A. (1994). Evaluation of a new Spectral Peak coding strategy for the Nucleus 22-channel cochlear implant system. *Am J Otol, 15 Suppl*(2), 15-27.

Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., & Demorest, M. E. (1999). Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies II. Consonants. *Ear Hear, 20*(6), 443-460.

Skinner, M. W., Holden, L. K., & Holden, T. A. (1995). Effect of frequency boundary assignment on speech recognition with the speak speech-coding strategy. *Ann Otol Rhinol Laryngol Suppl, 166*, 307-311.

Sladen, D. P., Tharpe, A. M., Ashmead, D. H., Wesley Grantham, D., & Chun, M. M. (2005). Visual attention in deaf and normal hearing adults: effects of stimulus compatibility. *J Speech Lang Hear Res, 48*(6), 1529-1537.

Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., & Irino, T. (2005). The processing and perception of size information in speech sounds. *J Acoust Soc Am, 117*(1), 305-318.

Sorensen, D., & Horii, Y. (1982). Cigarette smoking and voice fundamental frequency. *J Commun Disord, 15*(2), 135-144.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmuller, E., & Munte, T. F. (2009). Neural processing of vocal emotion and identity. *Brain Cogn, 69*(1), 121-126.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol, 19*(6), 498-502.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Brain Res Cogn Brain Res, 18*(2), 162-171.

Stevens, C., & Neville, H. (2006). Neuroplasticity as a double-edged sword: deaf enhancements and dyslexic deficits in motion processing. *J Cogn Neurosci, 18*(5), 701-714.

Stevens, S. S. (1935). The relation of pitch to intensity. *J Acoust Soc Am, 6*, 150-154.

Stevens, S. S., & Volkmann, J. (1940). The relation of pitch to frequency : a revised scale. *Am J Psychol, 53*, 329-353.

Stivalet, P., Moreno, Y., Richard, J., Barraud, P. A., & Raphel, C. (1998). Differences in visual search tasks between congenitally deaf and normally hearing adults. *Brain Res Cogn Brain Res, 6*(3), 227-232.

Strainer, J. C., Ulmer, J. L., Yetkin, F. Z., Haughton, V. M., Daniels, D. L., & Millen, S. J. (1997). Functional MR of the primary auditory cortex: an analysis of pure tone activation and tone discrimination. *AJNR Am J Neuroradiol, 18*(4), 601-610.

Strelnikov, K., Massida, Z., Rouger, J., Barone, P., & Belin, P. Brain network associated with vocoder degradation differs from that for natural speech and voice processing: evidence from fMRI study. *In preparation*.

Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Deguine, O., & Barone, P. (2009). Improvement in speech-reading ability by auditory training: Evidence from gender differences in normally hearing, deaf and cochlear implanted subjects. *Neuropsychologia, 47*(4), 972-979.

Stypulkowski, P. H., & van den Honert, C. (1984). Physiological properties of the electrically stimulated auditory nerve. I. Compound action potential recordings. *Hear Res, 14*(3), 205-223.

Stypulkowski, P. H., van den Honert, C., & Kvistad, S. D. (1986). Electrophysiologic evaluation of the cochlear implant patient. *Otolaryngol Clin North Am, 19*(2), 249-257.

Su, M. C., Yeh, T. H., Tan, C. T., Lin, C. D., Linne, O. C., & Lee, S. Y. (2002). Measurement of adult vocal fold length. *J Laryngol Otol, 116*(6), 447-449.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philos Trans R Soc Lond B Biol Sci, 335*(1273), 71-78.

Szelag, E., Kolodziejczyk, I., Kanabus, M., Szuchnik, J., & Senderski, A. (2004). Deficits of non-verbal auditory perception in postlingually deaf humans using cochlear implants. *Neurosci Lett, 355*(1-2), 49-52.

Takeuchi, A. H., & Hulse, S. H. (1991). Absolute pitch judgments of black- and white-key pitches. *Music Perception, 9*, 27-46.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychol Bull, 113*(2), 345-361.

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res, 150*(1-2), 225-244.

Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychol Rev, 61*(6), 401-409.

Tanzarella, S., & Mamecier, A. (2006). *Perception et communication chez les animaux*: De Boeck.

Terhardt, E. (1974). Pitch, consonance, and harmony. *J Acoust Soc Am, 55*(5), 1061-1069.

Terhardt, E. (1979). Calculating virtual pitch. *Hear Res, 1*(2), 155-182.

Terhardt, E., Stoll, G., & Seewann, M. (1982). Pitch of complex signal according to virtual-pitch theory: tests, examples and predictions. *J Acoust Soc Am, 71*, 671-678.

Titova, N., & Naatanen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett, 308*(1), 63-65.

Toivonen, M., & Rama, P. (2009). N400 during recognition of voice identity and vocal affect. *Neuroreport, 20*(14), 1245-1249.

Truy, E., Deiber, M. P., Cinotti, L., Mauguiere, F., Froment, J. C., & Morgon, A. (1995). Auditory cortex activity changes in long-term sensorineural deprivation during crude cochlear electrical stimulation: evaluation by positron emission tomography. *Hear Res, 86*(1-2), 34-42.

Tyler, R. S., & Lowder, M. W. (1992). Audiological management and performance of adult cochlear-implant patients. *Ear Nose Throat J, 71*(3), 117-122, 125-118.

Tyler, R. S., Parkinson, A. J., Woodworth, G. G., Lowder, M. W., & Gantz, B. J. (1997). Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *J Acoust Soc Am, 102*(1), 508-522.

Tyler, R. S., Preece, J. P., Lansing, C. R., Otto, S. R., & Gantz, B. J. (1986). Previous experience as a confounding factor in comparing cochlear-implant processing schemes. *J Speech Hear Res, 29*(2), 282-287.

Tyler, R. S., Preece, J. P., Lansing, C. R., Otto, S. R., & Grantz, B. J. (1986). Previous experience as a counfouding factor in comparing cochlear-implant processing schemes. *J. Speech Hear. Res., 29*(2), 282-287.

UKCISG, U. K. C. I. S. G. (2004). Criteria of candidacy for unilateral cochlear implantation in postlingually deafened adults I: theory and measures of effectiveness. *Ear Hear, 25*(4), 310-335.

Valimaa, T. T., Maatta, T. K., Lopponen, H. J., & Sorri, M. J. (2002). Phoneme recognition and confusions with multichannel cochlear implants: consonants. *J Speech Lang Hear Res, 45*(5), 1055-1069.

Van Borsel, J., Janssens, J., & De Bodt, M. (2009). Breathiness as a feminine voice characteristic: a perceptual approach. *J Voice, 23*(3), 291-294.

van Dijk, J. E., van Olphen, A. F., Langereis, M. C., Mens, L. H., Brokx, J. P., & Smoorenburg, G. F. (1999). Predictors of cochlear implant performance. *Audiology, 38*(2), 109-116.

Van Lancker, D., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829-834.

Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn, 1*(2), 185-195.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex, 24*(2), 195-209.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *J Clin Exp Neuropsychol, 11*(5), 665-674.

van Noorden, L. P. A. S. (1982). Two-channel pitch perception. In M. Clynes (Ed.), *Music, mind and brain* (pp. 251-269). New York: Plenum Press.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn Brain Res, 17*(1), 48-55.

von Kriegstein, K., Smith, D. R., Patterson, R. D., Ives, D. T., & Griffiths, T. D. (2007). Neural representation of auditory size in the human voice and in sounds from other resonant sources. *Curr Biol, 17*(13), 1123-1128.

von Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22*(2), 948-955.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Dev, 81*(2), 517-527.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *J Cogn Neurosci, 13*(7), 994-1005.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: the privileged status of speech for young infants. *Dev Sci, 7*(3), 270-276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev Sci, 10*(2), 159-164.

Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A., & Schwartz, D. M. (1978). Correlates of psychological dimensions in talker similarity. *J Speech Hear Res, 21*(2), 265-275.

Walker, K. M., Bizley, J. K., King, A. J., & Schnupp, J. W. (2010). Cortical encoding of pitch: Recent results and open questions. *Hear Res*.

Walliser, K. (1969). Uber die Abhängigkeit de Tonhöhenempfindung von Sinustönen vom Schallpegel, von überlagertem drosselndem Störschall und von der Darbietungsdauer. *Acustica, 21*, 211-221.

Walsh, S. M., & Leake-Jones, P. A. (1982). Chronic electrical stimulation of auditory nerve in cat: Physiological and histological results. *Hear Res, 7*(3), 281-304.

Waltzman, S. B., Fisher, S. G., Niparko, J. K., & Cohen, N. L. (1995). Predictors of postoperative performance with cochlear implants. *Ann Otol Rhinol Laryngol Suppl, 165*, 15-18.

Wang, X. (2000). On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A, 97*(22), 11843-11849.
Wang, X., & Kadia, S. C. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol, 86*(5), 2616-2620.
Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 265-298). San Diego: Academic Pres.
Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31*(3), 1389-1397.
Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A, 100*(17), 10038-10042.
Warren, J. D., Zielinski, B. A., Green, G. G., Rauschecker, J. P., & Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron, 34*(1), 139-148.
Warren, R. M., & Wrightson, J. M. (1981). Stimuli producing conflicting temporal and spectral cues to frequency. *J Acoust Soc Am, 70*, 1020-1024.
Weeks, R. A., Aziz-Sultan, A., Bushara, K. O., Tian, B., Wessinger, C. M., Dang, N., et al. (1999). A PET study of human auditory spatial processing. *Neurosci Lett, 262*(3), 155-158.
Wessinger, C. M., Buonocore, M. H., Kussmaul, C. L., & Mangun, G. R. (1997). Tonotopy in Human Auditory Cortex Examined With Functional Magnetic Resonance Imaging. *Human Brain Mapping, 5*(1), 18–25.
Whiteside, S. P. (1998). The identification of speaker's sex from synthesized vowels. *Percept Mot Skills, 87*, 595-600.
Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage, 15*(4), 856-869.
Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant future. *Hear Res, 242*(1-2), 3-21.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., & Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature, 352*(6332), 236-238.
Wilson, B. S., Lawson, D. T., Muller, J. M., Tyler, R. S., & Kiefer, J. (2003). Cochlear implants: some likely next steps. *Annu Rev Biomed Eng, 5*, 207-249.
Winter, P., & Funkenstein, H. H. (1973). The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (Saimiri sciureus). *Exp Brain Res, 18*(5), 489-504.
Wolfe, V. I., Ratusnik, D. L., Smith, F. H., & Northrop, G. (1990). Intonation and fundamental frequency in male-to-female transsexuals. *J Speech Hear Disord, 55*(1), 43-50.
Wong, D., Miyamoto, R. T., Pisoni, D. B., Sehgal, M., & Hutchins, G. D. (1999). PET imaging of cochlear-implant and normal-hearing subjects listening to speech and nonspeech. *Hear Res, 132*(1-2), 34-42.
Wong, P. C., Nusbaum, H. C., & Small, S. L. (2004). Neural bases of talker normalization. *J Cogn Neurosci, 16*(7), 1173-1184.
Wong, P. C., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., et al. (2008). Volume of left Heschl's Gyrus and linguistic pitch learning. *Cereb Cortex, 18*(4), 828-836.
Wu, C., Kirk, I. J., Hamm, J. P., & Lim, V. K. (2008). The neural networks involved in pitch labeling of absolute pitch musicians. *Neuroreport, 19*(8), 851-854.
Xin, L., Fu, Q. J., & Galvin, J. J., 3rd. (2007). Vocal emotion recognition by normal-hearing listeners and cochlear implant users. *Trends Amplif, 11*(4), 301-315.
Xu, L., Tsai, Y., & Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses. *J Acoust Soc Am, 112*(1), 247-258.
Yin, T. C., & Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol, 64*(2), 465-488.

Zaske, R., Schweinberger, S. R., Kaufmann, J. M., & Kawahara, H. (2009). In the ear of the beholder: neural correlates of adaptation to voice gender. *Eur J Neurosci, 30*(3), 527-534.
Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am, 84*(2), 566-572.
Zatorre, R. J. (2001). Do you see what I'm saying? Interactions between auditory and visual cortices in cochlear implant users. *Neuron, 31*(1), 13-14.
Zatorre, R. J. (2003). Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nat Neurosci, 6*(7), 692-695.
Zatorre, R. J., & Beckett, C. (1989). Multiple coding strategies in the retention of musical tones by possessors of absolute pitch. *Mem Cognit, 17*(5), 582-589.
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex, 11*(10), 946-953.
Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci, 6*(1), 37-46.
Zatorre, R. J., & Binder, J. R. (2000). Functional and structural imaging of the human auditory system. In *Brain Mapping: The Systems*: Academic Press.
Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is 'where' in the human auditory cortex? *Nat Neurosci, 5*(9), 905-909.
Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci, 363*(1493), 1087-1104.
Zatorre, R. J., & Penhune, V. B. (2001). Spatial localization after excision of human auditory cortex. *J Neurosci, 21*(16), 6321-6328.
Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proc Natl Acad Sci U S A, 95*(6), 3172-3177.
Zeng, F. G. (2004). Trends in cochlear implants. *Trends Amplif, 8*(1), 1-34.
Zilbovicius, M., Boddaert, N., Belin, P., Poline, J. B., Remy, P., Mangin, J. F., et al. (2000). Temporal lobe dysfunction in childhood autism: a PET study. Positron emission tomography. *Am J Psychiatry, 157*(12), 1988-1993.

14 ANNEXES

2 canaux	4 canaux	8 canaux	16 canaux	32 canaux	64 canaux
125	**125**	**125**	**125**	**125**	**125**
1374	503	275	192	157	141
8000	1374	503	275	192	157
	3381	849	377	232	174
	8000	1374	503	275	192
		2171	658	324	211
		3381	849	377	232
		5215	1085	437	253
		8000	1374	503	275
			1732	577	299
			2171	658	324
			2713	749	350
			3381	849	377
			4203	961	407
			5215	1085	437
			6463	1222	469
			8000	1374	503
				1544	539
				1732	577
				1940	617
				2171	658
				2428	702
				2713	749
				3029	798
				3381	849
				3770	904
				4203	961
				4683	1021
				5215	1085
				5807	1152
				6463	1222
				7192	1296
				8000	1374
					1457
					1544
					1635

1732
1833
1940
2053
2171
2296
2428
2567
2713
2867
3029
3200
3381
3570
3770
3981
4203
4436
4683
4942
5215
5503
5807
6126
6463
6818
7192
7585
8000

Tableau 14.1 : Fréquences de coupures en Hz des différentes conditions de vocoder : 64, 32, 16, 8, 4 et canaux.

Informations relatives au patient					Informations relatives à l'implant							Information relative aux tests					
N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
1	**MARNIO1**	F	53,32	> 40	D	1 jour	Freedom Implant (Contour advance) (Cochlear)	Freedom SP	ACE	18/22	O	X		X	X	X	
2	**CHEANN02**	F	48,67	> 30	G	7 ans et 1 mois	CI24M Nucleus24 (Cochlear)	Sprint	ACE	21/22	N				X		
3	**CULYVE05**	F	60,73	> 40	D	2 ans et 1 mois	HiRes 90K/HiFocus (Advance Bionic)	Auria	HiRes-S	15/16	N				X		
4	**PIEDAN07**	F	57,73	> 30	D	6 ans et 11 mois	CI24M Nucleus24 (Cochlear)	ESPrit	SPEAK	20/22	N				X		
5	**LAVJAC08**	M	73,52	= 2	G	9 mois	Freedom Implant (Contour advance) (Cochlear)	Freedom SP	ACE	19/22	O				X		
6	**JOLJOS09**	F	60,65	> 20	G	5 ans et 2 mois	CI24 R-CS (Cochlear)	ESPrit 3G	ACE	20/22	N				X		
7	**DOZHUG11**	F	63,03	> 20	G	6 mois	Freedom Implant (Contour advance) (Cochlear)	Freedom SP	ACE	20/22	O				X		

N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
8	**DOUMED12**	M	21,33	> 10	G	6 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	19/22	O		X		X		
9	**MAUJOS14**	F	64,32	> 10	D	3 mois	HiRes 90K/HiFocus ***(Advance Bionic)***	PSP	HiRes-S	16/16	N		X				
10	**DAGVIV15**	F	47,84	> 30	D	1 an et 9 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	22/22	O		X		X		
11	**LAGSAN16**	F	35,18	> 30	D	3 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	O		X		X		
12	**GALDEN17**	M	59,36	NR	D	6 ans et 9 mois	CI24M Nucleus24 ***(Cochlear)***	Sprint	ACE	21/22	N		X		X		
13	**GOMLOU18**	F	47,15	> 40	D	1 jour	SONATAti100 ***(MedEl)***	OPUS 2	FSP	10/12	O	X	X	X	X		
14	**CHAMAR19**	F	74,76	> 10	G	3 ans et 10 mois	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	18/22	N		X		X		
15	**GAUJEA20**	M	40,69	> 30	D	2 ans 1 mois	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	19/22	N		X		X		

N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
16	**GIMFRA21**	M	71,09	NR	G	1 an et 6 mois	Freedom Implant (Straight) ***(Cochlear)***	Freedom SP	ACE	22/22	N		X		X		
17	**DUBLIL22**	F	52,82	< 5	G	7 mois	HiRes 90K/HiFocus ***(Advance Bionic)***	Auria	HiRes-S	16/16	O		X		X		
18	**MARJAC23**	F	42,46	> 30	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	N	X		X	X	X	
19	**PASROG26**	M	58,90	> 10	D	4 ans et 2 mois	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	20/22	N				X		
20	**CERPAS27**	F	37,53	> 30	G	2 ans	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	19/22	N		X		X		
21	**PIGFRA28**	F	75,44	> 20	D	10 ans et 11 mois	Nucleus 22 Series ***(Cochlear)***	ESPrit 22	SPEAK	19/22	N		X		X		
22	**BARMAR29**	M	68,09	> 10	D	7 ans	CI24M Nucleus24 ***(Cochlear)***	Sprint	ACE	20/22	N				X		
23	**VIAMIC30**	M	63,02	> 20	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	21/22	N	X	X	X	X		

N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
24	**GIMMAR31**	F	47,94	> 40	D	1 jour	HiRes 90K/HiFocus ***(Advance Bionic)***	Auria	HiRes-S	16/16	N	X	X	X	X		
25	**BRIMAR32**	F	63,51	> 40	G	4 ans	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	18/22	N		X		X		
26	**MANGRA33**	F	59,63	> 40	G	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	N	X		X			
27	**BRUMAR34**	F	41,06	> 20	D	2 ans	CI24 R (CS) Nucleus 24 Contour ***(Cochlear)***	ESPrit 3G	ACE	20/22	O		X		X		
28	**RIGROB35**	M	66,81	> 30	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	N	X		X	X		
29	**ALAREN37**	F	66,94	> 10	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	O	X	X	X	X		
30	**BOUFRA39**	M	20,28	> 10	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	19/22	N	X	X	X	X	X	
31	**LEGPAT40**	M	51,09	> 30	G	6 mois	HiRes 90K/HiFocus ***(Advance Bionic)***	Auria	HiRes-S	16/16	O		X		X		

N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
33	**BRUDOM41**	M	44,83	= 3 ans	G	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	O	X	X	X	X		
34	**ANDJOS42**	F	47,87	> 30	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	22/22	N	X	X	X	X	X	X
35	**GENMAR43**	F	52,13	> 40	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	20/22	N	X	X	X		X	
36	**MONFEL44**	M	80,05	NR	G	1 jour	HiRes 90K/HiFocus ***(Advance Bionic)***	Auria	HiRes-S	16/16	O	X	X	X	X		
37	**DUBJOE45**	F	47,87	> 30	D	1 jour	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	19/22	N	X	X	X	X		
38	**CHEMAU46**	M	63,21	> 10	G	1 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom SP	ACE	21/22	N		X		X		
39	**FONMAR47**	F	56,01	> 50	G	3 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom	ACE	19/22	O		X				
40	**DEBCOR48**	F	47,66	> 30	D	14 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom	ACE	20/22	O		X		X		

N°	Code Sujet	Sexe	Age à l'inclus-sion (ans)	Durée de perte auditive avant implantation (ans)	Coté implan-té	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Pro-thèse contro-latérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
41	**SCHLAU49**	M	39,71	> 30	G	6 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom	ACE	18/22	O		X		X		X
42	**AVECLA50**	M	71,11	> 40	G	3 mois	HiRes 90K/HiFocus ***(Advance Bionic)***	Harmony	HiRes-S with fidelity 120	14/16	O		X		X		
43	**NOVCHR51**	M	51,85	> 40	D	9 mois	Freedom Implant (Contour advance) ***(Cochlear)***	Freedom	ACE	19/22	N		X		X		
44	**DELBRI52**	F	44,96	> 40	G	9 mois	HiRes 90K/HiFocus ***(Advance Bionic)***	Harmony	HiRes-S	16/16	O		X				
45	**FAVPHI53**	M	38,41	NR	G	3 mois	NR	NR	NR	NR	O						X
46	**MARCLA54**	F	74,52	NR	G	9 mois	NR	NR	NR	NR	N						X
47	**MERFAY56**	M	63,14	NR	G	4,3 ans	NR	NR	NR	NR	N						X
48	**RODJOS57**	M	48,59	NR	G	6,8 ans	NR	NR	NR	NR	O						X

N°	Code Sujet	Sexe	Age à l'inclusion	Durée de perte auditive avant implantation (ans)	Coté implanté	Durée d'activation de l'implant à l'inclusion	Type d'implant (fabricant)	Processeur	Stratégie	Electrodes (nb actives / nb disponible)	Prothèse controlatérale	V/NV (S)	V/NV (T)	Genre (S)	Genre (T)	Discri (T)	Pitch /Timbre (T)
49	**CALODI58**	F	69,42	NR	G	8,9 ans	NR	NR	NR	NR	N						X
50	**ANSCHR60**	F	44,76	NR	D	9 mois	NR	NR	NR	NR	O						X
51	**BLOMON61**	F	64,32	NR	D	11 ans	NR	NR	NR	NR	N						X
52	**MESMIC62**	M	56,16	NR	G	6 mois	NR	NR	NR	NR	N						X

Tableau 14.2 : Tableau d'information des patients qui ont participé aux tests de détection de la voix (V/NV), de catégorisation du genre en condition originale (Genre) et en conditions pitch constant et timbre constant (Pitch/Timbre), ainsi qu'au test de discrimination de voix (Discri). Les tests ont été différenciés selon que les patients étaient inclus en suivi (S) ou en transversal (T).

15 PUBLICATIONS

Abstracts :

Z. MASSIDA, J. ROUGER, C. JAMES, P. BELIN, P. BARONE, O. DEGUINE. (2008) Voice gender perception in cochlear implanted patients. 6thFENS, Geneva, july 12-16 2008.

Posters :

1) Z. MASSIDA, J. ROUGER, C. JAMES, P. BELIN, P. BARONE, O. DEGUINE. (2008) Voice gender perception in cochlear implanted patients. 6thFENS, Geneva, july 12-16 2008.

2) Z. MASSIDA ,P. BARONE, O. DEGUINE et P. BELIN. (2007) La reconnaissance de la voix chez les sujets sourds implantés cochléaires. Poster IFR jeune 96, Toulouse, 26 avril 2007.

3) Z. MASSIDA ,P. BARONE, O. DEGUINE et P. BELIN. (2008) La reconnaissance de la voix chez le sujet sourd implanté cochléaire et le sujet normo-entendant en condition de simulation d'implant. Poster IFR jeune 96, Toulouse, 29 avril 2008.

4) Z. MASSIDA, J. ROUGER, C. JAMES, P. BELIN, P. BARONE, O. DEGUINE. (2009) Voice perception in cochlear implanted patients. Poster in Rovereto workshop on crossmodal plasticity, Roveroto, Italy, 27-29 August 2009.

5) Z. MASSIDA, M. MARX, C. JAMES, P. BELIN, O. DEGUINE and P. BARONE. (2010) Voice discrimination through a cochlear implant. Poster at 11th International Conference on Cochlear Implants and other Auditory Implantable Technologies, Stockholm, Sweden, June 30 - July 3 2010.

6) M. MARX, Z. MASSIDA, P. BELIN, C. JAMES, P. BARONE and O. DEGUINE. (2010) Voice feature perception in cochlear implanted patients. Poster at 11th International Conference on Cochlear Implants and other Auditory Implantable Technologies, Stockholm, Sweden, June 30 - July 3 2010.

Communications orales :

1) Z. MASSIDA ,P. BARONE, O. DEGUINE et P. BELIN. (2007) La reconnaissance de la voix chez les sujets sourds implantés cochléaires. Communication au 114 ième Congrès National de la SFORL, Paris, 15 et 16 octobre 2007 (Communication de M le Pr Olivier DEGUINE)

2) Z. MASSIDA ,P. BARONE, O. DEGUINE et P. BELIN. (2008) Perception de la voix chez les patients implantés cochléaire. Communication au 115 ième Congrès National de la SFORL, Paris, 12, 13 et 14 octobre 2008.

3) Z. MASSIDA, J. ROUGER, C. JAMES, P. BELIN, P. BARONE, O. DEGUINE. (2008) Perception de la voix chez les patients implantés cochléaire. Cours supérieur – implant en otologie, Toulouse, 5 décembre 2008.

Articles en préparation dans une revue internationale :

1) Z. MASSIDA, C. JAMES, P. BELIN, P. BARONE, O. DEGUINE. Voice perception in cochlear implanted patients and normal-hearing subjects in cochlear implant simulation.

2) K. STRELNIKOV, Z. MASSIDA, J. ROUGER, P. BARONE, P. BELIN. Effects of speech intelligibility on auditory cortex response to sound vocoding

Elsevier Editorial System(tm) for Hearing Research
Manuscript Draft

Manuscript Number: HEARES-D-10-00154

Title: Voice detection in cochlear-implanted deaf subjects

Article Type: Research paper

Keywords: Voice; Cochlear implant; Vocoder; Voice area; Deafness; Speech

Corresponding Author: Dr. Pascal Barone, Ph.D.

Corresponding Author's Institution:

First Author: Zoé Massida

Order of Authors: Zoé Massida; Pascal Belin; Chris James; Julien Rouger; Bernard Fraysse; Pascal Barone; Olivier Deguine

Abstract: The human voice is important for social communication because voices carry speech and other information such as a person's physical characteristics and affective state. Further restricted temporal cortical regions are specifically involved in voice processing. In cochlear implanted deaf patients, the processor alters the spectral cues which are crucial for the perception of the paralinguistic information of human voices. The aim of this study was to asses the abilities of voice detection in CI users and in normal-hearing subjects (NHS) using a CI simulation (vocoder). In NHS the performances in voice detection decreased when reducing the spectral information by decreasing the number of channels of the vocoder. In CI patients with different delays of implantation we observed a strong impairment in voice detection at time of activation of the neuroprosthesis. No significant improvement can be detected in patients after two years of experience of the implant while they have reached a higher level of recovery of speech intelligibility suggesting a dissociation in the dynamic of functional recuperation of speech and voice processing. In addition to the lack of spectral cues due to the implant processor, we hypothesized that the origin of such deficit could derive from a crossmodal reorganization of the temporal voice areas in CI patients.

Dear Sir,

Please find enclosed a manuscript entitled “Voice detection in cochlear-implanted deaf subjects” we would like to submit to Hearing Research.

This is an original work that has not been published nor submitted to another journal

The present manuscript represents the continuity of the work we have engaged and aimed at understanding the mechanisms of auditory recovery in cochlear implanted deaf patients. While cochlear implantation allows most patients to understand speech, other features of auditory processing important for quality of life are still impaired such as voice recognition. This study is aimed at analyzing the performances of cochlear implanted patients (CIP) engaged in voice perception tasks and how this skill evolves following the experience by the patients in using the neuroprosthesis. To better understand how this deficit is linked to either the limitation of the implant or to a deficit of central origin, we have compared the patients performances to those obtained in normal-hearing subjects (NHS) with a simulation of a cochlear implant (vocoder).
Here we present evidences of a significant deficit in voice detection in patients with a weak improvement over time after several months or years post-implantation. In addition, the patients performances are dissociated from a significant recovery in speech comprehension. In relation to our previous brain imaging study performed in CI patients during a speech-reading task, our work suggest that the cross-modal reorganization that affects the cortical area involved in voice recognition (the Temporal Voice area described by Belin and colleagues) could participate in the observed deficit.

We believe our work provides crucial information about the neuronal mechanisms of the Person Identity Nodes as it brings theoretical information about the normal and hearing-impaired network of voice processing.
Altogether, we believe that our study will make our manuscript of interest for the readers of Hearing Research.

Suggested reviewers

- Christian LORENZI
  - Unit : Equipe Audition, LPP UMR CNRS-Paris Descartes, Département d'Etudes Cognitives, Ecole Normale Supérieure, 29 rue d'Ulm, 75005 Paris
  - Mail : lorenzi@ens.fr or lorenzi@psycho.univ-paris5.fr

- Robert V. SHANNON
  - Unit : House Ear Institute, 2100 West Third Street, Los Angeles, CA 90057
  - Mail : shannon@hei.org or bshannon@hei.org

- Robert J. ZATORRE
  - Unit : Department of Neuropsychology, Room 256, 3801 rue University, Montréal, Québec, Canada, H3A 2B4
  - Mail : robert.zatorre@mcgill.ca

***Research Highlights**

Research highlights

- Vocoder affects voice detection in normal hearing subjects.
- Speech voices detection are les affected by vocoder than non speech voices.
- Cochlear implanted patients have poor performance of voice detection compared to the normal hearing subjects, and do not improve with the experience of implant.

*Manuscript
Click here to view linked References

Massida Z. [1,2], P. Belin[3], C. James[4,5], J. Rouger[1,2], B. Fraysse[5], P. Barone[1,2]* and O. Deguine[1,2,5]

Formatted: From text: 0", Numbering: Continuous

Title: Voice detection in cochlear-implanted deaf subjects.

Shortened title: Voice perception in deaf

1., Université Toulouse, CerCo, Université Paul Sabatier.
2. CNRS, UMR 5549. Faculté de Médecine de Rangueil, Toulouse France
3. Voice Neurocognition Laboratory, Department of Psychology, University of Glasgow, Glasgow, UK.
4. Cochlear France SAS, Toulouse France
5. Service Oto-Rhino-Laryngologie et Oto-Neurologie, Hopital Purpan, Toulouse, France

* Corresponding author

Pascal Barone

e-mail: pascal.barone@cerco.ups-tlse.fr

CNRS-Universite Paul Sabatier Toulouse 3, Centre de Recherche Cerveau et Cognition UMR 5549, Faculté de Médecine de Rangueil, 31062 Toulouse CEDEX9, France.

Phone: +33 (0)5 62 17 37 79

Fax: +33 (0)5 62 17 28 09

Abstract.

The human voice is important for social communication because voices carry speech and other information such as a person's physical characteristics and affective state. Further restricted temporal cortical regions are specifically involved in voice processing. In cochlear implanted deaf patients, the processor alters the spectral cues which are crucial for the perception of the paralinguistic information of human voices. The aim of this study was to asses the abilities of voice detection in CI users and in normal-hearing subjects (NHS) using a CI simulation (vocoder). In NHS the performances in voice detection decreased when reducing the spectral information by decreasing the number of channels of the vocoder. In CI patients with different delays of implantation we observed a strong impairment in voice detection at time of activation of the neuroprosthesis. No significant improvement can be detected in patients after two years of experience of the implant while they have reached a higher level of recovery of speech intelligibility suggesting a dissociation in the dynamic of functional recuperation of speech and voice processing. In addition to the lack of spectral cues due to the implant processor, we hypothesized that the origin of such deficit could derive from a crossmodal reorganization of the temporal voice areas in CI patients.

Keywords.

Voice
Cochlear implant
Vocoder
Voice area
Deafness
Speech

1. **Introduction**

The human voice is a natural auditory stimulus with high ecological and social relevance. Human listeners possess an exquisite ability to extract information about a person's physical characteristics and affective state from their voice. From this perspective, the voice can be considered to be an "auditory face" (Belin et al., 2004): For example we can easily tell the gender of a person simply upon hearing his/her voice- even when there is no speech as in a cough or a laugh. We can often recognize a person simply by hearing a few words on the telephone. Voice perception in social interactions is most important in situations where cues from other modalities (e.g., visual) are absent such as on the radio or the telephone, or simply at a distance or when focus has shifted to another speaker in a meeting.

It is clear that any dysfunction of voice perception can have a large negative impact on the social interactions. This is particularly the case in profoundly deaf patients, with or without cochlear implants (CI). While cochlear implantation allows most patients to understand speech (UKCISG, 2004), other features of auditory processing important for quality of life are still deficient and generally patients have difficulty recognising some voice features such as the gender (Cleary and Pisoni, 2002, Fu et al., 2004, Fu et al., 2005).

Voice perception deficits probably arise from several different origins; among them the sound processing performed by the implant processor could be a major factor. A CI transforms the sound input into a series of electrical impulses that stimulate directly the auditory nerve via 12-22 electrodes implanted into the cochlea. Today, while this sound processing is implemented using a large variety of coding strategies, the auditory information delivered by this neuroprosthesis is generally spectrally degraded. Nevertheless CI coding strategies have been shown to provide acceptable levels of speech comprehension despite considerable spectral degradation (Shannon et al., 1995). However, the situation may be different for perceiving other, non-speech cues in voices. Studies using cochlear implant simulation (vocoding) in normal-hearing subjects have shown that linguistic information is not supported by the same spectrotemporal modulation as for paralinguistic information (Elliott and Theunissen, 2009). As a consequence, the coding strategies developed for speech comprehension may penalize the processing of paralinguistic cues such as occurs for gender discrimination in CI users (Fu et al., 2004, Fu et al., 2005).

The other source of deficit in voice processing in CI users could be dysfunction of the auditory pathway, i.e., a suboptimal reorganization of cortical regions normally devoted to sound processing. At the neuronal level, specific cortical regions mostly located along the

upper bank of the middle and anterior parts of the superior temporal sulcus (STS), are involved in processing voice information (Belin et al., 2002, Belin et al., 2000). The right anterior STS seems particularly involved in processing speaker identity (Belin and Zatorre, 2003), which suggests a neuronal dissociation between speech and voice processing. Recent work has shown that auditory STS regions sensitive to voice stimuli are weakly activated in CI deaf patients (Coez et al., 2008) while on the other hand, auditory presentation of words evokes crossmodal activity in the visual areas (Giraud et al., 2001). Thus one could argue that such cortical crossmodal reorganization induced by deafness could be deleterious to auditory recovery; this could equally apply to voice processing in CI patients.

Our first approach to understanding auditory processing in CI patients was to compare their performance with that observed in normal hearing subjects (NHS) using a noise band vocoder (Fu et al., 2004, Rouger, 2007). Previous studies have analyzed the impact of the vocoding performed by a cochlear implant processor on word recognition (Moore et al., 2008, Shannon et al., 1995); degrading signals in this way has been shown to affect differently the brain areas of the cortical network involved in speech comprehension and voice recognition (Obleser et al., 2008, Scott et al., 2000, Scott et al., 2006, Warren et al., 2006). But what is the effect of the implant processor on voice discrimination per se? To our knowledge no study has looked precisely into the discrimination of voice stimuli from natural environmental sounds in CI patients or in NHS listening via noise band vocoders. The aim of the present study was to compare the performance on a 'simple' voice/non-voice detection task of the two populations of subjects (normal hearing and deaf subjects) using a longitudinal approach in patients to reveal adaptive strategies developed following experience with a cochlear implant neuro-prosthesis.

Because speech comprehension and voice recognition are based on different spectral and temporal auditory cues (Elliott and Theunissen, 2009), we hypothesized that performance on the voice detection task would show a different profile of degradation with vocoding/CI than for speech perception.

2. Materials and Methods

2.1 Participants.

28 normally hearing (NHS) native French speakers (14 males; age: mean ±s.d.: 23.8 ± 2.4) with no self-reported history of auditory, neurological or psychiatric disorders

participated in the study. NHS were divided into two groups. The first group (Group A; n=14) was tested on a voice recognition task using original, untransformed stimuli and 4 vocoder conditions with 64, 32, 16 and 8 channels. Group B, (n=14) was tested using vocoder conditions of 16, 8, 4 and 2 channels.

30 cochlear implanted (CI) deaf adult patients (12 males, 18 females; age: 53.5 ± 15) took part in the study. The CI patients are on average older than the control hearing subjects but their age range from 20 to 80 years. In consequence we adopted the strategy of comparing the patient performance in voice discrimination with that obtained from an homogeneous group of young hearing subjects with no evidences of hearing loss. Performance in voice recognition was collected during regular visits to the ENT department following a standard rehabilitation program. Fifteen subjects (5 men) were right-implanted subjects. Post-implantation time varied from 1 day to 131 months and patients were divided into 4 different sub-groups according to duration of implantation. All patients had postlingually acquired profound bilateral deafness of diverse etiologies (meningitis, chronic otitis, otosclerosis, etc.) and durations. Only 1 patient presented a sudden deafness which occurred 3 years before the cochlear implantation. In all the other patients, the deafness was progressive, and the duration of hearing loss for each patient is shown in table 1. Because of this progressive hearing impairment, the duration of deafness could not be reliably defined and consequently we did not attempt to correlate this measure with any of the performance levels presented by the patients. Further, 15 patients had a hearing aid in the non-implanted ear and use it in daily life, but they were always tested without the prosthesis. All the information concerning the patients is provided in Table 1. All participants gave written informed consent prior to their inclusion in the study.

2.2 Stimulus material.

All stimuli used in our experiment came from a database of vocal and nonvocal sounds used in previous experiments (Belin et al., 2002, Belin et al., 2000). Two sets of 500-msec long stimuli were created: the first set contained 55 different human voice stimuli, including 29 speech stimuli (phonemes presented in a /h/vowel/d/ context, words in different languages) and 26 non-speech vocal stimuli (e.g., laughs, coughs). The second set contained 55 non-voice stimuli consisting of a wide variety of environmental sounds, including sounds from cars, telephones, bells, streaming water. Neither group contained animal vocalizations.

Based on these stimuli, we created 6 vocoder conditions simulating a cochlear implant with different numbers of channels, using MatLab 6.5 (Mathworks, Inc.) with a speech

spectrum shape noise (SSSN) procedure (Rouger et al., 2007). The sound was analyzed through 2, 4, 8, 16, 32 or 64 frequency bands by using sixth-order IIR elliptical analysis filters. The cutoff frequencies of these bands were calculated to ensure equidistance of the corresponding basilar membrane locations according to the human cochlear tonotopic map (Greenwood 1990). The total bandwidth was 125 and 8,000 Hz. For each filtered frequency band signal, the temporal envelope was extracted by half-wave rectification and envelope smoothing with a 500-Hz low-pass third-order IIR elliptical filter. The extracted temporal envelope was then used to modulate white noise delivered by a pseudorandom number generator, and the resulting signal was filtered through the same sixth-order IIR elliptical filter that was used for the frequency band selection. Finally, signals obtained from each frequency band were recombined additively, and the overall acoustic level was readjusted to match the original sound level.

We performed an analysis of the effect of the noise band-vocoder on the frequency representation of each stimulus by comparing the average power spectrum differences between categories of sounds (i.e. Voice vs. Non Voice and Speech vs. Non Speech). As shown in figure 1, there were no major differences in the spectra of the stimuli that might have been produced by the different vocoder conditions.

2.3 Stimulus presentation and procedure.

Subjects were tested in a sound-attenuated chamber with volume adjusted to 72 dB SPL. NHS were tested in the CerCo Laboratory and CIS in the Hospital of Purpan. Stimuli (16-bits, mono, 22050 Hz sampling rate) were presented binaurally to the control group via Sennheiser Eh 250 headphones in a pseudo-random order. The stimuli were presented to the CI users in free field conditions through loudspeakers (KINYO, model PS-240).

The NHS group A was tested with the original voice and non voice stimuli (OV) and with degraded stimuli vocoded with 64, 32 16 and 8 channels; Group B NHS were tested with degraded stimuli vocoded with 16, 8, 4 and 2 channels. The different vocoder conditions were presented in blocks in a pseudo-random order across subjects, except for the OV condition which was always presented last in the testing session. The task for NHS and CIS was a 2-alternative forced choice (2AFC) categorization: voice vs. non-voice. NHS were tested with a 1-s inter-trial delay and were instructed to respond as quickly and accurately as possible using the left or right control buttons of the computer keyboard corresponding to their answer (voice or non-voice). The response keys were counterbalanced across subjects. CIS were tested in original stimulus conditions with a 1.5-s inter-trial delay and were instructed to answer as

accurately as possible, with no reference to reaction time. The short duration of the stimuli makes the task difficult for the patients, and in few trials some patients did not provide a firm answer. Such cases were essentially present in subjects tested at the first day of activation of the implant and was observed in only 14% of the presentations. In experienced patients, such behavior occurred in less than 2% of the trials. In consequence, these trials were excluded from the analysis and only the positive responses were conserved. Each condition lasted 5 minutes for NHS and 7 to 10 minutes for CIS. Reaction times were recorded only for NHS.

We compared the speech recognition scores of CI patients to that obtained with NHS tested using the same stimuli vocoded with 4 channels. To do such comparison we used previously published data (Rouger et al., 2007), where subjects were tested on open-set recognition of French Fournier disyllabic words presented in auditory-only condition. The stimuli were vocoded using 2, 4, 8, or 16 channels, and these we used to compare CI users and NHS in the 4 channels condition.

Analysis. Using Signal Detection Theory (SDT, (Green and Swets, 1966, Tanner and Swets, 1954)) we measured d' which is a criterion of perception sensitivity independent from decision bias. Hit Rate (HR) and False Alarm Rate (FAR) were calculated assuming a detection task in which voices are the target. Similar analyses were performed for either the Speech stimuli or the Non-Speech Voices as the targets.

Analysis of reaction times was performed independently for Voice and Non-voice stimuli. RT for correct response and errors were similar to RT for correct response alone (ANOVA; $p < 0{,}005$). The analysis was performed only on correct responses (Hits and correct rejections).

To analyze the global effect of the vocoder on d' and reaction time measures, we used an ANOVA for repeated measures for ths NHS and an ANOVA for independent measures for CI patient. Bonferroni/Dunn test was used for post-hoc analysis. ANOVA was also used to compare d' between CI patient and NHS. A t-test was used to measure differences between voice and non voice (for RT) and between speech voices and non speech voices.

3. Results

3.1 Voice discrimination in normal hearing subjects.

Effect of vocoder on voice discrimination. When presented with the original sound (OV), the NHS showed good abilities to discriminate voice stimuli from environmental sounds as expressed by high d' values (Fig 1). With decreasing numbers of vocoder channels, voice detection performance strongly decreased to low values but above chance level for the 2-channel condition. In each group, our results showed that reducing the number of vocoder channels produced a significant reduction in voice detection performance (ANOVA on d': group A, $p<0.0001$; group B, $p<0.0001$). Compared to the original stimuli, even the smallest amount of spectral degradation (64 channels) already significantly reduced performance (Bonferroni/Dunn: $p = 0.0013$). Further, for both groups, all paired comparisons between conditions showed significant differences (Bonferroni/Dunn test all comparisons $p<0.05$) except in group A for which performance was similar between the 32 and 64 channel conditions (Bonferroni/Dunn: $p=0.5$) and the 8 and 16 channel conditions (Bonferroni/Dunn: $p=0.09$) and in group B for which performance was similar between the 16 and 8 channel conditions (Bonferroni/Dunn: $p=0.02$) and between the 4 and 2 channel conditions (Bonferroni/Dunn: $p=0.07$).

The16 and 8 channel conditions were tested in both groups and group A showed significantly better performances in both of these two conditions (t-test; $p<0.005$ in both cases).

Comparison between Speech and Non-Speech Voice Stimuli. Here we analyzed voice detection performance separately for Speech and Non-Speech Vocal sounds (Fig 2). This analysis confirmed that the ability to discriminate Speech or Non-Speech Voices from environmental sounds was strongly impaired when the number of vocoder channels was decreased. The decrease in d' was statistically significant for both groups for Speech Voices (ANOVA: group A: $p=0.0006$; group B: $p<0.0001$) and for Non-Speech Voices (group A: $p<0.0001$; group B, $p<0.0001$). However, the performances of detecting the two types of voice stimuli from environmental sounds were not comparable and were not similarly altered by vocoding. First of all, subjects were slightly better at detecting Speech Voices than at detecting Non-Speech Voices when listening to the original stimuli and in the vocoder condition 64 channels, but this tendency was not significantly different (d', t-test: $p>0.05$). Significantly higher performance in detecting Speech Voice was observed for all the other

vocoder conditions, from 32 to 2 channels (t-test, all comparisons $p<0.0026$) in addition to overall decreasing performance. Secondly, the rate of decrease in Drime values with reducing the number of channel was not the same for the two types of voice stimuli: For Non-speech Voices we observed a continuous and steep decrease in d' from 64 to 2 channels reduction, while a drop in performances for Speech Voices was only apparent following a reduction to 8 vocoder channels. When using original condition or 64, 32 or 16 channels of vocoder conditions, the d's values for speech stimuli were similar (Bonferroni/Dunn test all comparisons $p>0.05$). Further, the detection performance for Speech and Non-speech Voices was similar for the 16 and 8 channel and for the 4 and 2 channel conditions. But Non-speech Voices performance remained low and significantly below that of for Speech Voice stimuli.

Effect of acoustic parameters on performances. Firstly, we measured the mean pitch for Voices, SpeechVoices, Non-Speech Voices and for Non-Voice stimuli where possible in this later case (43/55 stimuli). The mean respective pitches were of 252.7 Hz (±88.9), 193.4 Hz (±50.1), 318.9 7 Hz (±75.1) and 257.6 7 Hz (±132.2). This analysis revealed no significant difference in mean pitch between Voice and Non-Voice stimuli (ANOVA: $p>0.05$). However, both Speech-Voices, and Non-Speech Voices were significantly different from the Non-Voices stimuli in term of pitch (Bonferroni/Dunn test all comparisons $p<0.016$). These differences cannot provide cues to help the subjects in detecting Voices stimuli because the mean pitch values of Speech-Voices and Non-Speech Voice are flanking at equal distances the pitch values of the Non-Voice stimuli. Further, because the task imposed to the subject was to categorize the stimuli with respect to a Voice/Non-Voice criteria, our results cannot be explained by a unique strategy of the subjects based on pitch which is strongly affected by the vocoding (Xu et al., 2002).

Secondly, we looked for a possible correlation between the performance in Voice/Non-voice discrimination and the temporal macroscopic structure that composed the stimuli. Speech voices were distinguished by the number of syllables (i.e. 1, 2 or 3), and Non-Speech Voices were distinguished by the number of elementary temporal elements (i.e. 1, 2 or 3). This analysis show that there was no statistical relation between the performance of detection and the number of temporal elements in both type while reducing the information of vocoder (ANOVA test $p>0.05$)

Measures of reaction time (RT). In Group A, analysis of the reaction times (RTs) showed that the reduction in the spectro-temporal information by the vocoder induced an increase in RT

for all types of stimuli (Fig 3). A statistically significant effect of vocoding was found on RTs for Speech and Non-Speech Voices as well as for Non-Voices stimuli (ANOVA, $p<0.0004$ all cases). For the group B, tested with the lower numbers of channels, no statistical effect of vocoding on RT was found suggesting that in these conditions the subjects have reached the longest RTs values with respect to the task instructions.

A first comparison of the reaction time for Voice and Non-voice stimuli (not shown) revealed that the values were not statistically different (t-test all comparisons $p>0.05$). Similarly, the distinction between Speech and Non Speech voice stimuli (fig 4) revealed that for both groups the subjects tended to respond more slowly to Non-Speech Voice stimuli than to any other type, but this tendency was only significant for the group A in 8 channels condition (t-test; $p=0.0313$).

3.2 Voice discrimination in cochlear implanted deaf subjects.

Voice discrimination performance. We have analyzed voice detection performance in a large cohort of CI deaf patients (see methods) according to the duration of experience with the neuroprothesis (Fig 4a). First we have included a group of patients at the time of the CI activation (T0, n = 9) i.e. without any adaptation to the new auditory stimulation. Three other sub-groups were established, from less than 6 months (n = 5), between 6 and 18 months (n = 8) or over 18 months (n = 8) of experience of the auditory stimulation through the implant.

CI patients had a strong deficit in discriminating Voice stimuli from environmental sound stimuli at the time of activation of the implant. This was expressed by performance levels near chance (Hit rates of about 60%) and low d' values (Fig 4a). After 18 months of CI experience, patients were better at voice detection but their performance remained lower than that observed in NHS (d': t-test: $p<0.0001$ for both speech or non-speech voice stimuli). In fact, when analyzing the performance of the 4 groups of patients (Fig 4a), while d' values seemed to increase with time after implantation, this tendency was not significant (ANOVA; $p>0.05$). However, as for NHS stimulated with the noise vocoded stimuli, CI patients had better voice detection performance for Speech than for non-speech vocal stimuli, as indicated by significantly different d' values after more than 18 months of implantation (t-test: $p= 0.0168$).

About half of the patients presented a residual hearing and consequently they were using an auditory prosthesis in their daily life. While the patients were tested only with the cochlear implant, we have investigated if this complementary hearing aid could improve the patient performances in voice detection. When comparing patients with or without a prosthesis, we

did not found a significative difference in d' values (respectively 1.702 and 1.133, t-test, p>0.05). Because of the small number of patients in each group, this analysis was performed by combining d' values of all the patients tested after more than a month of experience of the cochlear implant.

Next, we compared the performance of CI patients with that of normal hearing subjects tested with the vocoder conditions that mimic the processing of the cochlear implant. At the time of implant activation, patients performed not better than an untrained NHS tested with a 2-channels vocoder (t-test: $p<0.05$ for Speech and Non Speech). After more than a year and a half of experience, CI patients' ability to discriminate a human voice from environmental sounds was comparable to that of an untrained NHS tested with a 4 channel vocoder (t-test: $p>0.05$ for Speech and Non Speech Voice) and slightly above that of NHS listening to 2 channel vocoder.

Relation between speech intelligibility and voice detection. When the fine temporal cues within each spectral component are removed via vocoding, NHS have reduced speech recognition (Rouger et al., 2007, Shannon et al., 1995). Disyllabic word recognition through a 4 channel vocoder does not exceed 30% (see fig 4b, data from (Rouger et al., 2007). Thus we tried to assess if the auditory speech comprehension and voice detection are similarly affected by vocoding in NHS and CI patients. In agreement with our previous study on a large population of patients (Rouger et al., 2007), the present group of CI users presented a significant increase in word comprehension following implantation (ANOVA: $p<0,0001$). They reached the impressive level of over 70% word recognition after 6-18 month post implantation (Fig 4b). While not optimal, this performance level is comparable to that observed in untrained NHS with an 8 channels vocoder (68%). This level of speech recognition in CI users – even after only a few months of CI experience - is much higher that the score obtained by the untrained normal hearing subjects through a 4 channel vocoder (Fig 4b). We chose this comparison because during the voice detection task, CI patient performance was equivalent to that observed in NHS with a 4 channel vocoding stimulation. Thus these data showed that while impaired at voice detection, patients can reach a high level of speech recognition suggesting a possible dissociation between the two recovery processes.
In line with that, we searched for a correlation between performance levels in the two tasks: speech comprehension and voice detection (Fig 5). In the full set of CI patients (n = 30) that

encompass both novices and experienced users, the analysis revealed a low but significant correlation (Fisher test, $r^2 = 0.19$, $p = 0.0174$) between patient performance in word recognition (in %) and in voice discrimination task (d'). A significant correlation was also observed when we distinguish the performances in Speech and Non Speech Voice stimuli separately (Fisher test, r2 = 0.16 and 0.13 respectively, $p < 0.05$ both cases). Indeed, there was a tendency for patients with a significant recovery in speech recognition to present also a recovery in voice discrimination although in the latter case, performance levels remained far below that observed in NHS. However, when patient performance is considered at time of implant activation (first day), no significant correlation was obtained (Fisher test, r2 = 0.03, $p > 0.05$) in this limited subgroup (n = 8). Similarly in the experienced patients (over 18months post CI, n = 8) only a non-significant trend is observed (Fisher test, r2 = 0.48, $p > 0.05$). Altogether, these results show that progress in word recognition is also associated with progression in voice detection, but voice detection remains low. This suggests that the dynamics of the recovery of the two mechanisms in CI patients are different.

4. Discussion

Three main findings emerge from this study: First, we found that voice detection performance in NHS is strongly affected by vocoding and number of vocoder channels, and that even experienced CI patients never attain levels of performance comparable to NHS. Second, voice detection performance is consistently better for speech than for non-speech vocal sounds, both in NHS and in CI patients. Third, voice detection performance appears more affected than speech comprehension by either vocoding or CI use. These findings have important implications for CI coding strategies aimed at enhancing the processing of non-speech vocal information by CI patients.

Voice detection with vocoding or CI.

Previous studies have shown that when reducing the spectro-temporal information such as with a noise-band vocoder, speech recognition performance is strongly reduced (Elliott and Theunissen, 2009, Fu et al., 2004, Loizou et al., 1999, Lorenzi et al., 2006), although it remains relatively high even with very few vocoder channels (Shannon et al., 1995). Such reduction in spectral structure corresponds to the coding strategies implemented in cochlear implant processors and it can partly explain the reduced intelligibility of speech observed in

CI deaf patients (David et al., 2003, Friesen et al., 2001, Mulder et al., 1992, Proops et al., 1999, Spahr and Dorman, 2004).

When degrading the sound via vocoding, both gender and speaker identification and vocal emotion recognition are impaired in normal hearing subjects as in CI deaf users (Fu et al., 2004, Fu et al., 2005, Gonzalez and Oliver, 2005, Kovacic and Balaban, 2009, Luo et al., 2007). The present results generalize these previous finding and show that when the spectral information is reduced both NHS and CI patients present a strong impairment in discriminating voice stimuli from environmental sounds. In patients, voice detection performance remains low even after many months of CI experience, and the observed performance is in the same range (about 70% correct) as that observed for gender (Fu et al., 2004) or talker discrimination (Cleary et al., 2005).

The observed voice deficit in both NHS and CI users may be due to a lack of spectral information as speaker or gender-related features are based on the vocal tract anatomy that influences the acoustic structures of the voice, pitch and timbre (Belin, 2006, Fitch and Giedd, 1999, Roers et al., 2009, Smith et al., 2007, Story et al., 2001, Takemoto et al., 2006, Xue and Hao, 2006). Similarly, the ability to identify environmental sounds are also strongly impaired when the spectral resolution is reduced both in normal subjects (Shafiro, 2008) and in implanted deaf patients (Proops et al., 1999, Reed and Delhorne, 2005). In the latter group, in agreement with the present results, the capacity to recognize environmental sounds does not seem to improve greatly over time (Tye-Murray et al., 1992).

Voice detection for speech vs non-speech voice sounds with vocoding or CI

One of the most striking results from our study is a clear difference in rate of recognition between speech and non-speech voice stimuli. In both NHS listening to a noise-band vocoder and in CI patients, the d' values for the discrimination of speech voice stimuli are significantly higher than that obtained with the non-speech voice stimuli (laughs, coughs, …).

Several different but non-exclusive explanations can be suggested. First the two sets of stimuli present distinct mean pitch values (193 vs. 319 Hz for speech and non-speech voice stimuli respectively), and the subjects discrimination can be based on such differences. This explanation could stand for the patients' performances while we know that pitch recognition is impaired with the vocoder (Green et al., 2002) and the CI (Chatterjee and Peng, 2008, Donnelly et al., 2009, Gfeller et al., 2007). Furthermore, with respect to the frequency ranges

attributed to individual channels in each vocoding conditions, such pitch cues would not be available to the normal hearing subjects in the 2 and 4 channels conditions. In these particular cases the values of pitch that differentiate the two sets of stimuli fall into the same band of vocoder. However, in these conditions of reduced spectral information, NHS presents higher sensitivity to speech-voice than to non-speech voice stimuli, a result suggesting that the detection of these two sets of stimuli is not based on pitch soly.

More probably, we can consider that speech and non-speech are differently affected by vocoder. The speech and nonspeech vocal sounds are also characterized by partly different acoustical properties, which could be differentially affected by vocoding. In particular, speech sounds are characterized by fast, broadband changes related to articulation, which are relatively well preserved by vocoding. The presence of these fast, broadband changes in a vocoded stimulus could provide a cue to 'speechness', and hence 'voiceness', which survives vocoding well since they are essentially temporal cues. On the contrary, nonspeech vocal sounds: i) are much more heterogeneous in structure (e.g., coughs, laughs, screams, etc.) than speech sounds, thus they might be more difficult to identify based on a few salient acoustic cues such as the fast phonemic changes in speech; ii) contain identity or affective information largely in their fine spectral structure, which is disrupted by vocoding, and thus not available to the listener as a cue to identity a voice stimulus.

Voice detection vs. speech comprehension with vocoding or CI.

One interesting finding is that speech comprehension is less impaired than voice detection. Actually, CI patients were able to reach high levels of speech recognition with experience, while there was little or no evolution of their voice detection performance over time; experienced CI users were as poor as naïve NHS with 4-channel vocoding for voice detection. This difference can be explained first by the lexical contents of words which, even when the word is partially understood, will help the patient to complete recognition. Because there was no word in the voice detection task, there were no lexical cues to help patients increase their performance. Secondly, these two different tasks are supported by different acoustical cues; word comprehension leans more on temporal cues, which survive well after vocoding (Shannon et al., 1995), while voice (especially non-speech-voice) detection rely more on spectral cues which are very degraded by vocoding (Fu et al., 2004, Fu et al., 2005, Gonzalez and Oliver, 2005).

This would be in agreement with the previous analysis showing divergent effects of vocoding on speech features, for example recognition of voicing in a vowel requires more spectral information and is less affected by the reduction in temporal information (Fu et al., 2004, Loebach and Pisoni, 2008).

Further, studies in normal listeners have shown that a high level of speech recognition can be achieved even with only 4 channels (Loizou et al., 1999, Shannon et al., 1995) especially after training (Loebach and Pisoni, 2008). Similarly in deaf patients, high levels of speech intelligibility are observed after few months of experience of the cochlear implant (this study; (Rouger et al., 2007, UKCISG, 2004) probably because the semantic contents of the words provide complementary information in case of incomplete comprehension. Such effect could explain why in our group of deaf patients, we observed a good recovery of word recognition after implantation while the voice discrimination performance remains low. However, it has been proposed that speech is "special", an hypothesis which may explain better discrimination of speech-voice stimuli both in CI users and NHS. Previous reports have revealed that the capacity to perform some tasks on artificially degraded speech material are independent of the temporal or spectral abilities of subjects (Kishon-Rabin et al., 2009, Surprenant and Watson, 2001, Karlin, 1942) suggesting that speech recognition can involve some specific cognitive processes (Kidd et al., 2007). While environmental sounds can be categorized according to their acoustical features (Gygi et al., 2007), they can also be classified according to the sources or events that produce them (Gaver, 1993). Conversely, speech is produced by a unique source, the vocal human tract, a singularity that reduces the variability of spectra compared to environmental sounds. Such particularity might confer to speech sounds a specific familiarity feature used to built-up guessing strategies leading to more efficient recognition of speech sounds compared to non-speech voice or natural sounds (Kidd et al., 2007). Thus we cannot exclude that such mechanisms were involved at least partly in the performance of CI and NH subjects on the speech-voice discrimination test developed in the present study. It is noteworthy that a dissociation in processing voice-speech and non-speech voice stimuli is present in babies in paradigms of source identification (Vouloumanos et al., 2009).

Brain activity during voice discrimination through a cochlear implant.

The vocoding manipulation alters the spectral information (fine structure) while the temporal envelope is preserved. Such degradation of the sound alters the pattern of activity in the auditory system, but by increasing the number of frequency bands the auditory nerve

produces a temporal pattern of auditory response close to normal (Loebach and Wickesberg, 2006). At the cortical level and in agreement with a left-right hemispheric dissociation of processing spectral and temporal features (Zatorre and Belin, 2001), the alteration of speech through a vocoder revealed a sensitivity of the right temporal lobe to variations in spectral structure (Obleser et al., 2008, Scott et al., 2006). Such an effect is extended to the right STS when voice analysis mechanisms are involved (Warren et al., 2006). The superior temporal sulcus, and specifically the right STS has been shown to be sensitive to voice stimuli (speech, laugh, cough …) (Belin and Zatorre, 2003, Belin et al., 2002) and could be the cortical substratum of the lowest level of "voice structural analysis".

In consequence, it could be argued that the deficit in voice recognition we reported in CI patients could arise from a lack of functional activation of these temporal areas through the sound processor. Indeed, a recent brain imaging study reported a reduction in activation of the voice sensitive regions in experienced cochlear implanted subjects (Coez et al., 2008). Of importance, this absence in sensitivity of the temporal voice area (Belin et al., 2000), or TVA, to human voice stimuli is linked to the recovery of speech comprehension observed in the patients (Coez et al., 2008). In this study, patients were divided into 2 groups: the first group had good word comprehension (> 80%) and the second had poor performance (< 20%). The STS activation during passive listening of voice stimuli was significantly and bilaterally reduced in the first group (poor performance) compared with that of the control subjects while the second group (good performance) was not significantly different from the control subjects. While not explicitly tested in this study, it is highly probable that these patients with poor word comprehension recovery would be also impaired in the present task of voice/non-voice discrimination. These results could correspond to our longitudinal analysis as we showed that a recovery in speech comprehension is accompanied by higher levels of voice discrimination.

But how can CI patients present a relatively normal activation of the voice sensitive areas (as shown in the (Coez et al., 2008) and be deficient in a voice discrimination task (as shown in the present study)? Our hypothesis is that the lack of activation of the STS could reflect crossmodal reorganization induced by the prolonged period of deafness. We have shown that CI deaf patients rely strongly on visual information (speechreading) to compensate to the crude auditory information provided by the implant (Rouger et al., 2008, Strelnikov et al., 2009). Such a visual skill preserved years after implantation leads to near optimal performances of word comprehension during visuo-auditory presentation (Rouger et al., 2007). Further, using PET scan brain imaging in CI deaf patients performing a speechreading

task we reported at the initial stages post-implantation, an activation of the right STS by the visual cues. A year later, when the patients have recovered the auditory speech recognition these areas of the STS are no longer involved in speechreading. Thus we propose that in Coez et al (2008)'s study, the patients belonging to the "poor group" rely probably more on visual speechreading, in a situation and that implies a crossmodal involvement of the TVA and induces a poor responsiveness to voice stimuli. Conversely, the patients of the "good group" would be less dependent on speechreading which is no longer supported by the STS and in turn present a near-normal level of activation by voice stimuli. However, as suggested by our present results, this reorganization toward a normal functional specificity for human voice is probably not sufficient to support good voice recognition or detection when degraded by the sound processor of the implant.

Conclusion. Our study shows that while cochlear implantation allows most patients to understand speech, other features of auditory processing important for quality of life should now also be improved, such as voice perception. Indeed, algorithms optimized for speech processing and based on vocoding with a limited number of channels allowing good temporal resolution but poor spectral resolution, are not optimal for fine processing of the spectral structures of voices. However, it is also possible that performance levels in voice discrimination observed in patients could depend on the magnitude of crossmodal reorganization of the cortical regions in the STS which are sensitive to the human voice. Such role of crossmodal reorganization in cochlear implant outcomes has been previously reported when considering speech recognition (Lee et al., 2001). Similarly, further studies are needed to correlate individual performance to the activity pattern during speech recognition and voice discrimination tasks in CI users. Lastly, while CI user rehabilitation is essentially axed on speech recognition, a specific training on voice recognition should be proposed because it is probable that a better voice detection should help patients to better understand speech as shown when the voice speaker is familiar.

Acknowledgments.

We thank the cochlear-implanted and normally hearing subjects for their participation in this study, Marie-Laurence Laborde for help in collecting the data, C. Marlot for help in bibliography. This work was supported by a Cifre Convention to ZM (Cochlear France SAS-

ANRT N°979/2006), the ANR Hearing Loss (ANR-06-Neuro-021-04) and the recurrent funding of the CNRS.

References

Belin, P. 2006. Voice processing in human and non-human primates. Philosophical transactions of the Royal Society of London 361, 2091-107.

Belin, P., Zatorre, R.J. 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105-9.

Belin, P., Zatorre, R.J., Ahad, P. 2002. Human temporal-lobe response to vocal sounds. Brain research 13, 17-26.

Belin, P., Fecteau, S., Bedard, C. 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8, 129-35.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B. 2000. Voice-selective areas in human auditory cortex. Nature 403, 309-12.

Chatterjee, M., Peng, S.C. 2008. Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition. Hearing research 235, 143-56.

Cleary, M., Pisoni, D.B. 2002. Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. The Annals of otology, rhinology & laryngology 189, 113-8.

Cleary, M., Pisoni, D.B., Kirk, K.I. 2005. Influence of voice similarity on talker discrimination in children with normal hearing and children with cochlear implants. J Speech Lang Hear Res 48, 204-23.

Coez, A., Zilbovicius, M., Ferrary, E., Bouccara, D., Mosnier, I., Ambert-Dahan, E., Bizaguet, E., Syrota, A., Samson, Y., Sterkers, O. 2008. Cochlear implant benefits in deafness rehabilitation: PET study of temporal voice activations. J Nucl Med 49, 60-7.

David, E.E., Ostroff, J.M., Shipp, D., Nedzelski, J.M., Chen, J.M., Parnes, L.S., Zimmerman, K., Schramm, D., Seguin, C. 2003. Speech coding strategies and revised cochlear implant candidacy: an analysis of post-implant performance. Otol Neurotol 24, 228-33.

Donnelly, P.J., Guo, B.Z., Limb, C.J. 2009. Perceptual fusion of polyphonic pitch in cochlear implant users. The Journal of the Acoustical Society of America 126, EL128-33.

Elliott, T.M., Theunissen, F.E. 2009. The modulation transfer function for speech intelligibility. PLoS computational biology 5, e1000302.

Fitch, W.T., Giedd, J. 1999. Morphology and development of the human vocal tract: a study using magnetic resonance imaging. The Journal of the Acoustical Society of America 106, 1511-22.

Friesen, L.M., Shannon, R.V., Baskent, D., Wang, X. 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. The Journal of the Acoustical Society of America 110, 1150-63.

Fu, Q.J., Chinchilla, S., Galvin, J.J. 2004. The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. J Assoc Res Otolaryngol 5, 253-60.

Fu, Q.J., Chinchilla, S., Nogaki, G., Galvin, J.J., 3rd. 2005. Voice gender identification by cochlear implant users: the role of spectral and temporal resolution. The Journal of the Acoustical Society of America 118, 1711-8.

610 Gaver, W.W. 1993. What in the World Do We Hear?: An Ecological Approach to Auditory
611 Event Perception. Ecological Psychology 5, 1-29.
612 Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., Olszewski, C. 2007.
613 Accuracy of cochlear implant recipients on pitch perception, melody recognition, and
614 speech reception in noise. Ear and hearing 28, 412-23.
615 Giraud, A.L., Price, C.J., Graham, J.M., Truy, E., Frackowiak, R.S. 2001. Cross-modal
616 plasticity underpins language recovery after cochlear implantation. Neuron 30, 657-
617 63.
618 Gonzalez, J., Oliver, J.C. 2005. Gender and speaker identification as a function of the number
619 of channels in spectrally reduced speech. The Journal of the Acoustical Society of
620 America 118, 461-70.
621 Green, D., Swets, J. 1966. Signal Detection Theory and Psychophysics John Wiley and Sons,
622 New York.
623 Green, T., Faulkner, A., Rosen, S. 2002. Spectral and temporal cues to pitch in noise-excited
624 vocoder simulations of continuous-interleaved-sampling cochlear implants. The
625 Journal of the Acoustical Society of America 112, 2155-64.
626 Gygi, B., Kidd, G.R., Watson, C.S. 2007. Similarity and categorization of environmental
627 sounds. Perception & psychophysics 69, 839-55.
628 Karlin, J.E. 1942. A FACTORIAL STUDY OF AUDITORY FUNCTION.
629 PSYCHOMETRIKA 7.
630 Kidd, G.R., Watson, C.S., Gygi, B. 2007. Individual differences in auditory abilities. The
631 Journal of the Acoustical Society of America 122, 418-35.
632 Kishon-Rabin, L., Taitelbaum-Swead, R., Salomon, R., Slutzkin, M., Amir, N. 2009. Are
633 changes in pitch and formants enough to influence talker normalization processes in
634 children and adults? Journal of basic and clinical physiology and pharmacology 20,
635 219-32.
636 Kovacic, D., Balaban, E. 2009. Voice gender perception by cochlear implantees. The Journal
637 of the Acoustical Society of America 126, 762-75.
638 Lee, D.S., Lee, J.S., Oh, S.H., Kim, S.K., Kim, J.W., Chung, J.K., Lee, M.C., Kim, C.S. 2001.
639 Cross-modal plasticity and cochlear implants. Nature 409, 149-50.
640 Loebach, J.L., Wickesberg, R.E. 2006. The representation of noise vocoded speech in the
641 auditory nerve of the chinchilla: physiological correlates of the perception of
642 spectrally reduced speech. Hearing research 213, 130-44.
643 Loebach, J.L., Pisoni, D.B. 2008. Perceptual learning of spectrally degraded speech and
644 environmental sounds. The Journal of the Acoustical Society of America 123, 1126-
645 39.
646 Loizou, P.C., Dorman, M., Tu, Z. 1999. On the number of channels needed to understand
647 speech. The Journal of the Acoustical Society of America 106, 2097-103.
648 Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C. 2006. Speech perception problems
649 of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad
650 Sci U S A 103, 18866-9.
651 Luo, X., Fu, Q.J., Galvin, J.J., 3rd. 2007. Vocal emotion recognition by normal-hearing
652 listeners and cochlear implant users. Trends Amplif 11, 301-15.
653 Moore, B.C., Tyler, L.K., Marslen-Wilson, W. 2008. Introduction. The perception of speech:
654 from sound to meaning. Philosophical transactions of the Royal Society of London
655 363, 917-21.
656 Mulder, H.E., Van Olphen, A.F., Bosman, A., Smoorenburg, G.F. 1992. Phoneme recognition
657 by deaf individuals using the multichannel nucleus cochlear implant. Acta Otolaryngol
658 112, 946-55.

659 Obleser, J., Eisner, F., Kotz, S.A. 2008. Bilateral speech comprehension reflects differential
660 sensitivity to spectral and temporal features. J Neurosci 28, 8116-23.
661 Proops, D.W., Donaldson, I., Cooper, H.R., Thomas, J., Burrell, S.P., Stoddart, R.L., Moore,
662 A., Cheshire, I.M. 1999. Outcomes from adult implantation, the first 100 patients. The
663 Journal of laryngology and otology 24, 5-13.
664 Reed, C.M., Delhorne, L.A. 2005. Reception of environmental sounds through cochlear
665 implants. Ear and hearing 26, 48-61.
666 Roers, F., Murbe, D., Sundberg, J. 2009. Predicted singers' vocal fold lengths and voice
667 classification-a study of x-ray morphological measures. J Voice 23, 408-13.
668 Rouger, J. 2007. Perception audiovisuelle de la parole chez le sourd postlingual implanté
669 cochléaire et le sujet normo-entendant : étude longitudinale psychophysique et
670 neurofonctionnelle., Université de Toulouse 3 - Paul Sabatier, Toulouse.
671 Rouger, J., Fraysse, B., Deguine, O., Barone, P. 2008. McGurk effects in cochlear-implanted
672 deaf subjects. Brain Res 1188, 87-99.
673 Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P. 2007. Evidence that
674 cochlear-implanted deaf patients are better multisensory integrators. Proc Natl Acad
675 Sci U S A 104, 7295-300.
676 Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J. 2000. Identification of a pathway for
677 intelligible speech in the left temporal lobe. Brain 123 Pt 12, 2400-6.
678 Scott, S.K., Rosen, S., Lang, H., Wise, R.J. 2006. Neural correlates of intelligibility in speech
679 investigated with noise vocoded speech--a positron emission tomography study. The
680 Journal of the Acoustical Society of America 120, 1075-83.
681 Shafiro, V. 2008. Identification of environmental sounds with varying spectral resolution. Ear
682 and hearing 29, 401-20.
683 Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M. 1995. Speech recognition
684 with primarily temporal cues. Science 270, 303-4.
685 Smith, D.R., Walters, T.C., Patterson, R.D. 2007. Discrimination of speaker sex and size
686 when glottal-pulse rate and vocal-tract length are controlled. The Journal of the
687 Acoustical Society of America 122, 3628-39.
688 Spahr, A.J., Dorman, M.F. 2004. Performance of subjects fit with the Advanced Bionics CII
689 and Nucleus 3G cochlear implant devices. Archives of otolaryngology--head & neck
690 surgery 130, 624-8.
691 Story, B.H., Titze, I.R., Hoffman, E.A. 2001. The relationship of vocal tract shape to three
692 voice qualities. The Journal of the Acoustical Society of America 109, 1651-67.
693 Strelnikov, K., Rouger, J., Lagleyre, S., Fraysse, B., Deguine, O., Barone, P. 2009.
694 Improvement in speech-reading ability by auditory training: Evidence from gender
695 differences in normally hearing, deaf and cochlear implanted subjects.
696 Neuropsychologia 47, 972-9.
697 Surprenant, A.M., Watson, C.S. 2001. Individual differences in the processing of speech and
698 nonspeech sounds by normal-hearing listeners. The Journal of the Acoustical Society
699 of America 110, 2085-95.
700 Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., Honda, K. 2006. Acoustic roles of the
701 laryngeal cavity in vocal tract resonance. The Journal of the Acoustical Society of
702 America 120, 2228-38.
703 Tanner, W.P., Jr., Swets, J.A. 1954. A decision-making theory of visual detection.
704 Psychological review 61, 401-9.
705 Tye-Murray, N., Tyler, R.S., Woodworth, G.G., Gantz, B.J. 1992. Performance over time
706 with a nucleus or Ineraid cochlear implant. Ear and hearing 13, 200-9.

UKCISG, U.K.C.I.S.G. 2004. Criteria of candidacy for unilateral cochlear implantation in postlingually deafened adults I: theory and measures of effectiveness. Ear and hearing 25, 310-335.

Vouloumanos, A., Druhen, M.J., Hauser, M.D., Huizink, A.T. 2009. Five-month-old infants' identification of the sources of vocalizations. Proc Natl Acad Sci U S A 106, 18867-72.

Warren, J.D., Scott, S.K., Price, C.J., Griffiths, T.D. 2006. Human brain mechanisms for the early analysis of voices. NeuroImage 31, 1389-97.

Xu, L., Tsai, Y., Pfingst, B.E. 2002. Features of stimulation affecting tonal-speech perception: implications for cochlear prostheses. The Journal of the Acoustical Society of America 112, 247-58.

Xue, S.A., Hao, J.G. 2006. Normative standards for vocal tract dimensions by race as measured by acoustic pharyngometry. J Voice 20, 391-400.

Zatorre, R.J., Belin, P. 2001. Spectral and temporal processing in human auditory cortex. Cereb Cortex 11, 946-53.

Figure legends.

Figure 1. d' values for the voice detection test in normal hearing subjects. Mean d' values (±se) are presented for groups A (n = 14) and B (n = 14) as function of a reduction from 64 to 2 in the number of frequency channels of the vocoder. Vocoding the voice stimuli affected significantly the performances in both groups (***, $p < 0.0001$). OV : original stimuli.

Figure 2. d' values for the speech and non speech stimuli in normal hearing subjects. The mean d' (±se) values in voice detection illustrated in figure 1 are presented with a distinction of speech (dark blue) and non-speech stimuli (light blue) and compared to that obtained from non-voice stimuli (red). While in both groups the vocoder affects significantly the performances for speech and non speech stimuli (***, $p < 0.0001$ and **, $p < 0.001$) all subjects from 32 to 2 channels were better for detecting speech voice than non-voice stimuli (**, $p < 0.001$; *** $p<0.0001$). Conventions as in fig 1.

Figure 3. Reaction time (RT) values during the voice detection test in normal hearing subjects. Mean RT (±se) are presented for groups A and B while separating speech voice, non speech voice and non voice stimuli. The vocoding conditions caused a significant increase of RT for group A ($p < 0.0001$). We did not detect a significant difference in RTs between voice and non voice stimuli (not shown, see text), neither between speech and non speech except for group A in 8 channels condition of vocoder were subjects were significantly faster for speech stimuli ($p < 0.05$).

Figure 4. Performances in voice detection and speech comprehension in cochlear implanted patients. (A) Mean d' values (±se) in CI patients (green) are presented for speech voice and non speech voice stimuli according to the delay post-implantation. No significant improve of performance with the experience of the implant are observed, but CI patients showed higher abilities for detecting speech voice than non speech voice after 18 months of activation ($p < 0.05$). The patients' performance is compared to that obtained in NHS (blue) in original and vocoding condition. CIS were always lower than NHS in original condition ($p < 0.001$) and had statistically similar performance to NHS in 4 channels vocoder.
(B) Word recognition performances (±se) of CI deaf patients (green) and of NHS (blue). Performance of patients significantly improved while experiencing the implant to reach a plateau after the first semester post-implantation. In this case, the performances in speech comprehension of the patients are significantly higher than that presented by control subjects in a 4 channels vocoder condition.

Figure 5. Relationship between voice detection and word comprehension in cochlear implanted patients. The correlation between d' values in voice detection and % correct in word comprehension is weak but significant ($p < 0.05$) indicating that high performances in word recognition are associated to high performances in voice detection. This analysis is obtained by pooling all patients together with short or long period of experience of the neuroprosthesis. The correlation was significant when we consider Speech or Non Speech stimuli for the d'.

Table 1. Patients summary. In the individual columns, the age of the patient is provided in year. Delay activation corresponds to the delay (in days or month) between the day of activation of the implant and the day of testing the patients. The estimation of the duration of hearing loss is given in years : NA: non available; less than 5 years (<5): between 5 and 10 years (>5), between 10 and 20 years (>10), between 20 and 30 years (>20), between 30 and

807 40 years (>30), between 40 and 50 years (>30) and more than 50 years (>50). One patient
808 (#10) had a sudden deafness which occurred 3 years before the cochlear implantation. The last
809 column indicated which patient had a hearing aid in daily life.
810
811
812
813
814

Figure 1

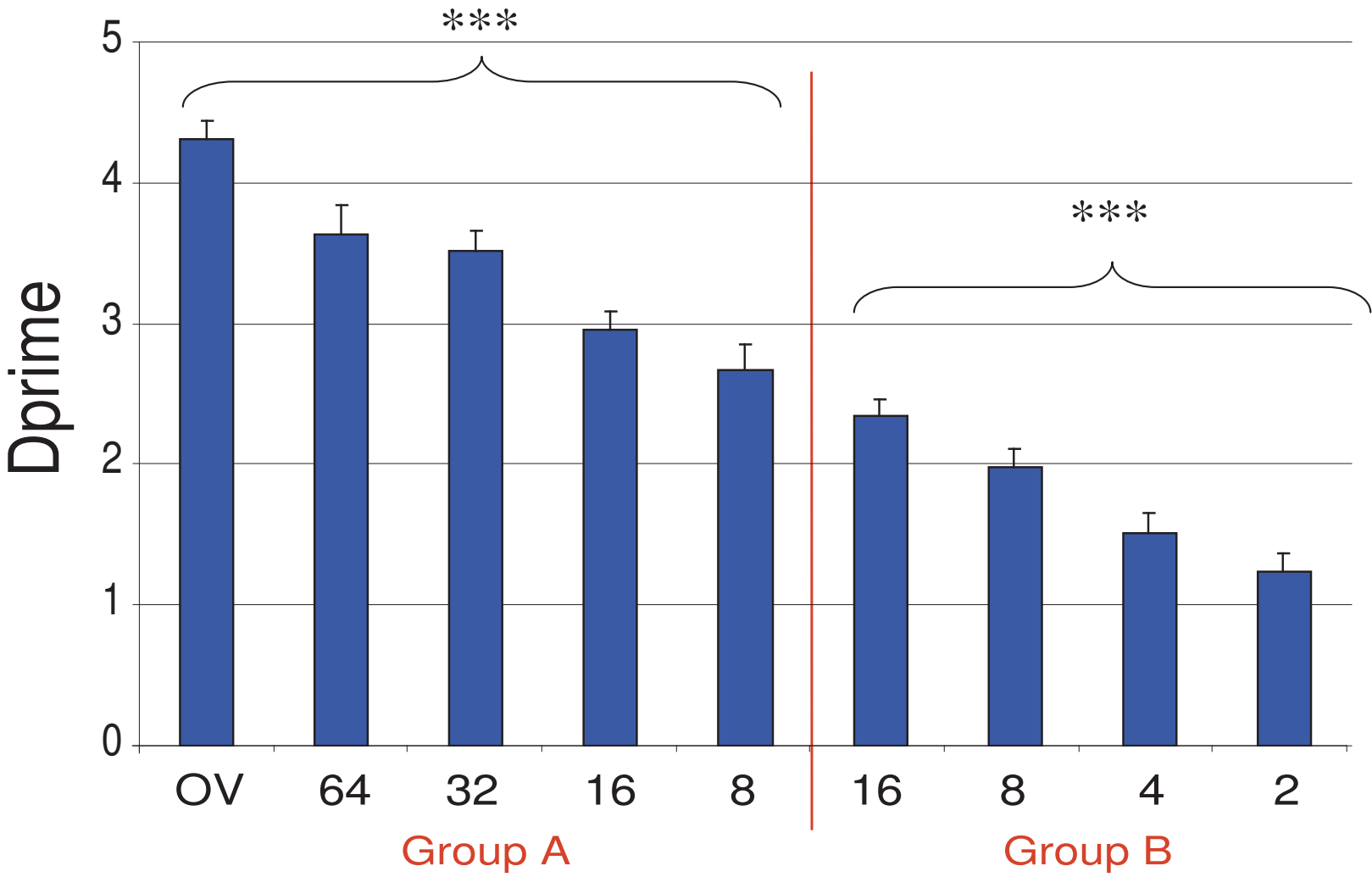


Figure 2

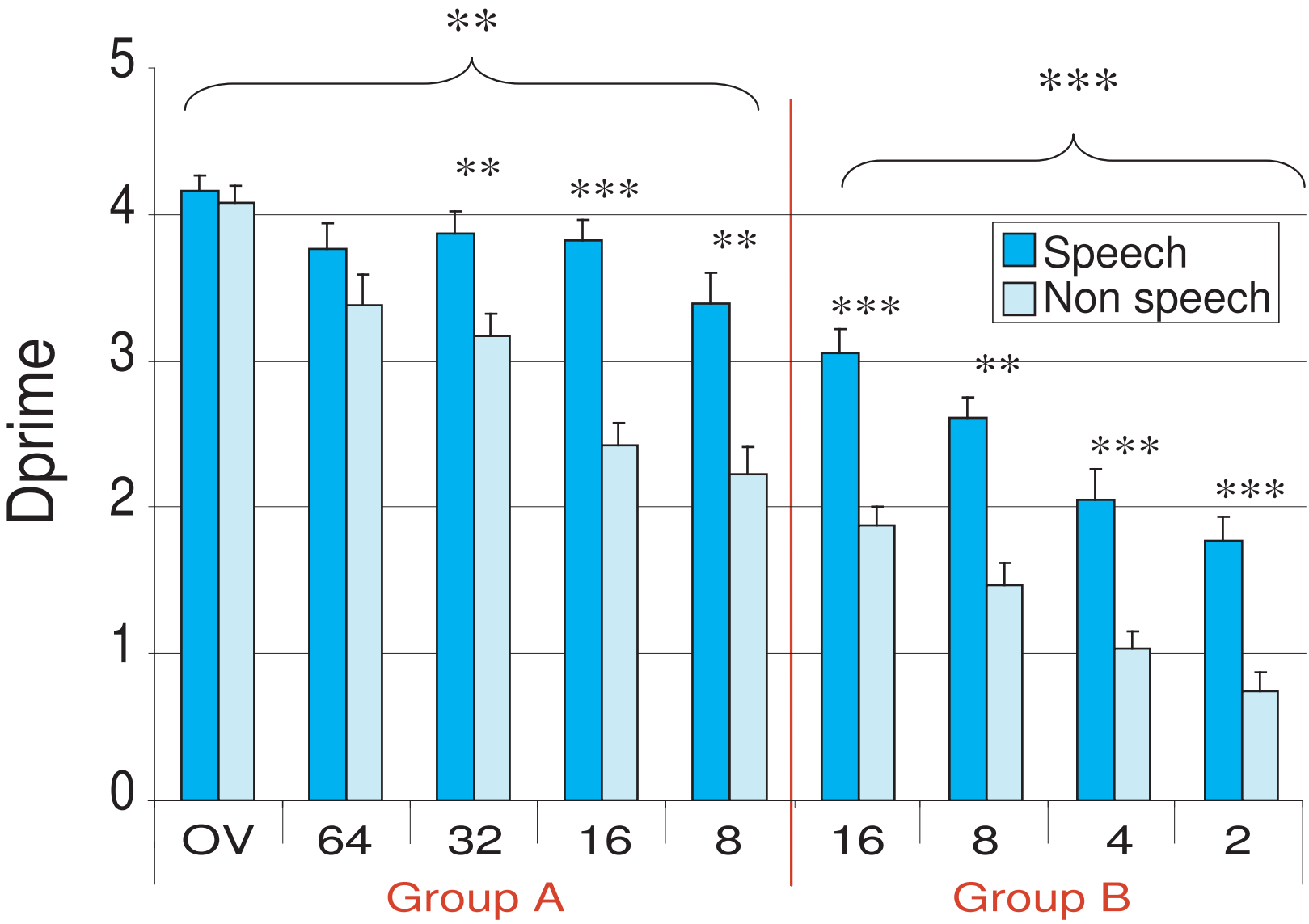

Dprime
5
4
3
2
1
0
**
***
**
***
**
***
**
***
***
OV
64
32
16
8
16
8
4
2
Group A
Group B
Speech
Non speech

Figure 3

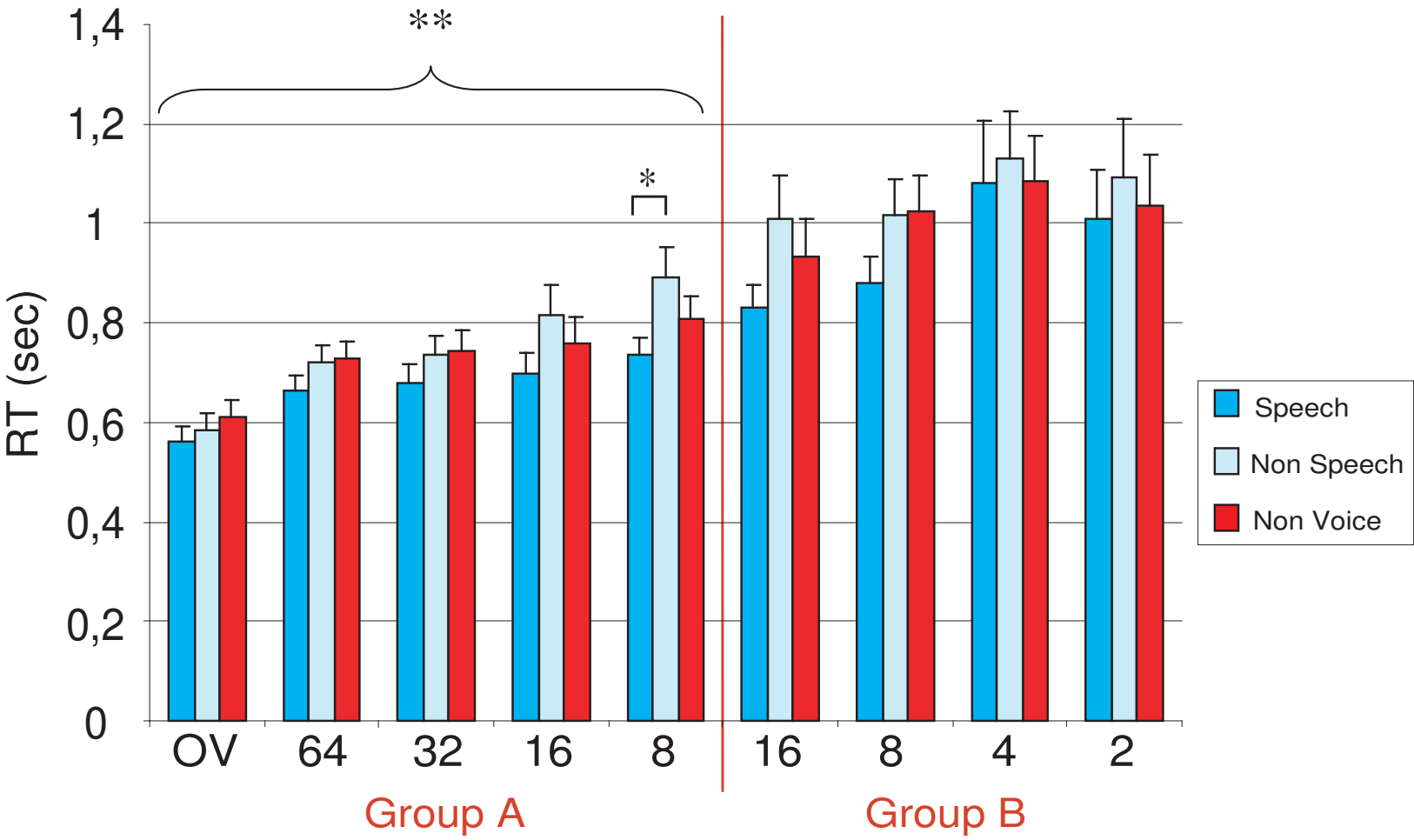
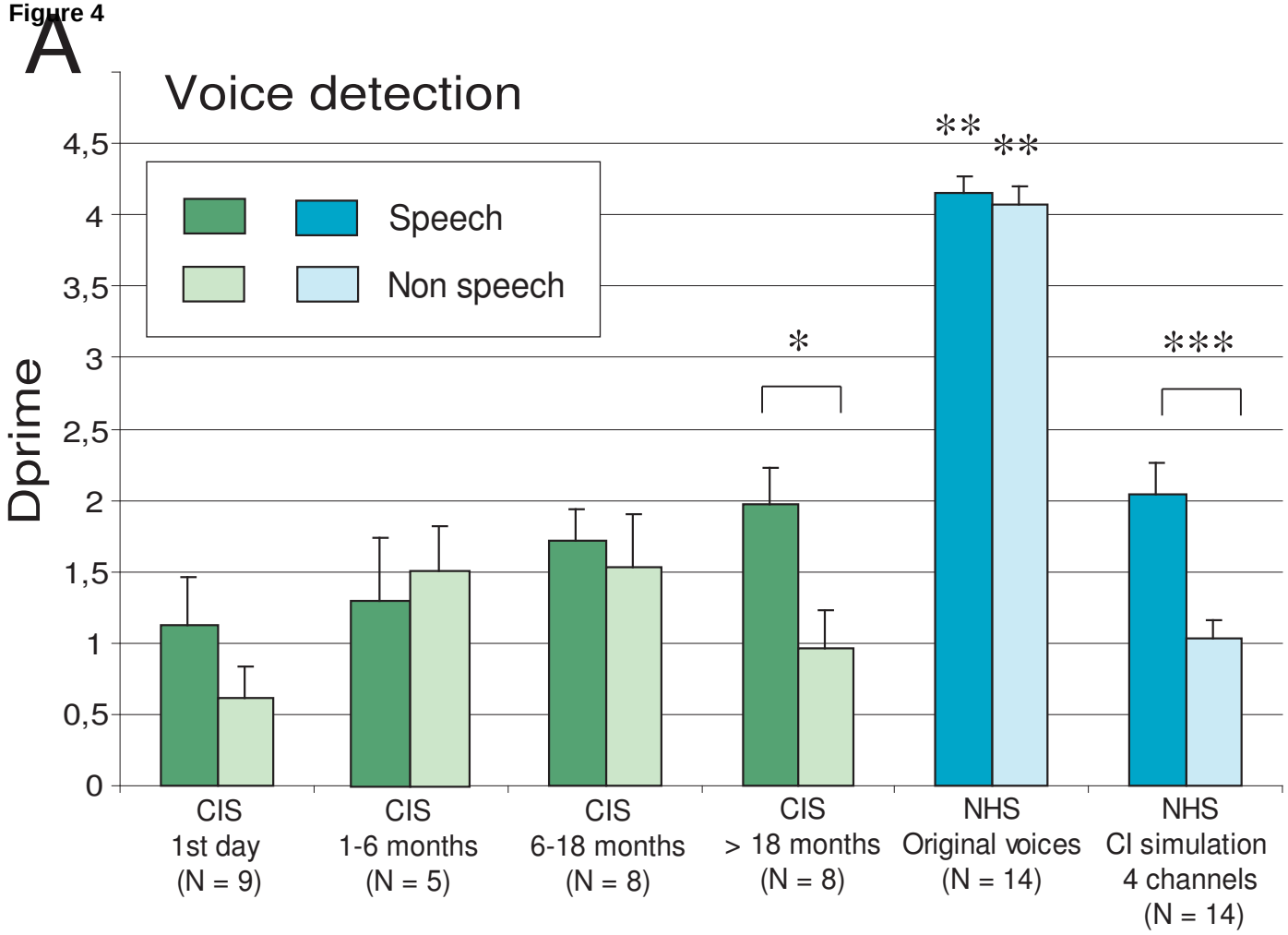

1,4
1,2
1
0,8
0,6
0,4
0,2
0
RT (sec)
**
*
OV
64
32
16
8
16
8
4
2
Group A
Group B
Speech
Non Speech
Non Voice

Figure 4


A
Voice detection
Speech
Non speech
Dprime
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
**
**
*
***
CIS 1st day (N = 9)
CIS 1-6 months (N = 5)
CIS 6-18 months (N = 8)
CIS > 18 months (N = 8)
NHS Original voices (N = 14)
NHS CI simulation 4 channels (N = 14)

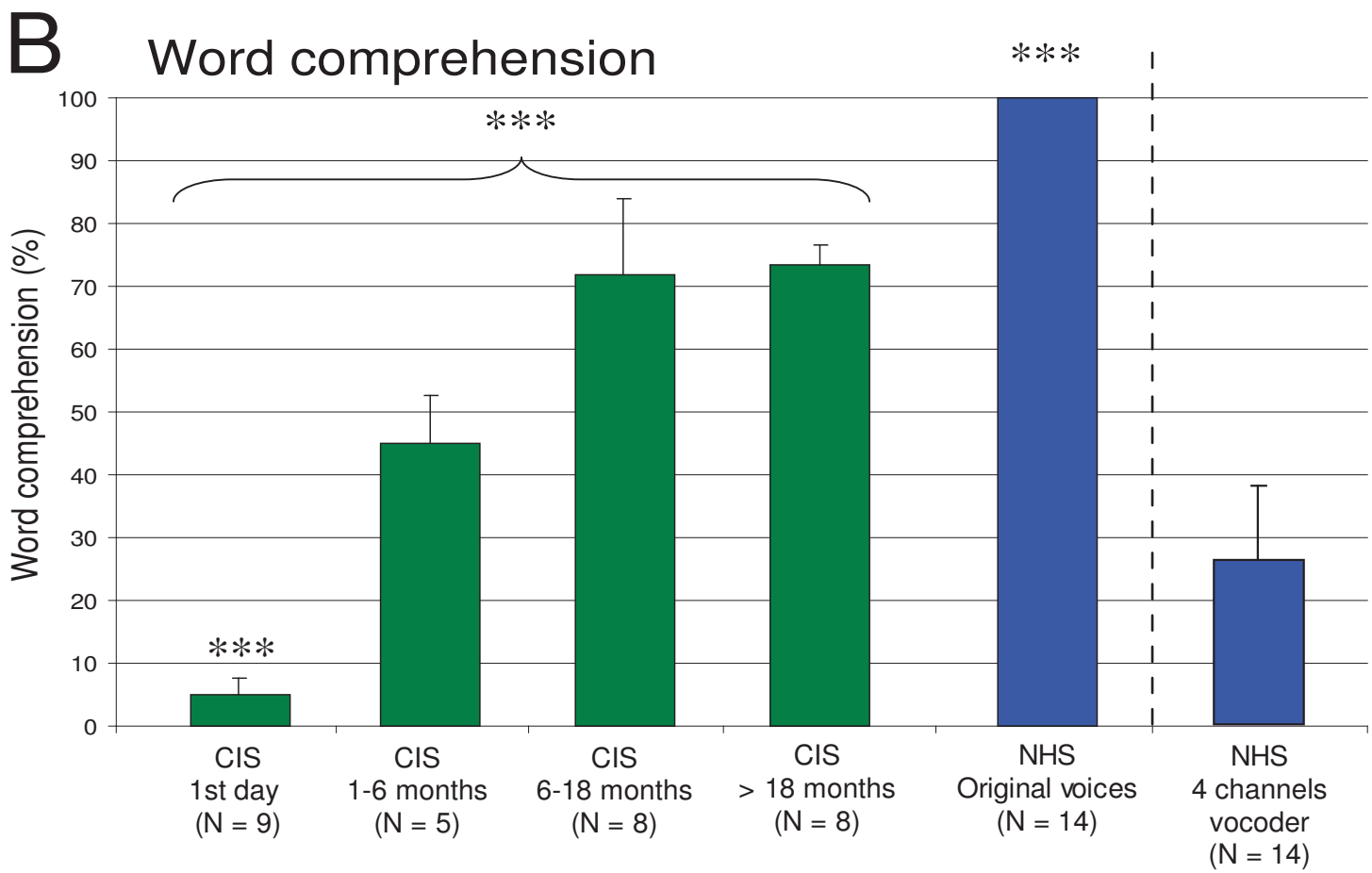

B
Word comprehension
Word comprehension (%)
100
90
80
70
60
50
40
30
20
10
0
***
***
***
CIS 1st day (N = 9)
CIS 1-6 months (N = 5)
CIS 6-18 months (N = 8)
CIS > 18 months (N = 8)
NHS Original voices (N = 14)
NHS 4 channels vocoder (N = 14)

Figure 5

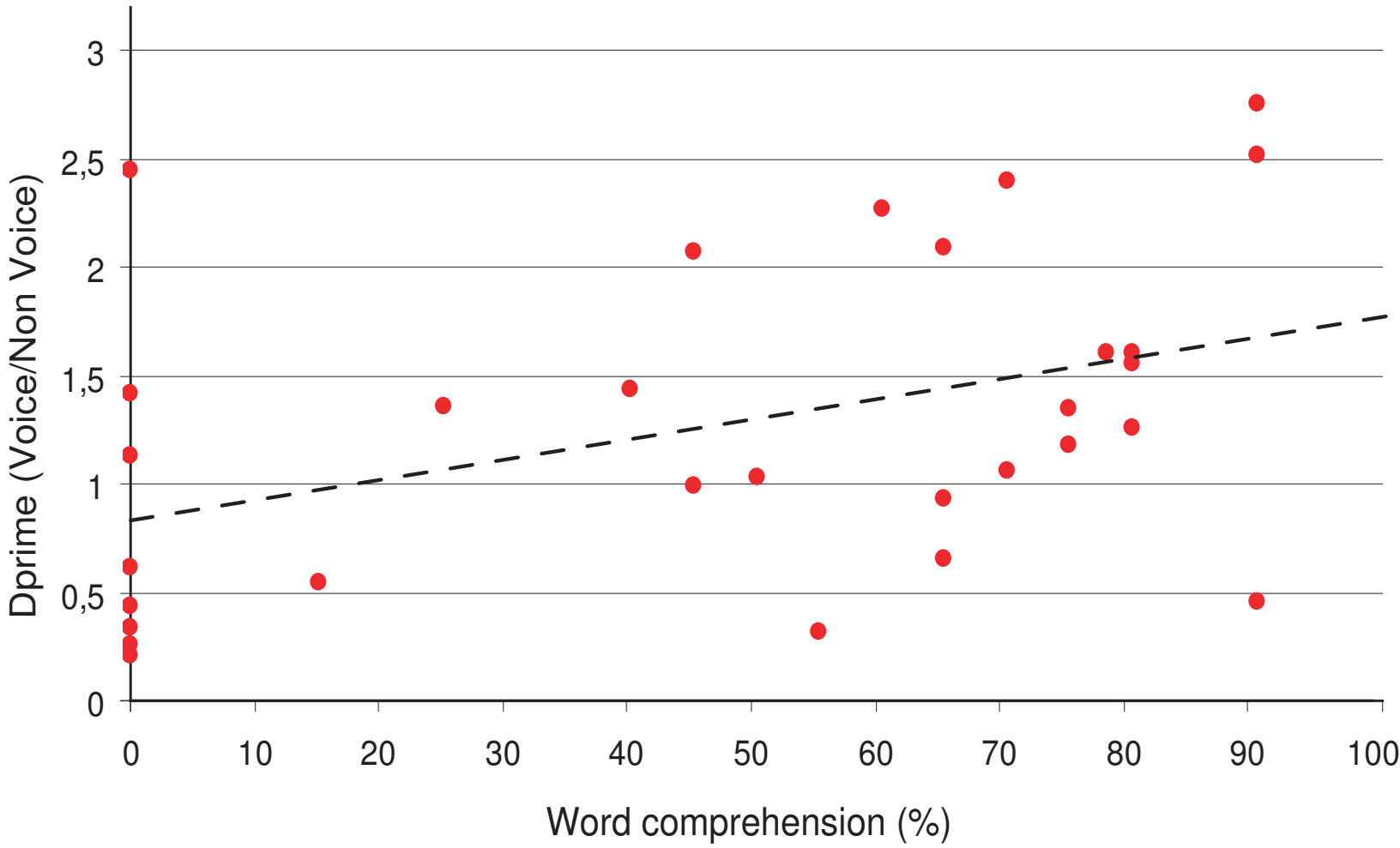

Dprime (Voice/Non Voice)
3
2,5
2
1,5
1
0,5
0
0
10
20
30
40
50
60
70
80
90
100
Word comprehension (%)

Table 1

Subject number	Group of delay activation	Age (year)	Estimation of delay of hearing loss (year)	Hearing aid Yes(Y)/No(N)
1	1st day	20.8	>10	N
2	1st day	80.05	NA	Y
3	1st day	63.02	>20	N
4	1st day	66.99	>10	Y
5	1st day	47.87	>30	N
6	1st day	47.87	>30	N
7	1st day	52.13	>40	N
8	1st day	47.94	>40	N
9	1st day	47.15	>40	Y
10	1 to 6 months	44.83	=3	Y
11	1 to 6 months	71.11	>40	Y
12	1 to 6 months	56.01	>50	Y
13	1 to 6 months	35.18	>30	Y
14	1 to 6 months	64.32	>50	N
15	6 to 18 months	21.23	>10	Y
16	6 to 18 months	51.09	>30	Y
17	6 to 18 months	39.71	>30	Y
18	6 to 18 months	52.82	< 5	Y
19	6 to 18 months	44.96	>40	Y
20	6 to 18 months	51.85	>40	N
21	6 to 18 months	47.66	>30	Y
22	6 to 18 months	71.09	NA	N
23	More than 18 months	47.84	>30	Y
24	More than 18 months	40.69	>30	N
25	More than 18 months	41.06	>20	Y
26	More than 18 months	37.53	>30	N
27	More than 18 months	74.76	>10	N
28	More than 18 months	63.51	>40	N
29	More than 18 months	59.36	NA	N
30	More than 18 months	75.44	>20	N

MASSIDA Zoé
Majouret
31 450 MONTGISCARD
Tel : 05 62 57 22 28
zoemassida@hotmail.com

29 ans
En couple
1 enfant
Permis B

FORMATION

Depuis octobre 2006 Université de Toulouse 3	**Doctorat Universitaire de Neurosciences**
2004-2005 Université de Bordeaux 2	**Master Professionnel de Neuropsychopharmacologie et Addictologie,** mention **B**
2003-2004 Université de Toulouse 3	**DEA National de Neuropsychologie**, mention **AB**
2002-2003 Université d'Aix-Marseille 1	**Maîtrise de sciences cognitives, mention neurosciences**, mention **AB**
1999-2002 Université de Toulouse 2	**Licence de psychologie**, mention **AB**
1999 Cahors (46)	**Baccalauréat S-SVT, spécialité SVT,** mention **AB**

EXPERIENCE PROFESSIONELLE

Depuis octobre 2006 CerCo (UMR 5549)	**Travail de recherche dans le cadre du doctorat (bourse CIFRE avec Cochlear) :** ➢ Etude de la **reconnaissance de la voix chez le sourd implanté cochléaire : psychophysique et imagerie cérébrale.**
Octobre 2006 à juin 2010	**Animation scientifique dans des collèges et lycées (Association inCOGnu)** ➢ Vulgarisation des sciences (thèmes : addiction, perception, mémoire) ➢ Collaboration avec l'Association « Les chemins buissonniers »
Octobre 2005 à octobre 2006 Laboratoire de psychiatrie (Université Bordeaux 2)	**Attachée de recherche** ➢ Etude de la **réactivité psychophysiologique** chez des patients dépendants des opiacés ➢ Entretient clinique ***Addiction Severity Index*** chez des patients dépendant d'une substance en demande de soins au département d'Addictologie de Bordeaux
2005 (février à juillet) Laboratoire de psychiatrie (Université Bordeaux 2)	**Stage d'attachée de recherche** sous la direction du Pr Marc AURIACOMBE ➢ Formation à l'entretient clinique ***Addiction Severity Index*** chez des patients dépendant d'une substance en demande de soins au département d'Addictologie de Bordeaux
2003 à 2004 CerCo (UMR 5549)	**Stage de recherche** sous la direction de Pascal BARONE ➢ Approche psychophysique de la perception multisensorielle audiovisuelle chez l'homme ➢ Validation d'une méthode d'analyse psychophysique : Théorie de détection du signal
2003 (janvier) INPC (CNRS)	**Stage de recherche** dans l'équipe de Marc JAMON ➢ Tests comportementaux sur des souris saines et des sourie KO du gène CHL1 (CAM).
2000 à 2005	**Garde d'enfants, Travaux saisonniers, Accompagnement de séjour linguistique**

INFORMATIONS COMPLEMENTAIRES

Informatique	**Bureautique (Mac et PC)** : Microsoft Office, OpenOffice, Adobe Illustrator, Filemaker Pro, LaTeX, EndNote **Statistiques** : SPSS, Statview, JMP **Programmation** : notions en langage C et MatLab
Langues	**Anglais** et **Espagnol** opérationnels
Séjours Linguistiques	**Ecosse** (2008 et 2006), **Angleterre** (2004), **Suède** (1997), **Espagne** (2006, 1996, 1995), **Pays de Galle** (1995)
Centres d'intérêt	**Sciences Cognitives** (Membre de l'association « InCOGnu »)**, cinéma, musique** (pratique du violon depuis 20 ans), **sport** (planche à voile, badminton)**, marche**

ZOE MASSIDA

ÉTUDE DE LA PERCEPTION DE LA VOIX CHEZ LE PATIENT SOURD POSTLINGUAL IMPLANTE COCHLEAIRE UNILATERAL ET LE SUJET NORMO-ENTENDANT EN CONDITION DE SIMULATION D'IMPLANT. PSYCHOPHYSIQUE ET IMAGERIE

Thèse réalisée sous la co-direction de Pascal Barone et Olivier Deguine

Résumé

Ce travail de thèse a consisté à étudier les mécanismes perceptif set neurofonctionnels impliqués lors de la perception de la voix chez des patients sourds postlinguaux implantés cochléaires unilatéralement, et chez des sujets normo-entendants en simulation d'implant.
Pour répondre à cet objectif, nous avons testé les performances comportementales des patients implantés dans des tâches de détection de la voix ainsi que dans des tâches de perception de l'information paralinguistique de la voix, comme le genre. Les patients ont été testés au cours d'un suivi ainsi qu'en mesures transversales. Nous avons comparé leurs performances à celles de sujets normo-entendants en condition de simulation d'implant cochléaire (vocoder). Nous avons également testé les sujets normo-entendants dans un protocole IRMf consistant à mesurer l'activité spécifique à la voix lors de la simulation d'implant.
Dans l'ensemble, ces travaux montrent qu'après implantation cochléaire, les patients sourds sont déficitaires en matière de perception de la voix, contrairement à la compréhension du langage. Ce déficit n'est pas uniquement lié à la dégradation du signal par le processeur de l'implant cochléaire, mais aussi certainement à des réorganisations corticales subséquentes à la surdité.

Mots-clés : voix, implant cochléaire, simulation d'implant, vocoder, aire de la voix

Laboratoire d'accueil :
Centre de recherche Cerveau et Cognition (CerCo)
UMR 5549 CNRS/UPS
Faculté de Médecine de Rangueil
31062 Toulouse Cedex 9

ZOÉ MASSIDA

STUDY OF VOICE PERCEPTION IN POSTLINGUALLY DEAF UNILATERALY COCHLEAR-IMPLANTED PATIENTS AND NORMAL-HEARING SUBJECTS IN COCHLEAR IMPLANT SIMULATION. PSYCHOPHYSIC AND IMAGERY

Abstract

This work consisted in studying perceptual and the underlying neuronal mechanisms involved during voice perception in postlingually deaf cochlear-implanted patients and normal-hearing controls stimulated through cochlear implant simulation.
We have analyzed behavioral performance of implanted patients during a voice detection task and other protocols perception tasks of paralinguistic information, such as gender. Two groups of patients were tested using either a longitudinal follow-ups or a transversal approach. We compared their performances to those of control normal-hearing subjects tested in cochlear implant simulation (vocoder). In addition, we have performed in normal-hearing subjects, a fMRI study, to reveal the effect of a cochlear implant simulation in the cortical activity of areas sensitive to human voice.
Results agree to point out, following cochlear implantation, a deficit in voice perception, unlike speech comprehension. This deficit is not only due to the degradation of the signal by the vocoder, but probably results from cortical reorganization induced by deafness.

Keywords: voice, cochlear implant, cochlear implant simulation, vocoder, voice area.