



**HAL**  
open science

# Approche psychophysique de la perception auditive para et extra linguistique chez le sujet sourd implanté cochléaire

Mathieu Marx

► **To cite this version:**

Mathieu Marx. Approche psychophysique de la perception auditive para et extra linguistique chez le sujet sourd implanté cochléaire. Sciences du Vivant [q-bio]. Université Paul Sabatier - Toulouse III, 2013. Français. NNT: . tel-00872077

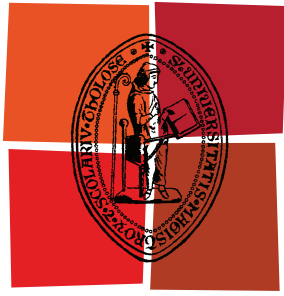
**HAL Id: tel-00872077**

**<https://theses.hal.science/tel-00872077>**

Submitted on 11 Oct 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.



Université  
de Toulouse

# THÈSE

En vue de l'obtention du  
**DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE**

**Délivré par :**

Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

**Discipline ou spécialité :**

Neurosciences, comportement, cognition

---

**Présentée et soutenue par :**

Mathieu Marx

**le :** 19 Septembre 2013

**Titre :**

Approche psychophysique de la perception auditive para et extra linguistique  
chez le sujet sourd post lingual implanté cochléaire

---

**Ecole doctorale :**

Comportement, Langage, Education, Socialisation, COgnition (CLESCO)

**Unité de recherche :**

Laboratoire CerCo - Cerveau Cognition, UMR 5549

**Directeur(s) de Thèse :**

Pascal BARONE

Olivier DEGUINE

**Rapporteurs :**

Paul AVAN

Nicolas GRIMAULT

**Membre(s) du jury :**

Paul AVAN

Nicolas GRIMAULT

Alain UZIEL

Bernard FRAYSSE

Olivier DEGUINE

Pascal BARONE



## Remerciements

Mes remerciements s'adressent en premier lieu aux patients implantés qui ont participé, souvent avec enthousiasme, toujours avec intérêt, aux travaux présentés dans ce mémoire.

Je remercie sincèrement M. Paul Avan et M. Nicolas Grimault d'avoir accepté d'évaluer ce mémoire, M. Alain Uziel d'avoir accepté de juger mon travail.

Je remercie Pascal Barone et Olivier Deguine qui ont dirigé cette thèse. Pascal Barone lui a donné ses principales orientations, m'a accompagné dans la rédaction du manuscrit et m'a apporté une aide précieuse dans la valorisation des résultats. Il s'est montré particulièrement compréhensif vis-à-vis d'un emploi du temps disons... dispersé. Olivier Deguine m'a fait partager son expertise clinique et bibliographique de l'implant cochléaire. Ses conseils m'ont permis de prendre le recul nécessaire à la synthèse de ce mémoire.

Je remercie mes Maîtres, le Pr Bernard Fraysse et le Pr Olivier Deguine, pour leurs encouragements répétés et leur bienveillance à mon égard. J'espère que ce travail sera à la hauteur de leurs attentes, de leur soutien et de leur confiance.

Je remercie Michèle Fabre-Thorpe de m'avoir accueilli au sein du laboratoire Cerveau et Cognition et d'avoir suivi mon parcours avec attention. Plusieurs rencontres y ont été déterminantes, depuis Jessica Foxton, qui m'a fait découvrir la prosodie, jusqu'à Pascal Girard, avec lequel des projets passionnants ont été initiés. Kuzma Strelnikov m'a toujours apporté son aide et je lui en suis sincèrement reconnaissant. Je remercie également Zoé Massida, qui m'a fait partager son expérience du d' et avec qui la collaboration s'est révélée facile et fructueuse. Je remercie Caroline Fonta, Joël Ferrer, Carmen Olmos, Luc Renaud, Maxime Rosito, Catherine Marlot et Claire Thouailles pour leur aide et leur accueil toujours chaleureux.

Un grand merci à Chris James pour son soutien actif et ses conseils tout au long de ces années. Chaque conversation est un enseignement!

J'adresse mes remerciements les plus sincères à mes collègues et amis du service d'ORL, du bloc opératoire, de la consultation, du centre implant, qui m'ont toujours soutenu et/ou supporté. Ils m'ont donné du temps et ont su composer avec l'idée approximative que je peux me faire de l'organisation.

Je remercie Pascal Gaillard, pour sa disponibilité et sa patience. Il nous a confiés un test prometteur et j'espère que ce travail saura gagner son intérêt.

Les échanges avec Marion Cousineau (post-doctorante chez Isabelle Peretz) ont été particulièrement simples et utiles. Je l'en remercie sincèrement, et ne doute pas que l'on concrétisera ce travail.

Je remercie Benoit Roby qui a, pour le coup, un sens de l'organisation infallible, et m'a généreusement transmis une partie de ses connaissances musicologiques.

Je dis merci aux amis. Je sais que je leur dois beaucoup, et pas seulement de l'argent. Ils savent que je le sais. Mais, est-on vraiment obligé de savoir ?

Je dédie cette thèse à mes parents Denis et Margot, à mes sœurs Raphaëlle et Claire, et à ma famille. Avec toute mon affection.

A la mémoire de mon grand-père, Paul Marx.

A la mémoire d'Ernestine Lagarde.

# Table des matières

Chapitre I.....	11
Perception auditive des sons linguistiques, para linguistiques et extra linguistiques .....	
1. LES INDICES ACOUSTIQUES FONDAMENTAUX.....	12
1.2 La durée .....	15
1.3 La fréquence .....	15
1.4 Le timbre.....	19
2. PERCEPTION DE LA PAROLE.....	21
2.1 Caractéristiques acoustiques des phonèmes .....	21
2.2. Structure temporelle du signal de parole .....	23
2.3 Traitement neuronal de la parole .....	25
3. PERCEPTION DE LA VOIX .....	27
3.1 Caractéristiques acoustiques de la voix .....	27
3.2 Traitement neuronal de la voix .....	28
4. PERCEPTION DE LA PROSODIE.....	30
4.1 Définitions et caractéristiques .....	30
4.2 Traitement neuronal de la prosodie .....	31
5. PERCEPTION AUDITIVE EXTRA LINGUISTIQUE : EXEMPLE DE LA MUSIQUE .....	32
5.1 Paramètres acoustiques spécifiques à la perception musicale .....	33
5.2 Traitement neuronal de la musique.....	35
Chapitre II .....	37
La perception auditive chez le sujet implanté cochléaire.....	
1. CODAGE ET RESTITUTION DU SIGNAL ACOUSTIQUE PAR L'IMPLANT COCHLEAIRE.....	38
1.2 Quel est le nombre de canaux nécessaires ? .....	39
1.2 Quel est le nombre d'électrodes effectives ? .....	45

1.3 Principales stratégies de codage .....	47
1.4 Les différents types d'implants.....	51
2. PERCEPTION DE LA PAROLE.....	53
3. PERCEPTION DES INFORMATIONS PARA ET EXTRA LINGUISTIQUES .....	55
3.1 Détection de la voix humaine .....	56
3.2 Perception des attributs de la voix liés à l'identité du locuteur .....	57
3.3 Perception des informations extra linguistiques : exemple de la musique .....	58
4. REORGANISATION NEUROFONCTIONNELLE CHEZ LE SUJET SOURD.....	60
4.1 Déficits neurosensoriels et plasticité cérébrale.....	60
4.2 Aspects comportementaux de la plasticité cérébrale des sujets sourds profonds .....	61
4.3 Réorganisation corticale chez le sujet sourd post lingual.....	64
4.4 Réactivation des aires auditives après implantation cochléaire.....	65
5. AUDITION RESIDUELLE ET IMPLANT COCHLEAIRE .....	69
5.1 Généralités-Définitions.....	69
5.2 Extension des critères audiométriques d'implantation cochléaire.....	70
6. METHODES ET PROBLEMATIQUES.....	72
6.1 Analyse des résultats .....	72
6.2 Problématique.....	74
Chapitre III.....	75
Catégorisation libre de sons linguistiques, para linguistiques et extra linguistiques par le sujet implanté cochléaire .....	
1. INTRODUCTION .....	76
1.1 La catégorisation libre : principes généraux.....	78
1.2 La catégorisation libre : méthodes d'analyse.....	79
2. MATERIELS ET METHODES .....	82
2.1 Catégorisation auditive de stimuli auditifs naturels .....	82
2.2 Population.....	84
2.3 Procédure .....	85
2.4 Analyse des résultats .....	86
3. RESULTATS.....	87
3.1 Caractéristiques des partitions .....	87
3.2 Analyses des distances entre stimuli dans les matrices de dissimilarités agrégées ....	89

3.3 Analyse de la catégorisation par arbre additif .....	92
3.4 Analyse de la catégorisation et de ses modalités par MDS .....	94
3.5 Existe-t-il des corrélats acoustiques aux modalités de catégorisation ? .....	96
4. DISCUSSION.....	97
4.1 Capacités de catégorisation chez les sujets implantés cochléaires .....	97
4.2 Principales catégories formées .....	100
4.3 Perspectives .....	101
Chapitre IV .....	103
Etude 2 : Perception de la prosodie affective par le sujet implanté cochléaire.....	
Reconnaissance et catégorisation des émotions .....	
1. INTRODUCTION.....	104
2. MATERIELS ET METHODES .....	105
2.1 Population.....	105
2.2 Appareillage et stimuli.....	108
2.3 Procédure .....	109
2.4 Analyse des résultats .....	110
3. RESULTATS.....	113
3.1 Description des groupes.....	113
3.2 Analyse de la pente de la courbe .....	114
3.3 Analyse du seuil C50 .....	117
3.4 Analyse des réponses aux extrémités « joie » et « tristesse » du continuum.....	118
4. DISCUSSION.....	120
4.1 Déficit de la catégorisation des émotions chez le sujet implanté cochléaire.....	120
4.2 Bases acoustiques de la perception des émotions.....	120
4.3 La réorganisation cérébrale secondaire à la surdité peut-elle limiter la perception de la prosodie affective chez le sujet implanté cochléaire ?.....	122



Chapitre V .....	125
Etude 3 : Perception de la prosodie linguistique par le sujet implanté cochléaire . .	
Intérêts d'une stimulation bimodale .....	
1. INTRODUCTION .....	126
2. MATERIELS ET METHODES .....	126
2.1 Population.....	126
2.2 Appareillage et stimuli.....	129
2.3 Procédure .....	133
2.4 Analyse des résultats .....	133
3. RESULTATS.....	135
3.1 Description des groupes.....	135
3.2 Perception de la prosodie linguistique.....	136
3.3 Discrimination de F0 .....	141
3.4 Impact du niveau d'audition résiduelle.....	142
3.5 Relation entre le F0DL et la perception de la prosodie linguistique .....	144
4. DISCUSSION.....	144
4.1 Perception de la prosodie linguistique et audition résiduelle .....	144
4.2 Perception de la prosodie linguistique par le groupe CI only.....	147
5. ETUDE PRELIMINAIRE DE LA PERCEPTION DES INDICES VISUELS DE LA PROSODIE LINGUISTIQUE.....	148
5.1 Méthodes .....	148
5.2 Résultats.....	150
5.3 Discussion préliminaire .....	153
 Chapitre VI.....	 157
Etude 4 : Etude EEG de l'intégration audiovisuelle du phénomène prosodique d'accentuation .....	
1. INTRODUCTION .....	158
2. MATERIELS ET METHODES .....	159
2.1 Population.....	159
2.2 Stimuli .....	159
2.3 Procédure .....	160

2.4 Enregistrement EEG et analyses.....	161
3. RESULTATS.....	162
3.1 Tâche comportementale de détection .....	162
3.2 Résultats EEG.....	162
4. DISCUSSION.....	167
4.1 Nature de la négativité mesurée.....	167
4.2 Les indices prosodiques visuels ont-ils un rôle prédictif ? .....	169
 Chapitre VII.....	 171
Etude 5 : Perception de la musique par le sujet implanté cochléaire .....	
Parallèles avec l'amusie congénitale.....	
1. INTRODUCTION .....	172
2. MATERIELS ET METHODES .....	173
2.1 Population.....	173
2.2 Appareillage et stimuli.....	175
2.3 Procédure et analyse des résultats.....	179
3. RESULTATS.....	180
3.1 Description des groupes.....	180
3.2 Perception de la fréquence fondamentale : F0DL .....	181
3.3 Perception de la musique par le test Amusia on line .....	182
4. DISCUSSION.....	189
4.1 La perception de la musique chez le sujet implanté cochléaire.....	189
4.2 Intérêts de l'audition résiduelle .....	190
4.3 Similitudes entre les sujets implantés cochléaires et sujets amusiques .....	192
4.4 L'étude de la perception musicale chez le sujet sourd implanté cochléaire peut-elle être utile à la compréhension de la perception musicale chez le sujet normo entendant ? .....	195

Chapitre VIII .....	199
Discussion générale et conclusion.....	
1. CODAGE ET RESTITUTION DU SIGNAL .....	200
2. INFLUENCES DE LA REORGANISATION CORTICALE SECONDAIRE A LA SURDITE .....	202
Références .....	205
Annexes .....	223
1. Publications en rapport avec la thèse.....	224
2. Communications orales en rapport avec la thèse (orateur).....	224
3. Communications affichées en rapport avec la thèse.....	225
4. Travaux encadrés pendant la période de la thèse.....	225
5. Tableau récapitulatif des sujets implantés cochléaires inclus.....	225

## Introduction générale

« Cochlear implants: a remarkable past and a brilliant future ». Le titre de la récente revue de la littérature publiée par Blake Wilson et Michaël Dorman [Wilson and Dorman, 2008] illustre parfaitement la situation actuelle de l'implantation cochléaire et le contexte dans lequel notre travail de thèse s'est déroulé. Nous sommes en effet arrivés aujourd'hui à la croisée des chemins, au moment où l'on peut se retourner en arrière et constater les progrès accomplis. L'implant cochléaire a permis de restaurer la discrimination de la parole chez plusieurs dizaines de milliers d'adultes devenus sourds et d'enfants sourds congénitaux. Les améliorations technologiques et le développement de stratégies de rééducation orthophonique spécifiques ont été accompagnés d'excellents résultats fonctionnels, tout du moins inédits dans le domaine des neuroprothèses. On considère ainsi que plus de 80% des patients implantés discriminent correctement plus de 80% des mots d'une phrase présentée dans le silence, que plus de 70% des enfants implantés pourront suivre une scolarité normale [Calmels et al., 2004].

Mais être à la croisée des chemins, cela implique aussi la nécessité de choisir la ou les futures directions à prendre pour améliorer la prise en charge des patients sourds concernés. Les rencontres entre cliniciens, chercheurs et industriels abordent régulièrement la question de « l'avenir » de l'implant et des priorités dans lesquelles s'impliquer avec conviction. Quatre grandes tendances se dégagent en général au terme de ces discussions :

1. Le nombre de patients implantés cochléaires est en augmentation constante, dépassant en 2010 les 200000 individus, alors que le nombre de professionnels de santé s'en occupant demeure relativement stable. Cette donnée démographique peut nécessairement retentir sur la qualité de la prise en charge des patients, qu'elle concerne l'activation de l'implant, ses réglages, ou la rééducation. Deux solutions peuvent être envisagées, vraisemblablement en complément l'une de l'autre, pour changer les termes de cette équation défavorable : favoriser le recrutement de professionnels impliqués dans cette prise en charge et faciliter la possibilité de réglages à distance, voire d'autoréglages, grâce aux outils informatiques qui sont à notre disposition.
2. Le deuxième point concerne la question des extensions d'indication de l'implant. La technique s'est en effet révélée efficace dans le traitement des surdités profondes à totales bilatérales, mais de nombreuses études récentes démontrent son efficacité dans

la prise en charge des surdités sévères, voire des surdités unilatérales. L'évaluation du bénéfice apporté doit être complétée sur le plan clinique et sur le plan médico économique pour affirmer l'intérêt de l'implant dans ces indications.

3. La troisième tendance a trait à un thème que nous ne ferons que citer tant il est controversé : l'implant cochléaire tout implantable.
4. L'amélioration des résultats auditifs après l'implantation constitue la quatrième priorité généralement évoquée. La restauration de la discrimination de la parole dans le silence a en effet largement été étudiée mais plusieurs domaines de la perception auditive demeurent insuffisamment explorés et/ou restaurés après implantation cochléaire. Parmi ces domaines, la discrimination de la parole dans le bruit, la discrimination de la parole au téléphone, la reconnaissance du locuteur, la perception de la prosodie et de la musique sont régulièrement cités.

L'amélioration des résultats auditifs dans leur globalité est la direction qui a motivé la réalisation de ce travail de thèse, par lequel nous avons souhaité apporter un nouvel éclairage sur les capacités auditives para et extra linguistiques des patients implantés. Cette phase préalable d'évaluation apparaît en effet nécessaire avant d'envisager d'éventuelles adaptations des stratégies de rééducation orthophonique ou des modifications technologiques au niveau de l'implant. L'extension de l'évaluation des capacités perceptives à d'autres domaines que la parole nous a permis de préciser la nature et l'importance des déficits présentés, et d'identifier certains facteurs expliquant potentiellement ces déficits.

L'exposé de nos travaux expérimentaux est précédé de deux parties introductives qui proposent une synthèse des connaissances actuelles sur la perception auditive chez le sujet normo entendant et chez le sujet implanté cochléaire. Cette synthèse n'a pas vocation à être exhaustive ; elle cherche plutôt à définir les bases acoustiques et les principaux processus cognitifs impliqués dans la perception des informations linguistiques, paralinguistiques et extra linguistiques. Une attention particulière sera portée à la situation des patients implantés cochléaires, dont la perception auditive est non seulement influencée par le traitement du signal réalisé par l'implant, mais également par la réorganisation cérébrale fonctionnelle secondaire à la surdité.

Dans le premier travail expérimental exposé, nous avons étudié les modalités de regroupement et de séparation de différents sons linguistiques, para linguistiques et extra linguistiques. Nous avons ainsi évalué les capacités de catégorisation auditive des sujets implantés cochléaires et cherché à déterminer les éventuelles stratégies sous-jacentes utilisées.

La suite de nos travaux expérimentaux s'intéresse à la perception de la prosodie affective et linguistique. Une de ces études a permis d'évaluer la contribution relative des différents indices acoustiques fondamentaux dans ces aptitudes perceptives et d'envisager l'intérêt d'une stimulation bimodale, associant l'audition résiduelle des patients à la stimulation par implant.

Notre dernière étude concerne la perception de la musique par les sujets implantés cochléaires, domaine pour lequel ils sont réputés présenter un déficit majeur. Le paradigme expérimental utilisé dans ce travail nous a permis d'établir un parallèle avec les individus atteints d'amusie congénitale, qui constituent une autre population de sujets handicapés pour la perception musicale.



## Chapitre I

# Perception auditive des sons linguistiques, para linguistiques et extra linguistiques



La diversité des sons auxquels nous sommes exposés quotidiennement rend difficile leur classification mais tous sont émis selon des règles physiques qui définissent des indices acoustiques fondamentaux. Ce sont la perception de chacun des indices acoustiques physiques, leur intégration et leur interprétation qui vont être à l'origine de la sensation auditive.

Nous nous intéresserons essentiellement aux sons complexes, qui constituent la majorité des sons rencontrés. Les sons complexes sont, à l'inverse des sons purs, composés de plusieurs fréquences et sont qualifiés d'harmoniques (ou périodiques) si les fréquences contenues dans son spectre sont multiples d'une fréquence dite fondamentale, ou de non harmoniques (non périodiques) si on ne peut les décomposer en fréquences liées à une fondamentale.

## **1. LES INDICES ACOUSTIQUES FONDAMENTAUX**

On distingue trois principales grandeurs physiques (intensité, durée et fréquence) responsables de la sensation auditive par trois attributs perceptifs leur correspondant. Le timbre est un attribut spécifique des sons complexes harmoniques que nous évoquerons également.

### **1.1 L'intensité**

L'intensité, ou amplitude, est liée à la puissance acoustique émise par une source sonore et se traduit sur le plan perceptif par une sensation appelée sonie. L'augmentation de la sensation de sonie ne croît pas de manière linéaire avec l'augmentation d'intensité, mais selon une échelle logarithmique graduée en décibels (dB). Le niveau acoustique d'intensité en dB est défini comme :  $N_I = 10 \log (I/I_0)$ , où  $I$  est l'intensité acoustique de l'onde considérée et  $I_0$  une intensité acoustique de référence correspondant au seuil d'audition, fixée à  $10^{-12} \text{W/m}^2$ .

Le seuil auditif correspond à la plus petite valeur d'intensité nécessaire pour évoquer une sensation de sonie (voir figure I.1). Le seuil différentiel ou de discrimination d'intensité caractérise la plus petite différence d'intensité susceptible d'être détectée entre deux stimuli d'intensité différente.

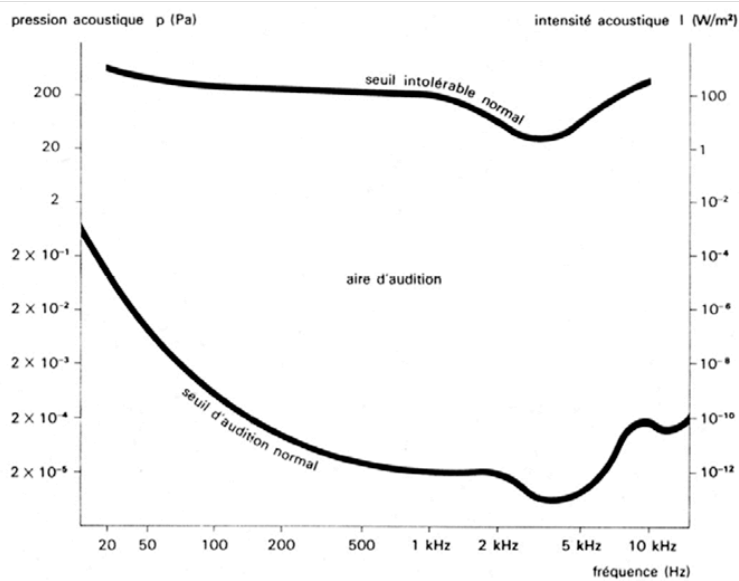


Figure I.1 Courbes des seuils d'audition chez le sujet normo entendant. D'après Galifret (1991)

La sonie est donc essentiellement liée à l'intensité du stimulus sonore mais également à sa fréquence. Les seuils auditifs ne sont en effet pas identiques d'une fréquence à l'autre et c'est pourquoi la notion d'isophonie a été introduite. Les courbes isophoniques permettent ainsi de représenter pour chaque fréquence le niveau d'intensité évoquant la même sonie (voir figure I.2). Deux stimuli positionnés sur une ligne isophonique donnée évoquent la même sensation d'intensité, mesurée en phones.

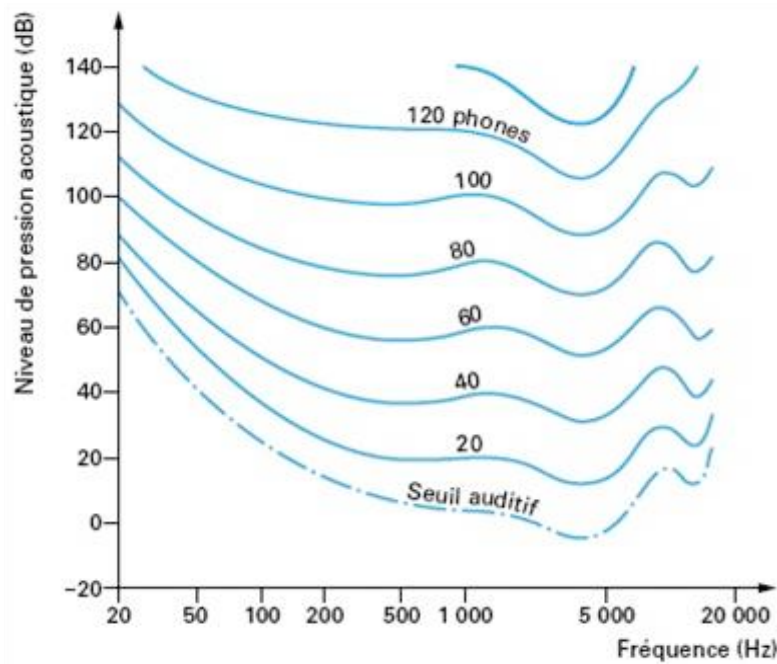


Figure I.2. Courbes isophoniques normalisées de l'audition humaine. D'après Avan (1998)

La sensation d'intensité évoquée par les sons complexes harmoniques, tels que la voix, est donc intimement liée au traitement de ses composantes spectrales. Les figures I.1 et I.2 illustrent comment le traitement de l'intensité acoustique varie en fonction de la région fréquentielle considérée. Pour une intensité donnée, le seuil d'audition est beaucoup plus faible pour les basses fréquences que pour les fréquences moyennes. Cette différence de sonie en fonction de la fréquence tend à diminuer avec l'augmentation de l'intensité acoustique (voir figure I.2).

Le traitement d'un son complexe fait ainsi intervenir un ensemble de filtres auditifs, chacun étant dédié à une bande fréquentielle donnée. La largeur de ces bandes, appelées bandes critiques a été évaluée en psycho acoustique fondamentale, au moyen de travaux utilisant le masquage de sons purs. La première description de ces bandes critiques remonte aux travaux de Fletcher (1940), qui a défini un filtre comme l'ensemble des fréquences adjacentes, centrées sur la fréquence d'un son pur, susceptible de le masquer. Au niveau cochléaire, le traitement des sons serait ainsi assuré par la juxtaposition de filtres, 36 filtres couvrant la gamme fréquentielle comprise entre 25 Hz et 11000 Hz [Moore, 1989]. Un filtre auditif est défini par sa fréquence centrale et par sa largeur de bande critique, évaluée par la largeur de bande rectangulaire équivalente (*ERB : Equivalent Rectangular Bandwidth*). Cette largeur de la bande critique augmente avec la fréquence centrale ( $f_c$ ) du filtre, selon une relation décrite par Moore [Moore et al., 1998] :  $ERB = 24.7 \times (4.37f_c + 1)$ . Les nombreux travaux réalisés par la suite ont permis de préciser la forme de ces filtres, plus triangulaires que rectangulaires, avec un sommet correspondant à leur fréquence centrale. La largeur de bande critique est plus réduite pour les basses fréquences inférieures à 500 Hz [Moore et al., 1990]. Il a également été mis en évidence une asymétrie de leur sélectivité en fréquence en faveur des fréquences basses lorsque l'intensité acoustique augmente [Moore, 1989].

La sonie est également influencée par le troisième indice acoustique fondamental, la durée du son, par les phénomènes d'intégration et d'adaptation. L'intégration dispose que pour des durées de présentation du stimulus sonore inférieures à 500 ms, la sonie diminue. L'adaptation traduit la diminution de sonie pour des stimuli dont l'intensité est inférieure à 30 dB au-dessus du seuil s'ils sont présentés sur une longue durée.

## 1.2 La durée

La définition de la durée d'un stimulus sonore est simple et traduit le temps écoulé entre son début et sa fin. Pour évaluer la sensation de durée, il est classiquement réalisé des tâches de détection de silence au sein d'un stimulus acoustique continu (*gap detection*), le seuil de détection correspondant à la durée minimale de silence perçue au sein d'un stimulus sonore entrecoupé d'un temps de silence donné. Ce seuil de détection est de l'ordre de 2 à 3 ms pour un bruit à large bande présenté à une intensité confortable (30 dB SPL), et augmente pour la présentation de sons purs. Entre 400 et 2000 Hz, il est ainsi d'environ 5 ms et augmente jusqu'à 18 ms pour les sons purs de fréquence inférieures à 400 Hz [Eddins et al., 1992]. Lorsque les fréquences des sons encadrant le silence sont différentes, ce seuil se dégrade, de la même manière que lorsque l'intensité du stimulus acoustique se rapproche du seuil auditif. Le seuil de détection entre deux silences peut également être déterminé (*gap duration detection*) [Grose et al., 2001b]. D'autres tests psycho acoustiques tels que la détection de la durée du son même ou l'évaluation des fonctions de détection de modulation [Grose et al., 2001a] permettent également d'évaluer la sensation de durée.

S'il est simple de définir la durée d'un son, on comprend donc que l'évaluation des capacités perceptives qui lui sont liées est complexe et pourtant fondamentale. Cet indice revêt en effet une importance toute particulière par sa contribution essentielle dans la transmission de l'information phonétique et prosodique dans le langage. C'est donc dans le contexte du langage que nous aborderons les processus impliqués dans son traitement.

## 1.3 La fréquence

La fréquence est la grandeur physique caractérisant un phénomène périodique et pour ce qui concerne les sons, trouve son corrélat perceptif dans la sensation de hauteur tonale. La hauteur tonale correspond à « l'attribut de la sensation auditive selon lequel les sons peuvent être rangés sur une échelle, telle qu'une échelle musicale, des graves vers les aigus » (American National Standards Institute, 1960), « selon que leur fréquence est basse ou haute », (précision de l'association française de normalisation, 1977).

Pour les sons purs, la sensation de hauteur tonale est en rapport direct avec la fréquence concernée. Pour les sons complexes harmoniques, la sensation de hauteur se rapporte à la fréquence fondamentale, dite  $F_0$ , fréquence de laquelle toutes les fréquences contenues dans le spectre du son sont multiples.

Si la fréquence fondamentale est absente du signal acoustique, comme c'est le cas au cours d'une conversation au téléphone, qui filtre les voix au sein d'une bande passante comprise entre 300 et 3500 Hz, elle peut être reconstruite à partir de ses fréquences multiples, les harmoniques, présentes dans le signal. La sensation de hauteur tonale rend compte des variations de hauteur dans une phrase (intonation) ou dans un extrait musical (mélodie) et les mécanismes de codage de la fréquence fondamentale sont donc de première importance.

### ***Mécanismes de codage de F0***

Ces mécanismes ont fait l'objet de multiples controverses, depuis la polémique autour de la F0 manquante impliquant Ohm et Seebeck, dont certaines se poursuivent encore aujourd'hui. Ces mécanismes sont donc incomplètement élucidés et nous n'évoquerons que certains des mécanismes supposés les plus consensuels.

Les sons complexes comportant un faible nombre d'harmoniques (<10) évoquent une sensation de hauteur tonale plus saillante que ceux en comportant un nombre plus important. Cette différence tient probablement au caractère résolu ou non des harmoniques compris dans le son. Certains sons, notamment ceux à F0 élevée, comportent des harmoniques suffisamment espacés pour être chacun traité, donc résolu, par un filtre auditif donné. A l'inverse, d'autres sons, notamment ceux à F0 basse, comprennent parfois plusieurs harmoniques non résolus, car traités par le même filtre auditif [Shera et al., 2002]. La saillance de F0 diminue donc en présence d'harmoniques non résolus, ce qui est susceptible de limiter la discrimination entre deux F0 basses chez le sujet normo entendant, ou entre deux F0 élevées chez le sujet présentant une surdité sur les fréquences aiguës [Grimault et al., 2001].

La qualité de perception de F0 d'un sujet peut être évaluée en mesurant le seuil de discrimination entre deux sons de F0 différentes présentés successivement [Shackleton and Carlyon, 1994, Demany and Semal, 2002, Bernstein and Oxenham, 2006]. Ce seuil est appelé F0 difference limen (F0DL) et caractérise donc la plus petite différence perceptible entre deux sons de F0 différentes. Les harmoniques résolus peuvent être codés par la région cochléaire qu'ils stimulent, répondant ainsi au principe de tonotopie cochléaire, qui veut que la base de la cochlée traite les fréquences aiguës et l'apex les basses fréquences, cette spatialisation du traitement fréquentiel étant par la suite conservée le long des voies auditives. Ils peuvent également être codés par un mécanisme temporel supporté par les neurones cochléaires, qui déchargent selon le principe de verrouillage de phase (*phase-locking*) de manière synchronisée à la structure temporelle fine du signal acoustique. Le verrouillage de phase peut

être efficace jusqu'à 4000 voire 5000 Hz [Moore, 1973]. Une association de ces deux mécanismes, tonotopique et temporel, a été proposée par plusieurs auteurs [Shamma and Klein, 2000, Oxenham et al., 2004]. Dans les modèles spectraux de perception de F0, les traitements de chaque harmonique sont combinés et intégrés pour déterminer la F0 du signal [Goldstein, 1973, Terhardt, 1974].

Dans le cas des harmoniques non résolus qui interagissent au sein d'un filtre auditif, F0 peut être extraite par le verrouillage de phase appliqué aux pics de la structure fine dans l'ensemble de l'onde acoustique, ou à l'enveloppe du signal après le premier filtrage auditif [Moore et al., 2008].

Au moins deux mécanismes seraient donc à l'œuvre pour coder F0 en fonction du caractère résolu ou non des harmoniques [Carlyon, 1994, Grimault et al., 2002]. L'association de ces mécanismes repose à la fois sur les propriétés tonotopiques cochléaires et sur l'analyse de la structure temporelle fine présente dans le signal acoustique, ou encore sur une analyse de la structure temporelle fine réalisée « au bon endroit » dans la cochlée [Cedolin and Delgutte, 2010, Carlyon et al., 2012].

### ***Traitement neuronal de F0***

Les études portant sur des sujets cérébro lésés ou des patients épileptiques traités par une cortectomie temporale ont apporté les premières informations concernant le traitement neuronal de F0, mais ces informations étaient variables d'une étude à l'autre et parfois contradictoires. La nature des stimuli employés est souvent critiquée a posteriori dans les études ultérieures, leur reprochant leur manque de spécificité pour F0. Des lobectomies partielles temporales droites emportant le gyrus transverse et le gyrus temporal supérieur droits peuvent ne pas avoir de conséquence sur le F0DL chez certains patients si l'énergie de F0 est faible voire absente [Warrier and Zatorre, 2004] mais l'inverse a également été montré, avec des lobectomies temporales droites provoquant une dégradation du F0DL [Johnsrude et al., 2000].

La plupart des études réalisées chez l'homme en IRMf pour déterminer les réseaux neuronaux impliqués dans le traitement de F0 sont également confrontées à la problématique du choix des stimuli auditifs, qui doivent être F0 spécifiques pour que les résultats puissent être interprétés correctement. Un certain nombre d'études ont utilisé un paradigme de soustraction par lequel l'activité neuronale en réponse à un stimulus véhiculant une F0 forte est comparée à l'activité observée pour le même stimulus sonore mais à F0 absente ou faible. Penagos et al. [Penagos et al., 2004] ont ainsi mis en évidence une activation dans une région latérale du

gyrus de Heschl, majorée quand F0 apparaissait saillante (tous harmoniques résolus) par rapport à l'activité mesurée en réponse à une F0 peu saillante (certains harmoniques demeurant irrésolus). D'autres travaux [Hall and Plack, 2007,2009,Staeren et al., 2009,Puschmann et al., 2010] ont également identifié cette région, avec des activations associées dans le planum tempore adjacent.

Certaines études chez l'animal sont également en faveur d'un traitement neuronal de F0 en dehors de l'aire auditive primaire, dans une région immédiatement antérieure et latérale par rapport à A1. Bendor et Wang [Bendor and Wang, 2005] ont ainsi mis en évidence dans cette région des neurones F0 spécifiques, à arrangement tonotopique, chez le marmouset vigile. Mais l'implication de A1 ne peut être écartée si l'on considère de récents résultats obtenus par Fischmann [Fishman et al., 2013] qui retrouvent, par des mesures multi unitaires réalisées chez le macaque en réponse à des sons complexes harmoniques un traitement de F0 présent dès A1 (voir figure I.3). Leurs résultats font état d'un traitement combiné de F0, tonotopique pour les F0 élevées, et par verrouillage de phase avec la structure temporelle fine du signal pour les F0 basses comprenant des harmoniques non résolus.

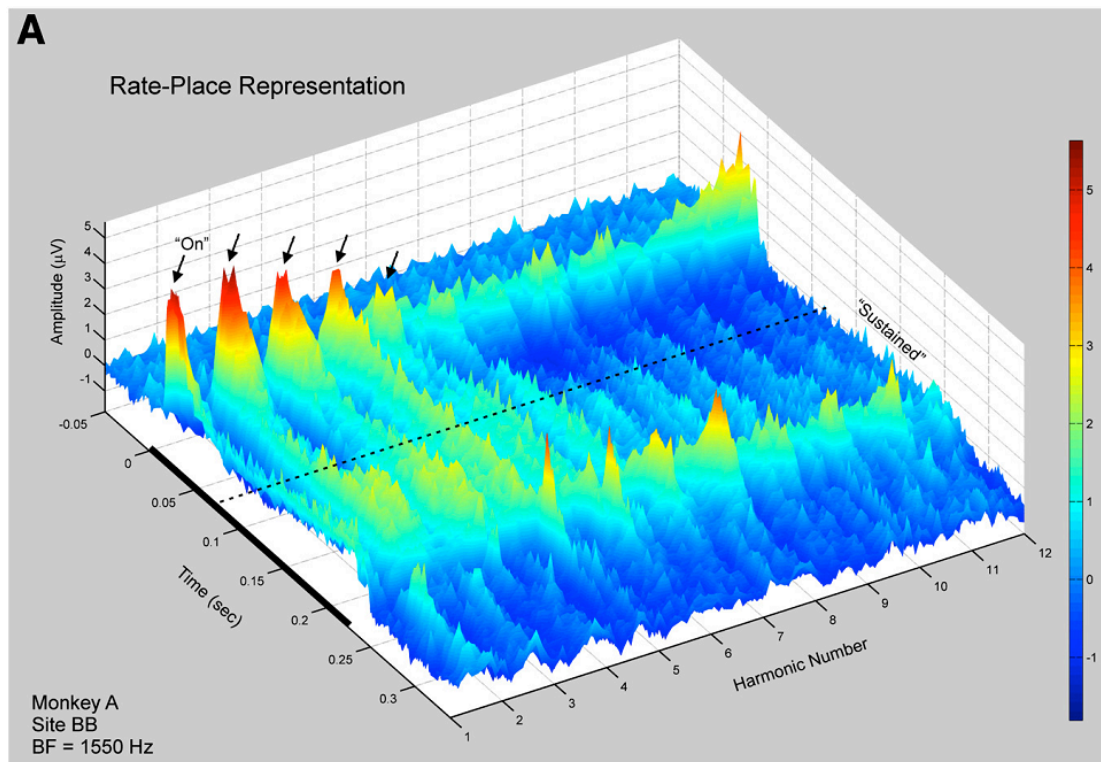


Figure I.3. Représentation des réponses évoquées par 89 sons complexes harmoniques, avec des pics d'amplitude répondant à la périodicité du signal, correspondant aux premiers harmoniques. Tiré de Fishman et al. (2013).

Le traitement neuronal de F0 serait donc initié dans A1 avant d'être complété dans les aires associatives adjacentes, dans la partie latérale du gyrus de Heschl et le planum tempore.

## 1.4 Le timbre

Le timbre est un paramètre acoustique multi dimensionnel difficile à définir, et repose plutôt sur un principe d'exclusion. L'ANSI le définit comme « l'attribut de la sensation sonore permettant de juger que deux sons complexes de même intensité, fréquence fondamentale et durée de présentation sont différents» et ajoute que «le timbre dépend essentiellement du spectre sonore, mais aussi de la forme de l'onde sonore, de son amplitude, de sa localisation et des caractéristiques temporelles du stimulus». (American National Standards Institute, 1960, 45). Le timbre permet ainsi de distinguer une même note jouée par deux instruments différents. Il s'agit donc un attribut essentiellement qualitatif du son, souvent défini par un adjectif (on parle de voix « claire », « voilée », d'une sonorité « ronde » caractérisant un jeu de trompette), qui est difficilement appréhendé quantitativement par ses propriétés acoustiques.

Le timbre perçu dépend ainsi des trois indices acoustiques fondamentaux que nous venons de décrire et de leurs interactions. Plusieurs travaux ayant porté sur ce paramètre se sont appuyés sur des analyses multi dimensionnelles (MDS pour Multi Dimensional Scaling) sur des matrices de dissimilarités constituées sur la base de jugements de similarité entre deux stimuli auditifs. Plusieurs espaces multi dimensionnels [Grey, 1977, McAdams et al., 1995] ont ainsi été proposés afin de représenter ces jugements et d'identifier les corrélats acoustiques qui les sous-tendent. L'étude de McAdams (1995) a permis de déterminer les dimensions sur lesquelles les sujets s'appuyaient pour juger de la proximité entre différents sons musicaux synthétiques. Le transitoire d'attaque de ces instruments, correspondant aux modifications spectro-temporelles dynamiques qui caractérisent les premières millisecondes d'un son, intervenait par exemple comme dimension discriminante. Dans d'autres études réalisées par J. Marozeau [Marozeau et al., 2003, Marozeau and de Cheveigne, 2007], F0 était identifiée comme discriminante pour une des dimensions de l'analyse MDS. L'adjectif « clair » est ainsi souvent attribué aux sons complexes harmoniques portant une F0 élevée. Plusieurs paramètres caractérisant le spectre du son ont également été identifiés comme facteurs acoustiques déterminant la perception du timbre :

- le centroïde spectral correspond à une moyenne des fréquences présente dans le spectre du son, pondérée par leur amplitude relative, qui est également corrélé à la clarté du son.
- la déviation spectrale caractérise la régularité de l'enveloppe spectrale et est associée à la sensation de richesse du son [Faure, 2000].



D'autres descripteurs temporels tels que l'impulsivité [Marozeau, 2004], qui correspond à la durée pendant laquelle l'enveloppe temporelle dépasse 40% de son amplitude maximale, ont été utilisés pour caractériser le timbre.

Nous avons à ce stade décrit trois indices acoustiques fondamentaux, à la base de tout phénomène perceptif sur le plan auditif, qui ont des caractéristiques intrinsèques mais également inter déterminées. Ces indices inter agissent au sein de l'espace acoustique multi dimensionnel qu'est le timbre. Les parties suivantes de ce chapitre sont consacrées aux modalités perceptives des sons structurés, linguistiques, para linguistiques ou non linguistiques.

## 2. PERCEPTION DE LA PAROLE

La perception de la parole repose sur des processus cognitifs permettant l'intégration d'unités linguistiques et la compréhension du signal reçu. Nous décrirons dans un premier temps les caractéristiques acoustiques des phonèmes avant d'évoquer les mécanismes cognitifs impliqués dans le traitement et l'intégration de la parole.

### 2.1 Caractéristiques acoustiques des phonèmes

#### *Structure acoustique générale*

On distingue deux grands types de phonèmes en fonction de leurs mécanismes de production. Les phonèmes voisés, comme toutes les voyelles et quelques consonnes, sont émis par une vibration des cordes vocales à l'origine de variations périodiques de pression acoustique. La fréquence fondamentale correspond à la fréquence de vibration des cordes vocales et le son transite ensuite par les cavités supra laryngées qui vont jouer le rôle de résonateurs acoustiques. Dans ces cavités se produit l'amplification sélective de certaines régions fréquentielles appelées formants. Ceux-ci sont aisément identifiés sur un spectrogramme (voir figure I.4) sous la forme de bandes fréquentielles à forte énergie acoustique. Ils sont numérotés F1, F2, F3, etc... Dans le signal de parole, F1 est en général compris entre 350 et 1000 Hz et F2 entre 500 et 2500 Hz.

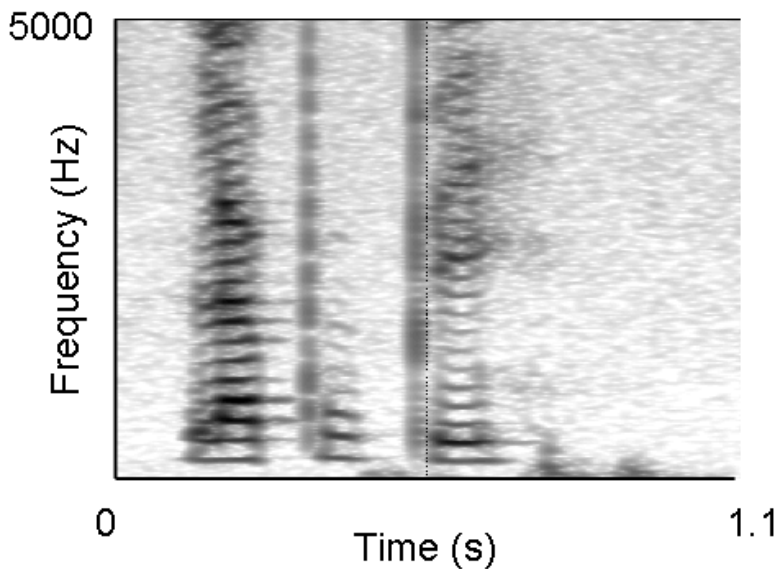


Figure I.4. Spectrogramme correspondant à l'énoncé de « la poupée », prononcée par un locuteur féminin. On distingue les trois syllabes prononcées, la fréquence fondamentale ainsi que les formants sous forme de bandes à haute énergie.

Les phonèmes non voisés sont quant à eux produits par des modifications du flux d'air provenant des poumons et génèrent des sons apériodiques. Certaines consonnes ont une structure acoustique mixte, caractérisée par un voisement associé des modifications du flux d'air.

### *Caractéristiques acoustiques des consonnes*

Les consonnes constituent le segment de parole essentiel à sa discrimination puisqu'elles supportent près de 80% de l'information transmise. Elles sont caractérisées par le lieu d'occlusion de la cavité orale et par les modalités de leur émission (consonnes voisées ou non voisées). Leur classification distingue cinq groupes constitués en fonction des modalités précises de leur articulation (voir tableau I.1a):

- voisement
- nasalité
- frication
- durée
- place de l'occlusion

Elles sont caractérisées sur le plan acoustique par les formants de bruit et les transitions formantiques. Les formants de bruit sont des bandes de fréquence relativement étendues dans lesquelles les intensités spectrales sont particulièrement élevées. Ces formants correspondent aux bruits d'explosion et de constriction générés lors de l'articulation des consonnes. Les transitions formantiques caractérisent les modifications apportées aux voyelles par les consonnes les entourant. La fréquence vers lesquelles ces transitions formantiques convergent pour le deuxième formant est appelée locus et permet de discriminer les consonnes occlusives. Les consonnes occlusives bilabiales /p/ et /b/ ont par exemple un locus situé dans les basses fréquences, aux environs de 700 Hz alors que le locus des labiodentales /t/ et /d/ est plus élevé, aux environs de 1800 Hz. Les formants de bruit interviennent dans la discrimination des consonnes constrictives. Les consonnes sifflantes comme /s/ et /z/ présentent des formants de bruit pouvant s'étendre jusqu'à 4000 Hz.

MODE D'ARTICULATION				LIEU D'ARTICULATION								
Type de consonne selon le mouvement	Passage de l'air		Vibration des cordes vocales	Bi-labiale	labio-dentale	Apico-dentale	Apico-alvéolaire	Pré-dorso-alvéolaire	Pré-dorso-pré-palatale	médio-palatale	Dorso-palatale ou vélaire	Post-dorso-uvulaire
OCCLUSIVE	ORAL		NON-VOISEE	p		t					k	
			VOISEE	b		d					g	
	NASAL		VOISEE	m		n				ɲ	(ŋ)	
CONSTRUCTIVE	ORAL	TYPÉ DE CONSTRUCTIVE										
		FRICATIVE	NON-VOISEE		f			s	f			
			VOISEE		v			z	ʒ			
		LATERALE	VOISEE				l					
	VIBRANTE	VOISEE									r	

Tableau I.1a : Classification des consonnes en fonction du mode de leur articulation et des positions respectives de la langue et des lèvres (lieu d'articulation). Tiré de Gezundhajt H.

### *Caractéristiques acoustiques des voyelles*

Les voyelles présentent en règle générale une durée plus longue que les consonnes avec une intensité également plus importante. Toutes les voyelles étant voisées, F0 et les différents formants qui les composent contribuent grandement à leur identification. La position respective de la langue et des lèvres permet leur articulation et la modulation des premier et deuxième formants, sur laquelle repose 70% des variations d'une voyelle à l'autre [Pols et al., 1969]. F1 peut évoluer entre 320 Hz pour la voyelle /u/ et 1000 Hz pour /a/. F2 peut varier quant à lui entre 800 Hz pour /u/ et 3200 Hz pour la voyelle /i/. Les fréquences caractéristiques d'une voyelle peuvent en effet être atteintes juste avant l'initiation de la consonne suivante par des effets dits co-articulatoires.

## **2.2. Structure temporelle du signal de parole**

La qualité des résultats obtenus par les premiers patients sourds bénéficiant d'une stimulation électrique par un implant cochléaire mono canal [House, 1978, Hochmair-Desoyer et al., 1980] a conduit à reconsidérer la structure purement temporelle des signaux de parole. Ces implants ne pouvaient en effet délivrer qu'une information temporelle, caractérisée par des fluctuations d'amplitude du signal en fonction du temps. Parallèlement, un intérêt croissant s'est développé pour le spectre des neuropathies auditives, dont la fréquence est vraisemblablement sous-estimée. Ces pathologies sont caractérisées par des anomalies de la synchronisation des impulsions du nerf auditif par rapport aux fluctuations d'amplitude plus ou moins rapides dans le signal acoustique [Starr et al., 1996, Giraudet and Avan, 2012, Narne, 2013]. Une attention particulière s'est donc portée sur la structure temporelle du signal acoustique.

Dans la publication référence de 1992, Rosen [Rosen, 1992] propose de décomposer la structure temporelle du signal de parole en trois types d'informations selon la fréquence des fluctuations d'amplitude (voir figure I.5):

- l'enveloppe du signal comprend les fluctuations d'amplitude à faible fréquence, entre 2 et 50 Hz. Ces fluctuations transmettent d'importantes informations relatives à l'évolution de l'intensité dans le temps. Elles rendent compte des quatre principaux temps d'un son de parole, compris dans la séquence ADSR : le transitoire d'attaque (A pour *Attack*), sa retombée (D pour *Decay*), la durée pendant laquelle le son est soutenu (S pour *Sustain*) et enfin le transitoire d'extinction (R pour *Release*). Ces fluctuations à basse fréquence permettent par ailleurs de transmettre le temps de voisement, utile à la discrimination de certaines consonnes.

- la périodicité du signal de parole comprend les fluctuations d'amplitude au cours du temps entre 50 et 500 Hz. Ces fluctuations peuvent permettre de représenter le voisement et donc la fréquence fondamentale de la voix du locuteur. L'identification des voyelles et des consonnes voisées est donc facilitée par ce type d'information. De même, ce phénomène temporel permet de véhiculer l'intonation, indispensable à la perception de la prosodie linguistique et des langues tonales, comme le mandarin ou le thaï.

- Rosen définit ensuite la structure temporelle fine comme les fluctuations rapides de l'amplitude dans le signal de parole, allant de 600 Hz à 10 kHz. Ce type d'information temporelle fournit des informations segmentales concernant la place de l'articulation, permettant la discrimination fine des consonnes et facilite la perception des formants et des transitions formantiques précédant ou suivant les voyelles, leur rendant ainsi l'ensemble de leurs qualités.

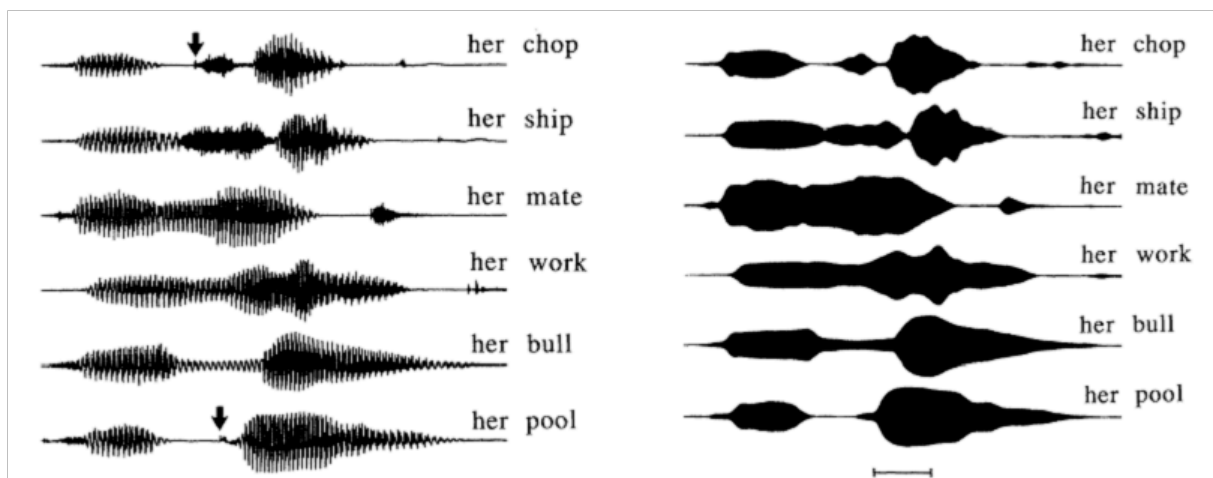


Figure I.5. ondes acoustiques correspondant à 6 mots précédés de leur article correspondant à l'enregistrement naturel (à gauche) et à l'enregistrement modifié après application d'un filtre passe-bas à 20 Hz, préservant ainsi la majeure partie des indices d'enveloppe, mais éliminant les indices de périodicité et liés à la structure fine. Tiré de Rosen (1992).

D'autres auteurs envisagent la décomposition temporelle du signal de parole différemment en distinguant deux types de phénomènes : les informations d'enveloppe à basse fréquence et les informations relatives à la structure fine à haute fréquence [Lorenzi et al., 2006, Moore, 2008]. Mais l'intérêt de la perception des informations temporelles à haute fréquence demeure souligné pour faciliter la discrimination de la parole dans le bruit [Schooneveldt and Moore, 1987, Lorenzi et al., 2006]. La structure temporelle fine permet ainsi de discriminer les éléments phonémiques lors des fluctuations de bruits environnants et de faciliter l'identification du locuteur d'intérêt par la transmission d'indices liés à la fréquence fondamentale de sa voix.

### **2.3 Traitement neuronal de la parole**

Le modèle classique de perception du langage, progressivement élaboré depuis les premières contributions de Broca (1861) et de Wernicke (1874), fait état de deux régions hémisphériques gauches, une temporale postérieure et une frontale inférieure, reliées par le faisceau arqué, impliquées dans le traitement de la parole. L'essor des nouvelles méthodes d'exploration non invasives et l'intérêt croissant pour les aspects neurobiologiques de la parole ont favorisé l'émergence de nouveaux modèles et de dresser de nouvelles cartes de la neuro anatomie fonctionnelle de la parole.

On identifie ainsi deux principaux réseaux dans le traitement de la parole [Hickok and Poeppel, 2004, Saur et al., 2008], qui ont une origine commune dans le gyrus temporal supérieur puis divergent sous la forme de deux voies : une voie dorsale et une voie ventrale. La voie dorsale projette sur des régions pariétales inférieures et frontales postérieures et traite l'aspect audio-moteur de la parole en faisant correspondre les entrées auditives aux représentations phonologiques. Cette voie est donc à l'œuvre dans des tâches phonologiques de répétition de mots. La voie ventrale est impliquée dans l'accès aux représentations sémantiques et donc dans la compréhension des mots parlés. Cette voie projette sur les gyri temporaux moyens et inférieurs (voir figure I.6).

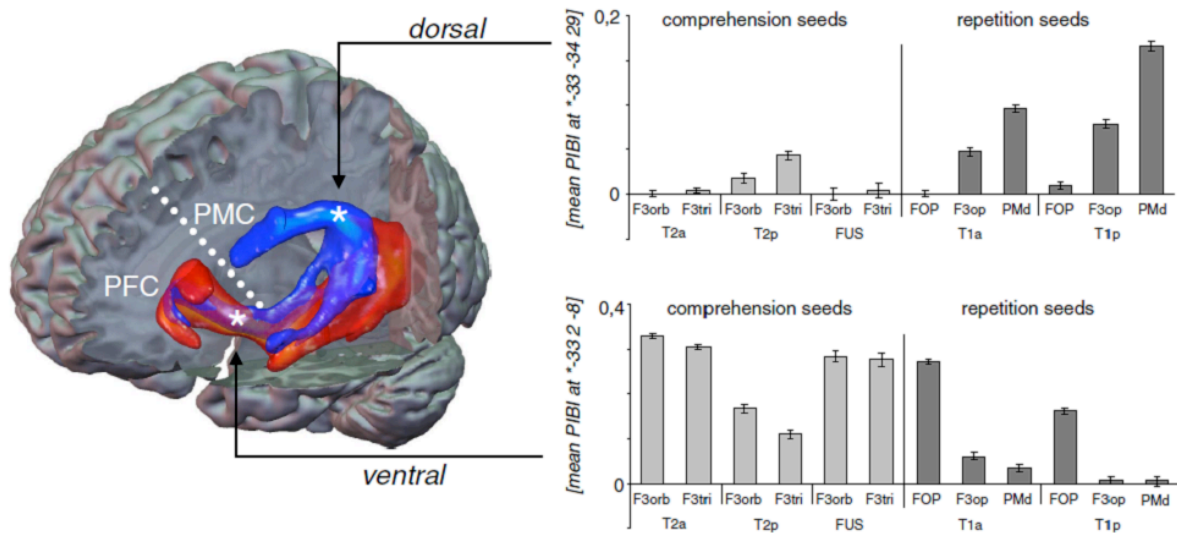


Figure I.6. Représentation des deux voies dorsale et ventrale du traitement de la parole, mises en évidence par IRM avec tracking de fibres. La voie dorsale est impliquée dans le traitement phonologique et la voie ventrale dans le traitement sémantique. Tiré de Saur et al. (2008).

Les corrélats neuronaux des différents aspects temporels de la parole, i.e les fluctuations d'amplitude à basse et à haute fréquence, pourraient être trouvés dans les oscillations neuronales intrinsèques (oscillations delta : 1-3 Hz ; theta : 4-8 Hz ; low gamma : 30-50 Hz ; high gamma : 75-115 Hz), ainsi que plusieurs travaux le suggèrent [Ghitza et al., 2012, Giraud and Poeppel, 2012, Luo and Poeppel, 2012, Kubanek et al., 2013]. Kubanek J et al (2013) ont ainsi montré que l'activité high gamma était en rapport direct avec l'enveloppe du signal de parole et que cette relation était particulièrement significative en réponse à un signal de parole dans les régions postérieures du gyrus temporal transverse.

Plusieurs travaux récents, qu'ils concernent la psycho acoustique ou les neurosciences, suggèrent donc l'importance des informations temporelles véhiculées par le signal de parole. Les contributions respectives de l'enveloppe et de la structure temporelle fine commencent ainsi à être relativement bien connues pour la discrimination de la parole. En revanche, leur implication potentielle dans le traitement des informations para linguistiques et de la musique demeure vraisemblablement en arrière-plan des informations spectrales.

### **3. PERCEPTION DE LA VOIX**

La voix humaine représente le support de la parole et si l'on s'attache à décrire ses caractéristiques acoustiques ou ses modalités de traitement neuronal, ce sont essentiellement ses aspects qualitatifs qui sont envisagés.

#### **3.1 Caractéristiques acoustiques de la voix**

La voix est caractérisée par une fréquence fondamentale  $F_0$ , correspondant à la fréquence de vibration des cordes vocales, et par un timbre, dont nous soulignons la nature multidimensionnelle sur le plan acoustique. Plusieurs caractéristiques physiques du locuteur peuvent faire varier ces deux paramètres.

$F_0$  dépend du genre du locuteur, avec des valeurs plus élevées (autour de 220 Hz) retrouvées chez les femmes par rapport aux hommes (autour de 110 Hz) [Gelfer and Mikos, 2005, Iseli et al., 2007]. Elle dépend de la même manière de son âge, les enfants présentant des valeurs de  $F_0$  plus élevées que les adultes alors les sujets âgés tendent à uniformiser ces valeurs entre hommes et femmes [Linville, 1996].

Le timbre d'une voix est essentiellement déterminé par la manière dont les harmoniques de  $F_0$  vont être traités dans les cavités de résonance supra laryngées. Des caractéristiques physiques, liées à la taille du tractus vocal, vont donc influencer le regroupement des harmoniques en formants et les mêmes différences de genre [Busby and Plant, 1995] et d'âge [Kent and Forner, 1979, Hillenbrand et al., 1995].

Les différences vocales entre le genre masculin et le genre féminin se manifesteraient dès l'âge huit ans par des modifications touchant les 4 premiers formants [Bennett, 1981]. Les énergies respectives contenues dans les deux premiers formants auraient vraisemblablement un rôle discriminant, avec une énergie acoustique majorée dans  $F_1$  par rapport à  $F_2$  chez les femmes, l'inverse étant observé chez les hommes [Klatt and Klatt, 1990].

Les principaux paramètres acoustiques permettant de définir la voix dépendent de caractères physiques du locuteur, mais également de son état émotionnel ou de l'information orale qu'il souhaite souligner. La prosodie, que nous aborderons dans le paragraphe suivant, influence donc également les caractéristiques de la voix.



### **3.2 Traitement neuronal de la voix**

Les premières évidences d'un traitement spécifique de la voix, dissocié de celui de la parole ont été apportées par les études de patients présentant une lésion cérébrale touchant le lobe temporal.

Ces études lésionnelles ont ainsi mis en évidence un trouble spécifique de reconnaissance des voix familières, la phonagnosie, sans atteinte des capacités de discrimination de la parole, par des lésions souvent d'origine vasculaire touchant le lobe temporal droit [Assal et al., 1976, Van Lancker and Canter, 1982, Van Lancker and Kreiman, 1987, Van Lancker et al., 1988]. Certaines aptitudes auditives des patients rapportés pouvaient toutefois être altérées dans ces premières études, sous la forme de troubles de perception de la musique ou de compréhension de la parole, l'hypothèse d'un centre régissant la perception de la demeure insuffisamment étayée.

Sous l'impulsion de P. Belin, cette hypothèse a été de nouveau explorée par des travaux en IRMf cherchant à identifier les possibles corrélats neuronaux de la perception de la voix humaine. Là encore, la question de la nature des stimuli s'est révélée problématique puisqu'il était difficile de conclure à l'existence d'une région neuronale spécifique après les premiers travaux d'Imaizumi [Imaizumi et al., 1997]. Ceux-ci avaient en effet mis en évidence une activation des deux pôles temporaux en réponse à des voix familières aux sujets testés. Il était dans ces conditions difficile de conclure à la spécificité « voix » de cette activation, la tâche engageant nécessairement d'autres réseaux, tels que la mémoire. P. Belin et son équipe [Belin et al., 2000, Belin et al., 2004] ont pu mettre en évidence trois régions neuronales localisées dans le sulcus temporal supérieur (STS) préférentiellement droit, sensibles et spécifiques de la voix humaine. Ces aires étaient activées dans une tâche d'écoute passive de stimuli vocaux et de stimuli correspondant à des bruits environnementaux.

Ces régions du STS droit, plus tard dénommées Temporal Voice Area, ont été identifiées par des études ultérieures [von Kriegstein et al., 2003, Kriegstein and Giraud, 2004] comparant les activations obtenues lors d'une tâche de reconnaissance de la voix et lors d'une tâche de discrimination de la parole. La spécificité absolue de ces régions pour la reconnaissance des voix demeure incomplètement déterminée puisque l'activation de ces régions a également été mise en évidence en fonction du nombre de locuteurs [Wong et al., 2004] ou en réponse à des stimuli vocaux porteurs d'émotion.

Belin a donc proposé un modèle hiérarchique de perception de la voix humaine [Belin et al., 2004, Campanella and Belin, 2007], inspiré du modèle de perception des visages décrit par Bruce et Young [Bruce and Young, 1986], et interagissant avec lui dans un processus cognitif

de reconnaissance de l'individu. Ce modèle prévoit un traitement modulaire des informations relatives à la voix, depuis l'analyse des caractéristiques acoustiques structurales de la voix jusqu'aux représentations liées à l'identité du locuteur et à l'analyse sémantique du message oral (voir figure I.7).

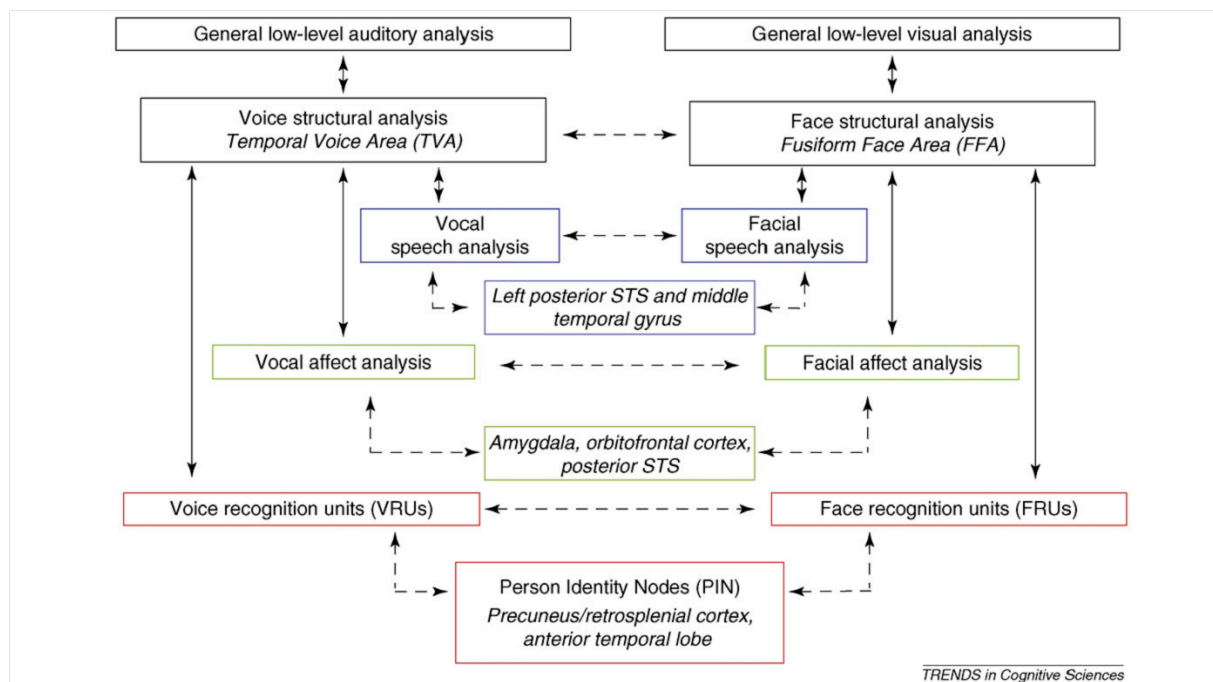


Figure I.7. Modèle d'intégration voix-visage décrivant les différents modules, leurs interactions, et les niveaux impliqués dans le processus visuo-auditif de reconnaissance de l'identité. Tiré de Campanella et Belin (2007).

## 4. PERCEPTION DE LA PROSODIE

### 4.1 Définitions et caractéristiques

La prosodie est communément définie comme la manière dont les mots sont prononcés et agencés au sein d'un énoncé oral. Elle regroupe l'ensemble des phénomènes supra segmentaux d'intonation, d'accentuation et de quantité syllabique et peut par ce biais être considérée comme le lien entre la parole et la voix. Elle est ainsi caractérisée par des modifications acoustiques de la voix qui porte la parole, modifications qui permettent de moduler le sens du contenu purement linguistique de l'énoncé.

Dans *Aurélien*, roman de Louis Aragon, la description de Bérénice illustre parfaitement ce lien entre voix et parole, que réalise la prosodie : « La seule chose qu'il aima d'elle tout de suite, ce fût la voix. Une voix de contralto chaude, profonde, nocturne. Aussi mystérieuse que les yeux de biche sous cette chevelure d'institutrice. Bérénice parlait avec une certaine lenteur. Avec de brusques emballements, vite réprimés qu'accompagnaient des lueurs dans les yeux comme des feux d'onyx. ».

Deux principales composantes prosodiques sont généralement identifiées :

- la prosodie linguistique par laquelle le sens des unités linguistiques présentes dans l'énoncé est modulé. Cette composante délivre ainsi les informations grammaticales quant au mode selon lequel la phrase est prononcée (interrogatif, affirmatif, exclamatif) mais également des informations para grammaticales, nécessaires à la perception de regroupements de mots sous la forme d'unités prosodiques, par l'intermédiaire des frontières prosodiques.
- la prosodie affective par laquelle l'état émotionnel ou l'intention du locuteur sont transmis à l'auditeur.

La prosodie entretient donc des relations étroites avec les caractéristiques qualitatives de la voix que nous avons évoquées dans le paragraphe précédent mais également avec le contenu linguistique de l'énoncé auquel elle se rapporte.

Si l'on considère par exemple l'énoncé suivant : « Vous voulez ce poste ? Mais vous êtes très...jeune. » . L'information prosodique qu'il contient permet d'identifier une question dans la première phrase, prononcée vraisemblablement avec étonnement. La deuxième phrase contient un silence traduisant la précaution que prend le locuteur à choisir le dernier mot, et selon la prosodie utilisée, on y percevra de la réticence ou une agréable surprise.

Les phénomènes prosodiques reposent sur les variations dans le temps des trois indices acoustiques fondamentaux : F0, la durée et l'intensité [Nagel et al., 1994]. L'accentuation d'un mot dans la phrase s'appuiera classiquement sur une augmentation de l'intensité parfois associée à une augmentation de F0 [Wingfield, 1975, Speer et al., 1996]. Le mode interrogatif est typiquement défini par une F0 ascendante en fin de phrase. De la même manière, la prosodie affective repose sur des changements de ces indices fondamentaux et on retrouve par exemple classiquement associés à la joie une F0 plus élevée, caractérisée par d'importantes variations de sa valeur, et un rythme de parole accéléré [Fonagy, 1958]. La colère se manifeste par une augmentation de l'amplitude sur les syllabes accentuées et une élévation moyenne de F0 [Fonagy, 1966, Williams and Stevens, 1972].

#### **4.2 Traitement neuronal de la prosodie**

Les liens étroits qui unissent la prosodie à l'information de parole rendent difficile la séparation des processus impliqués dans chacun de ces percepts. Les études sur sujets cérébro-lésés ont retrouvé une latéralisation inconstante du traitement de l'information prosodique. Certains auteurs ont en effet décrit une prédominance de l'hémisphère droit [Weintraub et al., 1981, Bradvik et al., 1991] alors que des lésions bilatérales peuvent également rendre compte de troubles de perception de la prosodie [Bryan, 1989]. Il semble toutefois que l'hémisphère droit intervienne de manière prédominante lorsque l'information linguistique est dégradée [Perkins et al., 1996].

Les études en neuro imagerie fonctionnelle renforcent l'hypothèse d'une spécialisation hémisphérique droite, et mettent en évidence des activations temporelles postéro-supérieures et fronto-operculaires si l'on compare les activations obtenues en réponse à l'information prosodique seule aux activations en réponse au signal auditif global [Meyer et al., 2002, Meyer et al., 2004] (voir figure I.8).

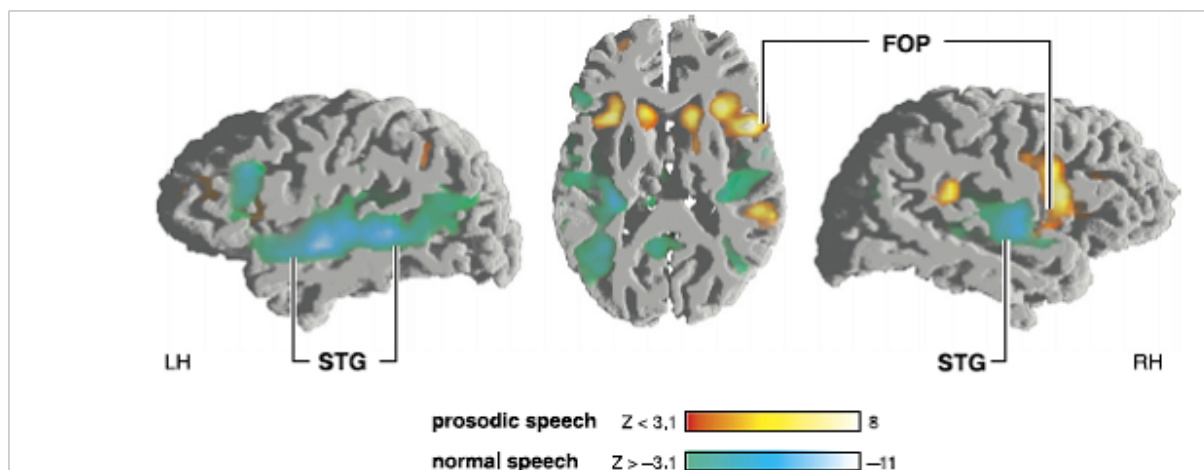


Figure 1.8. Représentation des activations en réponse à un signal auditif ne contenant que l'information prosodique (jaune-rouge) par rapport à un signal contenant l'information prosodique et l'information linguistique (bleu-vert). Tiré de Meyer et al. (2002).

La partie dorsolatérale du cortex préfrontal droit et des régions cérébelleuses droites ont également montré une activité augmentée en réponse à une information relative aux frontières prosodiques [Strelnikov et al., 2006].

Le traitement neuronal de la prosodie serait ainsi majoritairement assuré par l'hémisphère droit, par plusieurs régions neuronales incluant la région postérieure et supérieure du cortex temporal.

## 5. PERCEPTION AUDITIVE EXTRA LINGUISTIQUE : EXEMPLE DE LA MUSIQUE

Il est impossible de définir simplement la musique tant son application est large et ses représentations diverses. Le développement qu'elle a connu constitue une aptitude spécifique à l'espèce humaine, et sa diffusion est tout aussi universelle que le langage. Après une réflexion intense et coûteuse, mon collègue et ami Benoit Roby [Roby, 2012] l'a définie dans sa thèse de médecine comme étant « l'art de combiner les sons et des silences dans un but expressif ».

## 5.1 Paramètres acoustiques spécifiques à la perception musicale

La musique repose sur des variations des trois indices acoustiques fondamentaux que nous avons décrits, composées ou improvisées afin d'induire une réponse émotionnelle chez l'auditeur (les termes utilisés et les descriptions se limitent à la musique occidentale). Ces indices et certaines de leurs variations ont une dénomination particulière dans le domaine musical.

**La fréquence tonale** : qui correspond à la fréquence réelle d'un ton ou d'une note (voir tableau I.1). La suite de l'ensemble des notes naturelles, c'est-à-dire non altérées, prises à partir de Do, constitue la gamme de Do majeur. Elle est caractérisée par la succession des intervalles suivants : **ton, ton, demi-ton, ton, ton, ton, demi-ton**.

**L'intervalle** : c'est l'écart tonal entre deux notes, exprimé en tons et demi-tons. On distingue, par ordre croissant : l'unisson, la seconde, la tierce, la quarte, la quinte, la sixte, la septième et l'octave.

**Le timbre** : qui permet de distinguer un instrument d'un autre.

**Le volume** : ou puissance sonore. En musique, les nuances permettent de faire varier le volume. Les compositeurs utilisent les termes italiens suivants, du plus doux au plus puissant : pianississimo, pianissimo, piano, mezzo piano, sotto voce, mezza voce, poco forte, forte, fortissimo, fortississimo.

**Le rythme** : est la durée d'une série de notes et la manière dont elles se regroupent par unités. Dans le système de notation occidentale, on utilise les termes suivants, dans l'ordre décroissant de durée, chaque durée valant la moitié de celle qui la précède à tempo constant : carrée, ronde, blanche, noire, croche, double croche, triple croche, etc...De même, pour les silences : bâton de pause, pause, demi-pause, soupir, demi-soupir, quart de soupir, huitième de soupir, etc...

**Le tempo** : est l'allure générale d'un morceau. Les compositeurs utilisent principalement les termes italiens suivants, de la vitesse la plus lente à la plus élevée : largo, lento, adagio, andante, moderato, allegretto, allegro, vivace, presto, prestissimo. Pour mémoire, le tempo «largo» correspond à 40 à 60 pulsations par minute, «prestissimo» à plus de 188.

**La localisation spatiale** : La provenance dans l'espace du stimulus sonore.

Là où le concept de musique prend véritablement naissance, c'est lorsque ces éléments fondamentaux sont agencés de manière à créer du sens.

**La mesure** : désigne la manière dont les notes sont regroupées dans le temps, à partir du volume sonore et du rythme. Par exemple, la valse organise les tons par groupes de trois temps.

**La tonalité** : correspond à une hiérarchie entre les tons du morceau. Nous reviendrons ultérieurement sur le fait que dans une gamme, certaines notes sont plus importantes que d'autres.

**La mélodie** : est le thème principal du morceau, l'enchaînement de notes et de rythmes que l'auditeur va retenir.

**Le contour** : est la forme générale de la mélodie (si elle «monte» ou si elle «descend»).

**L'harmonie** : résulte de la combinaison de plusieurs notes jouées simultanément, ayant pour objectif d'instaurer un contexte tonal, permettant au compositeur de satisfaire ou non les attentes de l'auditeur pour les notes suivantes.

**La cadence** : est un enchaînement de mélodies ou d'harmonies utilisées pour structurer et ponctuer une phrase musicale. Les cadences se construisent par combinaison de succession de degrés musicaux.

Note	Base	Octave	Octave	Octave	Octave	Octave	Octave	Octave	Octave 8	Octave 9
La	27,5	55	110	220	440	880	1760	3520	7040	14080
Sib	29,14	58,27	116,54	233,08	466,16	932,33	1864,66	3729,31	7458,62	14917,24
Si	30,87	61,74	123,47	246,94	493,88	987,77	1975,53	3951,07	7902,13	15804,27
Do	32,7	65,41	130,81	261,63	523,25	1046,5	2093	4186,01	8372,02	16744,04
Do#	34,65	69,3	138,59	277,18	554,37	1108,73	2217,46	4434,92	8869,84	17739,69
Ré	36,71	73,42	146,83	293,66	587,33	1174,66	2349,32	4698,64	9397,27	18794,55
Mib	38,89	77,78	155,56	311,13	622,25	1244,51	2489,02	4978,03	9956,06	19912,13
Mi	41,2	82,41	164,81	329,63	659,26	1318,51	2637,02	5274,04	10548,08	
Fa	43,65	87,31	174,61	349,23	698,46	1396,91	2793,83	5587,65	11175,3	
Fa#	46,25	92,5	185	369,99	739,99	1479,98	2959,96	5919,91	11839,82	
Sol	49	98	196	392	783,99	1567,98	3135,96	6271,93	12543,85	
Sol#	51,91	103,83	207,65	415,3	830,61	1661,22	3322,44	6644,88	13289,75	

Tableau I.1. Valeur en fréquence (Hz) des notes de la gamme tempérée

## 5.2 Traitement neuronal de la musique

Le traitement de la musique, est assuré par un réseau soutenant vraisemblablement le système modulaire de reconnaissance musicale, décrit par I. Peretz (SRM) [Peretz and Coltheart, 2003] (voir figure I.9). Ce système met à l'œuvre deux principaux processus cognitifs, comparables à ceux impliqués dans le traitement de la parole. La première étape consiste en l'analyse de la structure acoustique globale de l'extrait musical par un module dédié au traitement du rythme et un module dédié au traitement de la mélodie. Les produits de cette première analyse sont ensuite confrontés aux représentations stockées dans le répertoire musical, cette étape s'apparentant à l'accès sémantique observé dans le traitement de la parole.



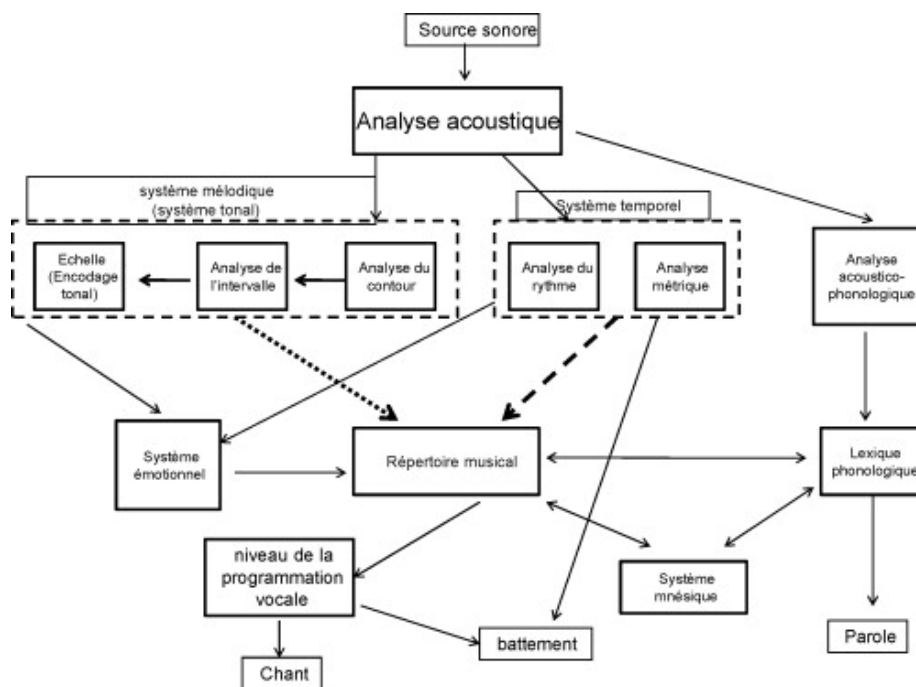


Figure I.9. Système de Reconnaissance Musicale, adapté de Peretz et al. (2010)

Ce système de reconnaissance musicale est vraisemblablement supporté par un vaste réseau de régions neuronales, dispersées mais prédominantes dans l'hémisphère droit, sauf chez certains sujets à forte pratique musicale. Les régions impliquées dans le traitement de la fréquence fondamentale, dans la partie latérale du gyrus de Heschl et le planum temporele montrent dans plusieurs études des activations en réponse à un signal musical [Patterson et al., 2002, Warren et al., 2003]. Le gyrus temporal supérieur droit assurerait ainsi le traitement du contour par une région antérieure et de la reconnaissance de la mélodie par une région postérieure [Liegeois-Chauvel et al., 1998]. Des régions frontales inférieures bilatérales et thalamiques médiales ont été identifiées comme intervenant dans le traitement de la tonalité et de son caractère majeur ou mineur.

## Chapitre II

# La perception auditive chez le sujet implanté cochléaire

Depuis l'avènement de l'implant cochléaire dans les années 1990, de multiples publications portant sur la restauration de la fonction auditive après implantation ont enrichi la littérature et l'on compte ainsi plus de 800 références concernant les résultats fonctionnels obtenus.

La technique chirurgicale, les populations concernées, le devenir de l'oreille controlatérale, l'extension des indications, l'impact médico-économique, la réorganisation des aires auditives corticales sont autant de thématiques différentes pouvant être reliées à la question des résultats post-implantation.

L'objectif de notre travail est d'étudier la perception auditive para et extra linguistique chez le sujet adulte, implanté cochléaire unilatéral, atteint d'une surdité neurosensorielle post linguale, bilatérale, sévère à totale. Dans cette partie introductive, nous nous intéresserons donc aux résultats perceptifs obtenus par cette population et aux modalités de la réorganisation cérébrale qui les accompagne. Les résultats sont caractérisés par deux observations presque constantes : l'efficacité globale de l'implant cochléaire dans la restauration d'une fonction auditive partielle, et la variabilité dans les performances obtenues, sous – tendue par des facteurs qui n'ont pas tous été élucidés. Avant d'aborder la question des performances obtenues, dans différents domaines auditifs, il nous apparaît nécessaire de faire le point sur le traitement du signal acoustique réalisé par le processeur vocal de l'implant. L'implant assure une stimulation électrique directe des fibres nerveuses afférentes du nerf auditif mais cette stimulation est précédée de traitements multiples et complexes, qui aboutissent à une dégradation du signal initial, tant sur le plan spectral que sur le plan temporel. De la même manière que nous avons précisé les différents paramètres acoustiques sur lesquels le sujet normo entendant s'appuyait dans plusieurs compétences auditives, il convient d'envisager le signal proposé au sujet implanté pour appréhender l'environnement sonore.

## **1. CODAGE ET RESTITUTION DU SIGNAL ACOUSTIQUE PAR L'IMPLANT COCHLEAIRE**

Le principe général du codage de l'information sonore en stimulations électriques est de se rapprocher autant que possible des conditions de stimulation naturelles des premiers neurones auditifs par les cellules ciliées. Il est bien entendu impossible d'obtenir une stricte

reproduction de ces conditions compte tenu du nombre de neurones afférents impliqués et de la subtile modulation de leur activité.

Les stratégies de codage ont donc été développées dans le but de restituer au nerf auditif le maximum d'informations relatives au signal de parole. L'intérêt pour les informations paralinguistiques et suprasegmentales, la question de la musique, sont venus dans un deuxième temps et constituent actuellement un des enjeux de l'amélioration des résultats auditifs après implantation. Schématiquement, le codage de ces deux grandes catégories de signaux acoustiques, la parole d'un côté et les informations extra/para linguistiques de l'autre, est réalisé en privilégiant deux types d'indices acoustiques fondamentaux distincts: les indices temporels pour la parole et les indices spectraux pour les informations extra/paralinguistiques. Nous avons évoqué les rôles de la structure temporelle fine chez le sujet normo entendant (voir chapitre I.2.2) qui peuvent faciliter la perception des deux types d'informations. L'accès à cette information par le sujet implanté cochléaire sera discuté.

La conception des stratégies de codage a ainsi toujours facilité la restitution de l'enveloppe temporelle du signal de parole, qui transmet les informations concernant le caractère voisé ou non du phonème, son temps de voisement, et sa durée globale.

## **1.2 Quel est le nombre de canaux nécessaires ?**

Le nombre minimal nécessaire de canaux stimulant le nerf auditif, soit le nombre d'électrodes insérées dans la cochlée, a fait l'objet de nombreux travaux, depuis la première description par Dudley d'un appareil capable d'exercer un traitement électrique d'un signal acoustique de parole (1939). Une inférence simple veut que ce nombre corresponde au nombre de filtres auditifs impliqués dans le traitement de la parole par le sujet normo entendant (14 à 19) mais c'est l'utilisation du vocoder dans de multiples études psycho physiques qui a permis d'évaluer les capacités de discrimination de la parole en fonction du type de signal (phonème, mot, phrase, dans le silence ou dans le bruit) et de l'importance de sa dégradation.

Le vocoder, ou speech spectrum-shaped noise permet de décomposer le signal acoustique naturel en plusieurs bandes de fréquences analysées séparément. Au sein de chacune de ces bandes, un traitement spectral et un traitement temporel du signal sont réalisés. Sur le plan spectral, toutes les fréquences présentes dans le signal naturel à des énergies acoustiques variables sont uniformisées en énergie. Cette manipulation rend compte du caractère uniforme de la stimulation réalisée par une électrode donnée de l'implant pour une bande fréquentielle. Sur le plan temporel, l'enveloppe du signal naturel est extraite et subit un filtrage passe-bas.

L'enveloppe ainsi obtenue est utilisée pour moduler un bruit blanc au sein de chaque canal. Le signal acoustique ainsi traité est délivré par un, deux, quatre, ou n canaux (voir figure II.1).

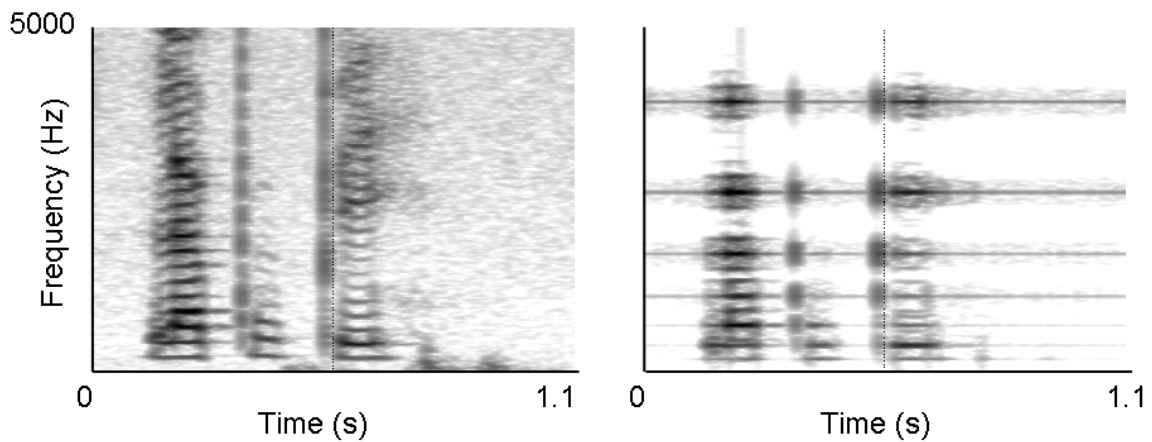


Figure II.1 Spectrogramme représentant le mot « poupée » précédé de son article, prononcé par un locuteur féminin en condition naturelle (gauche) et après un traitement par vocoder à 8 canaux.

Le vocoder a été utilisé dans de nombreuses études psycho physiques destinées à déterminer le nombre minimal de bandes fréquentielles et de canaux correspondant permettant de discriminer la parole. Dans l'article référence de Shannon [Shannon et al., 1995], la discrimination de la parole dans le silence était optimale dès que le signal était délivré par 4 canaux (voir figure II.2).

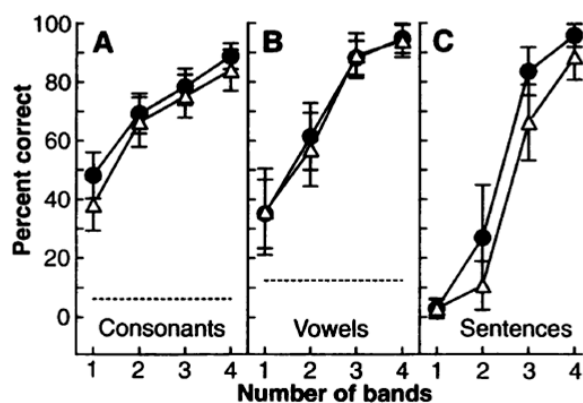


Figure II.2 : Pourcentage de reconnaissance correcte des consonnes (A), des voyelles (B) et des mots dans une phrase (C) en fonction du nombre de canaux délivrant le signal de parole après traitement préalable par le vocoder. Tiré de Shannon et al. (1995).

Dans une étude ultérieure, Shannon et al [Shannon et al., 2004] se sont intéressés aux différents traitements du signal de parole susceptibles d'être réalisés et à leurs conséquences éventuelles sur la discrimination de la parole. Pour ce qui concerne l'analyse du signal, ils ont

montré que les fréquences de coupure caractérisant la décomposition spectrale du signal acoustique naturel n'avaient que peu d'influence significative sur la discrimination de la parole. Considérons par exemple le traitement par vocoder de 4 bandes fréquentielles dont les fréquences de coupures sont définies les unes par rapport aux autres par un espacement linéaire (bande 1 : 0-1000 Hz ; bande 2 : 1000-2000 Hz ; bande 3 : 2000-3000 Hz ; bande 4 : 3000-4000 Hz). Considérons maintenant le traitement par vocoder de 4 bandes fréquentielles dont les fréquences de coupure sont définies par un espacement logarithmique de type tonotopique (bande 1 : 80-332 Hz ; bande 2 : 332-845 Hz ; bande 3 : 845-1886 Hz ; bande 4 : 1886-4000 Hz). La discrimination de la parole dans la décomposition spectrale linéaire n'était pas véritablement différente de la discrimination dans la décomposition de type tonotopique. De la même manière pour la synthèse du signal de parole après son analyse, le chevauchement des fréquences caractérisant les canaux n'avait pas d'impact sur la discrimination. En revanche, si l'enveloppe temporelle du signal extraite dans une bande fréquentielle donnée était utilisée pour moduler le bruit blanc dans un canal correspondant à une autre bande fréquentielle, les performances chutaient de manière significative.

La discrimination de la parole dans le bruit constitue une condition d'écoute difficile pour laquelle le nombre optimal de canaux délivrant le signal a également été évalué. Dans l'étude de Dorman et al [Dorman et al., 1998b], ce nombre augmentait significativement jusqu'à 12 pour obtenir une discrimination optimale des phrases présentées à un rapport signal/bruit de +2 dB, jusqu'à 20 canaux pour une discrimination maximale atteinte à un rapport signal/bruit de -2 dB (voir figure II.3).

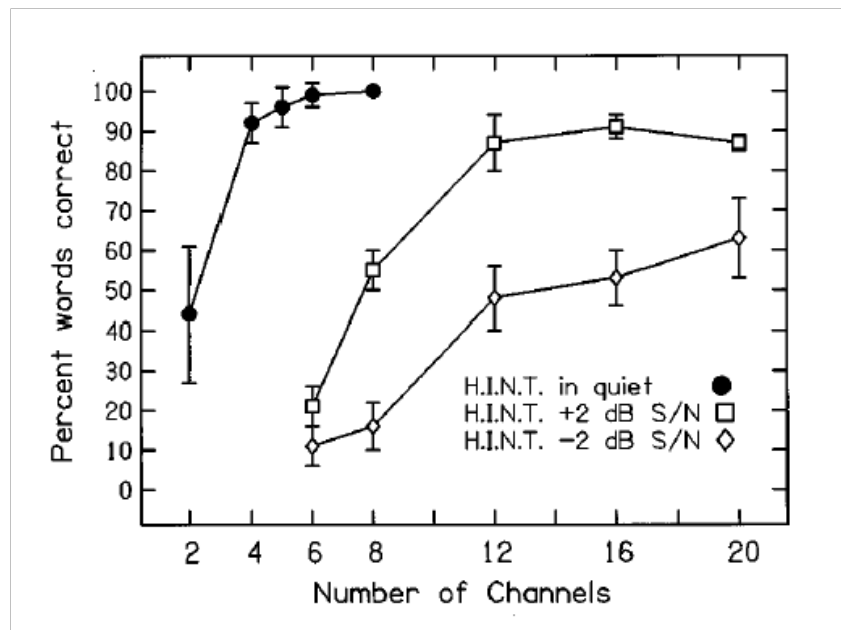


Figure II.3 : La discrimination des mots dans une phrase en présence d'un bruit concurrent augmente avec le nombre de canaux délivrant le signal, jusqu'à 12 pour un rapport signal/bruit de +2 dB, jusqu'à 20 pour un rapport signal/bruit de -2 dB. Tiré de Dorman et al (1998).

La validation de ces modèles acoustiques d'implant cochléaire a reposé sur plusieurs travaux comparant les performances des sujets normo entendant testés en condition d'implant aux performances de patients implantés cochléaires [Dorman and Loizou, 1997, Fishman et al., 1997, Dorman et al., 1998a]. Ces études ont démontré la validité de ces modèles, les patients et les sujets normo entendant obtenant des performances globalement comparables quelque soit le type de signal de parole testé (phonème, mot ou phrase, voir figure II.4).

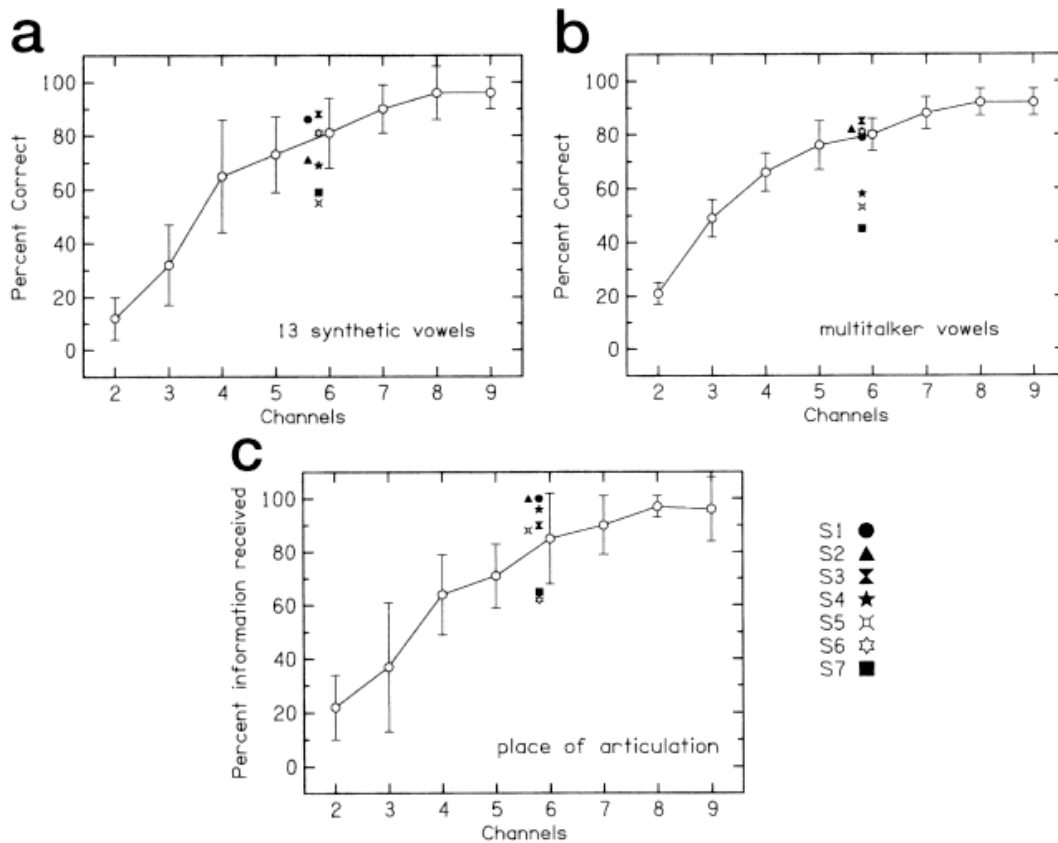


Figure II.4 : La reconnaissance des voyelles et des modalités articutoires sont comparables entre les patients implantés cochléaires et les sujets normo entendant testés avec un vocoder délivrant le signal par 2 à 9 canaux. Tiré de Dorman et Loizou, 1998.

Si l'on considère les différents résultats obtenus pour les simulations d'implant cochléaire par vocoder, la discrimination de la parole devient satisfaisante dans le silence dès que 4 canaux sont effectifs mais requiert 20 canaux pour être optimisée en présence d'un bruit concurrent.

Le même type d'étude a été réalisé chez les sujets implantés cochléaires afin de déterminer le nombre optimal d'électrodes actives. La plupart des études menées dans cet objectif ont montré que le nombre d'électrodes permettant d'atteindre le niveau maximal de discrimination de la parole était inférieur au nombre d'électrodes physiquement disponibles sur le porte électrodes, mais le caractère réduit des effectifs généralement évalués rend difficile la détermination précise de ce nombre. Kileny et al. [Kileny et al., 1998] ont ainsi montré que la discrimination de différents signaux de paroles dans le silence n'était que très légèrement affectée si le nombre d'électrodes activées était diminué de 20 à 10 chez des patients porteurs d'un implant Nucleus. De même pour la discrimination dans le bruit, Friesen



et al. [Friesen et al., 2001] ont montré un plateau de la reconnaissance des voyelles à différents rapports signal/bruit obtenu avec 7 électrodes activées (voir figure II.5).

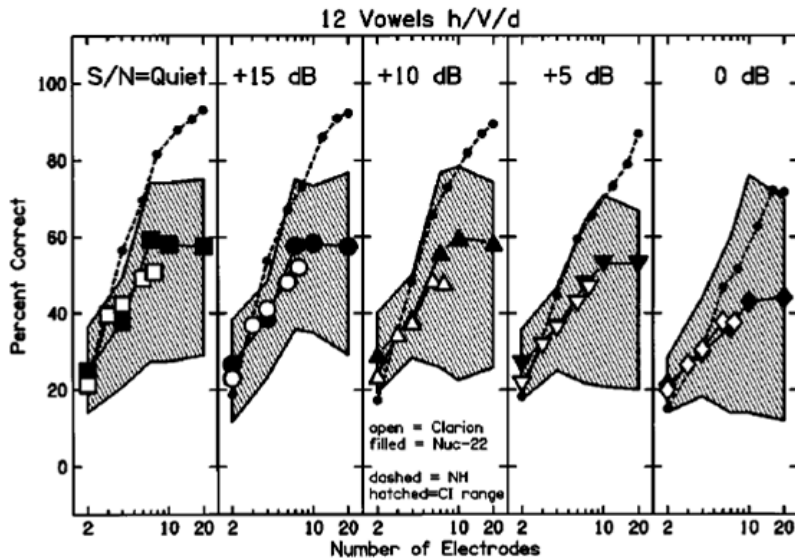


Figure II.5 La discrimination des voyelles dans le bruit atteint un plateau dès que 7 électrodes sont activées. Les performances des sujets normo entendant en condition de simulation d'implant par vocoder continuent de s'améliorer jusqu'à 20 canaux effectifs. Tiré de Friesen et al. (2001)

A l'inverse, il a été démontré qu'une plus grande quantité d'informations spectrales, par l'intermédiaire du nombre d'électrodes actives et de leur place dans la cochlée, avait un impact favorable sur la discrimination de la parole. Geier et Norton [Geier and Norton, 1992] ont par exemple évalué l'impact de la désactivation des 5 électrodes les plus apicales chez 5 sujets implantés utilisant une stratégie F0/F1/F2. Une diminution significative de la discrimination de la parole parmi 3 d'entre eux était retrouvée ainsi qu'une baisse subjective de la qualité globale du son pour les 5 sujets. Cette baisse serait attribuable au défaut de restitution de F1 normalement codé par les électrodes apicales. Dans l'étude de Blamey et al [Blamey et al., 1996] portant sur 64 sujets implantés, l'activation des 20 électrodes se révélait bénéfique pour la discrimination dans le bruit par rapport à une activation de 8 électrodes, mais cette étude visait à identifier l'ensemble des facteurs pronostiques des résultats post implantation et les performances n'étaient donc pas strictement évaluées en fonction du nombre d'électrodes activées.

S'il est difficile de déterminer avec précision le nombre optimal d'électrodes stimulant la cochlée, il apparaît probable que tous les sites de stimulation disponibles sur un porte

électrodes ne sont pas utilisés. Ce constat renvoie à la question de la spécificité de la stimulation électrique d'une population neuronale par une électrode.

## **1.2 Quel est le nombre d'électrodes effectives ?**

La spécificité de la stimulation dans la cochlée est nécessairement limitée par la diffusion du courant au sein des liquides labyrinthiques, qui constituent un milieu particulièrement conducteur. Le courant peut donc stimuler l'ensemble des ganglions spiraux adjacents à l'électrode concernée, ou bien encore être diffusé entre deux électrodes activées à un court intervalle. Ces phénomènes d'interférence inter électrodes ou de diffusion du courant limitent donc le nombre d'électrodes stimulant effectivement les ganglions spiraux.

Qu'elle concerne les phonèmes, les mots ou les phrases, dans le silence ou en présence d'un bruit concurrent, la discrimination moyenne atteint en effet un plateau à 8 électrodes activées et certains patients peuvent bénéficier de l'activation complémentaire de 2 ou plusieurs électrodes pour certaines conditions, considérées isolément (comme la discrimination des mots monosyllabiques à un rapport signal/bruit de 10 dB dans l'étude de Garnham) [Garnham et al., 2002].

La diffusion intra cochléaire du courant et son corollaire, la spécificité de la stimulation neuronale, pourraient également être influencés par la distance séparant l'électrode active de la population neuronale cible. En effet, une plus grande proximité entre l'électrode active et le modiolus où siègent les ganglions spiraux peut théoriquement permettre de cibler plus précisément les neurones à stimuler, d'augmenter la dynamique de stimulation et de réduire la quantité de courant nécessaire à cette stimulation [Merzenich et al., 1977, Frijns et al., 1996].

Le gain en spécificité de stimulation a pu être mis en évidence par des études psychophysiques utilisant le forward masking ou par des études électro physiologiques qui révèlent des seuils électriques abaissés pour les électrodes apicales lorsque le porte électrode est en position périmodiolaire [Wackym et al., 2004] (voir figure II.6).

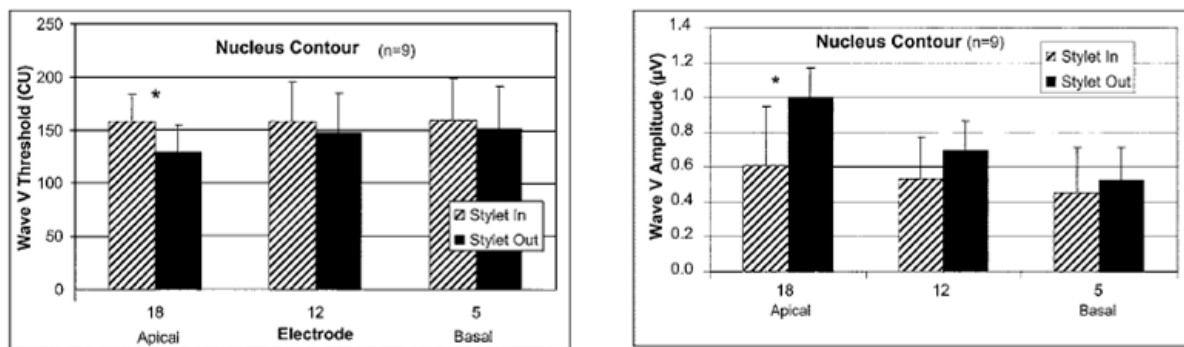


Figure II.6 Seuil auditif électrique mesuré par potentiel évoqué électrique (onde V, gauche) avec le porte électrodes positionné contre le mur latéral (histogrammes hachurés) et en position périmodiolaire (histogrammes noirs). Amplitude du potentiel évoqué électrique (onde V, droite) avec le porte électrodes positionné contre le mur latéral (histogrammes hachurés) et en position périmodiolaire (histogrammes noirs). Tiré de Wackym et al, (2004).

L'impact sur la discrimination de la parole de ce positionnement périmodiolaire est difficile à mettre en évidence, et la littérature rapporte généralement des résultats contradictoires. Gordin et al. [Gordin et al., 2009] ont retrouvé de meilleures performances chez les enfants porteurs d'un implant périmodiolaire (Nucleus 24RE) par rapport à des enfants appariés pour l'âge porteurs d'un implant droit (Nucleus 24M). Chez l'adulte, Bacciu A [Bacciu et al., 2004] et Fitzgerald [Fitzgerald et al., 2007] retrouvent des résultats favorables au design périmodiolaire pour la discrimination de la parole (Bacciu) ou pour la discrimination fréquentielle (Fitzgerald) mais dans des populations réduites. Marrinan MS [Marrinan et al., 2004] ne montraient aucune relation entre le contact avec le modiolus et les performances pour la discrimination de la parole ou les seuils électriques. Dans l'étude de Fitzgerald, les patients opérés avec un implant périmodiolaire n'obtenaient pas de meilleurs scores de discrimination que les patients porteurs d'un implant à porte électrodes droit. L'impact fonctionnel du design de l'implant demeure donc encore contesté.

En synthèse, la discrimination de la parole dans le silence est réalisable dès que 4 canaux transmettent effectivement le signal acoustique préalablement filtré. Pour atteindre ces mêmes niveaux de discrimination dans le bruit, environ 20 canaux effectifs seraient requis. Hors, chez le patient implanté, la diffusion du courant depuis l'électrode active et les interférences entre deux électrodes activées à court intervalle limitent le nombre de canaux effectifs entre 4 à 10 sites indépendants.

### 1.3 Principales stratégies de codage

On distingue classiquement deux grandes catégories de stratégies de codage : celles qui privilégient l'information temporelle et celles qui privilégient l'information spectrale. Cette distinction est en réalité historique dans la mesure où les stratégies actuellement utilisées visent à réaliser une combinaison de ces deux types d'informations.

Il est vrai que deux catégories distinctes de stratégies ont co existé dans les débuts de l'implant cochléaire, après les premières stratégies analogiques développées par Chouard et McLeod [Chouard and MacLeod, 1976]. La stratégie CIS, décrite juste après, continue d'être largement utilisée et restitue les informations relatives à l'enveloppe temporelle du signal. Les stratégies d'extraction des formants (F2 rate, F0-F2, puis F0/F1/F2), développées en Australie et principalement utilisées par Cochlear®, permettaient de détecter efficacement la durée des phonèmes, le voisement et certaines caractéristiques des consonnes (voir figure II.7).

Reconnaissance des voyelles				Reconnaissance des consonnes			
	F2	Strategy F0 F2	F0 F1 F2		F2	Strategy F0 F2	F0 F1 F2
Total	34%	56%	72%	Total	37%	43%	49%
Duration	83%	85%	94%	Voicing	35%	34%	50%
F1 grouping	12%	27%	81%	Nasality	86%	84%	98%
F2 grouping	25%	68%	55%	Affrication	31%	32%	40%
				Duration	62%	71%	81%
				Place	19%	28%	28%
				Amplitude envelope	47%	46%	61%
				High F2	48%	68%	64%

Figure II.7: Reconnaissance des voyelles et des consonnes selon la stratégie d'extraction des formants utilisée (d'après Blamey, 1985).

Ces stratégies ont actuellement laissé leur place aux stratégies de type SPEAK puis ACE.

#### *Stratégies de type CIS (Continuous Interleaved Sampling)*

La stratégie CIS, dont le principe a été décrit dès 1988, peut actuellement être utilisée par la plupart des processeurs d'implant cochléaire. Cette stratégie a fait la preuve de son efficacité dès 1991 dans la publication référence de Wilson [Wilson et al., 1991].

Le signal acoustique est filtré par plusieurs bandes de fréquence et au sein de chaque bande est extraite l'enveloppe temporelle. Cette enveloppe temporelle du signal est rectifiée et filtrée par un filtre passe-bas, typiquement réglé autour de 200 Hz afin de pouvoir représenter, dans

une certaine mesure, les variations de la fréquence fondamentale de la voix humaine. Elle subit ensuite une compression logarithmique pour pouvoir transmettre, dans une certaine mesure, l'importante dynamique d'amplitude des sons de l'environnement et de parole (jusqu'à 100 dB) à la dynamique bien plus réduite de l'activité du nerf auditif. Des modulations d'amplitude par des impulsions électriques biphasiques sont imprimées à l'enveloppe ainsi traitée et le signal résultant peut être transmis à l'électrode correspondante. En fonction des implants et du patient, entre 4 et 22 électrodes sont susceptibles d'être activées, distribuées dans la scala tympani selon un ordre pseudo tonotopique : le signal délivré par les canaux assignés aux basses fréquences sera transmis par les électrodes apicales, le signal traité pour les hautes fréquences par les électrodes basales. Deux électrodes ne peuvent pas être activées simultanément, afin de limiter le risque d'interférence décrit plus haut. Le nombre d'impulsions par seconde dépend du nombre d'électrodes susceptibles d'être activées et évolue typiquement entre 700 et 1600 par seconde.

La stratégie HiResolution (HiRes) est relativement proche de la stratégie CIS et peut permettre la stimulation de 16 électrodes à des vitesses de stimulation élevées.

### ***Stratégies de type n of m***

Parmi ces stratégies, on peut citer les stratégies n of m elle-même, la spectral peak (SPEAK), l'advanced combination encoder (ACE).

Ces stratégies reposent sur le principe général de sélection des canaux qui seront stimulés. Les enveloppes temporelles extraites des différentes bandes de fréquences sont analysées pour identifier les n enveloppes ayant la plus forte amplitude parmi les m canaux à stimuler par leur électrode associée. Le paramètre n est fixe dans les stratégies n of m et ACE et peut varier en fonction de la composition spectrale du signal dans la stratégie SPEAK. Les vitesses de stimulation (nombre d'impulsions par seconde) varient entre 700 et 1000 pulses par secondes dans les stratégies n of m et ACE et sont situés autour de 250 pulses par seconde dans les stratégies SPEAK. Les stratégies n of m et ACE sont comparables à la stratégie CIS en dehors de la sélection des canaux présentant la plus forte amplitude. Leur but est donc de restituer les composantes du signal acoustique ayant la plus forte énergie, tout en limitant une potentielle interférence inter électrode liée à des composantes acoustiques à plus faible énergie. L'avantage théorique d'une telle stratégie est donc de pouvoir transmettre préférentiellement les indices acoustiques liés à la parole.

### *Restitution de la structure temporelle fine*

La contribution de la structure temporelle fine pour la discrimination de la parole a été montrée dans le chapitre I, partie 2.2. Bien que la plupart des filtres passe-bas appliqués au signal acoustique écartent sa structure temporelle fine (>500 Hz selon Rosen), la fréquence de coupure utilisée peut être augmentée jusqu'à 400 Hz, voire au-dessus pour la stratégie HiRes, ce qui permet la transmission d'indices de périodicité se rapprochant la structure temporelle fine, au moins pour ses plus basses fréquences. La limite de ces traitements du signal acoustique réside dans les capacités d'évoquer des changements de hauteur tonale en augmentant la vitesse de stimulation, généralement inexistantes au-delà de 300 Hz [Shannon, 1983,Zeng, 2002]. De plus, le niveau de discrimination fréquentielle (FODL) en-dessous de 300 Hz en utilisant de tels indices temporels chez les patients implantés demeure très déficitaire par rapport à celui des normo entendant [Baumann and Nobbe, 2004,Zeng et al., 2005]. Le nombre réduit d'électrodes effectives par rapport au nombre d'électrodes activables limite plus encore la restitution tonotopique de ces informations temporelles fines.

Bien qu'il ne soit actuellement pas possible de déterminer avec précision la quantité de structure temporelle fine à laquelle les sujets implantés ont accès, deux des quatre compagnies ont développé des stratégies visant à améliorer cette transmission.

Une approche consiste à réaliser un traitement particulier du signal extrait de la bande de fréquences la plus basse généralement (ou des 2 à 4 bandes de fréquences les plus basses). Dans cette approche, développée par MedEl®, l'électrode ou les électrodes apicales correspondant à ces bandes de fréquences basses vont délivrer une stimulation différente de la stimulation délivrée aux autres électrodes. En effet, lorsque l'onde du signal extrait dans la bande de fréquences considérée franchit le zéro vers ses valeurs négatives, de courtes impulsions sont délivrées (voir figure II.8). Leur amplitude est déterminée par l'énergie acoustique globale détectée dans la bande de fréquence considérée. La stimulation des autres électrodes est réalisée selon la stratégie CIS classique, au moyen d'impulsions continues, décalées dans le temps d'une électrode à l'autre. Cette stratégie, appelée FSP pour Fine Structure Processing, se rapproche d'un traitement du signal décrit par Wilson en 1991, sous l'appellation « peak picker/CIS », dans lequel une seule impulsion était délivrée à l'électrode ou aux électrodes apicales.

Ce type de stratégies peut permettre de transmettre spécifiquement aux électrodes apicales des informations temporelles variables, pouvant rendre compte, dans une mesure encore indéterminée, de la structure fine du signal acoustique.

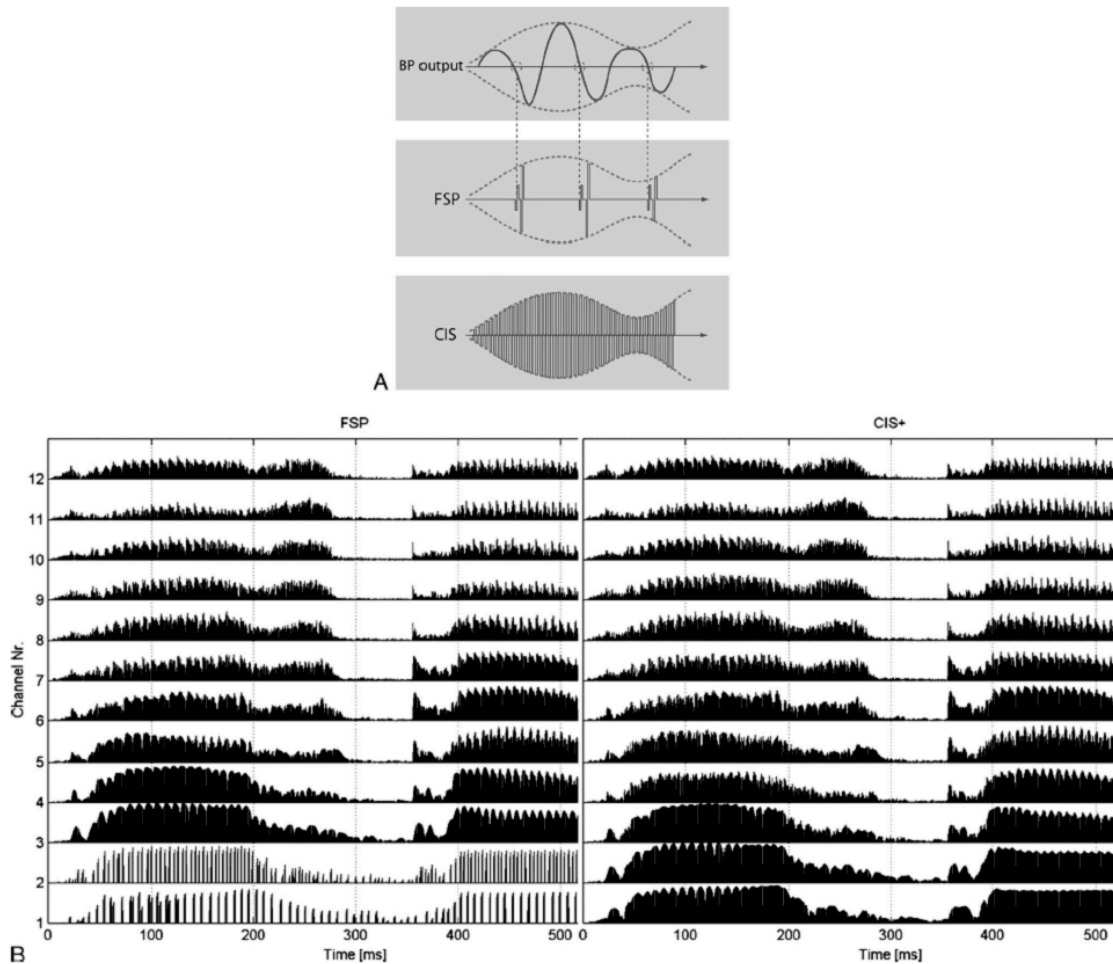


Figure II.8 : La stratégie FSP (Fine Structure Processing, à gauche) réalise une stimulation particulière pour l'électrode correspondant aux bandes de fréquences les plus basses, basée sur de courtes impulsions délivrées lorsque l'onde franchit le zéro vers les valeurs négatives. La stimulation des autres électrodes obéit à la stratégie CIS classique. Courtesy of Med-El®

L'autre approche destinée à augmenter les possibilités de représentation des informations temporelles fines dans le signal acoustique consiste à augmenter le nombre de sites effectifs dans la cochlée en utilisant la technologie dite « des canaux virtuels ». Les premières descriptions d'un tel traitement sont également le fruit des travaux de Wilson, développés par la suite [Donaldson et al., 2005, Firszt et al., 2007, Koch et al., 2007]. Cette technologie, également désignée sous l'appellation « current steering », repose sur la stimulation simultanée ou quasi simultanée de deux électrodes adjacentes. Le caractère concomitant de

cette double stimulation permet théoriquement d'évoquer une hauteur tonale intermédiaire, située entre les deux hauteurs correspondant à la stimulation isolée de chacune des électrodes. La stratégie HiRes avec l'option Fidelity 120. Les hauteurs tonales évoquées par ces stimulations simultanées dépendent de l'amplitude relative de l'impulsion délivrée à chacune des deux électrodes (voir figure II.9) et de leur polarité relative. Ainsi, une électrode donnée reçoit une impulsion. Dans le même temps, une impulsion à polarité inverse est délivrée à l'électrode adjacente. La hauteur tonale ainsi évoquée est théoriquement inférieure à celle évoquée si la première électrode est stimulée isolément.

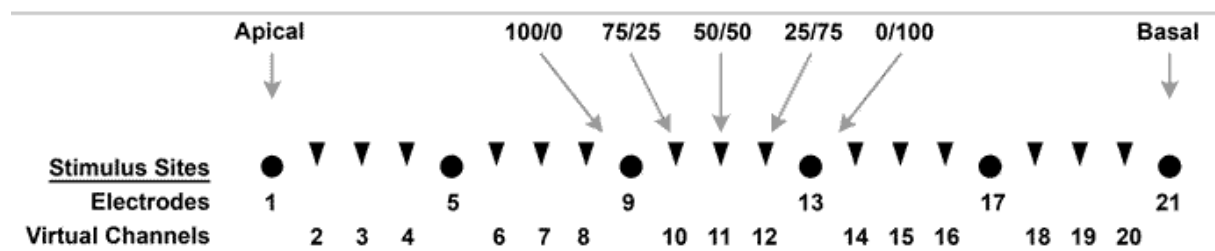


Figure II.9: Représentation des canaux virtuels attenants à 6 électrodes intra cochléaires, stimulés en fonction de l'amplitude relative des impulsions délivrées à deux électrodes adjacentes. Tiré de Wilson et al.(2007).

Dans la stratégie HiRes 120, 8 sites de stimulation sont théoriquement activés pour chacune des 15 bandes de fréquences, rendant donc disponibles 120 sites. Chacun de ces 120 sites peut être activé en fonction du rapport des amplitudes des impulsions délivrées à deux électrodes adjacentes. La multiplication ainsi obtenue des sites de stimulation pourrait permettre aux patients de mieux percevoir l'intonation dans une phrase ou dans un extrait musical, et de mieux cibler la F0 d'un locuteur en présence de bruit concurrent.

#### 1.4 Les différents types d'implants

Depuis l'avènement des implants multi canaux, de nombreux types d'implants cochléaires ont été développés. Le choix de l'implant mis en place pour un patient donné dépend de multiples facteurs parmi lesquels on peut citer l'existence éventuelle d'une fonction auditive résiduelle, les conditions anatomiques, le statut vestibulaire, l'expérience du chirurgien.

Le choix de l'implant se fait parmi plusieurs modèles fabriqués par plusieurs constructeurs, qui ont atteint un niveau de fiabilité comparable de l'un à l'autre [Venail et al., 2008]. Ces différents modèles varient pour la longueur du porte électrodes, son diamètre et sa flexibilité, le nombre d'électrodes, l'espacement inter électrodes, le caractère droit ou périmodiolaire.



Advanced Bionics® est un constructeur américain (ancien Clarion®) propose des implants périmodiolaires comportant 16 électrodes. Cochlear® est une entreprise australienne propose des implants droits et périmodiolaires comportant typiquement 22 électrodes. Les implants fabriqués par MedEl®, constructeur autrichien, sont droits et portent 12 électrodes. Neurelec® est une entreprise française fabriquant des implants droits porteurs de 20 électrodes.

Décrire chacun des modèles proposés par chaque constructeur n'est pas l'objet de notre travail mais nous précisons simplement que certains implants sont destinés à préserver la fonction auditive résiduelle. Ainsi qu'il l'est présenté ultérieurement, nous décrivons les performances de patients bénéficiant d'une stimulation électro-acoustique, l'implant cochléaire étant alors utilisé en combinaison avec l'audition résiduelle ipsi et controlatérale. Ces implants, développés par Cochlear®, MedEl® et Neurelec®, sont mis en place au décours d'une procédure chirurgicale dite atraumatique [Frayse et al., 2006], visant à préserver les structures intra cochléaires. Plusieurs modèles d'implant existent ou sont en cours d'évaluation (voir figure II.10) dans cette indication.

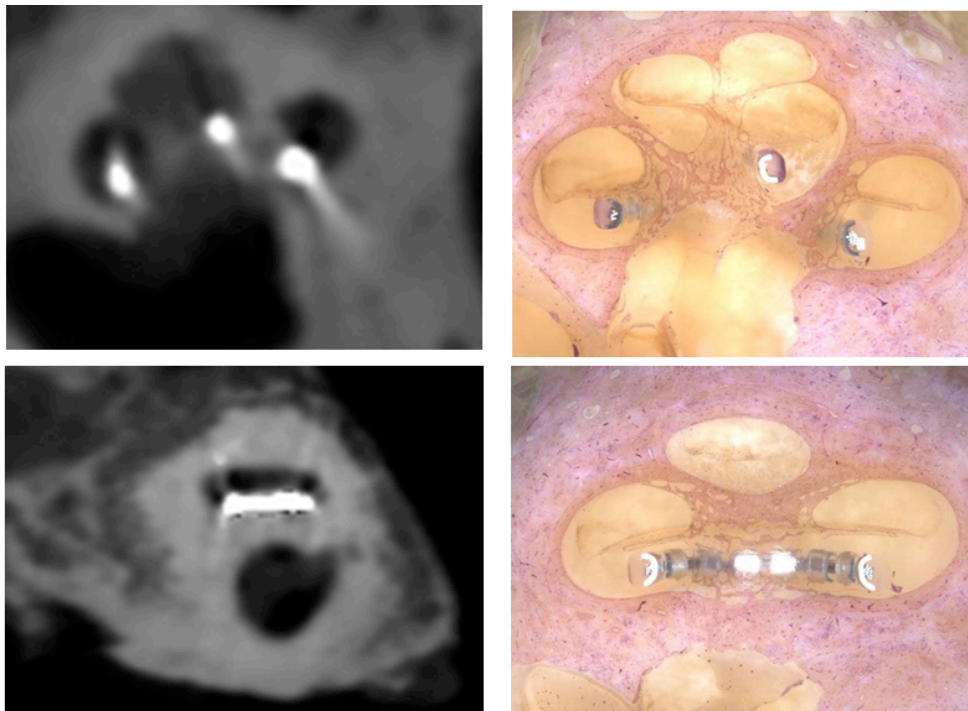


Figure II.10: Etude radio-histologique d'un prototype d'implant cochléaire destiné à préserver les structures intra cochléaires. Le porte électrodes est positionné dans la scala tympani, visualisé sur des coupes perpendiculaires au grand axe de la cochlée en cone beam computed tomography (gauche) et contrôlé par des coupes histologiques correspondantes (droite). (Tiré de Marx et al., 2013)

## 2. PERCEPTION DE LA PAROLE

La restauration de la discrimination de la parole est le principal objectif d'une réhabilitation par implant cochléaire et la réussite dans ce domaine est manifeste puisque l'on considère que plus de 80% des patients implantés comprennent plus de 80% des phrases dans le silence [Wilson and Dorman, 2008].

Si l'on s'intéresse à la perception de la parole à un niveau strictement phonétique, les performances des patients implantés apparaissent bien plus déficitaires. Rouger et al. [Rouger et al., 2008] ont ainsi retrouvé un taux de reconnaissance moyen des consonnes d'environ 30% (+/-18), avec des capacités de reconnaissance des bilabiales (/p/b/m) et des dentales (/t/d/n) relativement faibles (voir figure II.11). Le lieu d'articulation antérieur est donc mal restitué par l'implant avec une transmission d'information phonétique inférieure à 5% contre 20% pour les informations de voisement et de nasalité.

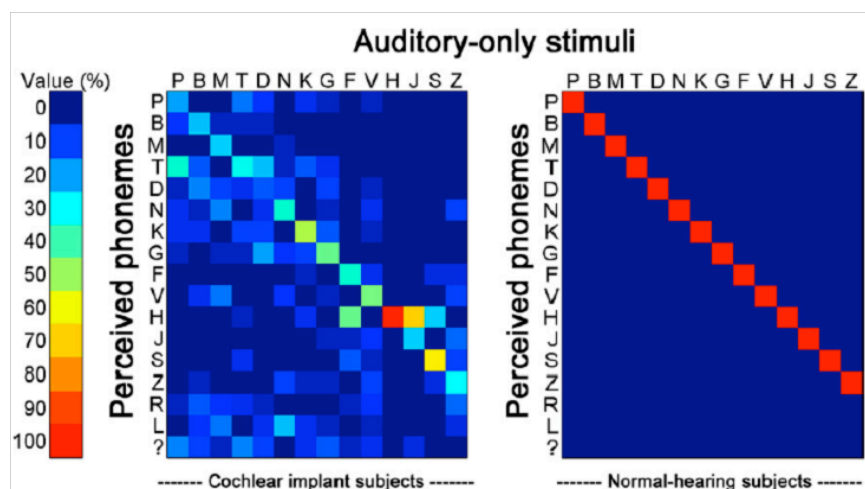


Figure II.11 : Matrices de confusion phonétique pour un groupe de sujets implantés cochléaires comparé à un groupe de sujets normo entendant. Les confusions sont particulièrement marquées pour les bilabiales et les dentales. Tiré de Rouger et al (2007).

Ce déficit dans la reconnaissance des phonèmes a été constaté dans une moindre mesure par Skinner [Skinner et al., 1999] et Doyle (AJO, 1995). Il s'améliore avec l'augmentation d'information dans le signal de parole. La discrimination moyenne des mots monosyllabiques est ainsi généralement rapportée supérieure à 50% [Kiefer et al., 1997, Bassim et al., 2005, Blamey et al., 2012] et les pourcentages moyens de discrimination correcte augmentent encore pour les mots dissyllabiques, avec des valeurs variant entre 70 % et 83% [Rouger et al., 2007, Marx et al., 2011].

Les principales études multicentriques [UKCISG, 2004, Blamey et al., 2012] témoignent de l'efficacité de l'implant pour permettre la discrimination de mots et de phrases, mais elles soulignent également la variabilité de ces résultats. La figure II.12 illustre ces deux caractéristiques dans l'étude multicentrique de P. Blamey portant sur plus de 2000 patients implantés.

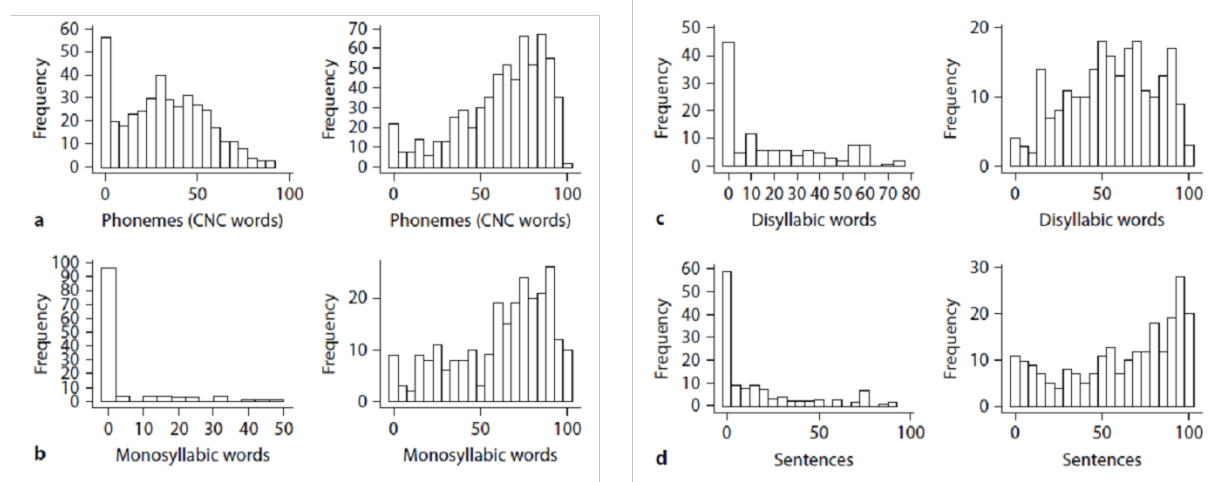


Figure II.12. Distribution des performances pour différents signaux de parole, chez 2251 patients avant implantation (gauche) et après implantation (droite). Tiré de Blamey et al, 2012.

Si l'on s'intéresse à la proportion de patients bénéficiant de l'implant, on considère que près de 80% des patients implantés ont une amélioration post opératoire supérieure ou égale à 20 points de pourcentage, pour les mots dans une phrase présentée dans le silence. Plus de 95% des patients implantés ont une amélioration de plus de 20 points de pourcentage pour la discrimination des phrases présentées dans le bruit (voire figure II.13 a et b).

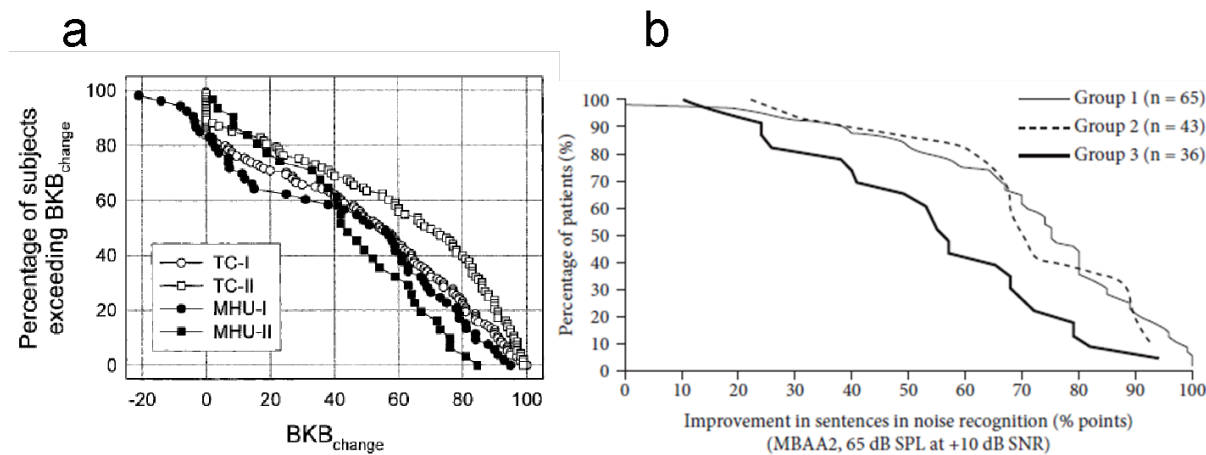


Figure II.13 a et b. a) Courbes de distribution cumulative pour plusieurs groupes de sujets implantés cochléaires en fonction du gain en discrimination des phrases dans le silence (Tiré de UKCISG, 2004). b) Courbes de distribution cumulative pour trois groupes de sujets implantés distingués selon leur niveau de discrimination pré-opératoire en fonction du gain en discrimination des phrases dans le bruit (Tiré de Marx et al, 2011).

La restauration de la discrimination de la parole est généralement rapide et atteint un plateau dès 6 mois [Rouger et al., 2007, Lenarz et al., 2012]. Ce plateau se maintient à long terme, avec des performances qui ont pu être évaluées jusqu'à 20 ans post implantation. L'important gain en discrimination avec l'augmentation de l'information dans le signal de parole traduit l'importance du contexte syntactique et lexical pour les sujets implantés, et l'optimisation des performances auditives par des processus de type top-down.

### 3. PERCEPTION DES INFORMATIONS PARA ET EXTRA LINGUISTIQUES

La restauration excellente de la discrimination de la parole contraste avec les résultats plus mitigés obtenus dans les autres domaines de la perception auditive. Notre équipe a contribué à la littérature sur la perception des informations para linguistiques, notamment par un travail sur la détection de la voix parmi les bruits de l'environnement et par une publication sur la perception du genre par les sujets implantés cochléaires, jointe en annexe de ce mémoire. Nous présenterons une synthèse des données disponibles en nous référant au modèle de traitement de la voix développé par l'équipe de Pascal Belin.

### 3.1 Détection de la voix humaine

En dehors de l'étude publiée par Zoé Massida [Massida et al., 2011], peu de travaux ont permis d'évaluer la capacité de discrimination de la voix humaine chez le sujet implanté cochléaire. Cette capacité est fortement altérée avec des performances inférieures de moitié aux performances des sujets normo entendant testés au titre de contrôle. Ces sujets normo entendant avaient également été testés en condition de simulation d'implant par vocoder et obtenaient des performances comparables à celles des sujets implantés pour une dégradation spectrale majeure du signal, équivalente à 4 canaux effectifs (voir figure II.14). Dans un sous-groupe de sujets implantés, un effet favorable lié à la présence de parole était observé. Les performances obtenues par les sujets implantés apparaissaient tout à fait indépendantes de leurs capacités de discrimination de la parole. Une autre spécificité notable de ce déficit résidait dans son caractère pérenne, ce qui constituait un élément de dissociation supplémentaire avec la restauration de la discrimination de la parole. Nous avons ainsi décrit les progrès réalisés par les patients dans cette compétence pendant la première année post implantation, absents pour la détection de la voix. Les raisons permettant d'expliquer un tel déficit tiennent d'abord à la dégradation spectrale réalisée par le processeur vocal de l'implant, limitant la perception de F0 et du timbre de la voix humaine. La réorganisation cérébrale secondaire à la surdité peut également être impliquée par la plasticité cross modale audition-vision mise en évidence au niveau des aires temporeles de la voix.

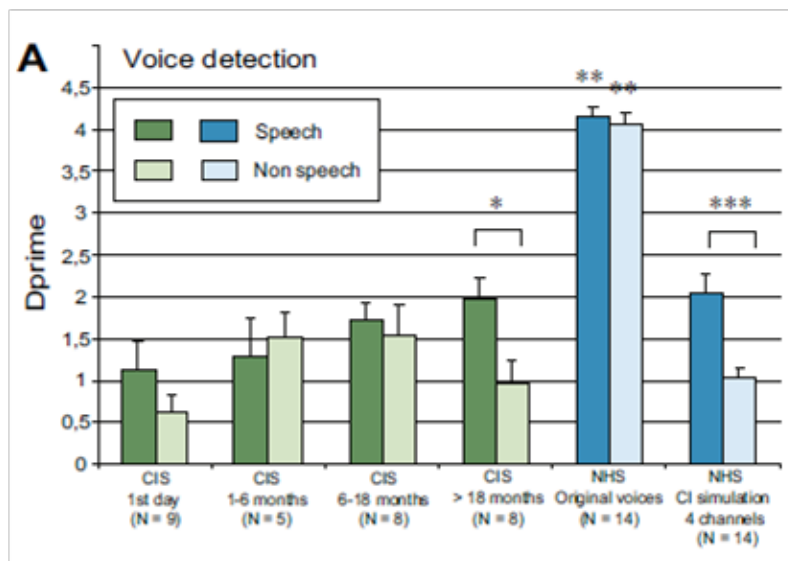


Figure II.14. Performances pour la détection de la voix parmi les bruits de l'environnement, particulièrement déficitaire chez les sujets implantés cochléaires, ne progressant pas avec l'expérience avec l'implant. Tiré de Massida et al (2011).

### **3.2 Perception des attributs de la voix liés à l'identité du locuteur**

Les travaux portant sur la perception des attributs de la voix font état d'un déficit constant, dissocié de la restauration des capacités de discrimination de la parole. Les performances pour la reconnaissance du locuteur si seule sa voix est présentée demeurent ainsi juste au dessus du niveau de la chance, même après un entraînement intensif, tel que proposé dans l'étude Vongphoe [Vongphoe and Zeng, 2005]. Ces performances sont comparables à celles obtenues par les sujets normo entendant en condition de simulation d'implant par vocoder à 8 canaux, traduisant ainsi l'impact de la dégradation spectrale sur les facultés de reconnaissance vocale. D'après nous, l'évaluation de la perception d'autres attributs de la voix conduit au même type de constat défavorable chez les sujets implantés, même si l'ensemble de la littérature n'est pas nécessairement concordant. Après avoir étudié la détection de la voix humaine, notre équipe s'est intéressée à une étape suivante dans le traitement hiérarchique modulaire de la voix : la perception du genre du locuteur.

Notre étude a porté sur deux groupes de sujets implantés cochléaires, dont un était évalué en longitudinal afin de déterminer l'impact de l'expérience avec l'implant sur les performances. Un continuum vocal masculin-féminin, mis au point par l'équipe de P. Belin à partir d'un mélange de 16 voix de chaque genre [Pernet and Belin, 2012], nous a permis d'évaluer les capacités de reconnaissance du genre du locuteur d'une part et les capacités de catégorisation d'autre part.

La reconnaissance était évaluée au niveau des deux extrémités du continuum, chacune correspondant à un genre masculin typique et un genre féminin typique. Les capacités de catégorisation étaient quant à elles déterminées au moyen d'une courbe psycho métrique sigmoïde, dont la valeur de la pente était choisie comme critère de jugement (voir figure II.15).

En accord avec les données de la littérature, nos résultats ont fait état d'une reconnaissance relativement peu déficitaire, avec des scores dépassant généralement les 80%, pour des voix d'homme et de femme typiques. Kovacic et al. [Kovacic and Balaban, 2009] ont ainsi montré que les capacités de reconnaissance du genre étaient directement liées à la valeur de F0 : plus F0 est basse, plus reconnaissance d'une voix masculine est facilitée.

Mais nous avons souhaité par cette étude insister sur le déficit de catégorisation du genre, en réponse à des stimuli ambigus, comme c'était le cas pour les éléments intermédiaires de notre continuum. Le déficit alors retrouvé était majeur, avec des performances de moitié inférieures

à celles des sujets normo entendant, et durable, ainsi qu'en témoignait l'absence de progression chez les sujets évalués longitudinalement.

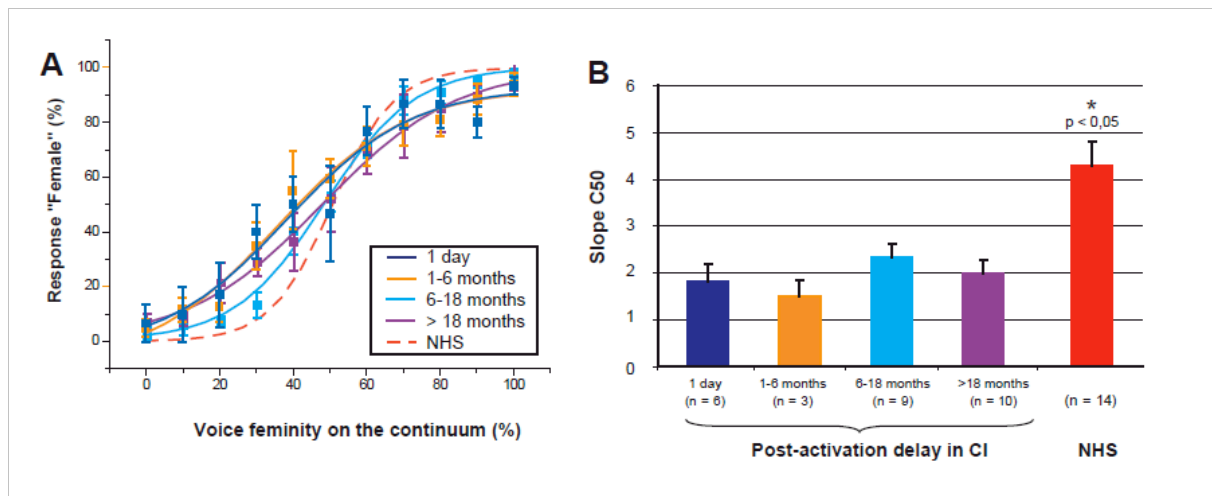


Figure II.15. Performances pour la reconnaissance et la catégorisation du genre du locuteur, évaluées au moyen d'une fonction sigmoïde (gauche). La valeur de la pente permettait de juger des capacités de catégorisation, fortement altérées chez les sujets implantés.

### 3.3 Perception des informations extra linguistiques : exemple de la musique

Les capacités perceptives musicales des patients implantés sont étroitement liées à leurs performances pour la discrimination de F0 et la perception du timbre.

La perception de F0 chez les sujets implantés est unanimement reconnue comme déficitaire avec des F0DL souvent très altérés par rapport aux sujets normo entendant, leurs valeurs variant entre 10 et 50%, [Vandali et al., 2005, Rogers et al., 2006, Cousineau et al., 2011] et demeurant supérieures à 5% même chez les implantés les plus performants [Geurts and Wouters, 2001, 2004]. Ce déficit pour la perception de F0 se manifeste pour la perception de la polyphonie, la majorité des implantés ne percevant généralement qu'une note lorsqu'un accord à deux ou trois notes est présenté [Donnelly et al., 2009]. Plus encore que le F0DL, le traitement de séquences de variations de F0 est déficitaire [Cousineau et al., 2011], ce qui majore la difficulté de perception des mélodies. Le déficit pour la perception du timbre altère les capacités de reconnaissance des instruments et le simple plaisir à l'écoute de la musique.

Il n'existe à notre connaissance qu'une étude en neuro imagerie fonctionnelle portant sur la perception musicale chez le sujet implanté cochléaire. Limb et al. [Limb et al., 2010] ont ainsi réalisé une étude en PET Scan à l'  $H_2O_{15}$ , permettant d'analyser les activations cérébrales corticales en réponse à 3 différents stimuli : des phrases parlées à discriminer, des mélodies

populaires jouées au piano à reconnaître et des rythmes joués par des percussions à reproduire. Les auteurs ont évalué les performances comportementales et les régions neuronales activées de 10 patients implantés cochléaires droits qu'ils ont comparées à celles de 10 sujets normo entendants. Les performances comportementales étaient comparables entre les deux groupes pour le rythme (100%) et la parole (98%+/- 6 pour les normo entendant contre 82%+/-32 pour les patients implantés). En revanche, les sujets normo entendant présentaient une meilleure reconnaissance des mélodies (88%+/-21,5 contre 46%+/- 28,4 ;  $p<0,001$ ). L'analyse inter groupe des activations cérébrales a permis de retrouver une hyperactivation des cortex temporaux chez les sujets implantés cochléaires (voir figure II.16), essentiellement localisée dans les deux pôles temporaux, dans les gyri temporal moyen gauche et supérieur droit, se manifestant de manière prédominante lors de la présentation des phrases. Cette hyperactivation, bien que significative sur le plan statistique, devenait marginale pour les mélodies. Ces résultats suggèrent que les réseaux neuronaux auditifs qui sont soit recrutés, soit hypersollicités chez les patients implantés cochléaires par rapport aux normo entendants ne le sont que lors de l'écoute de la parole. Une réhabilitation orthophonique intensive vis-à-vis de la musique pourrait permettre d'activer ces réseaux et d'améliorer les capacités de perception de la musique.

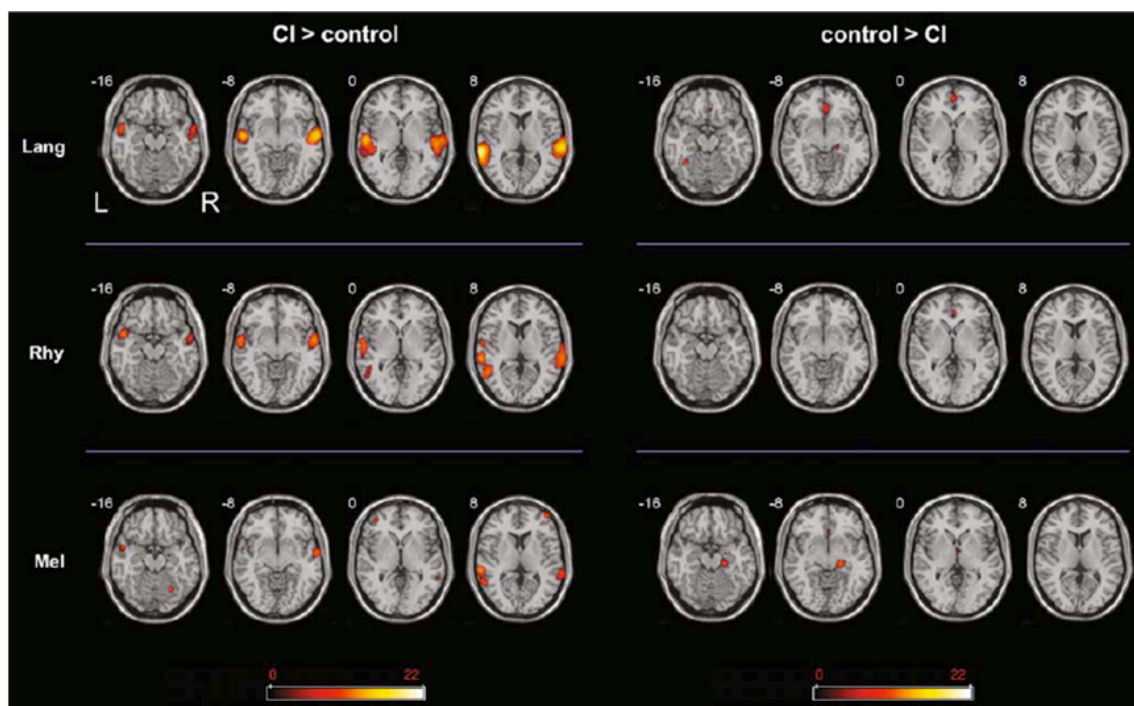


Figure II.16. Activations cérébrales en réponse à un signal de parole (haut), un signal rythmique (milieu) et une mélodie (bas). Limb et al observaient une hyperactivation de certaines régions temporales bilatérales, prédominante pour le signal de parole. Tiré de Limb et al (2010).



## **4. REORGANISATION NEUROFONCTIONNELLE CHEZ LE SUJET SOURD**

### **4.1 Déficits neurosensoriels et plasticité cérébrale**

L'absence ou la perte d'une modalité sensorielle s'accompagne d'une réorganisation cérébrale, destinée à faciliter le fonctionnement des sens épargnés. Les capacités de réorganisation apparaissent maximales lorsque le déficit est présent in utero ou à la naissance et la majorité des travaux portant sur la plasticité du système auditif central concernent donc la population des sujets sourds congénitaux. La réorganisation se produisant en cas de déficit auditif acquis implique un cerveau par définition moins plastique, dont les principales fonctions sont fixées.

La réorganisation secondaire à la perte d'un sens cherchera à optimiser l'accès du sujet à l'ensemble des informations de son environnement. Röder et Rosler [Röder and Rosler, 2004] ont décrit trois mécanismes pouvant rendre compte des différentes formes que la réorganisation plastique corticale peut prendre :

- des circuits neuronaux consacrés à une modalité préservée se réorganisent de façon interne pour optimiser leurs capacités de traitement (réorganisation intra modale)
- dans les aires multi sensorielles, une modalité peut se développer aux dépens de la modalité touchée (réorganisation méta modale)
- enfin, des circuits neuronaux normalement dédiés au traitement de l'information dans la modalité touchée sont recrutés par une ou plusieurs autres modalités (réorganisation inter modale).

La cécité et les stratégies d'adaptation à ce handicap constituent des exemples particulièrement représentatifs de la réorganisation cérébrale secondaire au déficit sensoriel. Depuis l'invention du Braille jusqu'aux succès de musiciens aveugles, on trouve ainsi de nombreuses illustrations de la compensation dans ce type de déficit. La facilitation de la perception des informations somato-sensorielles a permis aux sujets aveugles d'avoir accès au langage écrit, les informations auditives prenant également une dimension supérieure pour améliorer la localisation dans l'espace [Muchnik et al., 1991] ou développer les facultés musicales. Les travaux sur la transposition sensorielle de Bach y Rita ont bien avant souligné l'importance de ces informations somato sensorielles chez les sujets aveugles [Bach et al.,

1970]. Plusieurs travaux comportementaux, électro physiologiques et histologiques menés sur des modèles animaux de cécité ont cherché à identifier les corrélats neuronaux d'une telle compensation. Chez la souris devenue aveugle par exemple, une augmentation quantitative des récepteurs somato sensoriels (portés par la moustache) est accompagnée d'une extension des régions neuronales somato sensorielles correspondantes [Rauschecker et al., 1992], réalisant ainsi une compensation intra-modale du déficit sensoriel. Un chat aveugle développera des mouvements de tête et d'oreilles destinés à améliorer la localisation verticale de la source du son [Rauschecker, 2001], spécifiques de sa condition car non retrouvés chez le chat voyant. Au niveau neuronal, ces manifestations sont sous-tendues par une compensation méta-modale dans la région ectosylvienne antérieure. Cette région multi modale comporte en effet des neurones impliqués dans le traitement des informations visuelles, auditives et somato sensorielles. En réponse à la cécité, le nombre et la spécificité des neurones dédiés à la localisation de la source sonore étaient augmentés, les neurones à activité purement visuelle ayant quant à eux quasiment disparu [Rauschecker and Korte, 1993].

Comme les déficits visuels, les déficits auditifs profonds s'accompagnent de réorganisations cérébrales fonctionnelles favorisant l'intégration multi sensorielle afin de compenser le déficit présenté. La vision est probablement la modalité la plus sollicitée et son rôle devient essentiel pour l'accès à la parole chez le sujet sourd. Que ce soit par l'intermédiaire du langage signé ou de la lecture labiale, les performances visuelles développées par les sujets atteints d'une surdité profonde diffèrent des capacités visuelles normales et témoignent de la plasticité cérébrale secondaire au déficit auditif.

#### **4.2 Aspects comportementaux de la plasticité cérébrale des sujets sourds profonds**

Le développement des capacités visuelles chez les sujets sourds et leur investissement dans la compréhension du langage sont les manifestations comportementales les plus évidentes de leur compensation par réorganisation fonctionnelle. La plupart des travaux comportementaux dans ce domaine portent sur des sujets sourds congénitaux.

La compensation visuelle du déficit auditif se manifeste dans des secteurs spécifiques du traitement visuel et certaines fonctions basiques, telles que la perception du contraste [Finney and Dobkins, 2001], de la vitesse de mouvement [Brozinsky and Bavelier, 2004] ou de la lumière [Bross, 1979], ne sont pas différentes entre les sujets sourds et les sujets normo entendant. C'est dans l'exercice de tâches plus complexes, impliquant la vision périphérique comme le temps de réaction devant un mouvement dans le champ périphérique, que les

capacités visuelles des sourds vont se révéler supérieures [Neville and Lawson, 1987, Proksch and Bavelier, 2002]. C'est dans le domaine de la communication inter personnelle que le rôle de la modalité visuelle se manifeste de manière plus évidente car la compréhension de l'autre constitue l'objectif essentiel, en dépit du déficit auditif. La vision peut alors intervenir sous deux formes, soit par l'intermédiaire du langage signé, soit par la lecture labiale.

L'acquisition et l'utilisation des langues signées concernent principalement les sujets sourds profonds congénitaux. Comme le langage oral, elles sont organisées selon des règles phonologiques, syntaxiques et sémantiques [Sandler and Lillo-Martin, 2006]. Elles rendent compte de l'importance de la réorganisation corticale, par le rôle prépondérant pris par la modalité visuelle pour la perception des signes utilisés et par la modalité somato sensorielle pour la production des gestes. La compréhension de la langue signée fait en effet appel à l'intégration de mouvements séquentiels des mains et des doigts, associés à des mouvements du visage et de la bouche [Capek et al., 2008]. Sur ces bases, la phonologie des langues signées utilise la configuration des mains, le lieu où se produisent les mouvements (mains, corps ou visage) et enfin les mouvements eux-mêmes [Sandler, 2012]. Pour intégrer l'ensemble de ces mouvements, il a été montré que le regard des sujets sourds était essentiellement fixé sur une région faciale centrale inférieure (Muir LJ, 2005). La production de la langue signée repose sur des informations somato sensorielles optimisées afin d'assurer les placements successifs des mains et leur coordination avec les mouvements de l'extrémité céphalique.

Les travaux de Julien Rouger portant sur les capacités d'intégration multi sensorielle des sujets sourds implantés cochléaires ont permis de préciser l'importance des interactions audiovisuelles dans cette population. Les capacités de lecture labiale supra normales présentées par les sujets sourds, ont été chiffrées avec une discrimination moyenne des mots dissyllabiques de l'ordre de 35% contre 9% pour les sujets normo entendant. Elles se sont de plus révélées pérennes dans le temps, avec un maintien des performances jusqu'à 8 ans après implantation. Chez les sujets implantés, la discrimination des mots était toujours meilleure lorsque le signal de parole était présenté en modalité audiovisuelle que dans une seule des modalités auditive ou visuelle. Ce gain en discrimination était également observé en modalité audiovisuelle chez les sujets normo entendant quand le signal de parole était dégradé par un bruit concurrent ou par un vocoder. Cependant, le gain demeurait significativement plus élevé dans le groupe des sujets implantés, soulignant la qualité de l'intégration audiovisuelle dans cette population (voir figure II.17).

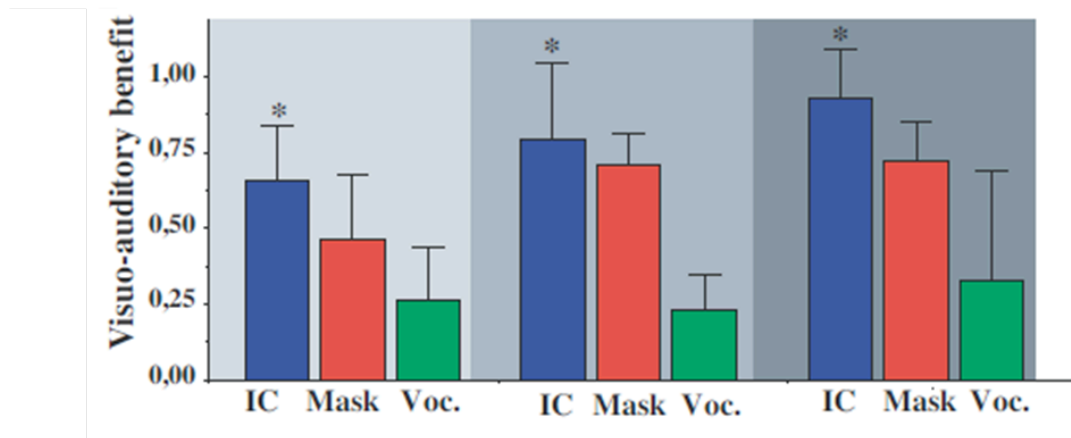


Figure II.17 : Gain en discrimination lié à la présentation des mots en condition audiovisuelle pour trois catégories de performances auditives : discrimination <30%, entre 30 et 60%, et > 60%. Le gain est toujours significativement plus élevé dans la population des sujets implantés cochléaires. Tiré de Rouger et al, 2007.

Le rôle de la vision chez les sujets sourds a plus rarement été évalué pour la perception d'autres signaux que la parole ou les signes, et son impact dans ce domaine reste incertain. Les études concernant la reconnaissance visuelle d'émotions faciales par des sujets sourds comparés à des sujets normo entendant portent essentiellement sur des populations pédiatriques et retrouvent majoritairement un déficit de perception visuelle des émotions [Peterson and Siegal, 1995, Ludlow et al., 2010]. L'identification de la joie a été retrouvée comme plus efficace chez les normo entendant à l'inverse du dégoût, plus aisément reconnu par les sujets sourds [Weisel, 1985]. Chez des enfants sourds implantés cochléaires, les capacités visuelles de discrimination entre deux émotions négatives (tristesse/colère) et d'identification d'une émotion cible parmi plusieurs émotions basiques (joie, tristesse, colère, peur) étaient significativement inférieures à celles d'enfants normo entendant appariés pour l'âge et le contexte socio-éducatif [Wiefferink et al., 2013].

L'impact de la compensation intermodale et de l'intégration multi sensorielle est donc démontré pour les aspects lexicaux et syntactiques du langage. En revanche, ces phénomènes ne sont pas retrouvés de manière reproductible pour la reconnaissance visuelle de l'émotion, autre indice de la communication inter personnelle.

### **4.3 Réorganisation corticale chez le sujet sourd post lingual**

Les capacités de réorganisation corticale sont, comme nous l'avons déjà souligné, plus limitées chez le sujet sourd post lingual que chez l'enfant sourd pré lingual en bas âge. Le traitement auditif de la parole demeure ainsi principalement réalisé dans le cortex temporal supérieur gauche, comme en témoignent les études électro physiologiques par potentiels évoqués tardifs réalisés après implantation cochléaire [Eggermont, 2007]. Ces études révèlent des réponses neuronales comparables entre sujets implantés et sujets normo entendant, provenant d'aires auditives associatives telles que le planum temporale gauche, impliquées dans la voie dorsale audio-motrice du traitement du langage.

Si certains circuits neuronaux temporaux gauches demeurent dédiés au traitement de la parole chez le sujet sourd post lingual, d'autres régions multi modales vont être le siège d'une réorganisation méta modale destinée à privilégier la modalité visuelle, susceptible d'apporter l'information de parole. Que ce soit par un maintien ou une réorganisation de son fonctionnement, le cerveau du sujet sourd post lingual cherchera ainsi toujours à optimiser le traitement de la parole. Plusieurs régions multi modales, telles que les régions frontales inférieures, normalement impliquées dans la planification de la production de parole, seront ainsi hyper activées chez les sujets sourds par rapport aux sujets normo entendant dans une tâche de lecture labiale traduisant l'importance prise par la modalité visuelle dans le traitement de la parole [Lazard et al., 2010].

L'autre caractéristique essentielle de la réorganisation corticale secondaire à la surdité est le recrutement de régions normalement dédiées au traitement de l'information auditive par d'autres modalités, au premier rang desquelles la modalité visuelle. Cette réorganisation inter modale est particulièrement significative dans le cortex temporal droit. Sa partie postérieure, normalement impliquée dans le traitement des sons environnementaux [Beauchamp et al., 2004, Bunzeck et al., 2005], est en effet recrutée par la vision pour réaliser la lecture labiale et le traitement phonologique de l'information écrite, donc recruté afin de faciliter le traitement de la parole. Sa partie antérieure, qui comprend les régions neuronales de l'aire de la voix, s'est également révélée activée dans des tâches de lecture labiale (voir figure II.18), traduisant une fois de plus le recrutement de régions par la vision afin de traiter la parole [Rouger et al., 2011].

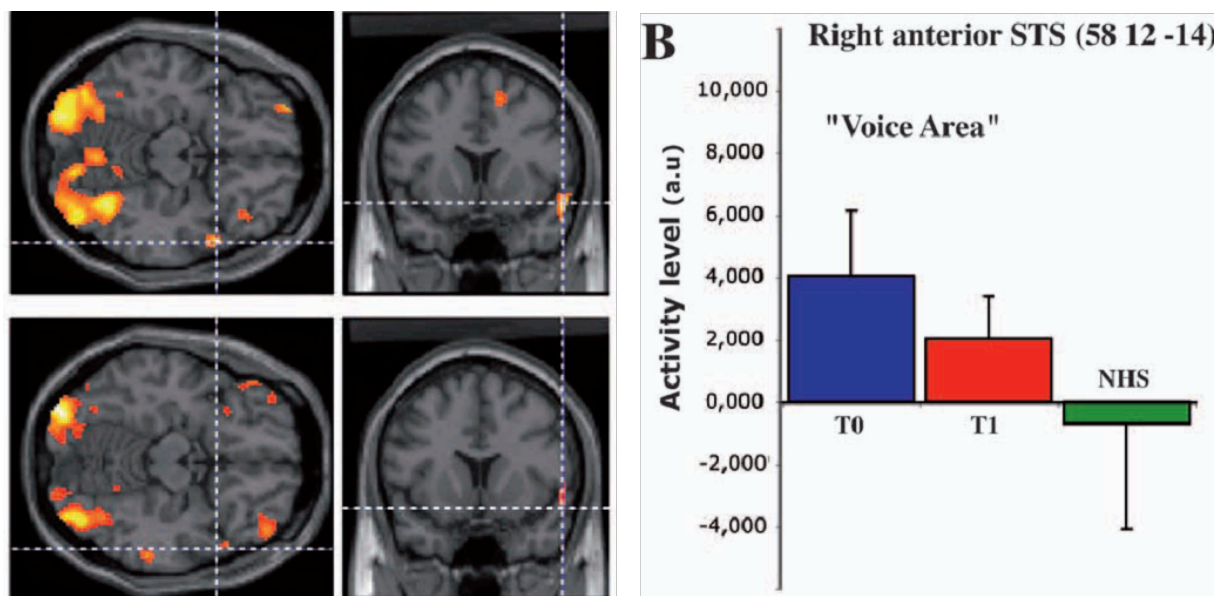


Figure II.18 : Activation des aires temporeles de la voix par une tâche de lecture labiale, observée chez des sujets implantés cochléaires après l'intervention (T0) en PET scan. Cette activation a tendance à diminuer à distance de l'intervention (T1). Tiré de Rouger et al. (2011).

Il est par ailleurs intéressant l'évolution des activations cérébrales à distance de la réhabilitation auditive par implant cochléaire. L'étude de J. Rouger (2011) révélait ainsi, à distance de l'implantation, une diminution de l'activation des aires temporeles de la voix en réponse à une tâche de lecture labiale, ce qui pouvait témoigner d'un processus de normalisation du fonctionnement des régions auditives corticales.

#### 4.4 Réactivation des aires auditives après implantation cochléaire

L'intérêt pour le devenir des régions corticales dédiées à l'audition après implantation cochléaire s'est manifesté dès les années 1990, dans le même temps que le développement de cette technique de réhabilitation [Ito et al., 1990, Herzog et al., 1991, Naito et al., 1995]. L'étude des réactivations cérébrales s'est principalement appuyée sur des études en neuro imagerie par PET scan. Le PET scan représente en effet la méthode de choix, l'IRM fonctionnelle étant évidemment proscrite une fois le matériel implanté en place et les explorations électro physiologiques nécessitant un traitement particulier du signal recueilli, en raison des artefacts liés au fonctionnement du processeur de l'implant. Comme la plupart des études comportementales, les travaux par PET scan ont essentiellement évalué les profils d'activation cérébrale des sujets implantés en réponse à des signaux de parole. Et comme la plupart des études comportementales, leurs auteurs ont été confrontés à l'hétérogénéité des

populations évaluées (quant à la cause de la surdité, la durée de surdité profonde, l'existence d'une audition résiduelle) et la variabilité dans les résultats obtenus.

Les premières études en PET scan retrouvaient ainsi des activations cérébrales siégeant dans les deux cortex auditifs. Cette activité prédominait dans le cortex auditif primaire controlatéral en réponse à un bruit blanc et dans les deux régions associatives en réponse à des stimuli de parole [Naito et al., 1995, Naito et al., 1997]. L'activation de ces régions auditives temporales associatives était plus importante chez les sujets implantés comparés à des sujets contrôles normo entendant [Naito et al., 2000], et plus précisément pour les gyri temporaux supérieurs et moyens. Cette hyper activation était alors attribuée à la plus grande sollicitation des aires auditives associatives chez les sujets implantés du fait de la dégradation du signal auditif présenté par l'implant. Dans leur première étude comparant des profils d'activation de sujets implantés et de sujets normo entendant, Giraud et al. [Giraud et al., 2000] retrouvaient une hypo activation globale de ces régions (gyri temporaux supérieur et moyen) chez les sujets implantés par rapport aux sujets normo entendant. En revanche, une hyper activation était observée chez les sujets implantés dans la partie postérieure du gyrus temporal supérieur droit et dans des réseaux attentionnels (régions préfrontale et pariétale). Ces résultats ont suggéré pour la première fois la spécificité du traitement du langage oral chez les sujets implantés cochléaires. L'hyper activation relative de la région postérieure du gyrus temporal supérieur droit rendrait compte d'une augmentation du traitement phonologique de bas niveau. Cette augmentation de l'analyse phonologique du signal de parole serait favorisée par des processus top-down attentionnels, et permettrait de compenser le déficit de traitement sémantique du signal de parole, reflété par les hypoactivations globales des gyri temporaux, supérieurs et moyens.

Parallèlement et suite à ces premiers travaux, l'attention s'est portée sur l'existence d'éventuelles relations entre le niveau d'activation des régions auditives et les performances obtenues par les sujets implantés pour la discrimination de la parole. Fujiki et al. [Fujiki et al., 1999] ont ainsi montré une corrélation positive significative entre le niveau d'activation de régions corticales auditives associatives, comprenant la partie moyenne des gyri temporaux moyens et supérieurs, et les performances pour la parole (évaluée par la reconnaissance de voyelles, de consonnes et speech tracking). Cette corrélation positive n'était en revanche pas retrouvée pour l'activation du cortex auditif primaire ou l'activation du cortex frontal inférieur gauche dans sa partie postérieure (aire de Broca). L'hypothèse d'une relation entre l'activité des aires auditives associatives (BA 21, 22 et 42) et la discrimination de la parole a été renforcée par plusieurs travaux ultérieurs [Fujiki et al., 2000, Green et al., 2005], la valeur

du débit sanguin cérébral dans les gyri temporaux moyens et supérieurs augmentant avec le pourcentage de discrimination correcte de différents signaux de parole.

Différentes entités anatomo-fonctionnelles des régions temporales ont par la suite été identifiées, permettant de caractériser plus précisément la réorganisation cérébrale chez le sujet sourd implanté cochléaire en fonction de ses performances [Mortensen et al., 2006, Rouger et al., 2011]. L'activation du gyrus temporal supérieur en réponse à un signal de parole signifiante a ainsi été localisée dans ses parties antérieure et postérieure pour l'hémisphère droit, et dans sa partie antérieure à gauche. Ce niveau d'activation était plus important chez des sujets implantés obtenant de bonnes performances pour la discrimination de la parole.

Les différences dans les profils d'activation entre les sujets présentant une discrimination de la parole satisfaisante et les sujets avec des résultats moins bons portent également sur le cortex préfrontal inférieur gauche. Après stimulation auditive par des phrases, les sujets dits « good performers » inclus dans l'étude de Mortensen (2006) présentaient en effet un niveau d'activité supérieur dans cette région par rapport aux sujets dits « poor performers » (voir figure II.19).

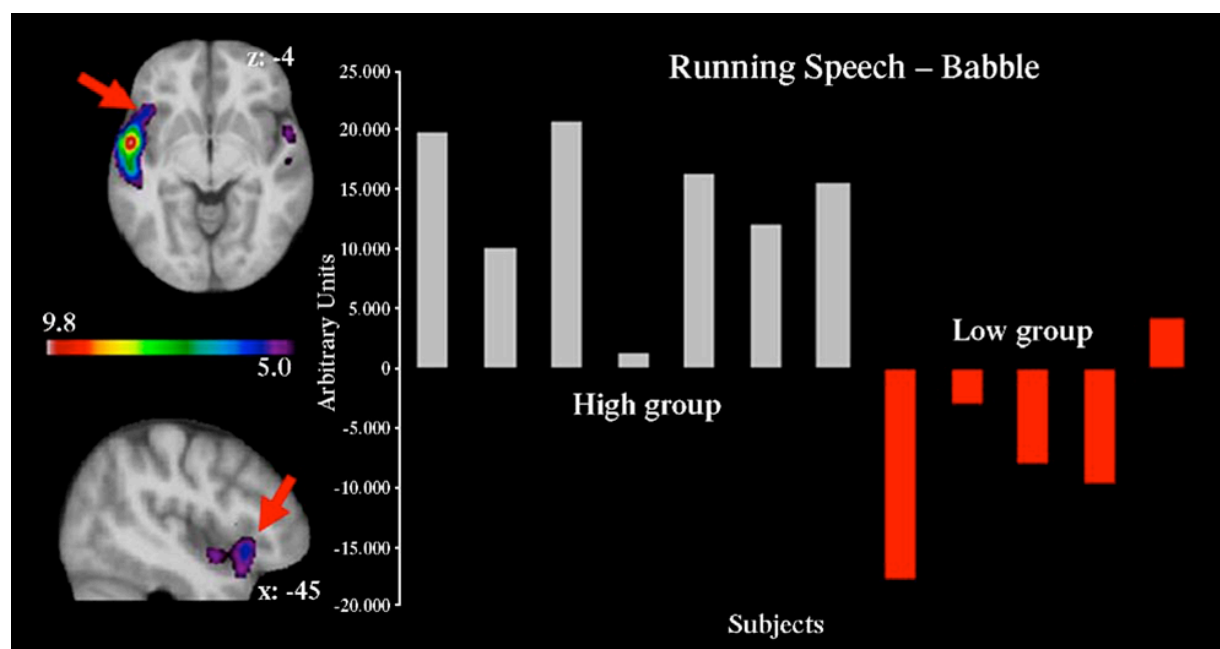


Figure II.19. Le niveau d'activation du cortex pré frontal inférieur gauche pourrait rendre, en partie, compte de la variabilité des performances des sujets implantés cochléaires. Il est ainsi plus activé en réponse à des signaux de parole chez les sujets dits « high performers » que chez les sujets dits « poor performers ». Tiré de Mortensen et al. (2006).



De plus, le cortex préfrontal inférieur gauche est l'objet d'une réactivation progressive après implantation, ainsi que l'ont montré Rouger et al. dans leur étude longitudinale [Rouger et al., 2011]. Un groupe de sujets implantés était évalué par PET scan une première fois juste après l'activation de l'implant (T0) puis une deuxième fois dès que la discrimination de la parole devenait satisfaisante (> 60% de discrimination des mots dissyllabiques). Chez ces sujets, dans une tâche de reconnaissance audiovisuelle de mots, le cortex préfrontal inférieur gauche n'était pas activé à T0. A T1 et pour la même tâche, une activation de cette région était retrouvée. La réactivation du cortex préfrontal inférieur gauche traduit un processus de normalisation des fonctions cérébrales avec le temps chez les sujets implantés obtenant de bonnes performances. En effet, chez le sujet normo entendant, après les traitements phonologiques bas niveaux réalisés par les régions temporales postérieures et supérieures, cette région permet la confrontation des représentations phonologiques aux commandes motrices correspondantes [Dufor et al., 2009, Shahin et al., 2009]. Cette région est également impliquée dans la planification de la production de parole et son activation pourrait donc rendre compte de la réactivation, au moins partielle, de la boucle audio phonatoire chez le sujet implanté obtenant de bons résultats pour la discrimination de la parole après implantation.

La réorganisation corticale secondaire à la surdit e profonde est donc principalement r ealis ee dans l'objectif d'optimiser le traitement de la parole, le r ole de la vision devenant ainsi pr epond erant, par r eorganisation m eta ou inter modale. Cet effet favorable de la r eorganisation pour le traitement de la parole pourrait  tre nuanc e face   ses cons equences potentielles sur le traitement d'autres informations. Ainsi, la r eorganisation touchant les aires de la voix ou le cortex temporal sup erieur et post erieur droit d edi e au traitement des sons environnementaux pourrait limiter le traitement des informations para et extra linguistiques.

## **5. AUDITION RESIDUELLE ET IMPLANT COCHLEAIRE**

### **5.1 Généralités-Définitions**

L'audition résiduelle est un terme générique qui recouvre plusieurs réalités cliniques dans les surdités sévères à profondes. En effet, cette audition résiduelle peut être unilatérale, bilatérale, présente sur toutes les fréquences ou prédominant sur les fréquences graves. Le niveau d'informations acoustiques que cette audition est susceptible d'apporter est extrêmement variable et la perception de l'environnement sonore se voit donc plus ou moins altérée.

L'intérêt de l'utilisation d'un implant cochléaire associé à la fonction auditive résiduelle a émergé à la fin des années 1990 pour s'imposer progressivement. Les termes de stimulation bimodale et de stimulation électro acoustique ont été consacrés dans la littérature scientifique mais leur formulation reste imprécise quand on considère les situations cliniques précises auxquelles elles renvoient.

Pour envisager correctement les différents modes d'association entre implant cochléaire et audition résiduelle, on doit distinguer :

- les implantés cochléaires unilatéraux avec audition résiduelle contro latérale, amplifiée ou non par une aide auditive externe. Le terme de stimulation bimodale désigne ce groupe de patients. C'est le mode d'association le plus fréquent.
- les patients implantés cochléaires unilatéraux, avec audition résiduelle ipsilatérale (ou bilatérale), amplifiée ou non par une aide auditive externe. Le terme de stimulation électro acoustique désigne ce groupe de patients.
- les patients implantés bilatéraux avec audition résiduelle unilatérale, amplifiée ou non par une aide auditive externe. Cette association n'est pas décrite comme un mode de stimulation en tant que tel dans la littérature concernant les implants bilatéraux.
- les patients implantés bilatéraux avec audition résiduelle bilatérale, amplifiée ou non par une aide auditive externe.

Stimulation bimodale et stimulation électro acoustique désignent la même réalité physiologique, dans le sens où elles combinent l'information acoustique résiduelle à la stimulation électrique apportée par l'implant. L'usage veut que la stimulation électro acoustique implique une audition résiduelle ipsi latérale à l'implant (ou bilatérale) et la stimulation bimodale, une audition résiduelle contralatérale à l'implant. Cette terminologie

étant bien ancrée dans la littérature, c'est elle que nous emploierons dans la suite de notre mémoire.

## **5.2 Extension des critères audiométriques d'implantation cochléaire**

Les premiers critères audiométriques officiels remontent à la conférence de consensus du National Institute of Health en 1988 [NIHCDC, 1988], et les candidats potentiels à l'implant cochléaire présentaient alors une surdité profonde à totale bilatérale avec une discrimination de phrases nulle (en condition auditive seule) dans les meilleures conditions d'appareillage. La deuxième conférence de consensus du National Institute of Health autour de l'implant cochléaire a eu lieu en 1995 [NIH, 1995] et s'est appuyée sur les nombreuses études publiées depuis 1988. Les performances post implantation étaient alors marquées par leur variabilité, mais des facteurs pronostiques de l'implantation ont progressivement été identifiés et ont permis de mieux cibler les indications. Les études s'intéressant à l'évolution des performances post opératoires mettaient en évidence une discrimination des mots souvent supérieure à 30% à plus d'un an après l'activation de l'implant. L'identification de facteurs pronostiques fiables et les publications récurrentes de performances en discrimination des mots supérieures à 30% ont ainsi permis d'orienter les indications d'implantation cochléaire.

Sur ces bases, la conférence de consensus de 1995 a proposé d'étendre les indications audiologiques aux surdités sévères ayant une discrimination de phrases inférieure ou égale à 30% (en condition auditive seule). Les travaux réalisés dans la continuité de ce consensus se sont donc plus particulièrement intéressés aux résultats de ces patients sourds sévères, obtenant un bénéfice marginal de l'aide auditive externe. Et parallèlement, les progrès réalisés, tant sur le plan technologique (amélioration des stratégies de codage) que sur le plan de la réhabilitation post implantation, ont permis d'améliorer encore les performances post opératoires des patients implantés.

Les résultats continuant de s'améliorer, la Haute Autorité de Santé a pris en compte la suggestion d'extension des critères dont elle précise les limites audiométriques de la manière suivante [HAS, 2007]: « Discrimination inférieure ou égale à 50 % lors de la réalisation de tests d'audiométrie vocale ... Les tests doivent être pratiqués à 60 dB, en champ libre, avec des prothèses bien adaptées». Cette extension progressive des critères d'implantation se

constate également aux Etats-Unis. En 2010, la Food and Drug Administration tolère jusqu'à 60% de discrimination de phrases dans le silence pour certains types d'implants. Il existe donc une tendance à tolérer des niveaux d'audition résiduelle de plus en plus importants, tendance légitimée par les organismes publics à caractère scientifique, qu'ils soient nord-américains ou européens. Par ailleurs, la plupart de ces organismes insistent aujourd'hui sur l'importance des données de l'audiométrie vocale, réalisée avec une réhabilitation audioprothétique réglée de manière optimale, plus que sur la perte auditive observée à l'audiométrie tonale.

L'esprit dans lequel ont été rédigés les rapports récents de la Food and Drug Administration de la Haute Autorité de Santé diffère cependant de l'argumentaire des conférences de consensus nord-américaines de 1988 et 1995. Le niveau d'audition résiduelle n'y est plus considéré comme une simple valeur seuil à dépasser pour pouvoir affirmer l'efficacité de l'implant cochléaire. Ces rapports ont en effet pris en compte les publications montrant les bénéfices d'associer l'implant cochléaire à l'audition résiduelle controlatérale. Cette audition résiduelle est progressivement considérée comme un élément utile, susceptible d'améliorer les résultats post opératoires, qu'il faut donc intégrer au fonctionnement des patients implantés.

La stimulation bimodale concerne aujourd'hui un nombre croissant de patients, du fait de l'extension progressive des indications d'implantation cochléaire vers les surdités sévères. L'association des deux types de stimulation présente plusieurs avantages théoriques. Elle permet aux patients de recouvrir des capacités de discrimination de la parole grâce à l'implant cochléaire tout en conservant certaines caractéristiques de leur audition naturelle binaurale antérieure. La composante acoustique de cette association permet aux patients de maintenir leur niveau de perception de la fréquence fondamentale. Elle autorise également une perception partielle du timbre par l'intermédiaire des premiers harmoniques. La perception du timbre, même partielle, est donc une représentation de la richesse spectrale du son et permet de reconnaître un locuteur ou d'identifier une mélodie. La composante acoustique de cette association joue donc un rôle considérable. La nécessité de l'amplifier par une aide auditive externe dépend du type de surdité puisque certains patients ont une perception des fréquences graves normale ou sub normale quand d'autres présentent une perception déjà altérée à ce niveau.

## **6. METHODES ET PROBLEMATIQUES**

### **6.1 Analyse des résultats**

Tout au long de notre travail, nous serons amenés à parler d'identification, de reconnaissance et de catégorisation d'objets. Ces termes désignent des réalités qui peuvent se recouvrir et bien que celle-ci soit variable d'un domaine de connaissance à l'autre, chacun a une définition qui lui est propre. Un objet est identifié une fois que ses différents attributs ont été perçus et intégrés, ce processus impliquant donc une étape de décodage. La reconnaissance d'un objet suppose une exposition antérieure, qui autorise une réduction du nombre d'attributs nécessaires à percevoir pour le nommer. La catégorisation d'un objet repose également sur la perception de certains de ses attributs, suffisante pour pouvoir le placer mentalement dans une classe d'objets avec lesquels il possède plus ou moins d'attributs communs. Dans le cadre de la catégorisation, les classes d'objets à constituer peuvent être prédéfinies (comme dans notre étude sur la perception de la prosodie affective) ou laissées au libre choix des sujets (comme dans notre étude sur la catégorisation auditive). De la même manière qu'un objet, une catégorie peut être identifiée ou reconnue, quand le nombre requis d'attributs communs aux objets la composant est atteint.

Les études portant sur les sujets implantés cochléaires se heurtent à des limites méthodologiques récurrentes. La première concerne les propriétés des groupes étudiés, sur le plan des effectifs et de la distribution des valeurs. Les groupes de sujets implantés sont souvent de taille réduite et caractérisés par une grande variabilité des performances. L'application d'analyses paramétriques peut être envisagée si de multiples mesures pour une variable ont été réalisées. S'il s'agit d'études comparatives, s'intéressant par exemple à un groupe de sujets implantés cochléaires et à un groupe de sujets normo entendant contrôle, comme c'est le cas dans notre mémoire, les méthodes d'analyses paramétriques peuvent être difficilement applicables. Travailler sur un faible nombre de valeurs distribuées de manière hétérogène peut rendre en effet ces méthodes inappropriées.

Nous avons donc utilisé dans notre mémoire des méthodes d'analyse comparative non paramétrique, basées sur un ré échantillonnage par bootstrapping pour les comparaisons entre deux groupes, sur un test de Kruskal-Wallis suivis d'un bootstrapping avec comparaison deux à deux si plusieurs groupes étaient considérés. Les deux derniers travaux réalisés par l'équipe ont toutefois été acceptés avec des analyses supposant une distribution normale (modifications demandées par les reviewers) et c'est pourquoi le manuscrit actuellement en révision « Speech prosody perception in cochlear implant users with and without residual hearing » comportait également ce type d'analyses.

Lorsque la théorie de détection du signal était applicable, nous avons utilisé des mesures du  $d'$  pour représenter les performances des sujets. Le  $d'$  est un index de sensibilité permettant de dissocier les processus perceptifs des processus décisionnels dans une tâche de comportementale de détection (stimulus présent ou absent) ou de discrimination (stimulus cible discriminé par rapport au stimulus concurrent). En effet, si l'on considère le simple pourcentage de réponses correctes, on mesure des capacités perceptives (sensibilité du sujet) associées à des facteurs décisionnels, qui vont déterminer l'attitude du sujet pendant le test. Celle-ci est classiquement qualifiée de libérale si le sujet prend des risques en détectant/discriminant le stimulus fréquemment. A l'inverse, cette attitude est considérée comme conservatrice si la détection/discrimination du stimulus est peu fréquente, témoignant de la prudence du sujet pendant la réalisation du test.

La théorie de détection du signal prévoit 4 catégories en fonction de la réponse du sujet, et de la présence ou l'absence du stimulus (voir tableau II.1).

		STIMULATIONS	
		BRUIT	SIGNAL + bruit
REPONSES	OUI	Fausse Alarme FA	Détection Correcte DC
	NON	Rejet correct	Omission
		Nombre de présentations du stimulus BRUIT	Nombre de présentations du stimulus SIGNAL + BRUIT

Figure II.17. Principes généraux qui sous-tendent la théorie de détection du signal, avec les 4 types de réponses possibles du sujet selon que le signal est jugé présent ou absent.

Les valeurs de l'indice de sensibilité  $d'$  représentent les capacités perceptives du sujet. Sa valeur corollaire, le critère  $\beta$ , représente les facteurs décisionnels. Ces valeurs sont calculées en fonction de la probabilité de détection correcte et la probabilité de fausse alarme, par l'équation de Creelman [McMillan and Creelman, 1991], dans laquelle  $z$  (DC) est le  $z$  score correspondant à la probabilité de détection correcte et  $z$  (FA) est le  $z$  score correspondant à la probabilité de fausse alarme :

$$d' = z(\text{DC}) - z(\text{FA})$$

$$\beta = e^{(-1/2)(z^2(\text{DC}) - z^2(\text{FA}))}$$

## 6.2 Problématique

Les derniers travaux réalisés dans notre équipe et présentés en introduction font état d'un déficit de reconnaissance de la voix et d'un de ses attributs chez le sujet implanté cochléaire. Ce déficit est dissocié des capacités de discrimination de la parole dans le silence, généralement restaurées après implantation, et persiste même plusieurs années après implantation. Les facteurs expliquant ce déficit sont incomplètement élucidés et nos dernières études, en accord avec la littérature, nous avaient essentiellement fait envisager le défaut de restitution de F0 par l'implant, éventuellement associé aux caractéristiques de la réorganisation corticale secondaire à la surdité.

Les études comportementales portant sur une aptitude perceptive particulière (parole, détection de la voix, perception de la musique) permettent de déterminer l'importance du déficit chez les sujets implantés et éventuellement d'en identifier les mécanismes explicatifs. Une des limites qui peut leur être opposée est l'absence d'évaluation globale des processus perceptifs, quand plusieurs types de signaux auditifs sont présentés, et éventuellement de manière concomitante. Le cadre théorique dans lequel s'inscrit ce type d'évaluation est celui de l'analyse de scène auditive, théorie décrite par A. Bregman [Bregman et al., 1990], qui nous tend nos capacités de décomposer les environnements sonores complexes pour sélectionner le signal considéré comme pertinent. Pour chaque scène auditive, les différents objets sonores sont ainsi regroupés ou séparés sous la forme de flux auditifs, permettant d'optimiser la perception de l'objet concerné comme pertinent. Avant d'évaluer l'analyse de scène auditive chez les sujets implantés cochléaires, donc de les confronter à de multiples sources sonores concomitantes ou rapprochées dans le temps, nous avons souhaité au préalable démontrer leurs capacités de catégorisation de différents types de sons, linguistiques, para linguistiques et extra linguistiques.

Notre travail était également destiné à compléter l'évaluation des capacités des sujets implantés pour les informations para et extra linguistiques que nous avons initiée. Nous avons ainsi étudié leur perception de la prosodie (affective et linguistique) et de la musique en nous intéressant notamment à la contribution apportée par leur niveau de perception de F0.

## Chapitre III

Catégorisation libre de sons linguistiques,  
para linguistiques et extra linguistiques par  
le sujet implanté cochléaire



## **1. INTRODUCTION**

La plupart des patients récemment implantés rapportent des expériences singulières à la redécouverte de leur environnement sonore. Et à chacun son histoire : M. Im. « entend les piafs » dans son jardin montalbanais, Mme Ma. perçoit le bip de son alarme de voiture, et sur son parcours de golf, M. Dec reconnaît le bruit de la balle qui tombe dans le trou. Ces expériences auditives après implantation sont souvent vécues avec excitation et l'identification de nouvelles sources sonores constitue un des tout premiers défis chez les patients dont l'implant vient d'être activé.

Loin d'être anecdotiques, elles contrastent pourtant avec les performances généralement faibles obtenues par les patients dans des tâches d'identification des bruits environnementaux. Le déficit présenté par les patients dans de telles tâches est en effet majeur, avec des pourcentages d'identification correcte généralement inférieurs à 50% [Proops et al., 1999, Inverso and Limb, 2010, Shafiro et al., 2011], pouvant approcher le niveau de la chance. De la même manière que la discrimination de la parole, la reconnaissance d'instruments de musique, ou la reconnaissance d'une mélodie ont été évalués séparément, la perception des bruits de l'environnement a donc fait l'objet d'études cherchant à déterminer les capacités d'identification des patients implantés. Pour évaluer les performances auditives des patients dans leur globalité, les premiers travaux des années 80 ont considéré les résultats obtenus par les patients pour un test d'identification de phonèmes, puis un test de discrimination de mots, puis un test de discrimination de phrases, puis un test d'identification de bruits environnementaux, etc.. [Edgerton et al., 1983, Eisenberg et al., 1983, Schindler and Kessler, 1987, Tyler et al., 1989]. Ces études permettent de rendre compte de l'ensemble des capacités auditives post implantation des patients, mais seulement dans une certaine mesure. Elles n'évaluent pas le comportement des patients dans des conditions auditives naturelles, caractérisées par la co occurrence de multiples signaux acoustiques linguistiques, paralinguistiques et non linguistiques.

Dans une configuration naturelle, le patient se retrouve confronté à de multiples sources sonores. Il sera naturellement amené à « faire le tri » entre les signaux acoustiques pour sélectionner le signal considéré pertinent. Ce signal peut être un signal de parole émis par un locuteur au sein d'une conversation ou un bruit de l'environnement parmi d'autres bruits environnementaux comme dans le cas rapporté par M. Im. Face à un environnement sonore

complexe, le patient va donc s'attacher à séparer les différentes sources sonores et déterminer leur importance relative.

Ces capacités de groupage et de séparation des sources sonores renvoient à la notion d'analyse de scène auditive. Initialement décrite par A. Bregman en 1990 [Bregman, 1990], l'analyse de scène auditive est inspirée de la théorie gestaltiste selon laquelle les processus perceptifs traitent les phénomènes globalement, comme des ensembles structurés et non comme une simple addition d'éléments isolés. Bregman a ainsi proposé l'existence de flux auditifs que le système auditif sépare pour se concentrer sur l'information à percevoir. Au sein d'un courant auditif, des objets ou événements sonores sont regroupés en fonction de leurs similarités, du moment auquel ils sont émis ou de leur localisation spatiale. La séparation d'un environnement sonore complexe en courants auditifs et leur intégration fait donc intervenir plusieurs étapes depuis l'analyse acoustique des indices acoustiques fondamentaux jusqu'aux processus d'attention sélective. Les intervalles séparant les fréquences fondamentales des différentes sources sonores, ou les différences de spectre caractérisant leur timbre sont utilisés. Sur le plan temporel, la séparation en courants peut reposer sur les intervalles de temps séparant les sons émis par les différentes sources sonores ou dans une moindre mesure sur l'aspect temporel de la fréquence fondamentale. A niveau supérieur, la localisation spatiale des différentes sources sonores repose notamment sur les capacités centrales d'intégration binaurale des sujets. Enfin, des processus attentionnels de haut niveau permettent de se concentrer sur le signal acoustique cible.

Les modalités selon lesquelles les patients implantés cochléaires regroupent et séparent différentes sources sonores en courants auditifs sont actuellement peu connues. Certaines études psychophysiques ont évalué ces capacités en utilisant deux sons purs et ont montré que la séparation en courants auditifs reposait sur la fréquence, que celle-ci soit codée par la zone stimulée (information tonotopique) ou par les modulations d'amplitude [Chatterjee et al., 2006, Cooper and Roberts, 2007]. Pour les sons complexes, tels que des voyelles utilisées dans l'étude de Gaudrain et al [Gaudrain et al., 2008], les capacités de séparation en courants auditifs des patients implantés cochléaires seraient liées à leurs capacités de perception de la fréquence fondamentale. Cette étude portait sur des sujets normo entendant en condition de simulation d'implant par vocoder et retrouvait en effet une diminution des capacités de séparation en courants, secondaire à la dégradation spectrale et non compensée par les modulations d'amplitude présentes dans le signal.

Les capacités des patients implantés cochléaires pour l'analyse de scènes auditives naturelles, comprenant de multiples objets sonores (linguistiques, non linguistiques ou

environnementaux) n'ont à notre connaissance pas encore été évaluées. L'étude que nous présentons ne porte pas à proprement parler sur l'analyse de scènes auditives naturelles complexes mais elle constitue une étape préalable à ce genre d'études. Notre étude cherche en effet à caractériser les modalités de groupage et de séparation de stimuli auditifs naturels isolés par les patients implantés en posant deux principales questions :

-l'implant cochléaire leur transmet-il une information auditive suffisante pour réaliser ces processus de groupage ou de séparation ?

-le cas échéant, quelles sont les catégories auditives formées et selon quelles stratégies ?

Pour répondre à ces questions, une tâche de catégorisation libre a été proposée à un groupe de sujets implantés cochléaires dont les réponses ont été comparées à un groupe de sujets normo entendant. La question des stratégies utilisées dans les processus de catégorisation est complexe dans la mesure où plusieurs stratégies peuvent co exister, déterminées par autant de facteurs, en fonction de la nature des stimuli auditifs. Ce travail constitue donc une étude pilote, impliquant une tâche dont il fallait évaluer la faisabilité chez les sujets implantés, et pour laquelle les méthodes d'analyse spécifiques devaient être développées. Nos premières conclusions serviront de point de départ au travail de thèse engagé sur le sujet par E. Collett, doctorant en 1<sup>ère</sup> année au CerCo.

### **1.1 La catégorisation libre : principes généraux**

La perception, comme d'autres fonctions cognitives telles que la mémoire ou le langage, s'appuie implicitement sur les processus de catégorisation pour réduire la quantité d'informations à traiter. En observant un moineau au rebord d'une fenêtre, on « sait » ainsi qu'un moineau fait partie de la catégorie « oiseaux », et que la catégorie « oiseaux » fait partie de la catégorie « animaux ». On sait d'ailleurs qu'il s'agit d'un oiseau avant de reconnaître le moineau. Notre système de connaissance du monde le divise ainsi en catégories/classes/concepts entretenant des relations hiérarchiques entre eux [Rosch, 1973], depuis les éléments infra ordonnés jusqu'aux catégories supra ordonnées : une guitare (élément infra ordonné) est un instrument (catégorie de base) servant à jouer de la musique (catégorie supra ordonnée). La représentation interne d'une catégorie peut être plus facilement sollicitée par un de ses éléments infra ordonnés, considéré comme plus typique de la catégorie, comme élément dit prototype : un moineau est plus typique de la catégorie « oiseaux » qu'une poule et peut donc être utilisé comme élément de référence au sein de cette catégorie.

Les épreuves de catégorisation libre, ou tri libre, visent à étudier les modalités selon lesquelles des stimuli sensoriels sont classés par une ou plusieurs populations de sujets. Ce type de tâche est régulièrement utilisé en recherche industrielle, dans les domaines agro alimentaires ou textiles, car il permet d'appréhender de manière globale la perception de produits par les consommateurs ou de mettre en évidence des différences de perception entre populations de consommateurs.

La catégorisation libre est sous-tendue par le principe général suivant : le sujet testé doit constituer des classes à partir de produits/stimuli qui lui sont présentés sans critère de classification imposé. La catégorisation libre revient donc à partitionner des produits/stimuli en fonction de leur degré de similitude perçue. Les stimuli perçus comme similaires ou proches sont ainsi regroupés au sein d'une même classe et le sujet a la liberté de former autant de classes qu'il l'estime nécessaire. Le sujet peut ensuite intituler chacune des classes qu'il a constituées. Pour chaque sujet évalué, l'expérimentateur recueille ainsi une partition en différentes classes éventuellement intitulées. Au sein de chaque classe, un son peut être identifié comme son caractéristique de la catégorie, et qualifié dès lors de son prototype de cette catégorie. Dans une épreuve de catégorisation libre, on ne mesure pas de performances mais on évalue les stratégies utilisées par les sujets pour partitionner les stimuli d'une part, et la reproductibilité de ces stratégies entre sujets d'autre part. En d'autres termes, l'étude d'une épreuve de catégorisation libre dans un groupe de sujets donné cherche à répondre à deux principales questions :

- quels sont les stimuli proches ou distincts d'un point de vue perceptif dans un groupe de sujets?
- selon quels critères ces stimuli sont-ils classés ?

Il est également possible de rechercher d'éventuelles différences dans les partitions et les stratégies qui les sous-tendent d'un groupe de sujets à l'autre. L'extraction de l'ensemble des informations permettant de répondre à ces questions est donc difficile et peut requérir des méthodes d'analyse des résultats particulières.

## **1.2 La catégorisation libre : méthodes d'analyse**

La première étape de l'analyse des résultats d'une épreuve de catégorisation libre fait appel à des statistiques descriptives simples. On peut ainsi déterminer le nombre de classes constituées par chaque sujet, le nombre de classes médian constituées par groupe de sujets et le nombre de classes à intitulé identique ou semblable que les sujets ont constituées.

Si l'on s'intéresse plus précisément aux jugements de proximité inter stimuli et aux stratégies méthodes d'analyse des résultats sont multiples et l'on en distingue deux principales familles : les méthodes de classification et les méthodes factorielles. Ces familles de méthodes sont souvent utilisées en complément l'un de l'autre et nous décrivons une méthode, issue de chaque famille, que nous avons utilisée pour l'analyse des résultats. Ces deux méthodes reposent sur le traitement d'une matrice de dissimilarités agrégées. Cette matrice correspond à la somme des matrices de dissimilarités individuelles, obtenues pour chacun des sujets, dans lesquelles la proximité entre deux stimuli est codée 0 ou 1 selon qu'ils sont classés dans la même catégorie ou non. Cette matrice peut également être appelée matrice de distances ou de proximité dans certains logiciels de traitement statistique.

### ***Une méthode de classification : l'arbre additif***

La première méthode est issue des méthodes de classification et correspond à la méthode des arbres additifs. Elle permet de représenter les proximités entre stimuli sous la forme d'un arbre, composé de sommets correspondant aux stimuli à classer et de nœuds internes (ou virtuels) reliés par des arêtes. Deux stimuli reliés au même nœud interne sont, dans une certaine mesure, jugés similaires (voir figure III.1). Sur un arbre additif, ce niveau de proximité est représenté par la somme des longueurs des arêtes les séparant : plus cette longueur est faible, plus les stimuli sont perçus comme semblables. Les arbres sont constitués à l'aide de l'algorithme Addtree, développé par Sattah et Tversky [Sattah and Tversky, 1977] et adapté par Barthélémy (*in* Dubois, 1993) qui cherche à optimiser « la condition des quatre points » selon laquelle deux stimuli rattachés au même nœud interne sont plus proches que deux objets rattachés à deux nœuds différents. L'arbre additif qui est construit comprend ainsi un certain taux de quadruplets (« quatre points ») bien représentés par rapport aux données originales et un certain taux de quadruplets mal représentés. Le taux de quadruplets bien représentés est optimisé par l'algorithme AddTree. On parle d'accentuation ou de stress dont l'importance peut être évaluée au moyen d'une valeur numérique, obtenu par le calcul de Kruskal [Kruskal, 1978]. Plus la valeur du stress tend vers 0, plus la représentation par l'arbre additif est correctement ajustée aux données originales.

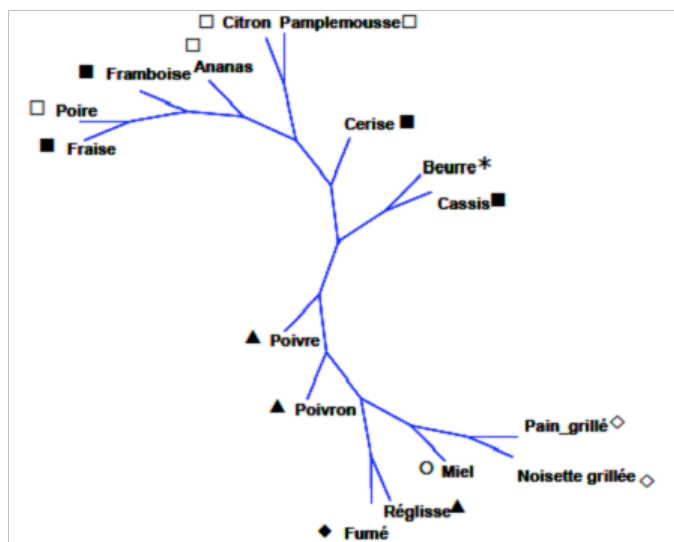


Figure III.1 : Arbre additif représentant les jugements de proximité ou de distance entre seize arômes de vin. L'utilisation de cette méthode est en effet relativement répandue en recherche agro alimentaire. Certains arômes fruités sont clairement regroupés dans une catégorie alors que des arômes tels que le poivre ou le poivron sont isolés, n'intégrant aucune catégorie. Tiré de Faye et al. (2011).

L'évaluation de la qualité de la représentation des données par l'arbre additif constitue en effet une étape importante de l'analyse des résultats. La valeur du stress, témoin de l'optimisation du taux de quadruplets bien représentés, est l'un des critères essentiels du contrôle de l'adéquation de la représentation par rapport aux données originales.

Les autres critères utilisés dans cet objectif sont les suivants [Poitevineau, 2009] :

- le degré d'arboricité qui doit tendre vers 1

- le taux de quadruplets supportant chaque arête interne. Les arêtes internes séparent deux groupes de stimuli. Plus le taux de quadruplets est proche de 1, plus la séparation représentée par l'arbre correspond aux jugements rapportés par les sujets.

### ***Une méthode factorielle : l'analyse MDS***

La deuxième méthode est factorielle et s'apparente au MultiDimensional Scaling (MDS) et vise à représenter dans un espace de faible dimension les distances entre stimuli en fonction de leur degré de similitude perçue par les sujets. Au sein de cet espace, deux stimuli jugés régulièrement semblables par les sujets apparaissent proches l'un de l'autre, deux stimuli jugés dissemblables sont à l'inverse éloignés (voir figure III.2). La représentation de ces similarités/dissimilarités est également réalisée sur la base de matrices de dissimilarités agrégées. De même que la méthode des arbres additifs, la qualité de la représentation par rapport aux données originales est évaluée au moyen du stress de Kruskal décrit plus haut.

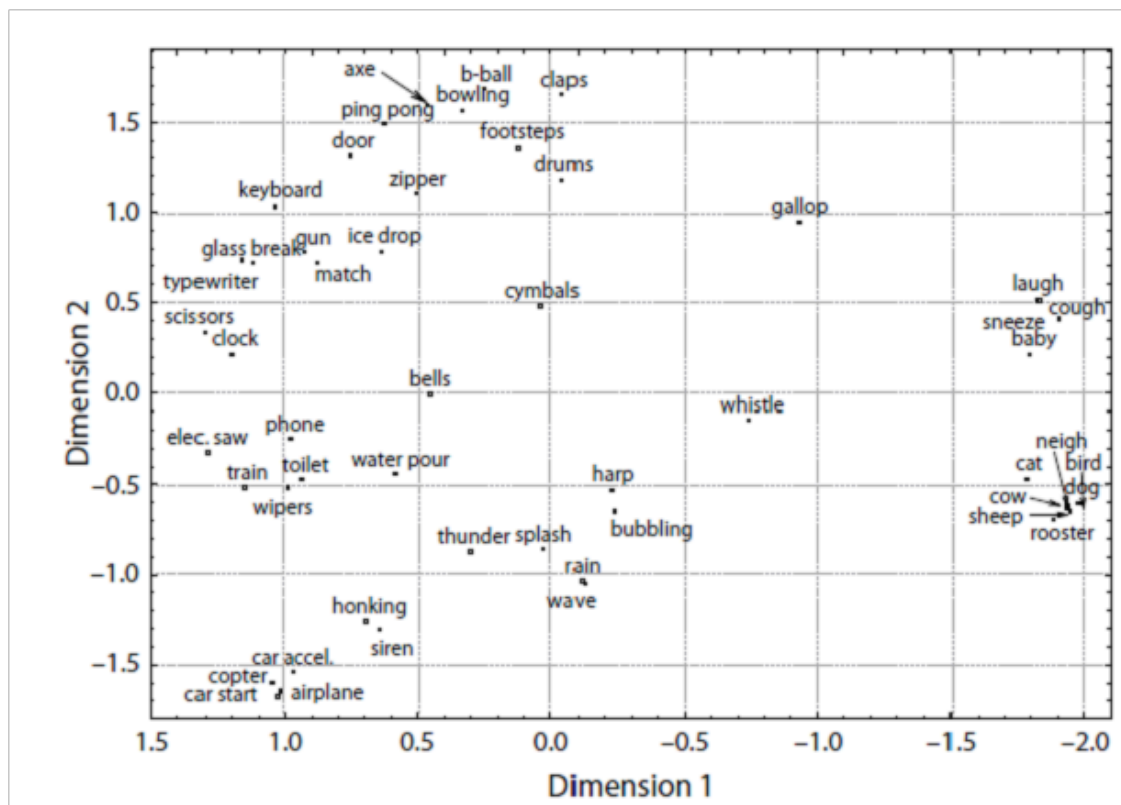


Figure III.2 : Catégorisation de 50 séquences sonores selon deux dimensions. On identifie clairement un cluster correspondant à des bruits provenant d’animaux, un cluster correspondant à des sons de voix humaine et un cluster de bruits d’engins/ de machine. Tiré de Gygi et al. (2007).

## 2. MATERIELS ET METHODES

### 2.1 Catégorisation auditive de stimuli auditifs naturels

La tâche de catégorisation libre réalisée dans notre étude a été conçue à partir du logiciel TCL-LabX [Gaillard, 2009]. Dans cette tâche, les sujets devaient catégoriser librement 16 stimuli auditifs correspondant à une séquence sonore écologique, comme par exemple le bruit d’une porte ou une mélodie jouée par un violon. Les stimuli sont représentés dans le tableau III.1.

Nous avons considéré quatre catégories d’appartenance a priori en fonction de la nature des stimuli. Ces catégories formées a priori nous ont permis d’étudier les distances entre les stimuli qui les composaient après réalisation de la tâche. Nous avons ainsi pu comparer les capacités de catégorisation entre les deux groupes de sujets, pour ces catégories formées a priori. Nous n’avons pas informé les sujets de l’existence ni de la nature de ces catégories a

priori afin qu'ils puissent réaliser la tâche de catégorisation librement, sans référence à une liste fermée de classes.

N° Stimulus	Séquence sonore	Catégorie d'appartenance a priori
S1	Porte qui s'ouvre	Environnement
S2	Mélodie jouée par une guitare	Musique
S3	Toux émise par un homme	Voix
S4	Rire émis par une femme	Voix
S5	Moteur de voiture	Objet
S6	Battements de cloche	Objet
S7	Syllabes prononcées par une femme	Voix
S8	Bruits de pas	Environnement
S9	Bruit d'eau qui coule	Environnement
S10	Note jouée par un vibraphone	Musique
S11	Mélodie jouée par un violon	Musique
S12	Moteur d'hélicoptère	Objet
S13	Sonnerie de réveil	Objet
S14	Bruit de verre qui se casse	Environnement
S15	Note jouée par un hautbois	Musique
S16	Syllabes prononcées par un homme	Voix

Tableau III.1: Nature et catégorie d'appartenance a priori des séquences sonores présentées aux deux groupes de sujets.

Ces 16 stimuli appartenait donc à 4 catégories formées a priori :

- voix humaine : parlée ou non parlée
- musique
- objets
- bruits environnementaux

La durée moyenne d'un stimulus était d'environ 2 secondes (1.54 seconde pour S14, 3.75 pour S13). Les principales caractéristiques acoustiques d'un son représentatif de chaque catégorie a priori sont illustrées dans la figure III.3.



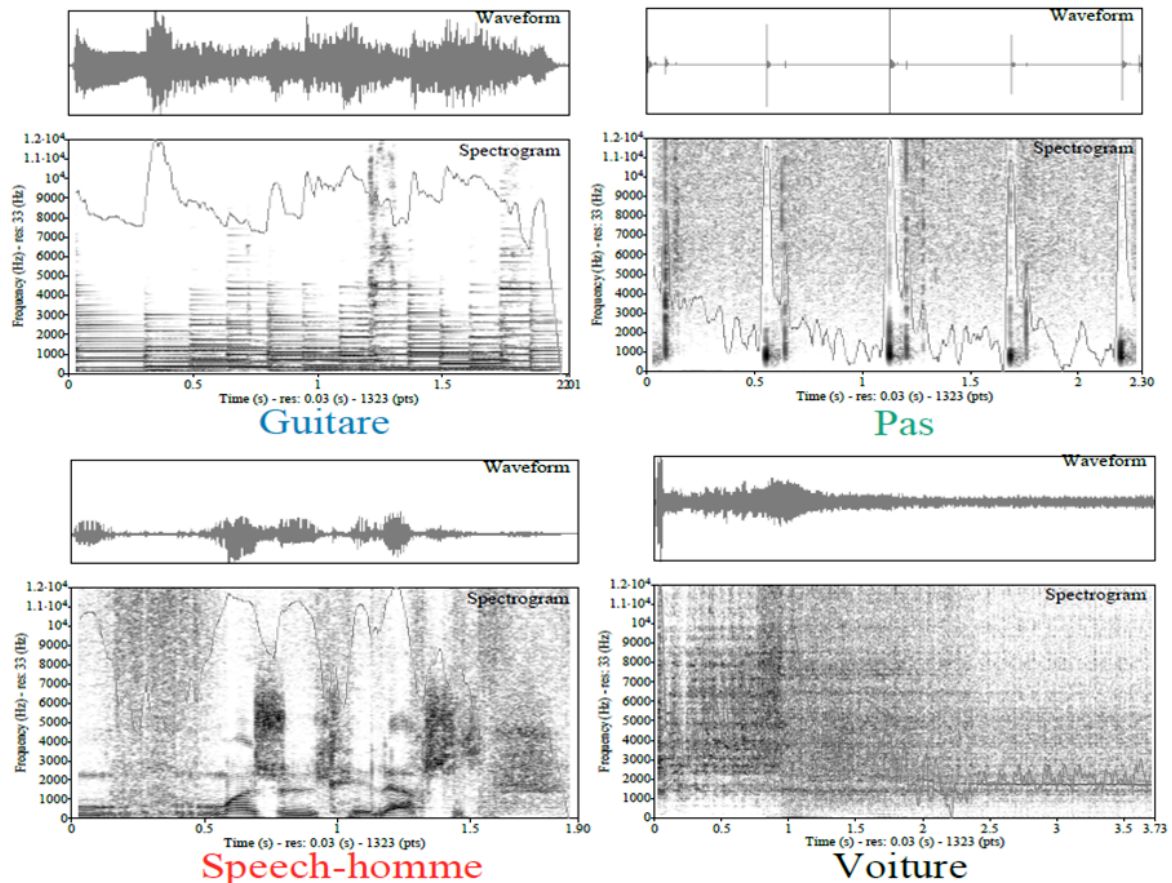


Figure III.3. Représentation de l'onde acoustique, du spectrogramme et des valeurs d'amplitude pour quatre sons compris dans les stimuli présentés aux sujets, chacun appartenant à une catégorie formée a priori (musique, voix, environnement, et objet).

## 2.2 Population

Deux groupes de sujets ont été étudiés dans ce travail. Le premier groupe était composé de 16 sujets implantés cochléaires, tous opérés dans le service et suivis à l'Unité d'Implant Auditif (voir tableau III.2). Ces sujets étaient âgés entre 45 et 76 ans (médiane : 55 ans) et étaient considérés comme expérimentés dans la mesure où leur implant avait été activé depuis plus de 18 mois. Les délais post activation étaient en effet compris entre 23 et 129 mois (médiane : 68 mois). Tous les sujets présentaient un niveau de discrimination de la parole dans le silence satisfaisant avec une discrimination correcte des mots dissyllabiques supérieure à 70%. Tous ces patients étaient porteurs d'un implant cochléaire unilatéral, à l'exception d'un sujet implanté de manière bilatérale.

Sujet	Age	Cause de la surdit�	Genre	D�lai post activation (mois)	Mod�le d'implant	Processeur	Strat�gie de codage	C�t� implant�	Audition r�siduelle
CI 1	72	Inconnue	F	71	Cochlear CI 24	Freedom	ACE	droit	non
CI 2	53	Inconnue	F	68	Cochlear CI 24	Freedom	ACE	droit	non
CI 3	48	G�n�tique	F	53	Cochlear CI 24	Freedom	ACE	droit	oui
CI 4	54	Otospongieuse	M	36	Cochlear CI512	Freedom	ACE	bilat	non
CI 5	56	Inconnue	F	77	Cochlear Hybrid L	Freedom	ACE	droit	oui
CI 6	70	Lobstein	F	84	Advanced Bionics HiRes	Fidelity	HiRes 90 k	gauche	oui
CI 7	52	Lobstein	F	55	MedEl sonata	Opus II	CIS	droit	non
CI 8	48	Inconnue	M	134	Cochlear CI 24	Freedom	ACE	droit	non
CI 9	52	Inconnue	M	90	Cochlear CI 24	Freedom	ACE	gauche	non
CI 10	64	Otospongieuse	F	42	Cochlear CI 24	CP810	ACE	droit	non
CI 11	62	Otite chronique	F	33	Cochlear CI512	CP810	ACE	droit	oui
CI 12	47	Inconnue	M	129	Cochlear CI22	Esprit3G	ACE	droit	non
CI 13	66	Inconnue	F	36	Advanced Bionics HiRes	Auria	HiRes90k	gauche	oui
CI 14	76	Inconnue	F	23	Cochlear CI512	CP810	ACE	droit	non
CI 15	70	Inconnue	M	68	Advanced Bionics HiRes	Auria	HiRes90k	droit	oui
CI 16	45	Inconnue	F	83	Cochlear CI 24	Freedom	ACE	droit	non

Tableau III.2 : Donn es d mographiques, cliniques et techniques relatives aux 16 sujets implant s cochl aires inclus et   l'implant utilis .

Le deuxi me groupe de sujets  tait compos  de 16 sujets normo entendant, sans ant c dent m dical de type otologique ou neurologique. Ces sujets  taient  g s entre 21 et 32 ans (m diane : 22 ans). Ce groupe nous a permis d' valuer les modalit s de cat gorisation par un groupe contr le sans d fici nce auditive. Il  tait n cessaire d'inclure le m me nombre de sujets dans chaque groupe pour pouvoir comparer les distances sur les matrices de dissimilarit s agr g es.

## 2.3 Proc dure

Les principes g n raux qui sous-tendent la passation d'une  preuve de cat gorisation libre diff rent des tests psychophysiques classiques. Il n'y a ainsi aucune contrainte de temps, qu'elle concerne le nombre de pr sentation des stimuli ou la dur e globale du test. Les sujets ont ainsi la possibilit  d' couter les stimuli autant de fois qu'ils l'estiment n cessaire avant et pendant la r alisation de la partition. Une fois la partition termin e, l'exp rimentateur demande au sujet d'intituler les classes qu'il a constitu es et d'identifier un son caract ristique de la classe au sein de chaque cat gorie, appel  son prototype.

Dans cette  tude, la pr sentation des stimuli et l'enregistrement des donn es ont  t  r alis s   l'aide de TCL-LabX, logiciel d velopp  sp cifiquement pour les  preuves de cat gorisation

libre (Gaillard, 2009). Chacun des seize stimuli auditifs était représenté à l'écran de l'ordinateur par une icône de couleur numérotée (S1 à S16). Pour écouter le son souhaité, le sujet devait simplement cliquer sur l'icône correspondante. L'expérimentateur donnait au sujet la consigne orale et écrite suivante : le sujet a pour tâche de constituer des classes de sons en fonction des critères qui leur semblaient les plus pertinents, en faisant glisser et en regroupant les icônes correspondantes. Lors de l'exposition de la consigne, il était précisé aux sujets qu'on ne mesurait pas de performances auditives mais que l'on s'intéressait à la manière dont ils regroupaient les sons. Le nombre de classes n'était pas limité et pouvait être compris entre 2 et 15. Une fois que le sujet avait constitué les classes de sons, il lui était demandé d'intituler les différentes catégories ainsi formées et d'identifier un son prototype au sein de chaque classe. Les stimuli étaient présentés en champ libre à l'aide de haut-parleurs (KINYO, modèle PS-240) à une intensité d'environ 65 dB, intensité à laquelle les sujets implantés sont évalués pour la discrimination de la parole.

Tous les sujets implantés cochléaires ont été évalués dans une cabine insonorisée du service d'ORL, Purpan. Le test durait entre 15 et 30 minutes en fonction des sujets. Tous les sujets étaient régulièrement suivis par l'Unité d'Implant Auditif du service et habitués aux évaluations des capacités de discrimination de la parole. Aucun des sujets n'était donc familier avec ce type de test et c'est pourquoi tous se sont étonnés de l'absence de mesure de performance. Ils interrompaient ainsi régulièrement leur tâche pour demander s'ils répondaient correctement. La consigne était alors répétée, en insistant sur le fait qu'on ne mesurait pas de performances mais « une manière de regrouper les sons » et les sujets reprenaient la tâche où ils l'avaient laissée.

## **2.4 Analyse des résultats**

Ainsi que nous l'avons décrit plus haut, nous avons réalisé dans un premier temps des analyses descriptives simples, en étudiant le nombre de classes constituées par sujet et la reproductibilité des intitulés de classes. Pour chaque sujet, les matrices de dissimilarités entre les stimuli ont été enregistrées. Les distances entre stimuli dans la matrice de dissimilarités agrégées ont fait l'objet d'analyses complémentaires visant à étudier et comparer leurs valeurs au sein de chaque catégorie formée a priori (voix, musique, objets, bruits environnementaux), pour chacun des deux groupes de sujets. Une recherche de corrélation entre les matrices agrégées des deux groupes par un test de Mantel a également été effectuée. Ces analyses ont

été réalisées avant que les matrices de dissimilarités agrégées soient traitées par l'algorithme AddTree.

Nous avons donc par la suite évalué les jugements de proximité inter stimuli au sein du groupe des sujets implantés à l'aide d'un arbre additif, en déterminant la qualité de sa représentation par les critères cités plus haut (stress, degré d'arboricité, taux de quadruplets bien représentés, taux de quadruplets supportant les arêtes internes). Nous avons considéré que les groupes de stimuli étaient constitués de manière reproductible entre les sujets lorsque le taux de quadruplets supportant les arêtes internes était supérieur à 0.8.

Enfin, nous avons réalisé une analyse MDS afin de représenter les distances originales entre les stimuli classés par les sujets implantés. Cette analyse a été réalisée à l'aide des logiciels R® et Xlstat®. L'adéquation de cette représentation a été évaluée au moyen de la valeur du critère de stress de Kruskal [Kruskal, 1978]:

- Stress > 0.2 : mauvais
- Stress entre 0.1 et 0.2 : acceptable
- Stress entre 0.05 et 0.1 : bon
- Stress < 0.05 : excellent

### **3. RESULTATS**

#### **3.1 Caractéristiques des partitions**

Les résultats des sujets normo entendant ont servi de référence pour étudier les partitions des sujets implantés cochléaires. Dans le groupe de sujets normo entendant, le nombre de classes constitué variait entre 1 et 8, avec un nombre médian de 5. Les principales catégories constituées étaient la catégorie intitulée « musique, instruments, mélodie » (14 sujets sur 16) et la catégorie « voix, parole, être humain » (13/16). Une catégorie intitulée « bruits de moteur, engins » était constituée par 6 sujets. L'identification d'un son prototype de chaque catégorie formée n'a pas été systématiquement recherchée ni déterminée par les sujets mais certains stimuli apparaissaient comme prototypes récurrents. Ainsi, les stimuli les plus fréquemment identifiés dans la catégorie « musique » étaient la guitare et le violon (3 sujets chacun) ; et dans la catégorie « voix, parole », les stimuli comprenant un signal de parole, prononcé par un homme ou une femme étaient considérés comme prototypes par 3 sujets chacun. Le comportement du groupe des sujets normo entendant dans cette tâche de catégorisation était considéré comme référence et les réponses des sujets implantés tendaient à

s'en rapprocher, tant au niveau du nombre et de l'intitulé des catégories que dans l'identification du son prototype.

Chez les sujets implantés, le nombre de classes constituées variait entre 4 et 10, avec une valeur médiane de 6. Plusieurs catégories étaient régulièrement formées par les patients (voir tableau III.2). La catégorie « musique, mélodie, instruments » était ainsi constituée par 13 des 16 sujets et la catégorie « voix humaine, parole » également par 13 sujets. L'identification d'un son prototype au sein de chaque catégorie apparaissait en revanche plus difficile et tous les sujets ne parvenaient pas à déterminer un stimulus caractéristique des classes constituées. Pour la catégorie « musique », la séquence sonore correspondant à la mélodie jouée par une guitare était la plus régulièrement considérée comme son prototype (4 sujets). Parmi les sons classés dans la catégorie « voix, parole », l'extrait correspondant à un signal de parole prononcé par un homme était identifié comme prototype de la catégorie par 5 sujets implantés.

La plupart des sujets implantés semblait donc constituer une catégorie supra ordonnée en fonction de ce qu'ils avaient identifié dans la séquence sonore. L'observation de leurs réponses pendant le test rendait également compte de cette étape d'identification du stimulus préalable à sa catégorisation. La plupart des sujets demandaient en effet à l'expérimentateur s'ils avaient bien reconnu le stimulus ou certaines de ses caractéristiques avant de l'intégrer à une classe.

Deux autres classes étaient constituées de manière répétée mais dans une moindre mesure. Plusieurs sons étaient ainsi regroupés dans une catégorie « bruits de moteur » pour 5 sujets et dans une catégorie « sonneries » chez 4 sujets. Parmi ces sons, l'identification d'un son prototype n'a pas pu être réalisée par les sujets implantés.

Catégorie	Nombre de sujets NHS	Nombre de sujets CI	Son prototype
Musique, Mélodie, Instruments	14	13	Guitare
Voix, Parole, Etres humains	13	13	Parole homme
Sonneries		5	x
Bruits de moteurs engins	6	5	x

Tableau III.2. Catégories générales les plus fréquemment formées par les sujets normo entendant et les sujets implantés cochléaires. L'identification d'un son prototype n'était pas possible pour la catégorie « sonneries » ou « bruits de moteur ».

Cette première étape descriptive met donc en évidence deux principales catégories formées par les sujets implantés sur la base d'un traitement sémantique des stimuli. Les sujets ont ainsi majoritairement identifié l'objet sonore, tout du moins ses attributs essentiels, avant de le classer dans une catégorie générale. En plus de ce traitement sémantique, un sujet parmi les 16 a eu recours à une stratégie narrative pour catégoriser certains sons, les intégrant à deux saynètes (par exemple : « quelqu'un qui s'éclaircit la voix pour chanter et jouer de la guitare »). Un seul sujet (CI 14) a utilisé une stratégie uniquement acoustique basée sur la hauteur tonale ou l'intensité perçue des stimuli pour constituer les classes de sons.

### **3.2 Analyses des distances entre stimuli dans les matrices de dissimilarités agrégées**

Les distances entre les stimuli partitionnés telles que représentées par les matrices de dissimilarités agrégées pouvaient varier entre 0, si les deux stimuli concernés avaient toujours été classés dans la même catégorie, et le nombre maximal de sujets ajusté en fonction du nombre de classes constituées (ici 19.15) si les deux stimuli avaient toujours été classés dans une catégorie différente. Ces distances ont été analysées dans leur totalité, puis au sein de chacune des quatre catégories qui avaient été considérées a priori (voix, musique, objets, bruits environnementaux). Ces analyses quantitatives sur les distances permettent d'évaluer les capacités de catégorisation d'un groupe de sujets, dans leur globalité ou par rapport à une des 4 catégories formées a priori. Ces valeurs témoignent en effet de la variabilité des jugements de proximité inter stimuli au sein d'un groupe de sujets. Plus les distances sont élevées, plus la catégorisation a été variable dans le groupe de sujets considéré. Plus ces distances sont faibles, plus la catégorisation a été reproductible d'un sujet à l'autre au sein de ce groupe. Nous avons assimilé des distances élevées au sein d'une catégorie formée a priori à un déficit de catégorisation dans la mesure où elles traduisaient alors l'insuffisance du regroupement de stimuli considérés, a priori, comme proches.

Nous avons tout d'abord comparé les distances séparant tous les stimuli entre le groupe des sujets implantés cochléaires (CI) et le groupe contrôle des sujets normo entendant (NHS) (voir figure III.3) par une analyse en bootstrap. Ces distances reflètent la dispersion globale des stimuli tels qu'ils ont été catégorisés par les sujets. Il existait une différence significative entre les deux groupes, avec une distance moyenne de 13,1 (intervalle de confiance à 95% :12.2-13.9) dans le groupe CI contre 9,6 (intervalle de confiance à 95% :8.1-11.2) dans groupe

NHS. Évaluées par les distances séparant tous les stimuli, les capacités de catégorisation étaient donc légèrement déficitaires dans le groupe des sujets implantés.

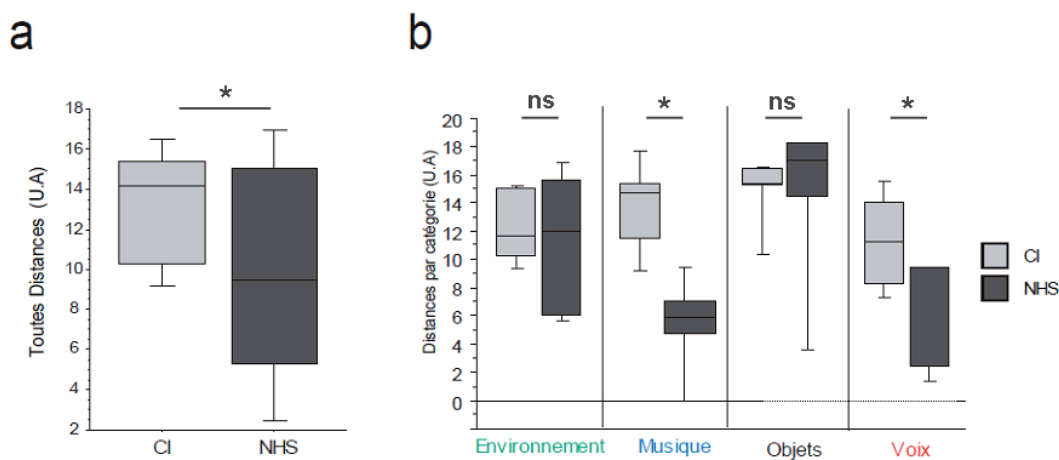


Figure III.3 : Analyse comparative des distances séparant tous les stimuli (a) et des distances séparant les stimuli au sein d'une catégorie formée a priori (b) entre le groupe des sujets implantés (CI) et le groupe contrôle des sujets normo entendant (NHS).

L'analyse des distances séparant les stimuli au sein de chacune des catégories formées a priori retrouve également des différences entre les sujets implantés et les sujets normo entendant. Nous avons en effet retrouvé des distances significativement plus faibles dans le groupe des sujets normo entendant pour la catégorie formée a priori « musique » avec une valeur moyenne respective de 5.5 (intervalle de confiance à 95% :3.5-6.8) contre 13,9 (intervalle de confiance à 95% :12.1-15.3) dans le groupe des sujets implantés cochléaires (voir figure III.3b). De même pour la catégorie formée a priori « voix », les distances entre stimuli étaient plus faibles dans le groupe NHS (moyenne : 6.9 ; intervalle de confiance à 95% : 4.5-8.8) que dans le groupe CI (moyenne : 11.3 ; intervalle de confiance à 95% : 9.5-13). Il n'existait pas de différence significative entre les groupes CI et NHS pour les distances originales séparant les stimuli des autres catégories formées a priori (« objet » et « environnement »).

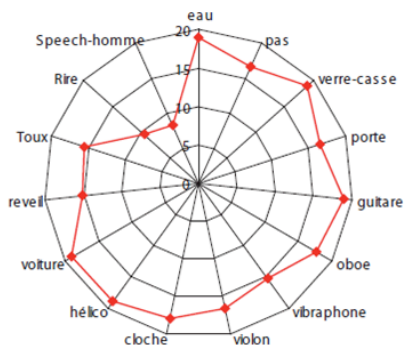
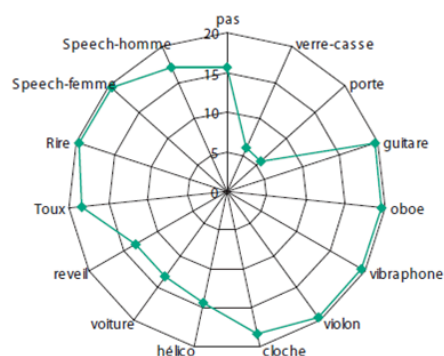
Ces différences témoignent de meilleures capacités de catégorisation dans le groupe NHS si l'on considère les catégories formées a priori « musique » et « voix ». Les modalités de catégorisation de chaque stimulus par rapport aux autres peuvent également être illustrées à l'aide de polar plots (voir figure III.4). Ces graphes permettent de représenter dans chaque groupe les distances séparant un stimulus des autres stimuli de sa catégorie d'appartenance a priori, mais également des extraits sonores correspondant aux autres catégories.

**Distances stimulus speech femme**



**Groupe NHS**

**Distances stimulus eau**



**Groupe CI**

**Distances eau**

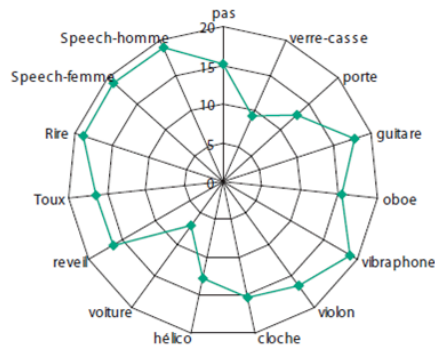


Figure III.4 : Polar plots représentant les distances sur matrices de dissimilarités agrégées du groupe NHS (haut) et du groupe CI (bas) pour un stimulus issu de la catégorie formée a priori « voix » (gauche) et un stimulus issu de la catégorie formée a priori « environnement » (droite). Ces graphes illustrent les relatives similitudes entre sujets CI et NHS pour la catégorisation de la parole, et les différences de catégorisation inter groupes pour un stimulus de type « bruit de l'environnement ».

Ces premières analyses démontrent l'existence de capacités de catégorisation auditive chez le sujet implanté cochléaire. Comme le sujet normo entendant, il utilise un traitement sémantique de l'information auditive pour réaliser la tâche de partition : les stimuli sont identifiés pour être partitionnés. Deux principales catégories étaient constituées par les sujets implantés cochléaires (une catégorie « voix » et une catégorie « musique »).

Les analyses comparatives des distances séparant les stimuli sur les matrices de dissimilarités agrégées ont retrouvé une différence significative pour les catégories formées a priori « voix » et « musique ». Dans ces catégories, le déficit d'identification des stimuli expliquait probablement les distances plus importantes retrouvées dans le groupe des sujets implantés. L'analyse des modalités de catégorisation par arbre additif nous a permis de compléter l'évaluation qualitative des modalités de partition des deux groupes de sujets.



### **3.3 Analyse de la catégorisation par arbre additif**

Un arbre additif correspondant aux réponses de chaque groupe a pu être généré en appliquant l'algorithme AddTree (Sattath et Tversky, 1977 ; version de Barthélémy et Guenoche, 2001) à la matrice de dissimilarités agrégées de chaque groupe.

Ces arbres (figure III. 4a et b) ont pu être construits avec des critères d'adéquation satisfaisants par rapport aux données originales. Le critère de stress de Kruskal était en effet calculé à 0.084 dans le groupe des sujets implantés et à 0.074 dans le groupe des sujets normo entendant, ce qui témoigne d'une optimisation limitée du taux de quadruplets bien représentés. Dans chaque groupe, l'arbre additif a ainsi permis de représenter les distances entre les stimuli partitionnés, la distance entre deux objets correspondant à la somme des arêtes qui les séparent.

Dans le groupe des sujets implantés cochléaires, l'observation de l'arbre généré (voir figure III.5 a) permet d'identifier deux principales catégories au sein desquelles les distances entre objets sont relativement faibles, regroupant donc des séquences sonores jugées proches. Ces catégories correspondaient aux deux principales classes intitulées « voix, parole » et « musique, mélodie », constituées par la majorité des sujets implantés (13/16). La catégorie « voix, parole » comprenait 3 stimuli jugés proches avec deux de signaux de parole prononcés par deux locuteurs de genre différent, dont le son identifié comme prototype et un signal paralinguistique (rire de femme). La catégorie « musique, mélodie » était également composée de 3 stimuli (guitare, violon et cloche) dont le stimulus identifié comme prototype de la classe. Une troisième catégorie, comprenant les extraits sonores S05 (« moteur voiture ») S9 (« eau qui coule »), et S14 (« verre cassé ») a également regroupé des sons jugés proches. Ceux-ci pourraient correspondre à la catégorie « bruits de moteurs, engins » ainsi qu'elle a été intitulée par 5 sujets.

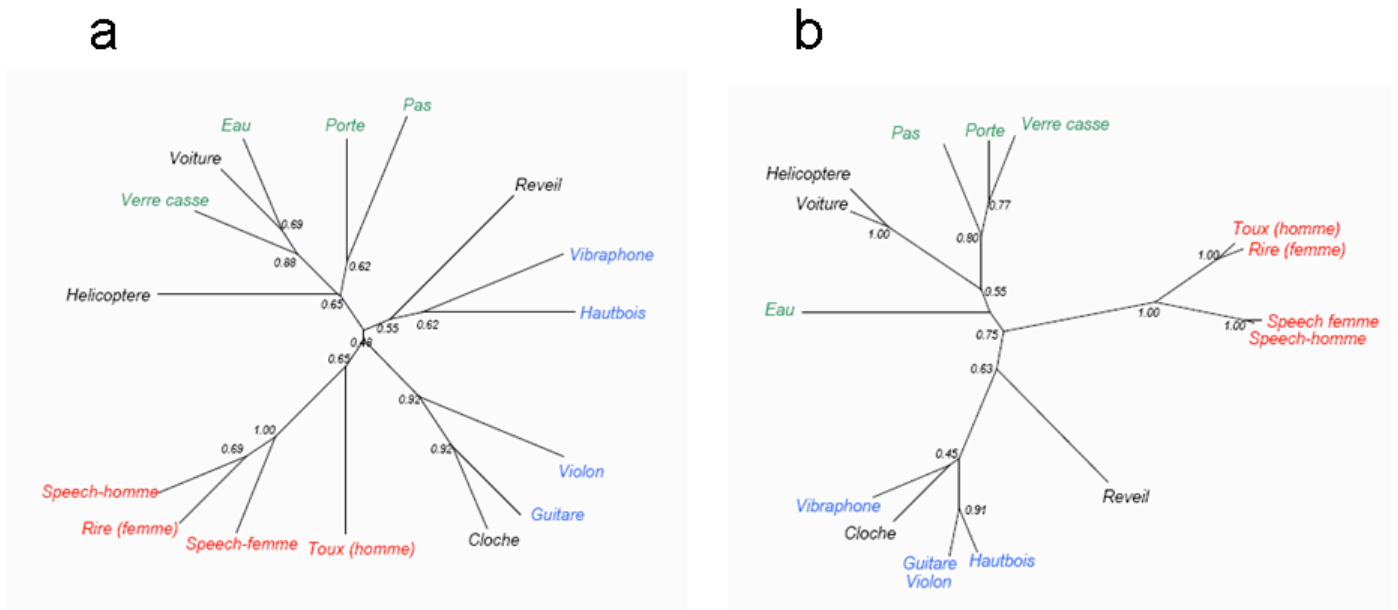


Figure III.5 a et b : Arbres additifs représentant les distances entre les stimuli classés par les sujets implantés cochléaires (a) et normo entendant (b). Le taux de quadruplets supportant chaque arête interne est indiqué et représente la fiabilité de la séparation en groupes d'objets suivant le nœud interne. Cette valeur est supérieure à 0.8 pour les catégories « voix » et « musique » dans le groupe CI et pour les catégories « voix », « musique », « véhicules » et « environnement » dans le groupe NHS.

Le traitement des distances originales entre stimuli par l'algorithme AddTree a permis de souligner les différences de distance inter stimuli entre le groupe des sujets implantés et le groupe des sujets normo entendant. L'observation de l'arbre additif généré à partir des réponses du groupe NHS (figure III.5 b) révèle en effet des distances plus faibles entre stimuli partitionnés dans la même catégorie. Ainsi, si l'on compare la somme des arêtes reliant les stimuli classés dans la catégorie « voix » par les sujets implantés à la même somme dans le groupe des sujets normo entendant, la différence entre les deux groupes est manifeste.

La représentation par arbre additif a donc permis de confirmer ce que la description des classes formées par les sujets suggérait. Deux principales catégories de sons sont formées : une catégorie correspondant à la voix humaine parlée ou non parlée et une autre catégorie correspondant à la musique. La deuxième méthode spécifique d'analyse que nous avons appliquée relève de l'analyse MDS. Comme l'arbre additif, cette méthode permet de représenter les jugements de proximité entre stimuli et leur reproductibilité entre les sujets. Elle cherche en plus à définir les stratégies qui ont été utilisées pour catégoriser les stimuli.

### 3.4 Analyse de la catégorisation et de ses modalités par MDS

L'analyse d'une catégorisation de stimuli par MDS vise à représenter ces stimuli par des points dans un espace euclidien de faible dimension (ici deux dimensions) de telle sorte que les distances entre les points reflètent au mieux la matrice de dissimilarités agrégées.

Nous avons représenté ces distances dans un espace à deux dimensions, témoins de la stratégie utilisée par les sujets pour catégoriser les stimuli. L'analyse MDS était réalisée sur la base de la matrice de dissimilarités agrégées du groupe CI et du groupe NHS. Pour chacune des analyses MDS, le critère de stress était  $>$  à 0.2 (0.28 dans le groupe CI et 0.21 dans le groupe NHS). La représentation des données originales sous la forme d'un espace à deux dimensions était donc imparfaite et à interpréter avec précaution. Nous nous sommes appuyés sur les résultats des sujets normo entendant pour interpréter l'analyse MDS correspondant aux sujets implantés. Le résultat de l'analyse MDS pour chacun des groupes illustré figure III.6 et 7.

Dans le groupe des sujets normo entendant, l'analyse par MDS permettait de retrouver globalement les quatre catégories formées a priori (musique, voix, environnement, objet).

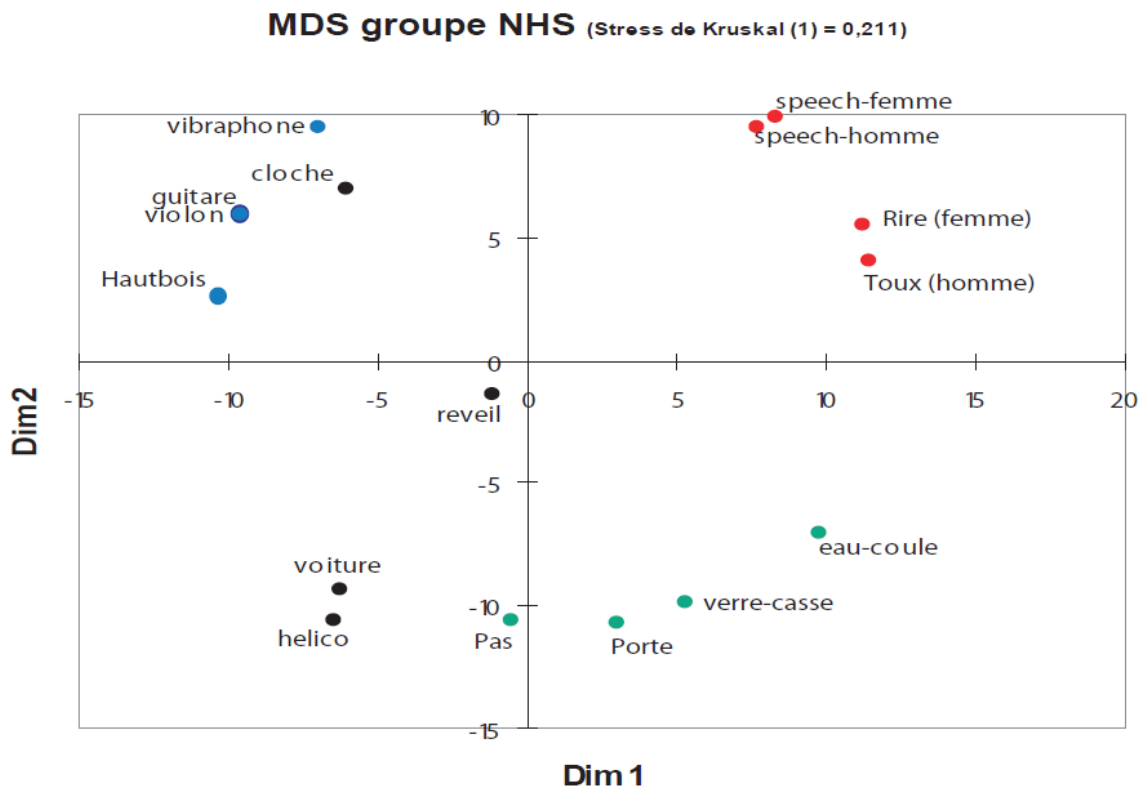


Figure III.6. Analyse factorielle par MDS sur matrice de dissimilarités agrégées dans le groupe NHS. Les distances entre stimuli sont plus faibles et 4 clusters différents peuvent être identifiés : un cluster « voix », un cluster « musique », un cluster « bruit d'engins » et un cluster « bruits de l'environnement ».

Les clusters regroupant les stimuli apparaissent en effet plus regroupés et plus nombreux dans le groupe NHS. Les catégories « voix » et « musique » étaient également représentées et clairement séparées ici par la dimension 1. La dimension 2 permettait de les séparer de deux autres clusters composés de bruits d'engins pour l'un et de bruits environnementaux pour l'autre. L'extrait sonore correspondant à la cloche était également classé dans la catégorie « musique ».

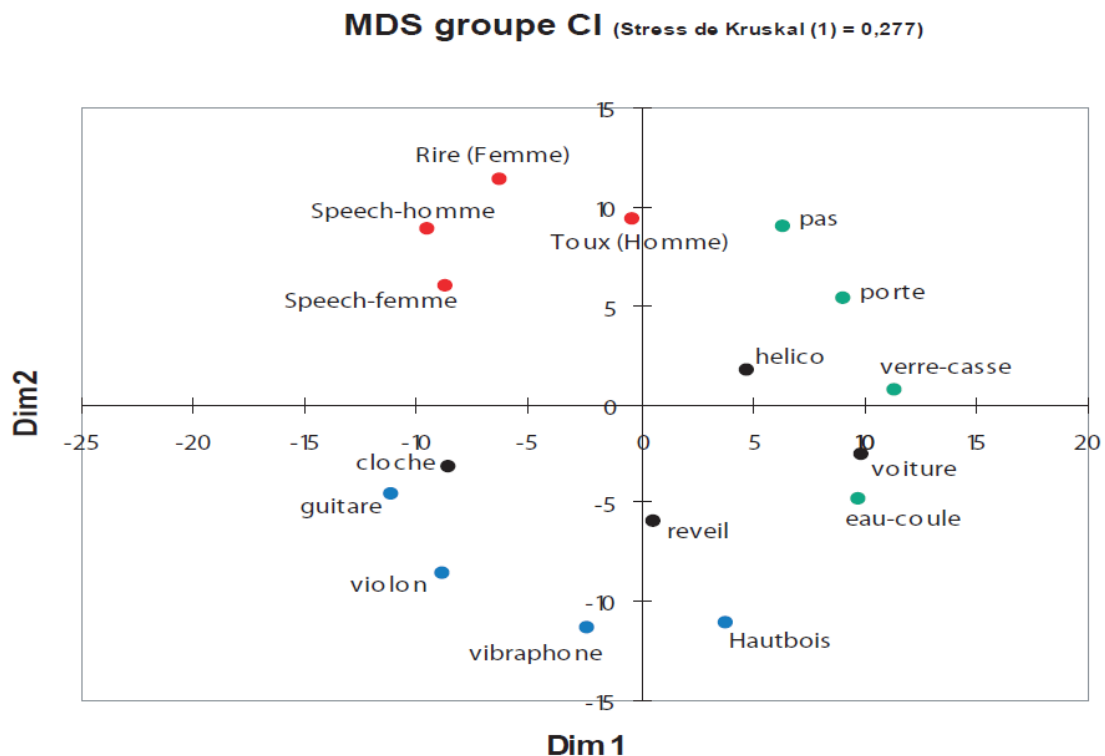


Figure III.7. Analyse factorielle par MDS sur matrice de dissimilarités agrégées dans le groupe CI. Cette analyse permet l'identification de clusters de stimuli selon les groupages et séparation réalisés par les sujets. Deux clusters se dégagent et sont séparés par la dimension 1 : un cluster « voix » et un cluster « musique ». Les autres stimuli apparaissent plus dispersés

L'observation de l'analyse MDS de la catégorisation des stimuli dans le groupe des sujets implantés cochléaires nous a permis d'identifier deux clusters de stimuli, correspondant aux deux principales catégories formées précédemment visualisées sur l'arbre additif : un cluster « voix » comprenant les stimuli « paroles homme et femme » et « rire (femme) » ; un cluster « musique » comprenant les extraits de guitare, de violon et la cloche. Ces deux clusters étaient séparés entre eux par la dimension 2 et des autres stimuli par la dimension 1. Ces clusters apparaissaient toutefois moins clairement constitués que dans le groupe des sujets

normo entendant, et cette méthode de représentation illustre donc la variabilité des réponses des sujets implantés.

Qu'elles soient formées par le groupe des sujets implantés ou par le groupe des normo entendant, les catégories « voix » et « musique » étaient donc toujours séparées des autres stimuli par une dimension (dimension 1 pour le groupe CI, dimension 2 pour le groupe NHS). Des analyses complémentaires portant sur les caractéristiques acoustiques des stimuli ont été réalisées afin de rechercher d'éventuels corrélats acoustiques à la catégorisation des sujets implantés cochléaires. S'il apparaît clair que la catégorisation s'opère selon un traitement sémantique dans nos deux groupes de sujets, les sujets implantés pourraient s'appuyer sur certaines caractéristiques acoustiques spécifiques, corrélées aux dimensions de l'analyse MDS, pour réaliser la catégorisation des stimuli.

### **3.5 Existe-t-il des corrélats acoustiques aux modalités de catégorisation ?**

Plusieurs caractéristiques acoustiques des stimuli ont été étudiées plus avant afin de rechercher une éventuelle relation entre leurs valeurs et les dimensions 1 et 2 retrouvées lors de l'analyse MDS. Nous avons utilisé certains des paramètres acoustiques déterminés par Gygi et al. dans leur étude portant sur les facteurs acoustiques spectro temporels à la base de l'identification de bruits environnementaux [Gygi et al., 2004]. Certains de ces paramètres concernaient l'enveloppe du stimulus, comme le nombre de pics d'intensité, le nombre de bursts (augmentation de l'intensité d'au moins 4 dB, tenue pendant plus de 20 ms), la durée des bursts rapportée à la durée totale du stimulus. L'importance de la fréquence fondamentale du stimulus était également étudiée au travers de la F0 moyenne, de la F0 maximale et de la saillance moyenne de F0 (illustre la facilité d'extraction de F0 par le niveau d'harmonicité dans le stimulus). Le spectre du stimulus était également évalué par la centroïde spectrale (moyenne des fréquences représentées dans le stimulus) qui est un moyen de caractériser son timbre.

Aucun de ces paramètres acoustiques n'était corrélé de manière significative (test de Spearman,  $r > 0.05$ ) avec les dimensions 1 et 2 mises en évidence par l'analyse MDS dans le groupe des sujets implantés. Le degré d'harmonicité, évalué par la saillance de F0, ne semble donc pas en rapport avec une quelconque des dimensions de la catégorisation chez les sujets implantés. Il est évident que ceci n'exclut pas formellement l'existence de corrélats acoustiques à la catégorisation. D'autres paramètres, non testés dans cette première analyse, sous-tendent vraisemblablement le processus global de catégorisation et d'identification des

objets sonores. L'évaluation d'autres paramètres caractérisant le spectre (spectral skew, spectral kurtosis), l'impulsivité [Marozeau, 2004] ou l'aspect temporel de l'enveloppe du signal (mesures d'autocorrélation) est nécessaire et susceptible de retrouver les facteurs acoustiques explicatifs de la catégorisation. Cette évaluation sera poursuivie dans le cadre du travail de thèse d'E. Collett (doctorant 1<sup>ère</sup> année).

## **4. DISCUSSION**

### **4.1 Capacités de catégorisation chez les sujets implantés cochléaires**

Notre étude a permis d'apporter des réponses à deux des trois principales questions que nous nous étions posés. Nos résultats démontrent tout d'abord que les patients implantés cochléaires ont les capacités de classer des séquences sonores écologiques en différentes catégories. Avant de réaliser cette étude, nous doutions effectivement de l'existence même de ces capacités, en nous appuyant notamment sur les travaux de Zoé Massida, ancienne doctorante au CerCo (PhD, 2010). Ceux-ci [Massida et al., 2011] comprenaient notamment une étude comportementale de détection de la voix humaine parmi des bruits environnementaux, réalisée chez 30 sujets implantés cochléaires et des sujets contrôle normo entendant. Le déficit retrouvé chez les patients était important, avec des performances approchant le niveau de la chance, et inférieures à 50% des capacités des normo entendant. Considérés dans le même temps que le faible niveau de reconnaissance du timbre des instruments de musique [Gfeller et al., 1998], ces résultats laissaient ainsi penser que la catégorisation des sujets implantés serait particulièrement perturbée. S'il est vrai que nos analyses comparatives des distances inter stimuli sur les matrices de dissimilarités agrégées montrent l'existence d'un déficit de catégorisation chez les sujets implantés, celle-ci demeure tout à fait possible avec deux catégories de stimuli majoritairement formées identiques entre implantés et normo entendant. D'une manière générale pour l'ensemble des jugements de proximité inter stimuli, il existait une relative similitude entre les réponses des sujets implantés et des sujets normo entendant, comme en témoigne le résultat du test de Mantel corrélant les matrices de dissimilarités agrégées des deux groupes ( $r = 0.59$  ;  $p < 0.0001$ ). Les catégories « voix » et « musique » demeuraient toutefois les deux classes formées par la

majorité des sujets dans chaque groupe (13/16 chez les implantés, 13 et 14 sujets normo entendant) même si elles ne comprenaient pas nécessairement les mêmes stimuli.

Les sujets implantés, comme les sujets normo entendant, catégorisaient donc les stimuli après les avoir identifiés, ou après avoir reconnu certaines de leurs caractéristiques, selon un traitement majoritairement sémantique. Les classes formées correspondaient à des catégories générales ou supra ordonnées (« voix », « parole », « mélodies », « musique ») d'objets suffisamment identifiés (« gens qui parlent », instruments ») pour pouvoir les regrouper. Un seul sujet implanté a utilisé une stratégie explicitement acoustique, basée sur la comparaison de paramètres tels que l'intensité ou la hauteur tonale. Ce mode de catégorisation sémantique est régulièrement rapporté dans la littérature portant sur la catégorisation auditive en conditions normales. Dans une étude de catégorisation libre de plus de 120 bruits environnementaux [Marcell et al., 2000], des catégories générales avaient été formées par deux juges après catégorisation par les sujets évalués. La majorité des catégories y était constituée à partir de l'identification de l'objet sonore (animal, instrument de musique). De rares stimuli étaient regroupés selon une stratégie plutôt narrative, comme des éléments intégrant une suite d'événements. Si ce traitement sémantique semble indéniablement être à la base des modalités de catégorisation des sujets implantés, la notion d'identification préalable des objets sonores pose problème : comment les sujets implantés peuvent-ils catégoriser ce qu'ils sont réputés avoir du mal à identifier ?

La première raison expliquant cette discordance tient au type de tâche qui est proposé aux sujets implantés cochléaires. Dans les études précédemment citées, les expérimentateurs mesurent leurs performances (pourcentage de réponses correctes ou critère assimilé comme le d') pour reconnaître des stimuli courts (<500 ms dans l'étude de Massida), présentés une seule fois, dans des conditions potentiellement anxiogènes pour les sujets. Les épreuves de catégorisation libre ne prévoient quant à elles aucune contrainte de temps, avec des stimuli écologiques dont l'écoute peut être répétée indéfiniment. Ici, le nombre d'écoutes d'un stimulus demeurait toujours inférieur à 5 dans le groupe des sujets implantés cochléaires. L'identification des objets sonores a ainsi pu être facilitée par les conditions expérimentales dans lesquelles notre étude s'est déroulée.

La deuxième raison susceptible d'expliquer l'apparente discordance entre la littérature et nos résultats réside dans la définition même qu'on donne à l'identification. Celle-ci est le résultat d'un processus rapide mais dynamique, par lequel la perception des différents attributs caractérisant un objet est progressivement intégrée. Dans notre étude, les sujets implantés reconnaissaient vraisemblablement « suffisamment » les attributs de certains sons pour

pouvoir les classer dans une catégorie générale, sans nécessairement pouvoir finaliser l'identification de l'objet concerné. L'étude d'Inverso et al. [Inverso and Limb, 2010] à propos de l'identification des sons non linguistiques par les sujets implantés cochléaires renforce cette hypothèse. Les auteurs y ont évalué les capacités d'identification d'extraits sonores classés dans 5 catégories formées a priori (animal, humain, machines, environnement, musique). Contrairement à notre étude, les sujets étaient informés de l'existence de ces classes a priori. L'identification précise de l'objet sonore était difficile pour la majorité des sujets, qui parvenaient toutefois à les catégoriser (voir figure III.8). Cette dissociation était particulièrement nette pour les sons musicaux, classés dans la catégorie « musique » dans près de 90% des cas, mais identifiés pour seulement 30% d'entre eux.

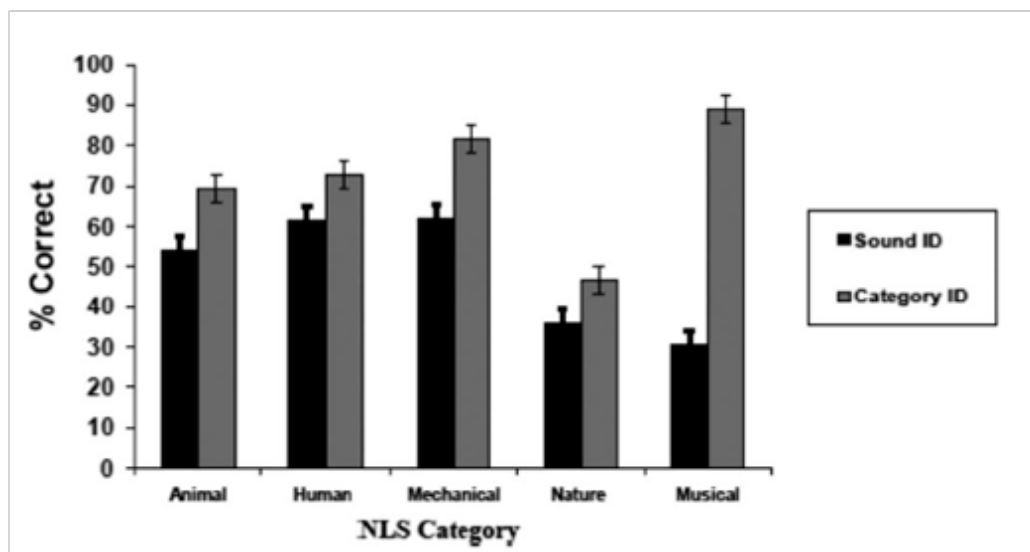


Figure III.8. Les performances pour la catégorisation des sons sont significativement supérieures aux capacités d'identification de sujets implantés cochléaires. La différence est particulièrement significative pour la musique. Tiré de Inverso et al. (2010).

La question qui demeure tient aux caractéristiques des sons suffisamment reconnues par les sujets pour le regroupement et la séparation des stimuli. Les paramètres acoustiques que nous avons évalués dans notre analyse préliminaire n'étaient pas corrélés aux dimensions déterminées par l'analyse MDS, mais ils ne peuvent être écartés du processus global puisqu'ils ont nécessairement contribué à bas niveau à l'identification de la catégorie à laquelle l'objet appartenait.



## 4.2 Principales catégories formées

Dans le groupe des sujets implantés comme dans le groupe des normo entendant, deux catégories « voix » et « musique » étaient majoritairement formées. Bien que l'adéquation de l'analyse MDS aux données originales ne soit pas parfaite, nous pouvons nous attarder sur les dimensions ayant supporté la catégorisation de ces stimuli. Dans le groupe des sujets implantés, la catégorie « voix » et la catégorie « musique » étaient séparées des autres stimuli par la dimension 1 (voir figure III.6). De la même manière dans le groupe NHS, les catégories « voix » et « musique » étaient séparées des autres stimuli par la dimension 2, ce qui indique que ces deux catégories présentaient des caractéristiques communes.

Chez les sujets implantés, les stimuli intégrés dans la catégorie « voix » étaient les deux signaux de parole prononcés par un homme et une femme, ainsi que le rire de femme ; la catégorie « musique » regroupait quant à elle les deux mélodies jouées par la guitare et le violon ainsi que les battements de cloche. Des similitudes rythmiques dans la structure temporelle de ces stimuli pourraient avoir servi de base à la séparation d'avec les autres extraits sonores. En effet, et ainsi que l'illustre la figure III.5, l'autre stimulus de la catégorie « voix » (toux d'homme) et les deux autres stimuli de la catégorie « musique » (note unique jouée par un hautbois ou un vibraphone) n'étaient pas partitionnés de manière robuste. Cette hypothèse semble valide si l'on considère les capacités connues des patients implantés pour la détection de l'enveloppe du signal acoustique, principal paramètre transmis par le processeur, à la base des capacités de discrimination de la parole. Cette hypothèse d'une catégorisation sur la base d'indices temporels est renforcée par l'existence d'interactions démontrées entre les représentations mentales du rythme pour la parole et la musique [Patel and Daniele, 2003] [Patel and Daniele, 2003, Hannon, 2009]. Le sujet normo entendant est par exemple capable de déterminer la langue dans laquelle une chanson a été composée à partir de son seul motif rythmique, quand les indices liés au timbre et aux variations de F0 ont été enlevés de l'extrait musical [Hannon, 2009]. Les similitudes rythmiques rapprochant les stimuli musicaux des stimuli de parole pourraient donc avoir été utilisées pour les séparer des autres stimuli dans le groupe de sujets implantés cochléaires.

Les modalités de catégorisation des sons environnementaux et des objets sont plus difficiles à évaluer chez les sujets implantés dans la mesure où leurs capacités d'identification de ces stimuli auditifs sont médiocres et compte tenu de l'importante diversité de ces signaux. Plusieurs travaux menés chez le sujet normo entendant ont cherché à mettre en évidence les facteurs acoustiques spectraux et temporels nécessaires à l'identification des bruits

environnementaux. Dans la littérature, la catégorie « bruits environnementaux » peut regrouper des stimuli très variés, tels que les vocalisations animales, les bruits de scène urbaine, rurale ou domestique. Les régions fréquentielles contribuant le plus à leur identification ont été évaluées à l'aide de filtres passe-haut et passe-bas et sont situées entre 1200 et 2400 Hz [Gygi et al., 2004], de manière assez comparable à la discrimination de la parole [French and Steinberg, 1947]. Cermak et Cornillon [Cermak and Cornillon, 1976] ont étudié les modalités de catégorisation de bruits de véhicule par de sujets normo entendant et ont identifié l'intensité de l'extrait sonore comme facteur essentiel de la partition. La structure temporelle rythmique, au travers de son percussifs et répétés, a également été impliquée par Vanderveer [Vanderveer, 1980] et al qui ont montré une tendance à regrouper les sons rapidement répétés comme un bruit de marteau, de porte frappée ou de tir en rafale.

### **4.3 Perspectives**

Il existe donc vraisemblablement des indices acoustiques bas niveau supportant la séparation en catégories que notre analyse préliminaire n'a pas permis de mettre en évidence dans le groupe des sujets implantés cochléaires. Cette analyse doit être complétée car les stimuli au sein des catégories « voix » et « musique » partagent vraisemblablement des caractéristiques acoustiques communes, notamment au niveau de leur structure rythmique et certains bruits environnementaux, aussi variés soient-ils, peuvent être identifiés, donc catégorisés, par les sujets implantés cochléaires.

L'évaluation de la perception des bruits environnementaux, et plus généralement de l'analyse de scène auditive, ont longtemps été délaissées, chez le sujet implanté cochléaire comme chez le sujet normo entendant. Elles peuvent pourtant faire l'objet de recherches fondamentales psycho acoustiques et de recherches clinique en « vie réelle », nécessaires chez le sujet implanté cochléaire par des études transversales mais également longitudinales.

L'effet d'un réentraînement spécifique à l'écoute de bruits environnementaux a pu être mis en évidence, notamment par Shafiro et al. [Shafiro et al., 2012] (2012), avec des performances améliorées d'un facteur 3 à 4. La plasticité du système auditif central du sujet implanté observée pour la discrimination de la parole existe donc vraisemblablement pour la perception des bruits environnementaux et mériterait d'être explorée par neuro imagerie fonctionnelle ou en électro-physiologie. Certaines études en IRMf réalisées chez le sujet normo entendant montrent en effet que les régions temporelles postérieures sont activées en réponse à des changements de timbre. La réorganisation de ces régions chez le sujet sourd pourrait intervenir comme facteur pronostique de la perception de scènes auditives complexes, de la

même manière qu'elle le fait pour la restauration de la discrimination de la parole (Lazard, 2011).

## Chapitre IV

### Etude 2 : Perception de la prosodie affective par le sujet implanté cochléaire. Reconnaissance et catégorisation des émotions

## 1. INTRODUCTION

Cette étude a été réalisée dans le prolongement des travaux réalisés par l'équipe sur la perception de la voix et de ses attributs par les sujets implantés cochléaires, en référence au modèle hiérarchique de perception de la voix proposé par Pascal Belin (voir introduction). Notre première étude a démontré les capacités des sujets implantés cochléaires de regrouper des sons issus de la voix humaine. Dans cette catégorie de sons figuraient des signaux de parole mais également l'enregistrement d'un rire, témoin de la proximité ressentie entre les informations linguistiques et para linguistiques. Le regroupement de ces informations, par les sujets implantés comme par les sujets normo entendant, traduit l'importance de leur co occurrence et de leurs interactions dans la compréhension générale du message oral.

La manière dont les mots sont agencés et prononcés permet en effet de moduler le sens apporté par la simple succession des unités linguistiques. Ces phénomènes supra segmentaux d'intonation, d'accentuation (ou emphase) caractérisent l'information prosodique. Ils permettent au locuteur de souligner l'information considérée comme pertinente et reposent sur de multiples variations de certains indices acoustiques fondamentaux (F0, amplitude, durée).

Lors de notre description du traitement du signal acoustique par l'implant, nous avons insisté sur la qualité de la restitution de ces indices acoustiques fondamentaux, relativement satisfaisante pour l'amplitude et la durée, certainement insuffisante pour F0. Le défaut de transmission de F0 pourrait limiter la perception de la prosodie, de la même manière qu'il impacte la détection de la voix humaine [Massida et al., 2011] ou la reconnaissance de certains de ses attributs comme le genre [Kovacic and Balaban, 2009].

Notre première étude portant sur la perception de la prosodie s'intéresse plus spécifiquement à une de ses composantes, la prosodie affective (ou prosodie émotionnelle). Celle-ci a déjà fait l'objet de plusieurs travaux chez le sujet implanté cochléaire sourd pré lingual [Straatman et al., 2010, Most et al., 2012, Nakata et al., 2012] et post lingual [Fu and Galvin, 2008, Cullington and Zeng, 2011]. Le paradigme expérimental majoritairement utilisé repose sur la reconnaissance d'une émotion dans une phrase neutre, prononcée par un acteur/une actrice, parmi plusieurs émotions cibles proposées, comme la joie, la peur, la colère, ou la tristesse. La joie ou la colère peuvent être identifiées par rapport à une prosodie neutre en se basant sur un rythme d'élocution accéléré, de plus grandes variations de F0, une F0 moyenne plus haute et une intensité plus élevée. A l'inverse, la tristesse est caractérisée par un rythme ralenti, un champ de variations de F0 resserré, et une F0 globalement plus basse [Petruhin, 2000, Scherer, 2003, Yildirim et al., 2004]. Les résultats des différentes études font état d'un

déficit chez les sujets implantés avec des pourcentages de reconnaissance correcte autour de 50% [House, 1994,Pereira, 2000,Luo and Fu, 2007] pour 70 à 90% de bonnes réponses chez les sujets normo entendant. L'étude de Pereira et al suggérait que la reconnaissance reposait sur les variations d'amplitude dans la mesure où les performances des sujets implantés chutaient significativement (38% de réponses correctes) si cet indice était supprimé. Ces études permettent d'évaluer les capacités de reconnaissance d'une émotion quand celle-ci est typique mais ne rendent pas compte de leur devenir en cas d'ambiguïté.

Nous avons cherché à évaluer les capacités de reconnaissance de deux émotions caractéristiques (joie et tristesse) des sujets implantés mais également leurs capacités de catégorisation quand les stimuli correspondaient à des émotions intermédiaires entre ces deux extrêmes. Nous avons utilisé pour ce faire un paradigme similaire à celui développé pour l'étude sur la perception du genre que nous avons présentée en introduction. L'utilisation d'un continuum vocal joie-tristesse nous a en effet permis de déterminer d'une part les capacités de reconnaissance d'une émotion typique et d'évaluer d'autre part les capacités de catégorisation face à des stimuli plus ambigus.

## **2. MATERIELS ET METHODES**

### **2.1 Population**

Seize sujets porteurs d'un implant cochléaire unilatéral ont été inclus dans cette étude. Cette population comprenait 8 sujets masculins pour 8 sujets féminins, 9 implants cochléaires droits pour 7 implants cochléaires gauches. La médiane d'âge était de 66.5 ans (intervalle de confiance à 95%:57-68 ans).

L'ensemble des données démographiques, cliniques et techniques relatives à l'implant utilisé est illustré dans le tableau IV.1. Leurs performances ont été comparées à un groupe contrôle de 12 sujets normo entendant (7 femmes, 5 hommes ; âge médian : 28.5 ans avec un intervalle de confiance à 95% :27-37).

Tous les sujets ont été inclus après avoir reçu une information orale et écrite concernant l'étude, et signé un consentement écrit, ainsi que le prévoyait l'autorisation de réalisation de l'étude délivrée par le Comité de Protection des personnes du Sud-ouest et d'Outre-mer (n° : 08 161 03).

Nous avons déterminé les critères d'inclusions suivants :

- âge supérieur à 18 ans
- Acceptation du protocole et signature du consentement
- Absence de pathologie neurologique associée

Cette étude représente le prolongement de travaux réalisés par l'équipe sur la perception de la voix et de ses attributs par les sujets implantés cochléaires. Comme dans ces précédentes contributions, nous avons évalué des sujets implantés à différents délais post activation et obtenant différents niveaux de performances pour la discrimination de la parole, sans critère d'inclusion particulier vis-à-vis de ces paramètres.

Notre hypothèse était que, de la même manière que la reconnaissance de la voix parmi les bruits environnementaux ou la catégorisation du genre, la perception de l'émotion dans la voix serait déficitaire dans le groupe de sujets implantés cochléaires par rapport aux sujets normo entendant.

N° Sujet	Sous Groupe	Age (ans)	Etiologie de la surdité	Genre	Durée IC (mois)	Modèle d'implant	Processeur	Stratégie	Côté implanté
1	Bimodal	38	Ototoxicité	M	2	Cochlear MRA	Freedom	ACE	G
2	IC seul	56	Labyrinthite	F	85	Cochlear CI 24	Freedom	ACE	D
3	IC seul	73	Inconnue	M	85	Cochlear CI 24	Freedom	ACE	D
4	Bimodal	61	Otospongiose	F	12	Cochlear CI 24	Freedom	ACE	D
5	IC seul	53	Inconnue	F	24	Cochlear CI 24	Freedom	ACE	D
6	IC seul	72	Ototoxicité	M	42	Cochlear CI 24	Freedom	SPEAK	G
7	Bimodal	49	inconnue	F	24	MedEl Sonata	Opus II	CIS	D
8	IC seul	67	Inconnue	M	109	Cochlear CI 24	Freedom	Hi-Res 90	G
9	Bimodal	62	Inconnue	F	6	Advanced Bionics	Harmony	Hi-Res 90	G
10	Bimodal	76	Inconnue	M	36	Advanced Bionics	Harmony	Hi-Res 90	D
11	Bimodal	70	Otospongiose	M	10	Advanced Bionics	Harmony	Hi-Res 90	D
12	Bimodal	72	Inconnue	M	12	Advanced Bionics	Harmony	Hi-Res 90	G
13	Bimodal	66	Inconnue	F	6	Cochlear CI 24	Freedom	ACE	D
14	IC seul	85	Inconnue	F	10	MedEl Sonata	Opus II	CIS	D
15	IC seul	52	Traumatisme	M	109	Cochlear CI 22	Esprit 3G	ACE	D
16	IC seul	70	Inconnue	F	84	Cochlear CI 24	Esprit 3G	ACE	G

Tableau IV.1 : données démographiques et techniques relatives aux seize sujets implantés cochléaires inclus et à l'implant utilisé



## 2.2 Appareillage et stimuli

### *Continuum vocal joie-tristesse*

Ce continuum fait partie de la batterie Voice Perception Assessment, série de tests de perception des attributs de la voix développés par P. Belin et son équipe (Voice Neurocognition Laboratory, Univ. Glasgow) [Pernet and Belin, 2012]. Ce continuum a été généré à partir de l'enregistrement d'un acteur professionnel, chargé de prononcer une interjection répétant la voyelle /a/, sur un mode joyeux (rire) puis sur un mode triste (sanglot). A partir de ces enregistrements originaux, un morphing a été réalisé en utilisant STRAIGHT (Hideki Kawahara, Université de Wakayama, 1999) [Kawahara H, 1999] dans Matlab (the Mathworks, Inc., Natick, MA, USA). STRAIGHT décompose les stimuli vocaux en cinq paramètres distincts : la fréquence fondamentale (F0), les fréquences correspondant aux différents formants, la durée, la densité spectro-temporelle et la périodicité. Ces cinq paramètres sont manipulés afin de synthétiser des éléments acoustiques hybrides entre les deux enregistrements originaux. La synthèse du morphing aboutit ainsi à la création de deux extrémités (joie-tristesse) et de neuf intermédiaires hybrides évoluant depuis la joie jusqu'à la tristesse par paliers de 10% (90% joie-10% tristesse ; 80% joie-10% tristesse ; ..., 10% joie-90% tristesse). Les analyses des caractéristiques acoustiques de chacun de ces éléments sont représentées dans le tableau IV.2, la figure IV.1.

	F0 moyenne (Hz)	F1 moyenne (Hz)	F2 moyenne (Hz)	Pics d'intensité	Intensité moyenne (dB)
Position 1	269,783	916,34	1568,9	89	72,78
Position 2	270,18	894,1	1557,7	93	73,9
Position 3	266,5	860,84	1500,91	96	75,4
Position 4	259,82	852,88	1495,72	99	75,07
Position 5	258,93	850	1494,8	103	74,36
Position 6	271,41	832,59	1450,7	106	73,62
Position 7	268,31	804	1417,96	110	73,12
Position 8	268,7	798,92	1413,51	113	72,34
Position 9	255,85	802,33	1385,21	117	72,2
Position 10	261,54	788,5	1389,26	120	71,58
Position 11	272,4	799,5	1370,99	123	71,57

Tableau IV.2 : Caractéristiques acoustiques de chacun des onze éléments composant le continuum vocal depuis l'extrémité « joie », position 1, jusqu'à l'extrémité « tristesse », position 11.

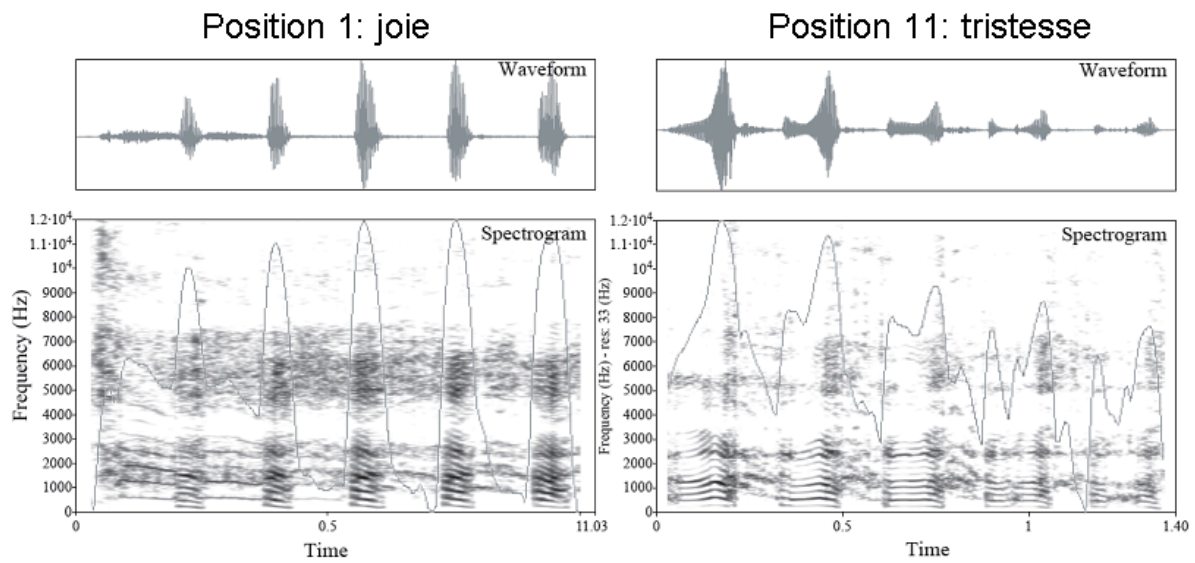


Figure IV.1 : Représentation graphique de l'onde acoustique (en haut), du spectrogramme et des pics d'intensité de l'élément occupant la position 1 (joie) et de l'élément occupant la position 11 (tristesse) dans le continuum.

### *Discrimination de la parole*

Les performances des sujets pour la discrimination des mots dissyllabiques de la liste de Fournier, présentée à 65 dB et pour les phrases dans le bruit à un rapport signal/bruit de +10 dB ont également été évaluées. Ces évaluations sont classiquement utilisées au sein des bilans orthophoniques avant et après implantation cochléaire.

### **2.3 Procédure**

Tous les sujets implantés cochléaires et normo entendant ont été évalués dans la même cabine insonorisée du service d'ORL, CHU Purpan. Les stimuli étaient présentés en champ libre, par deux haut-parleurs (KINYO, model PS-240), à une intensité mesurée de 65 dB au niveau du sujet, valeur qui correspond à l'intensité à laquelle sont évalués les sujets implantés pour la discrimination de la parole. Les sujets implantés étaient tous évalués dans leurs meilleures conditions auditives, c'est-à-dire avec les réglages utilisés quotidiennement au niveau du processeur vocal et éventuellement avec leur audioprothèse controlatérale.

La présentation des stimuli et l'enregistrement des réponses des sujets étaient assurés par le logiciel Matlab 8®. Chacun des 11 éléments du continuum était présenté 10 fois, pour un total de 110 stimuli présentés, dans un ordre pseudo-randomisé. Pour les sujets implantés comme

pour les sujets normo entendant, la tâche reposait sur une procédure de choix forcé ayant comme objectif de catégoriser le stimulus présenté entre deux choix, « joie » ou « tristesse ». Les sujets devaient ainsi répondre « joie » ou « tristesse » en utilisant la touche contrôle correspondante droite ou gauche sur un clavier d'ordinateur. La signification des touches contrôle droite et gauche était inversée d'un sujet à l'autre pour éviter les effets liés à la main dominante.

Les sujets normo entendant étaient évalués avec un délai de 1 seconde entre chaque stimulus et les sujets implantés cochléaires avec un délai de 1.5 secondes. Aucune consigne de rapidité n'était donnée. Dix secondes après la présentation du stimulus, si le sujet n'avait répondu, le stimulus suivant était présenté. L'ensemble du test durait environ 7 minutes pour les sujets normo entendant, 10 à 15 minutes pour les sujets implantés cochléaires.

## **2.4 Analyse des résultats**

Le pourcentage de réponse « tristesse » pour chaque élément du continuum, depuis l'extrémité « joie » à l'extrémité « tristesse » a été utilisé comme valeur permettant l'estimation d'un profil de réponses pour l'ensemble du continuum. Pour cette estimation, nous avons appliqué une fonction sigmoïde de Boltzmann (Levenburg-Marquart, Origin v6.1) aux pourcentages de réponse « tristesse » de chaque sujet pour l'ensemble des éléments du continuum, obtenant ainsi une courbe psychométrique de type sigmoïde (voir exemple, figure IV.2), représentant le pourcentage de réponse « tristesse » en fonction de la position de l'élément sur le continuum. Une reconnaissance normale de l'émotion se traduirait donc par 0% de réponse tristesse pour la position 1 du continuum et 100% de réponse tristesse pour la position 11. La catégorisation des éléments occupant les positions intermédiaires sur le continuum est sujet dépendante, mais pour que cette catégorisation soit considérée comme efficace, elle doit être reproductible d'une présentation à l'autre d'un même élément du continuum.

Plusieurs informations peuvent être extraites de ce type de courbe et nous renseigner ainsi sur les modalités de catégorisation de l'émotion par les sujets implantés cochléaires.

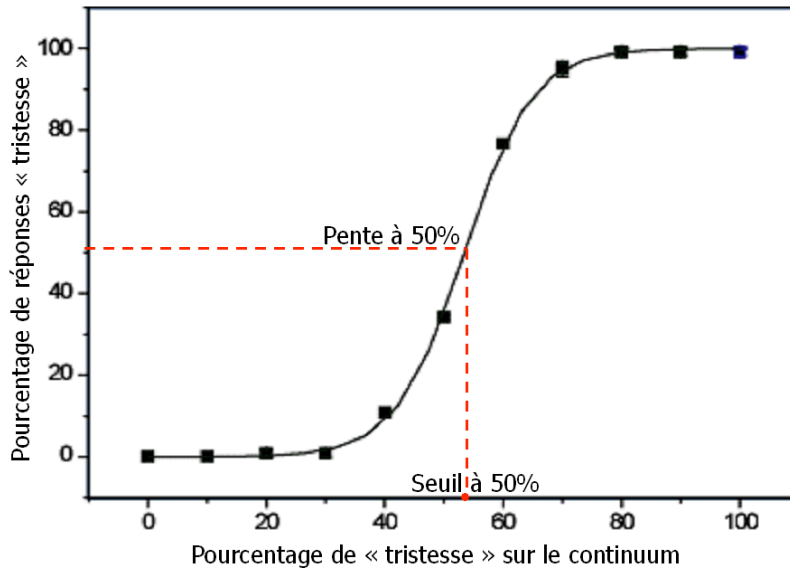


Figure IV.2. Exemple de courbe sigmoïde, représentant l'estimation par la fonction sigmoïde des réponses du groupe des sujets normo entendant et représentation des critères étudiés (pente et seuils à 50%)

Le premier élément d'information à considérer est le niveau d'inclinaison de la courbe, qui constitue l'indice essentiel des capacités de catégorisation des sujets évalués. Cette inclinaison est mesurée en calculant la pente de la fonction sigmoïde appliquée aux réponses des sujets. La pente d'une courbe sigmoïde est maximale en son centre de symétrie, qui caractérise le point d'équivalence subjective, c'est-à-dire le point sur la courbe correspondant ici à 50% de réponses « joie » et 50% de réponses « tristesse ». Une pente à valeur élevée est synonyme d'une catégorisation efficace. Un élément du continuum situé près de l'extrémité « joie » sera identifié comme tel par le sujet, et la catégorisation d'un élément donné du continuum est reproductible d'une présentation de ce stimulus à l'autre. A l'inverse, une pente de faible valeur reflète les erreurs du sujet dans la catégorisation des stimuli par rapport à leur position sur le continuum, ainsi que la variabilité intra sujet des réponses pour un élément donné du continuum. La valeur de la pente constitue donc un critère sensible permettant d'évaluer les capacités de catégorisation d'un sujet.

La deuxième information apportée par l'analyse de la courbe concerne l'éventuelle tendance des sujets à catégoriser préférentiellement les stimuli présentés dans une catégorie, ici « joie » ou « tristesse ». Cette information est extraite en mesurant la valeur retrouvée en abscisse pour le point d'équivalence subjective (50% de réponse « joie », 50% de réponses « tristesse ») correspond au seuil C50, qui correspond donc à la position dans le continuum

pour lequel les sujets répondent 50% « joie » et 50% « tristesse ». Si le seuil C50 est décalé vers l'extrémité « joie », alors les sujets ont tendance à majoritairement catégoriser les stimuli dans la catégorie « tristesse ». Si le seuil C50 est décalé vers l'extrémité « tristesse », alors les sujets ont majoritairement tendance à classer les stimuli dans la catégorie « joie ».

Le troisième critère de mesure que nous avons étudié est le pourcentage de reconnaissance correcte aux extrémités du continuum, qui reflète les capacités de reconnaissance de l'émotion quand celle-ci est typique.

L'analyse des résultats a dû prendre en compte l'effectif de chacun des groupes étudiés et la distribution non normale de leurs différents résultats (pente, seuil à 50%, pourcentage de réponse aux extrémités « joie » et « tristesse »). Nous avons ainsi réalisé un ré échantillonnage par bootstrapping permettant le calcul et la comparaison des intervalles de confiance à 95% de distribution des valeurs obtenues dans chacun des groupes. Si ces intervalles de confiance se chevauchent, alors il n'existe pas de différence entre les deux groupes. Si ces intervalles de confiance sont distincts, alors la différence entre les deux groupes peut être considérée comme significative. Les résultats du groupe de sujets implantés cochléaires et du groupe de sujets normo entendant ont ainsi été étudiés et comparés pour les trois critères cités précédemment. Nous avons ensuite réalisé des analyses complémentaires pour déterminer si les capacités de catégorisation de l'émotion étaient corrélées aux performances obtenues pour la discrimination de la parole en calculant le coefficient de corrélation  $r$  et son niveau de significativité (test de  $r$  en  $z$ , significatif si  $p < 0.05$ ). De même, l'impact éventuel de l'expérience avec l'implant a également été évalué en cherchant une corrélation entre la valeur de la pente et la durée d'expérience avec l'implant. Enfin, les performances pour la catégorisation de deux sous-groupes de sujets implantés cochléaires ont été comparées pour rechercher un éventuel impact de l'existence d'une audition résiduelle controlatérale à l'implant. Nous avons ainsi constitué un sous-groupe dit bimodal de sujets utilisant une audio prothèse controlatérale à l'implant et un sous-groupe dit IC seul de sujets n'ayant pas d'audition résiduelle controlatérale à l'implant.

### **3. RESULTATS**

Nous présentons dans un premier temps les résultats des deux groupes et leurs analyses comparatives pour les trois critères de mesure que nous avons établis (pente de la courbe, seuilC50 et pourcentage de réponses correctes aux extrémités du continuum). Les analyses complémentaires portant sur l'éventuelle relation avec le niveau de discrimination de la parole et l'impact potentiel d'une audition controlatérale à l'implant sont présentées dans la deuxième partie des résultats.

#### **3.1 Description des groupes**

Dans le groupe des sujets implantés cochléaires, la discrimination moyenne des mots dissyllabiques dans le silence était de 65.6% (intervalle de confiance à 95% :51.3-75.3) et la discrimination moyenne des phrases dans le bruit à un rapport signal/bruit de +10 dB était de 69.9% (intervalle de confiance à 95% :45.3-79.3). Dans le sous – groupe bimodal, la discrimination moyenne était de 66.25% (intervalle de confiance à 95% :55-80.6) pour les mots dissyllabiques et de 68.6% (intervalle de confiance à 95% :50.75-86) pour les phrases dans le bruit. Dans le sous – groupe IC seul, la discrimination moyenne était de 68.3% (intervalle de confiance à 95% :40.6-79.4) pour les mots dissyllabiques et de 71.7 (intervalle de confiance à 95% :24.7-86.4) pour les phrases dans le bruit. Il n'y avait pas de différence significative entre les deux sous – groupes pour ces variables.

Les délais post activation, correspondant à la durée d'expérience avec l'implant, étaient également distribués de manière hétérogène (2-109 mois, médiane : 39 mois). Il n'y avait pas de corrélation entre la durée d'expérience avec l'implant et les performances pour la discrimination de la parole ( $r = -0.13$ ,  $p = 0.64$  pour la corrélation avec la discrimination des mots dissyllabiques ;  $r = -0.12$ ,  $p = 0.64$  pour la corrélation avec la discrimination des phrases dans le bruit). Il est classiquement retrouvé dans la littérature que la discrimination de la parole s'améliore progressivement en post implantation pendant les 6 à 7 premiers mois au-delà desquels on observe un plateau des performances [Lenarz et al., 2012].

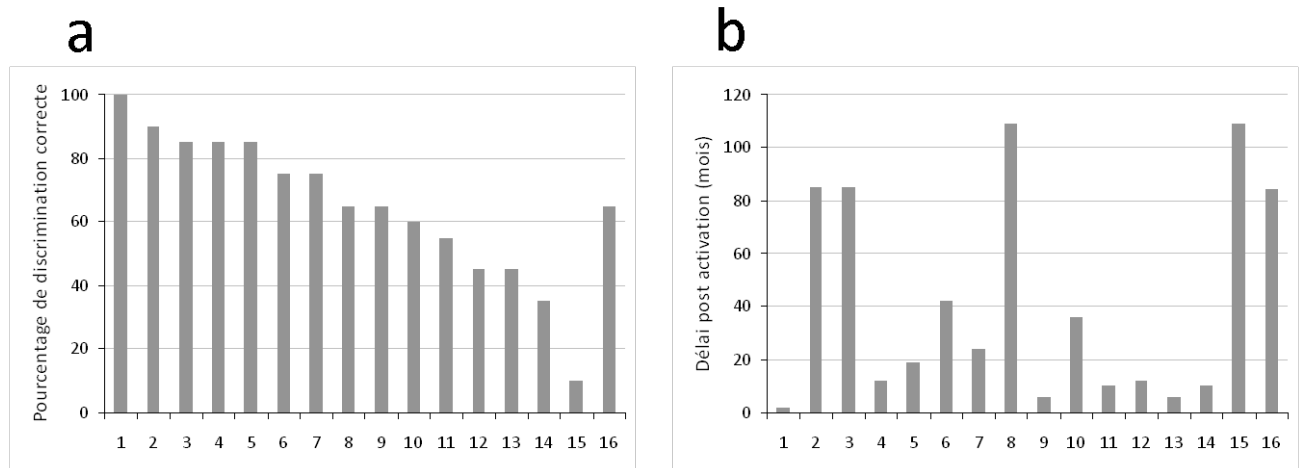


Figure IV.3 a et b. a) Pourcentage de discrimination correcte des mots dissyllabiques dans le silence par chacun des 16 sujets implantés cochléaires. b) Délais post activation de chacun des 16 sujets implantés cochléaires.

### 3.2 Analyse de la pente de la courbe

#### *Comparaison sujets implantés cochléaires vs sujets normo entendant*

L'estimation des réponses des sujets des deux groupes par une courbe sigmoïde nous permet d'extraire deux premières informations simples (voir figure IV.4b). Les réponses des sujets implantés cochléaires ont pu être estimées au moyen de la fonction sigmoïde, ce qui traduit l'existence de capacités de catégorisation de l'émotion dans ce groupe. Par ailleurs, il existe une variabilité des réponses pour chacun des éléments du continuum, prédominant pour les éléments intermédiaires ambigus, dans les deux groupes de sujets.

L'analyse comparative de la pente entre le groupe de sujets implantés cochléaires et le groupe de sujets normo entendant a mis en évidence une pente significativement plus élevée dans le groupe des sujets normo entendant. La figure IV.4a illustre la différence des valeurs de pente entre le groupe de sujets implantés (moyenne : 3.19 ; intervalle de confiance à 95% : 2.35 – 4.72) et le groupe de sujets normo entendant (moyenne : 16.8 ; intervalle de confiance à 95% : 5.74 – 43.7).

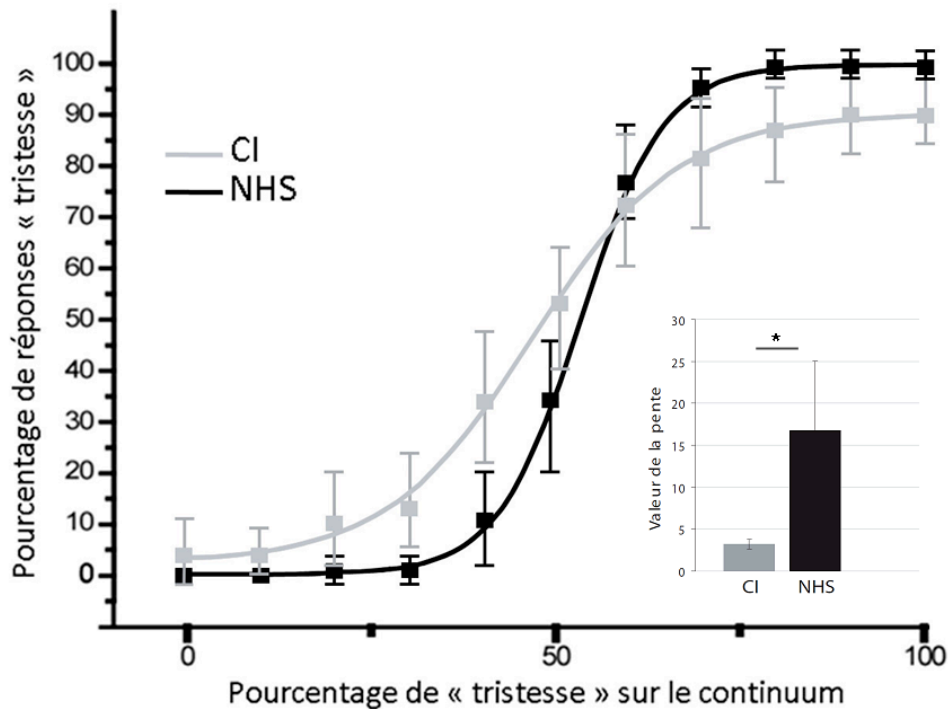


Figure IV.4. Comparaison de la courbe psychométrique sigmoïde estimée dans le groupe des sujets implantés cochléaires à la même courbe estimée dans le groupe des sujets normo entendant. La variabilité des réponses est majorée pour les éléments intermédiaires du continuum, et ce dans les deux groupes (gauche). Comparaison des valeurs moyennes de la pente et l'erreur-type dans chaque groupe. On note une variabilité des réponses plus importante dans le groupe des sujets normo entendant (droite)

Les sujets implantés cochléaires présentaient donc un déficit de catégorisation de l'émotion par rapport aux sujets normo entendant mais dans les deux groupes, les réponses étaient marquées par une forte variabilité inter - sujet.

***Relation avec la discrimination de la parole. Impact de l'expérience avec l'implant***

Nos analyses se sont ensuite portées sur l'existence d'une éventuelle relation entre les capacités de catégorisation et les performances en terme de discrimination de la parole dans le groupe des sujets implantés cochléaires. Cette analyse a été réalisée après exclusion de 2 sujets implantés cochléaires (sujet n°14 HE.J, sujet et sujet n°7 GO.L) dont les valeurs de la pente étaient mesurées plus de deux écart-types au-dessus de la moyenne présentée dans le groupe. Le coefficient de corrélation  $r$  calculé était égal à  $-0.135$  ( $p=0.62$ ) pour l'analyse de la relation entre la valeur de la pente et la discrimination des mots dissyllabiques et égal à  $0.1$



( $p=0.74$ ) pour l'analyse de la relation entre la valeur de la pente et la discrimination des phrases dans le bruit (rapport signal/bruit +10 dB). Que cette recherche concerne la discrimination des mots dissyllabiques dans le silence ou la discrimination des phrases dans le bruit, elle n'a donc pas permis de retrouver de relation avec les capacités de catégorisation de l'émotion (voir figure IV.5a). Il n'existait donc pas d'effet auditif global caractérisant des sujets implantés obtenant de bons résultats et des sujets avec des résultats plus médiocres.

La distribution des délais post implantation nous également permis de rechercher un éventuel impact de la durée d'expérience avec l'implant. Mais, comme pour le niveau de discrimination de la parole, nous n'avons pas mis en évidence de relation significative entre ces deux variables ( $r = -0.2$ ,  $p=0.5$ ) (voir figure IV.5b).

Les capacités de catégorisation de l'émotion étaient donc indépendantes des capacités de discrimination de la parole et ne tendaient pas à s'améliorer malgré l'expérience acquise avec l'implant.

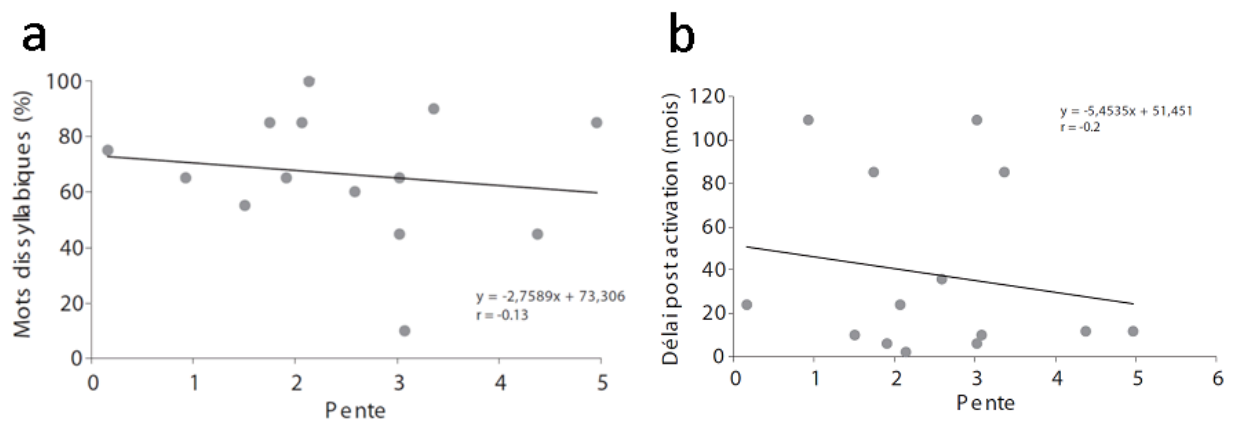


Figure IV.5a et b. a) Étude de la relation entre la discrimination des mots dissyllabiques et la pente de la courbe sigmoïde. b) Étude de la relation entre le délai post activation et la pente de la courbe sigmoïde.

Les informations complémentaires que l'audition résiduelle controlatérale à l'implant est susceptible de véhiculer (F0, premier formant, structure temporelle fine aux basses fréquences) peuvent améliorer les performances globales des sujets implantés. Nous avons donc étudié et comparé les performances de la population de sujets implantés divisée en deux sous-groupes : un groupe avec audition résiduelle controlatérale à l'implant dit « bimodal », un groupe sans audition résiduelle controlatérale dit « implant seul ».

Chacun de deux sous – groupes de sujets implantés comptait 8 sujets. Dans le groupe bimodal, la valeur moyenne de la pente était de 3.77 (intervalle de confiance à 95% : 2.54 - 6.39) et de 2.6 (intervalle de confiance à 95% : 1.59 - 4.15) dans le groupe implant seul. Il n’existait donc pas de différence significative entre les deux sous – groupes de sujets implantés, mais la catégorisation tendait à être meilleure chez les sujets avec une audition résiduelle controlatérale à l’implant (voir figure IV.6).

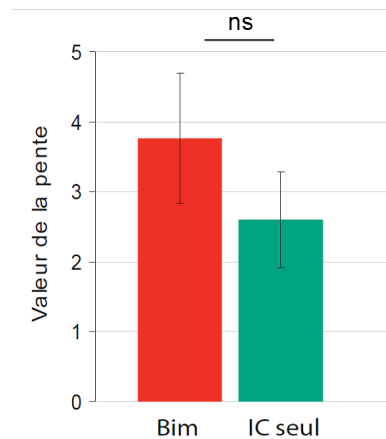


Figure IV.6. Comparaison des pentes moyennes dans le sous – groupe bimodal et dans le sous – groupe IC seul (absence de différence significative).

### 3.3 Analyse du seuil C50

Le seuil C50 représente la position sur le continuum pour laquelle les sujets catégorisent les stimuli à 50% comme de la joie et à 50% comme de la tristesse. C’est donc l’hypothèse d’une différence qualitative dans les modalités de catégorisation entre les deux groupes de sujets que nous évaluée.

Le seuil C50 moyen était de 52.68% (intervalle de confiance à 95% : 46.7-61.5) dans le groupe des sujets implantés cochléaires et de 51.86 (intervalle de confiance à 95% : 48.3-55) dans le groupe des sujets normo entendant. Il n’existait pas de différence significative entre les deux groupes, qui « situaient » donc la limite entre joie et tristesse dans la même région du continuum, sans biais particulier en direction de l’une ou de l’autre.

Comme pour la pente de la courbe, la recherche d’une relation entre la valeur du seuil C50 et la discrimination de la parole ou la durée d’expérience avec l’implant n’a pas mis en évidence de corrélation significative. Il n’y avait donc pas d’impact de l’apprentissage avec l’implant ni d’effet auditif global, Dans le premier cas, le coefficient de corrélation  $r$  était de -0.097

( $p=0.72$ ). Pour la durée d'expérience avec l'implant, le coefficient de corrélation  $r$  était de  $-0.13$  ( $p=0.63$ ).

L'éventuelle influence de l'audition résiduelle sur la valeur du seuil C50 a également été testée par une analyse en sous – groupes comparant le sous – groupe bimodal au sous – groupe implant seul. Le seuil C50 n'était pas différent d'un sous – groupe à l'autre à la lumière de comparaison des intervalles de confiance à 95% après bootstrapping (intervalle de confiance à 95% : 46%-60.4% pour le sous – groupe bimodal versus 43.8%-68.9% dans le sous – groupe implant seul). Si le seuil C50 n'était pas significativement modifié en présence d'une audition résiduelle, la variabilité de sa position sur le continuum était réduite dans le sous – groupe bimodal.

En synthèse, l'analyse des valeurs du seuil C50 n'a pas révélé de différence significative entre le groupe de sujets implantés cochléaires et le groupe des sujets normo entendant. Ces deux groupes ne présentaient une catégorisation biaisée en direction d'une émotion particulière. Au sein du groupe des sujets implantés cochléaires, l'existence d'une éventuelle audition résiduelle n'avait pas d'impact sur sa valeur.

### **3.4 Analyse des réponses aux extrémités « joie » et « tristesse » du continuum**

Nous nous sommes ensuite intéressés à la reconnaissance d'émotions typiques sur le plan acoustique, correspondant aux deux éléments situés chacun à une extrémité du continuum (extrémité « joie » ou extrémité « tristesse »). Le pourcentage moyen de réponses « tristesse » des sujets du groupe implant cochléaire a donc été comparé à celui des sujets du groupe contrôle normo entendant, pour chacun de ces deux éléments (voir figure IV.7a).

Pour l'élément situé à l'extrémité « joie », il n'y avait pas de différence entre les réponses des sujets implantés cochléaires (moyenne : 3.75% ; intervalle de confiance à 95% : 0-17.3) et des sujets normo entendant (0% de réponses « tristesse » pour tous les sujets). Sur l'ensemble du groupe implantés cochléaires, il n'y avait donc pas de déficit de reconnaissance de la joie par rapport aux sujets normo entendant.

Pour l'élément situé à l'extrémité « tristesse », on retrouvait une différence significative entre le pourcentage des réponses des sujets implantés cochléaires (moyenne : 90.6% ; intervalle de confiance à 95% : 81.25-95) et les réponses des sujets normo entendant (99.2% ; intervalle de

confiance à 95% : 95.8-100). Les sujets implantés cochléaires présentaient donc un déficit de reconnaissance de la tristesse par rapport aux sujets normo entendant.

Nous avons ensuite réalisé une analyse en sous – groupes au sein du groupe des patients implantés cochléaires (voir figure IV.7b) afin de rechercher un éventuel bénéfice lié à l’audition résiduelle. Nous n’avons pas retrouvé de différence significative entre le sous – groupe bimodal et le sous – groupe IC seul pour la reconnaissance de la joie (intervalles de confiance à 95% : 95-100 pour le sous – groupe bimodal versus 75-100 pour le sous – groupe IC seul) ni pour la reconnaissance de la tristesse (intervalles de confiance à 95% :88.75-98.75 pour le sous – groupe bimodal versus 68.75-93.75 pour le sous – groupe IC seul). De la même manière que pour la valeur de la pente, les performances tendaient à être meilleures chez les sujets présentant une audition résiduelle.

Il existait toute fois une tendance à obtenir de meilleures performances dans le sous- groupe bimodal.

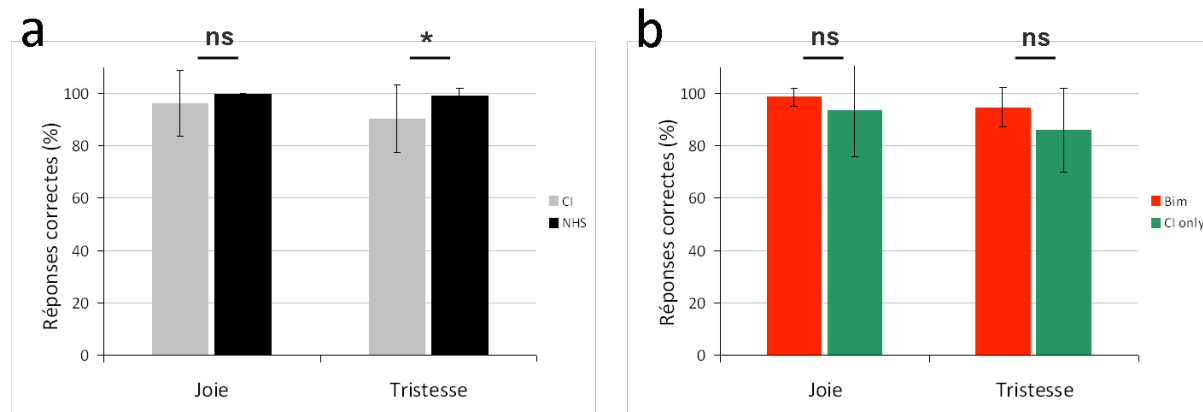


Figure IV.7a et b. a) Pourcentage de réponses correctes pour l’extrémité joie et l’extrémité tristesse du continuum dans le groupe des sujets implantés cochléaires et le groupe des sujets normo entendant. Il existe une différence significative entre les deux groupes pour la reconnaissance de la tristesse. b) Pourcentage de réponses correctes pour l’extrémité joie et l’extrémité tristesse du continuum dans le sous – groupe bimodal et le sous – groupe IC seul. Il n’y a pas de différence significative entre ces deux sous – groupes.

Les sujets implantés présentaient donc des scores élevés de reconnaissance de deux émotions typiques (joie et tristesse), avec un léger déficit de reconnaissance de la tristesse. Mais leur déficit se manifestait principalement en réponse aux éléments intermédiaires du continuum avec une valeur moyenne de la pente de la courbe significativement altérée par rapport aux sujets normo entendant.

## **4. DISCUSSION**

### **4.1 Déficit de la catégorisation des émotions chez le sujet implanté cochléaire**

Les précédentes études réalisées par notre équipe sur la détection de la voix et la catégorisation de son genre avaient mis en évidence un déficit important, durable et dissocié de la restauration de la discrimination de la parole chez les sujets implantés cochléaires. Leurs performances pour la perception de la prosodie affective apparaissent dans la continuité de ces premiers résultats avec un important déficit pour la catégorisation, leur valeur de pente étant inférieure de plus de la moitié par rapport à celle des normo entendant.

En revanche, la reconnaissance des émotions caractéristiques, évaluée aux extrémités « joie » et « tristesse », était moins déficitaire que ce que la littérature le laissait attendre. Avec plus de 90% de reconnaissance correcte et malgré le léger déficit présenté pour la tristesse, ces performances excèdent largement les 50% précédemment rapportés [House, 1994, Pereira, 2000, Luo and Fu, 2007]. Ceci tient essentiellement au type de tâche proposé dans les études citées, cherchant à évaluer la reconnaissance parmi plusieurs émotions cibles, divisant par autant de cibles la probabilité de réponse correcte. Par ailleurs, la joie et la tristesse sont plus facilement reconnues que d'autres émotions comme la peur, ou la colère [Johnson et al., 1986, Murray and Arnott, 1993].

L'utilisation du continuum vocal nous a donc permis de distinguer un important déficit de la catégorisation de l'émotion d'un déficit plus relatif de la reconnaissance, pour deux émotions indiquées. Le continuum rend en revanche plus difficile la détermination des indices acoustiques sur lesquels les sujets se sont basés pour réaliser la catégorisation.

### **4.2 Bases acoustiques de la perception des émotions**

Dans ce paragraphe, nous évoquons les bases plutôt que les corrélats acoustiques. Il serait en effet caricatural de rapprocher strictement une émotion d'un ou plusieurs indices acoustiques fondamentaux tant les paramètres acoustiques (F0, rythme de l'énoncé oral, jitter, shimmer, intensité) mesurés et la nature des stimuli (phrases, mots, interjections mono ou polysyllabiques) peuvent être variables.

Dans notre étude, plusieurs indices varient progressivement au sein du continuum vocal, ainsi que l'a illustré le tableau IV.2, depuis la position 1 du continuum (extrémité « joie ») jusqu'à

la position 11 (extrémité « tristesse »). Sur le plan spectral, les deux premiers formants diminuent en effet progressivement (F1 passant de 916 Hz à 799 Hz ; F2 de 1568 Hz à 1370 Hz). Dans le même temps, le nombre de pics d'intensité augmente de 89 à 123, caractérisant des changements progressifs dans le rythme des différents éléments du continuum. Si l'on tient compte des capacités connues des sujets implantés cochléaires, il apparaît vraisemblable qu'ils se soient appuyés sur le nombre de pics d'intensité, et donc la structure rythmique du stimulus pour le catégoriser. Les sujets implantés peuvent en effet percevoir sans difficulté les modulations d'enveloppe inférieures à 50 Hz, comme c'est le cas pour la parole ou pour l'interjection utilisée dans notre étude. Mais il est également possible que les variations mesurées pour le premier formant aient pu être utilisées par les sujets implantés pour la reconnaissance de l'émotion correspondant à chaque extrémité. Fitzgerald et al. [Fitzgerald et al., 2007] ont en effet montré que le seuil de discrimination de F1 était compris entre 50 et 100 Hz dans un test d'identification de voyelles chez des sujets implantés porteurs d'un implant Nucleus®. De même, les variations de F2 ont pu être perçues puisque les mêmes auteurs ont rapporté une discrimination possible de ses variations quand elles excèdent 10% de sa valeur. Il est en revanche exclu que les sujets se soient appuyés sur F0, qui n'est pas significativement modifié entre les différentes positions du continuum.

Le fait que F0 n'ait pas été discriminant trouve son explication dans le type de stimuli utilisés. Chaque élément du continuum correspondait à une interjection (phonème « a » répété sous la forme d'un rire ou d'un sanglot), ne comportant aucune information lexicale. Les variations de F0, perçues sous la forme d'intonation, se retrouvent essentiellement dans des phrases ou des mots. Les contributions relatives des différents indices acoustiques pour la reconnaissance des émotions rapportées par la littérature varient d'ailleurs en fonction du matériel lexical de test utilisé (phrases, mots, interjections) et interagissent fortement avec lui. Chez les sujets normo entendant, la reconnaissance de l'émotion repose autant sur les phénomènes prosodiques que sur l'information lexicale quand celle-ci est suffisante et congruente, comme dans le cas de phrases ou d'interjections connotées [Dietrich et al., 2006]. Si l'on considère ce matériel de test, certains corrélats acoustiques des principales émotions exprimées sont régulièrement retrouvés dans la littérature.

F0, le nombre et l'importance de ses variations sont souvent, et depuis longtemps, identifiés comme les principaux facteurs modulant l'émotion dans la phrase. D'autres facteurs acoustiques, tels que le rythme de parole, l'intensité sont également impliqués dans la transmission d'une émotion donnée. Les principales modifications des paramètres mesurés dans la littérature jusqu'aux années 90 sont illustrées dans le tableau IV.3.

	Anger	Happiness	Sadness	Fear	Disgust
Speech rate	slightly faster	faster or slower	slightly slower	much faster	very much slower
Pitch average	very much higher	much higher	slightly lower	very much higher	very much lower
Pitch range	much wider	much wider	slightly narrower	much wider	slightly wider
Intensity	higher	higher	lower	normal	lower
Voice quality	breathy, chest tone	breathy, blaring	resonant	irregular voicing	grumbled, chest tone
Pitch changes	abrupt, on stressed syllables	smooth, upward inflections	downward inflections	normal	wide, downward terminal inflections
Articulation	tense	normal	slurring	precise	normal

Tableau IV.3. Principales modifications de plusieurs paramètres acoustiques en fonction de l'émotion véhiculée. Ce récapitulatif insiste sur les modifications de F0, dans sa valeur moyenne, l'importance quantitative et qualitative de ses variations. Tiré de Murray et Arnott (1993).

### **4.3 La réorganisation cérébrale secondaire à la surdité peut-elle limiter la perception de la prosodie affective chez le sujet implanté cochléaire ?**

A notre connaissance, il n'existe actuellement aucune étude s'intéressant aux corrélats neuronaux de la perception de la prosodie affective chez le sujet implanté cochléaire. Nous ne pouvons donc qu'émettre des hypothèses à partir des données connues sur la réorganisation cérébrale chez le sujet sourd ou à partir d'études réalisées chez le sujet normo entendant en condition auditive dégradée par vocoder.

En ce qui concerne le deuxième point, Agrawal et al. [Agrawal et al., 2012] ont réalisé une étude électro physiologique par potentiels évoqués auditifs en réponse à trois émotions (joie, tristesse et colère) après que les stimuli aient été vocodés. Une onde positive significative P200, témoin de la perception émotionnelle [Kotz and Paulmann, 2007], était identifiée en

réponse à la joie par rapport à une stimulation neutre (voir figure IV.8). Il est admis que l'onde P200 est obtenue en réponse à des stimuli comportant une émotion par l'intégration de différents indices acoustiques, tels que F0 [Alain et al., 1997] l'intensité [Picton et al., 1970] ou du timbre [Shahin et al., 2003]. Les auteurs concluaient donc à la possibilité pour les sujets implantés de percevoir, dans une certaine mesure, l'émotion présente dans le signal acoustique.

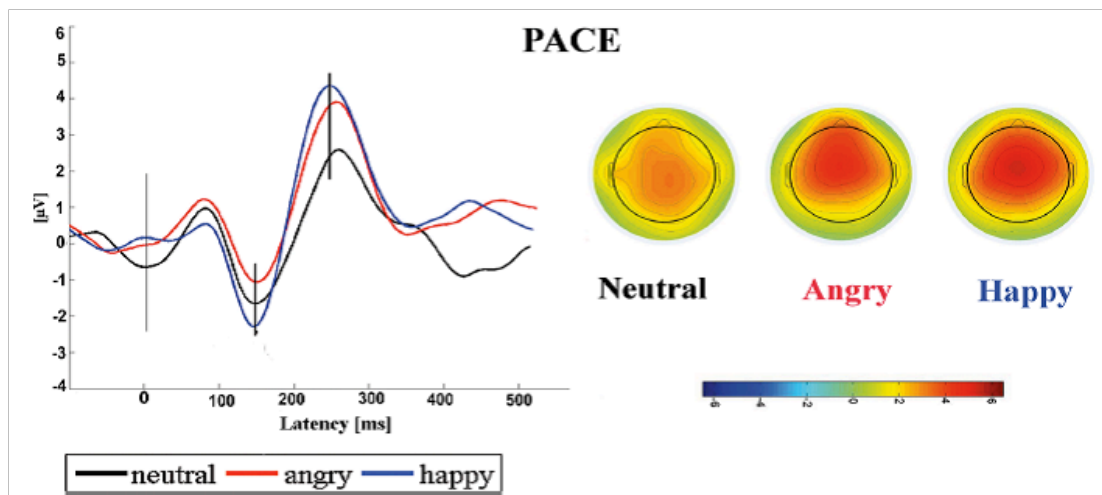


Figure IV.8. En condition de simulation d'implant par vocoder, une réponse neuronale auditive de latence tardive à un stimulus « joie » peut être enregistrée. Tiré de Agrawal et al. (2012).

Cette conclusion pourrait être nuancée à l'examen des régions auditives impliquées dans le traitement de certains de ces indices. Ces régions sont en effet précisément parmi les plus sensibles à la réorganisation (voir introduction) secondaire à la surdité et à la recherche d'optimisation du traitement de la parole. Les études en neuro imagerie fonctionnelle réalisées chez le sujet normo entendant ont ainsi démontré que le traitement neuronal de l'émotion dans la voix était principalement réalisé, en dehors de l'amygdale, dans l'hémisphère droit, par des réseaux neuronaux localisés dans la partie postérieure du STS et du gyrus temporal moyen droits ainsi qu'au niveau du planum temporale. Le recrutement de ces régions pour l'intégration audiovisuelle réalisée à visée phonologique pourrait rendre compte d'une partie des déficits présentés par les sujets implantés cochléaires dans l'identification auditive des émotions.

Le déficit significatif et durable de catégorisation de l'émotion observé chez les sujets implantés cochléaires apparaît en cohérence avec les déficits précédemment observés pour la détection de la voix humaine et la catégorisation du genre de la voix. Certains de nos résultats laissaient toutefois penser que certains patients implantés pouvaient bénéficier des



informations apportées par l'audition résiduelle controlatérale à l'implant au sein d'une stimulation bimodale, à la fois électrique par l'implant et acoustique par l'oreille controlatérale. Cet avantage potentiel a été pris en compte dans nos études suivantes sur la perception de la prosodie linguistique et la perception de la musique. Nous avons ainsi systématiquement émis l'hypothèse que ces sujets présenteraient de meilleures performances grâce aux informations acoustiques apportées par leurs restes auditifs.

## Chapitre V

### Etude 3 : Perception de la prosodie linguistique par le sujet implanté cochléaire.

Intérêts d'une stimulation bimodale

## **1. INTRODUCTION**

Notre étude précédente a mis en évidence un déficit présenté par les sujets implantés cochléaires pour la composante affective de la prosodie, se manifestant dans la catégorisation de l'émotion. Un sous –groupe de sujets, présentant une audition résiduelle associée à l'implant, tendait à obtenir de meilleures performances, probablement grâce à la transmission d'indices acoustiques sur les basses fréquences, permettant de transmettre l'intonation, via les variations de F0, et une partie du timbre.

Notre étude sur la prosodie linguistique a donc cherché à évaluer deux éléments essentiels : la contribution relative de différents indices acoustiques fondamentaux et l'impact éventuel d'une audition résiduelle complémentaire, combinée à la stimulation par implant. Nous avons étudié trois groupes de sujets implantés, distingués selon la présence/l'absence d'audition résiduelle. Cette étude a fait l'objet d'un article actuellement en révision à *Ear & hearing*.

## **2. MATERIELS ET METHODES**

### **2.1 Population**

Vingt et un sujets porteurs d'un implant cochléaire unilatéral ont été inclus dans cette étude. Quatre présentaient une audition résiduelle bilatérale et ont été inclus dans un groupe appelé Electric Acoustic Stimulation (EAS). Sept sujets implantés présentaient une audition résiduelle controlatérale à l'implant et ont été inclus dans un groupe appelé bimodal (Bim). Les dix autres sujets implantés ne présentaient de résidus auditifs acoustiques et ne bénéficiaient que d'une stimulation électrique par l'implant. Ces sujets ont intégré un groupe appelé implant cochléaire seul (CI only). Les caractéristiques des sujets implantés et les éléments techniques concernant l'implant utilisé sont reportés dans le tableau V.1. La cause de la surdité était inconnue pour 15 sujets, la pathologie se manifestant sous la forme d'une surdité neurosensorielle bilatérale progressive. L'expérience avec l'implant, déterminée par le délai post activation, était variable d'un sujet à l'autre, ce qui nous a permis d'étudier son impact éventuel sur la perception de la prosodie. Les seuils auditifs mesurés en audiométrie tonale oreilles nues sont représentés dans le tableau V.2 pour les groupes EAS et Bim. Certains de ces sujets (sujets du groupe EAS, et 2 sujets du groupe Bim) présentaient des seuils auditifs assez peu altérés sur les basses fréquences, et avaient donc théoriquement un meilleur accès aux informations liées à F0 et ses variations. Ces sujets, atteints d'une surdité classée sévère, mais en réalité asymétrique et touchant principalement les fréquences aiguës.

Pour les sujets du groupe EAS, une préservation de l'audition résiduelle a été assurée au moyen d'une chirurgie atraumatique [James et al., 2005, Fraysse et al., 2006], et ces sujets bénéficiaient donc d'une audition résiduelle bilatérale associée à l'implant.

L'objectif de notre étude était d'étudier la contribution relative de différents indices acoustiques pour la perception de la prosodie linguistique en fonction de la présence ou de l'absence d'audition résiduelle. Dans cette étude, pour focaliser nos analyses sur les indices qui sous-tendent les capacités perceptives de la prosodie et l'impact éventuel d'une audition résiduelle, les groupes de sujets implantés devaient être comparables pour les capacités de discrimination de la parole. Nous avons ainsi sélectionné des sujets présentant une discrimination de la parole satisfaisante (discrimination des mots dissyllabiques dans le silence > 60%). Pour éviter une sous-estimation de leurs performances en rapport avec une configuration auditive inhabituelle, tous les sujets implantés cochléaires ont été évalués dans les conditions auditives optimales, c'est-à-dire en utilisant les réglages habituels de leur processeur, et éventuellement une audioprothèse controlatérale ou bilatérale pour les sujets présentant une audition résiduelle.

Sujet	Groupe	Age (années)	Cause de la surdit�	Genre	D�lai post Activation (mois)	Mod�le d'implant	Processeur	Strat�gie de codage	C�t� implant�
1	EAS	43	Inconnue	F	68	<b>Cochlear Contour</b>	Freedom	ACE	droit
2	EAS	51	Inconnue	F	60	<b>Cochlear Hybrid L</b>	Freedom	ACE	gauche
3	EAS	47	Inconnue	F	51	<b>Cochlear Hybrid L</b>	Freedom	ACE	droit
4	EAS	70	Inconnue	M	7	<b>Cochlear MRA</b>	Freedom	ACE	gauche
5	Bimodal	68	Inconnue	M	60	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	gauche
6	Bimodal	63	Otospongiose	F	24	<b>Cochlear CI 24</b>	Freedom	ACE	droit
7	Bimodal	62	Inconnue	M	7	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
8	Bimodal	42	Inconnue	M	18	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
9	Bimodal	43	Inconnue	F	50	<b>Cochlear CI 24</b>	Freedom	ACE	droit
10	Bimodal	21	Inconnue	F	10	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
11	Bimodal	68	Inconnue	M	13	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	gauche
12	CI only	76	Traumatique	M	15	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	droit
13	CI only	54	Inconnue	M	44	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	gauche
14	CI only	73	Meniere	F	7	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	droit
15	CI only	76	Meniere	M	26	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	droit
16	CI only	44	Inconnue	M	60	<b>Cochlear CI 24</b>	Freedom	ACE	droit
17	CI only	70	Inconnue	F	84	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
18	CI only	69	Otospongiose	M	36	<b>Cochlear CI 24</b>	Freedom	ACE	droit
19	CI only	66	G�n�tique	F	84	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
20	CI only	69	Inconnue	M	55	<b>Cochlear CI 24</b>	Freedom	ACE	gauche
21	CI only	72	Inconnue	M	43	<b>Advanced Bionics</b>	Harmony	Hi-Res 90	droit

Tableau V.1 : donn es d mographiques et techniques relatives aux 21 sujets implant s cochl aires inclus et   l'implant utilis 

Sujet	125 Hz	250 Hz	500 Hz	750 Hz	1 kHz	2 kHz	4 kHz	8 kHz
eas1	20	30	40	90	90	95	100	100
eas2	40	50	65	75	85	x	x	x
eas3	10	30	55	55	100	110	x	x
eas4	45	50	60	65	80	90	110	105
Bim1	na	50	40	45	35	45	x	x
Bim2	70	75	70	na	65	75	80	x
Bim3	65	85	85	90	90	80	80	100
Bim4	65	90	90	95	95	95	95	x
Bim5	25	25	45	na	110	105	x	x
Bim6	45	80	95	na	100	100	115	x
Bim7	65	85	85	85	85	85	110	x

Tableau V.2 : seuils auditifs mesurés en audiométrie tonale oreilles nues, pour la meilleure oreille, chez les sujets implantés inclus dans les groupes de sujets avec audition résiduelle (groupes EAS et Bim). Na : non renseigné. X : absence de réponse du sujet.

Quinze sujets normo entendant ont été inclus dans un groupe dit contrôle (NHS) permettant de définir le profil de performance normal dans chacune des conditions auditives testées. Chez ces quinze sujets, il n’y avait aucun antécédent de pathologie otologique ou neurologique.

Tous les sujets ont été inclus après avoir reçu une information orale et écrite concernant l’étude, et signé un consentement écrit, ainsi que le prévoyait l’autorisation de réalisation de l’étude délivrée par le Comité de Protection des personnes du Sud-ouest et d’Outre-mer (n° : 08 161 03).

## 2.2 Appareillage et stimuli

### *Evaluation de la perception de la prosodie linguistique*

Pour évaluer les capacités de perception de la prosodie dans chacun des quatre groupes de sujets (EAS pour Electric Acoustic Stimulation, Bim pour bimodal, CI only pour implant cochléaire seul et NHS pour le groupe contrôle normo entendant), nous avons utilisé un paradigme souvent utilisé dans les études portant sur la perception de l’intonation dans une phrase. Il s’agit d’une tâche simple de discrimination entre question et affirmation : une phrase, neutre sur le plan syntactique, est présentée au sujet qui doit déterminer s’il reconnaît une phrase interrogative ou affirmative.

Vingt quatre phrases ont été extraites à partir d’un enregistrement de trente phrases, réalisé auprès de deux acteurs professionnels (un homme, une femme), qui devaient répéter la phrase

une fois sur un mode interrogatif, une fois sur un mode affirmatif. La liste des phrases a été établie d'après l'étude de Patel réalisée auprès de sujets amusiques. Trois sous – tests ont été réalisés, permettant d'évaluer la contribution relative des différents indices acoustiques (F0, amplitude, durée).

Dans la première condition, les enregistrements originaux étaient présentés aux sujets et les performances mesurées étaient celles présentées par les patients en « vie réelle » (voir figure V.1).

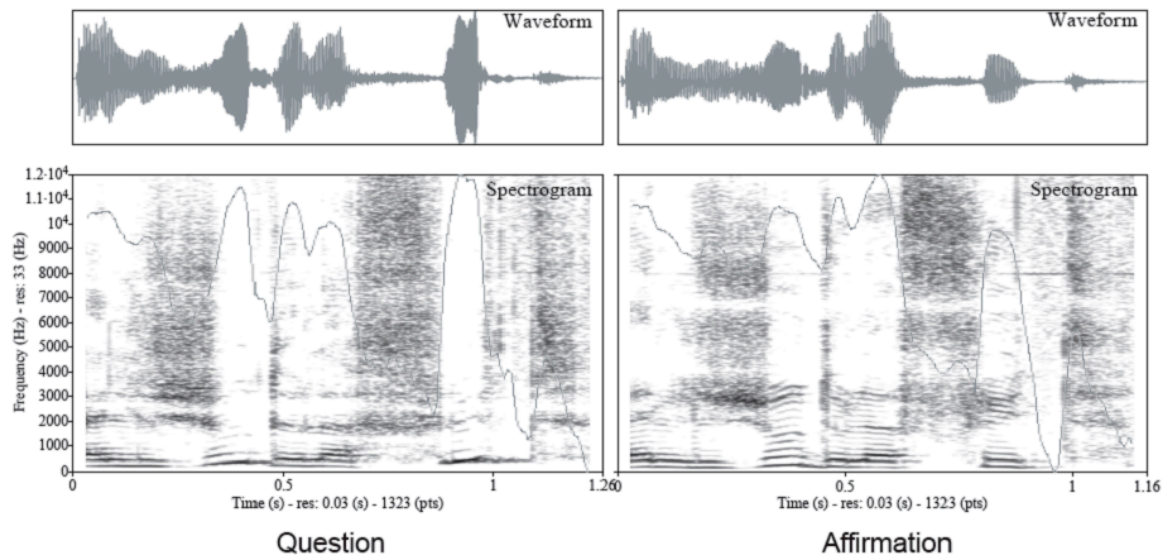


Figure V.1 : Représentation graphique de l'onde acoustique (en haut), du spectrogramme et des valeurs d'intensité (en bas) pour la forme interrogative (gauche) et la forme affirmative (droite) de la phrase suivante: « Il joue de la flûte », en condition originale. Les principales différences acoustiques dans l'intensité et le caractère ascendant (question) ou descendant (affirmation) de F0 et des formants pour la dernière syllabe.

Dans la deuxième condition, l'enregistrement original a été manipulé à l'aide du logiciel PRAAT® [Boersma and D, 2009] afin que la fréquence fondamentale du locuteur ou de la locutrice soit constante tout au long de la phrase (voir figure V.2). L'évaluation des performances dans cette condition permettait de mesurer les performances des sujets implantés cochléaires et de sujets normo entendant lorsque l'indice essentiel qui sous – tend la prosodie linguistique, i.e les variations de F0 dans la phrase, est supprimé. Pour obtenir de bonnes performances dans cette condition, les sujets devaient donc s'appuyer sur d'autres informations présentes dans le stimulus, comme les variations d'intensité ou le rythme correspondant donc aux indices d'amplitude et de durée.

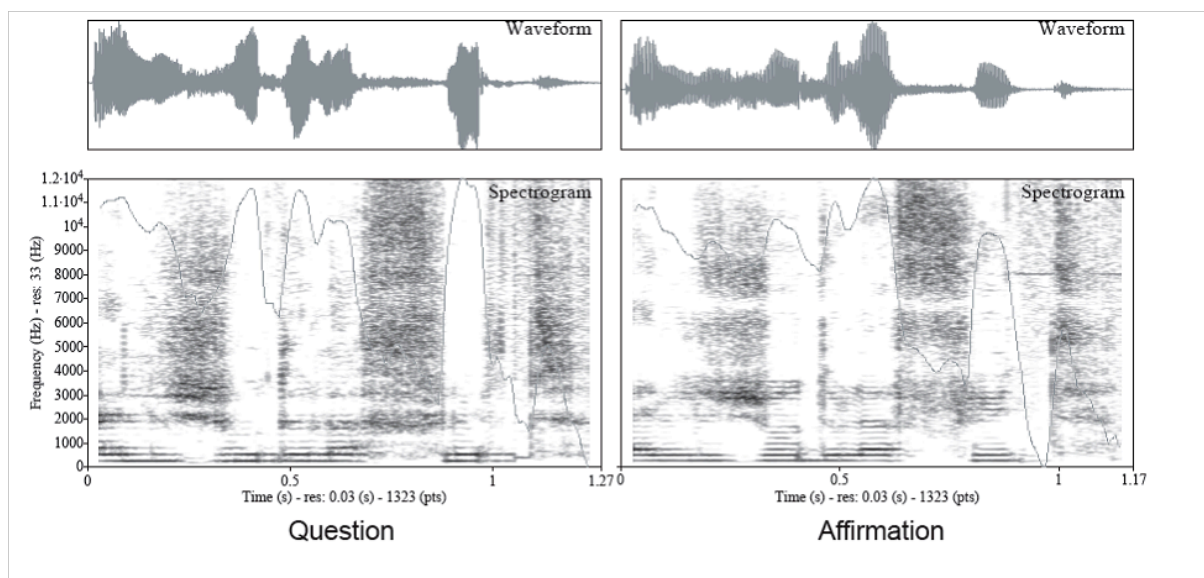


Figure V.2 : Représentation graphique de l'onde acoustique (en haut), du spectrogramme et des valeurs d'intensité (en bas) pour la forme interrogative (gauche) et la forme affirmative (droite) de la phrase suivante: « Il joue de la flûte », après manipulation de F0. F0 et les formants apparaissent constants sur le spectrogramme. Les principales différences acoustiques entre les deux formes résident dans les variations d'intensité et la durée de la dernière syllabe.

Dans la troisième condition, nous avons également utilisé des stimuli manipulés sur le plan acoustique, de la même manière que Patel et al. [Patel et al., 1998] (voir figure V.3). Ainsi, les deux versions d'une même phrase, interrogative et affirmative, sont similaires sur le plan acoustique, jusqu'à la dernière syllabe. L'amplitude et la durée de cette dernière syllabe sont égalisées d'une version à l'autre, si bien que la variation de F0 demeure le seul indice existant sur lequel les sujets peuvent s'appuyer. Dans cette condition, si les sujets utilisent les indices d'amplitude et de durée, leurs performances diminueront par rapport à celles obtenues pour l'enregistrement original. En revanche, si les sujets s'appuient essentiellement sur les variations de F0, la manipulation acoustique n'aura pas d'impact significatif sur leurs performances.



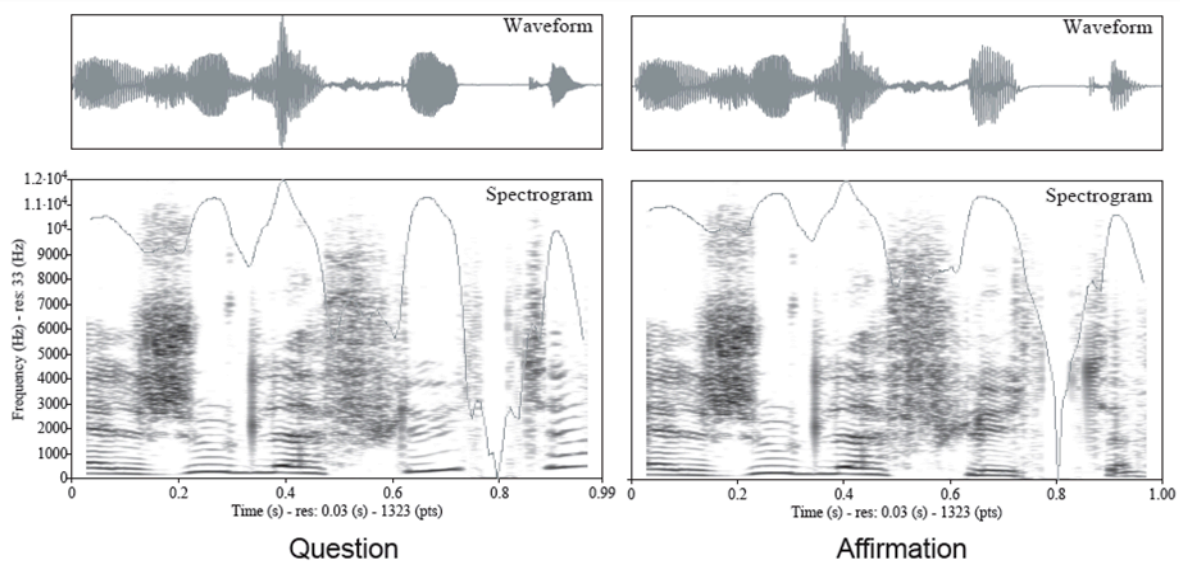


Figure V.3 : Représentation graphique de l'onde acoustique (en haut), du spectrogramme et des valeurs d'intensité (en bas) pour la forme interrogative (gauche) et la forme affirmative (droite) de la phrase suivante: « Il joue de la flûte », après manipulation de l'intensité et de la durée de la dernière syllabe. Ces paramètres ont été égalisés entre les deux formes interrogative et affirmative. La principale différence acoustique réside dans les variations de F0 et des formants au niveau de la dernière syllabe.

Pour chacun des trois sous – tests, les 24 phrases étaient présentées de manière randomisée (12 phrases prononcées par l'acteur, 12 phrases prononcées par l'actrice). Toutes les phrases étaient présentées écrites sur l'écran de l'ordinateur avant d'être présentées sur le plan auditif. Cette précaution a été prise afin que les sujets implantés puissent se concentrer sur la tâche de discrimination question/affirmation, sans avoir d'effort attentionnel à fournir pour la discrimination des mots de la phrase. Quatre phrases d'entraînement étaient présentées avant la réalisation du test à proprement parler.

### ***Test de discrimination de la fréquence fondamentale (F0)***

Ce test a pour objectif de mesurer la distance minimale perceptible entre deux sons complexes de fréquences fondamentales différentes, comparables par ailleurs sur le plan acoustique. La littérature psycho acoustique utilise le terme de F0 différence limen (F0DL) pour faire référence à ce seuil de discrimination et nous emploierons cette appellation dans la suite de notre travail. Nous avons utilisé un test disponible on line, développé par Grassi et Soranzo, qui permet de déterminer le F0DL pour plusieurs régions fréquentielles. Le F0DL a ainsi été mesuré pour les fréquences suivantes : 110-220-400-500-750 Hz. Ces fréquences correspondent aux principaux seuils concernés par l'audition résiduelle chez les sujets implantés cochléaires.

Ce test utilise une méthode adaptative non paramétrique basée sur la maximum likelihood procedure (mlp), et est présenté sous Matlab® (the Mathworks, Inc., Natick, MA, USA). Les stimuli étaient des sons complexes à 5 composantes (1 F0 et 4 harmoniques), durant 250 ms chacun. Le test repose sur une procédure de choix forcé entre 3 stimuli, présentés 30 fois. Parmi ces trois stimuli, deux ont la même F0 et sont appelés F0 standards. Le troisième a une F0 plus élevée et est appelé F0 cible. Si le sujet identifie correctement le F0 cible deux fois de suite, la distance entre F0 standard et F0 cible est réduite. Si le sujet se trompe, la distance est augmentée. La procédure adaptative permet de déterminer le F0DL au terme des 30 présentations.

### *Discrimination de la parole*

Les performances des sujets pour la discrimination des mots dissyllabiques de la liste de Fournier, présentée à 65 dB et la discrimination des phrases dans le silence ont également été étudiées. Ces paramètres intègrent le bilan orthophonique réalisé avant l'implantation cochléaire et dans le suivi post implantation.

## **2.3 Procédure**

Tous les sujets implantés cochléaires et normo entendant ont été évalués dans la même cabine insonorisée du service d'ORL, CHU Purpan. Les stimuli étaient présentés au moyen d'un casque SENNHEISER HD 280 pro ® à une intensité de 65 dB, qui correspond à l'intensité à laquelle les sujets implantés sont évalués pour l'évaluation de la discrimination de la parole.

Les trois sous – tests de perception de la prosodie linguistique étaient présentés à l'aide du logiciel Presentation® et le F0DL était mesuré au moyen de la fonction mlp dans Matlab (the Mathworks, Inc., Natick, MA, USA). L'ordre des différents sous – tests pour la prosodie linguistique et pour le F0DL était modifié d'un sujet à l'autre afin d'éviter les éventuels effets d'entraînement ou de fatigue. L'expérimentateur était également installé dans la cabine insonorisée et enregistrait les réponses des sujets. L'ensemble de la procédure durait entre 1 heure et 1 heure 30 selon les sujets en incluant les items d'entraînement.

## **2.4 Analyse des résultats**

Dans chacun des trois sous – tests concernant la perception de la prosodie linguistique, nous avons mesuré le pourcentage de réponses correctes. Dans cette tâche, la difficulté pour les sujets réside dans la détection des phrases prononcées sur un mode interrogatif, qui constitue donc le signal cible. Pour réaliser cette détection, le sujet s'appuie sur ses capacités perceptives propres mais il existe également un facteur décisionnel dépendant de son attitude libérale ou conservatrice. La théorie de détection du signal peut donc être appliquée à cette tâche et nous avons converti les pourcentages de réponses correctes en  $d'$ , indice de sensibilité.

Pour chaque sujet dans chacun des sous – tests, le  $d'$  a été calculé en utilisant l'équation de Creelman (voir chapitre II.6, *Problématique et méthodes*).

Le F0DL a été déterminé au moyen de la fraction de Weber (plus petite différence en Hz entre la F0 cible et la F0 standard rapportée à la F0 standard) et exprimé en pourcentage.

Les analyses statistiques descriptives et comparatives ont été réalisées à l'aide des logiciels Excel® et Statview®. L'étude comparative pour le F0DL a été conduite sur l'ensemble des 6 régions fréquentielles testées et sur un F0DL moyen, correspondant à la moyenne des différents F0DL obtenus pour chacune des F0 standards évaluées. Pour l'analyse comparative des performances obtenues aux différents sous – tests de perception de la prosodie, nous avons réalisé des comparaisons portant sur le pourcentage de réponses correctes, puis sur le  $d'$ . La recherche d'un effet lié au groupe a été réalisée au moyen d'un test de Kruskal-Wallis compte tenu de la distribution non normale des valeurs. Les valeurs des groupes ont été comparées *post hoc* deux à deux afin de déterminer l'existence d'une supériorité d'un groupe par rapport à l'autre, à l'aide d'un ré échantillonnage par bootstrapping. La recherche de corrélation entre les performances obtenues et le niveau d'audition résiduelle a été réalisée dans les groupes EAS et bimodal en calculant le coefficient de corrélation  $r$  et son niveau de significativité par un test de  $r$  en  $z$ . La recherche d'une relation entre les performances obtenues et la durée d'expérience avec l'implant a été réalisée dans les trois groupes de sujets implantés cochléaires en calculant le coefficient de corrélation  $r$  et son niveau de significativité. Les résultats pour les tests de discrimination question - affirmation ont été corrélés aux seuils de discrimination de la fréquence tonale et au niveau d'audition résiduelle. Des analyses complémentaires par le test de Kruskal-Wallis et bootstrap deux à deux ont été réalisées afin de s'assurer de la comparabilité des groupes pour plusieurs critères, qu'ils soient d'ordre pronostique (l'âge, la durée d'expérience avec l'implant), ou qu'ils concernent l'efficacité de l'implant (discrimination des mots dissyllabiques dans le silence, discrimination des phrases dans le bruit avec un rapport signal/bruit de +10 dB).

### **3. RESULTATS**

Après avoir décrit les caractéristiques de chacun des 4 groupes étudiés, nous présentons leurs résultats pour les sous – tests de perception de la prosodie et pour la discrimination de F0, puis les analyses concernant l’impact de l’audition résiduelle sur leurs performances. Pour chaque score, nous présentons le résultat moyen et les valeurs correspondant à l’intervalle de confiance à 95% entre parenthèses.

#### **3.1 Description des groupes**

Vingt et un sujets implantés cochléaires adultes unilatéraux ont été inclus dans 3 groupes différents en fonction de l’absence (CI only, n = 10) ou de la présence d’audition résiduelle (groupe EAS pour le groupe présentant une audition résiduelle bilatérale, n = 4 ; groupe Bim pour le groupe présentant une audition résiduelle controlatérale à l’implant, n = 7). Nous avons présenté les valeurs moyennes pour les différentes caractéristiques des groupes étudiées, ainsi que leur intervalle de confiance à 95% après ré échantillonnage par bootstrapping (entre parenthèses).

La moyenne d’âge était de 52.7 ans (45-65) dans le groupe EAS, de 52.4 (36.8-61.1) dans le groupe Bim, de 66.9 ans (58.6-71.5) dans le groupe CI only et de 27.9 (24.1-35.7). Il n’existait donc pas de différence significative pour l’âge entre les différents groupes de sujets implantés.

Le délai moyen post activation était de 46.5 mois (18-62) dans le groupe EAS, 26 mois (14-43.7) dans le groupe Bim et de 45.4 mois (30.4-61.2) dans le groupe CI only. L’analyse comparative de ces délais post activation ne révélait donc pas de différence significative entre les différents groupes. Toutefois, le délai tendait à être moins important dans le groupe Bim, facteur susceptible d’influencer négativement leurs performances, que nous interpréterons donc en connaissance de cause. Le délai tendait également à être inférieur dans le groupe EAS par rapport au groupe CI only.

Cependant, les tendances vers une différence entre les groupes n’avaient pas d’impact sur les performances pour la discrimination de la parole dans le silence. En effet, le score moyen pour la reconnaissance des mots dissyllabiques était de 83.7% (77.5-87.5) dans le groupe

EAS, 87.1% (76.4-94.3) dans le groupe Bim et 79.5% (72.8-94.1) dans le groupe CI only. Le score moyen pour la reconnaissance des phrases dans le bruit était de 93.5% (91-98.2) dans le groupe EAS, 87.4% (69.4-94.4) dans le groupe Bim et 77.1% (68.6-86.6) dans le groupe CI only. Les sujets du groupe EAS présentaient donc une meilleure discrimination dans le bruit par rapport aux sujets CI only, ce qui constitue une donnée déjà connue dans la littérature (voir chapitre II.5). Cependant, nous avons pu étudier trois groupes comparables pour la discrimination de la parole dans le silence et centrer nos analyses sur la perception de la prosodie et la discrimination de F0. Les performances des groupes Bim et EAS ont toutefois été interprétées avec précaution compte tenu de la durée d'expérience avec l'implant réduite par rapport aux sujets du groupe CI only. La suite des résultats nous a révélé que ces précautions n'étaient pas nécessaires puisque ces groupes de sujets obtenaient généralement de meilleures performances par rapport aux sujets CI only.

Quinze sujets normo entendant ont été inclus dans un groupe contrôle (groupe NHS). La moyenne d'âge était de 27.9 ans (24.1-35.7). Un profil de réponses normales a ainsi pu être déterminé dans chacun des sous – tests de perception de la prosodie linguistique et pour la discrimination de F0.

### **3.2 Perception de la prosodie linguistique**

La figure V.4 illustre les résultats obtenus pour la discrimination question-affirmation dans les trois différentes conditions auditives originale, à F0 constante, modifiée de type Patel (amplitude et durée égalisée sur la dernière syllabe) pour tous les sujets implantés comparés aux sujets normo entendant. La figure V.5 illustre les résultats une fois les groupes de sujets implantés en fonction de leur audition résiduelle (EAS, Bim et CI only).

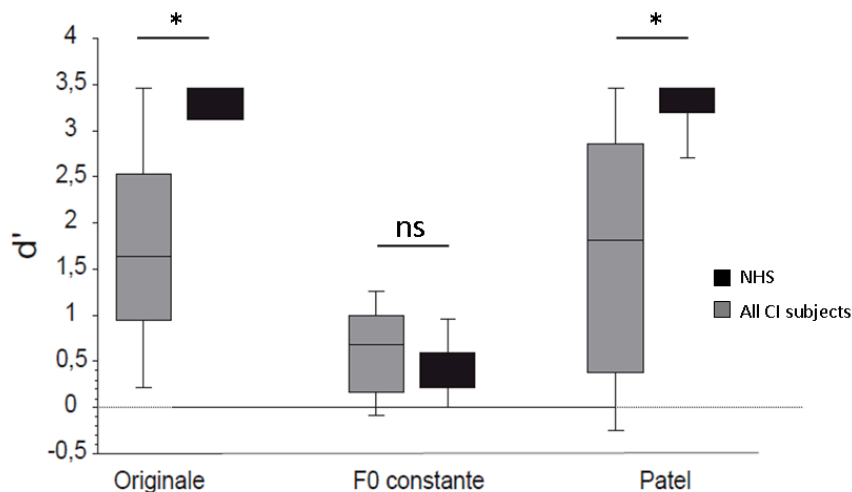


Figure V.4 : Synthèse des valeurs de  $d'$  moyennes obtenues dans chaque condition auditive pour tous les sujets implantés cochléaires versus les sujets normo entendant. On note l'importante variabilité des performances des sujets implantés cochléaires.

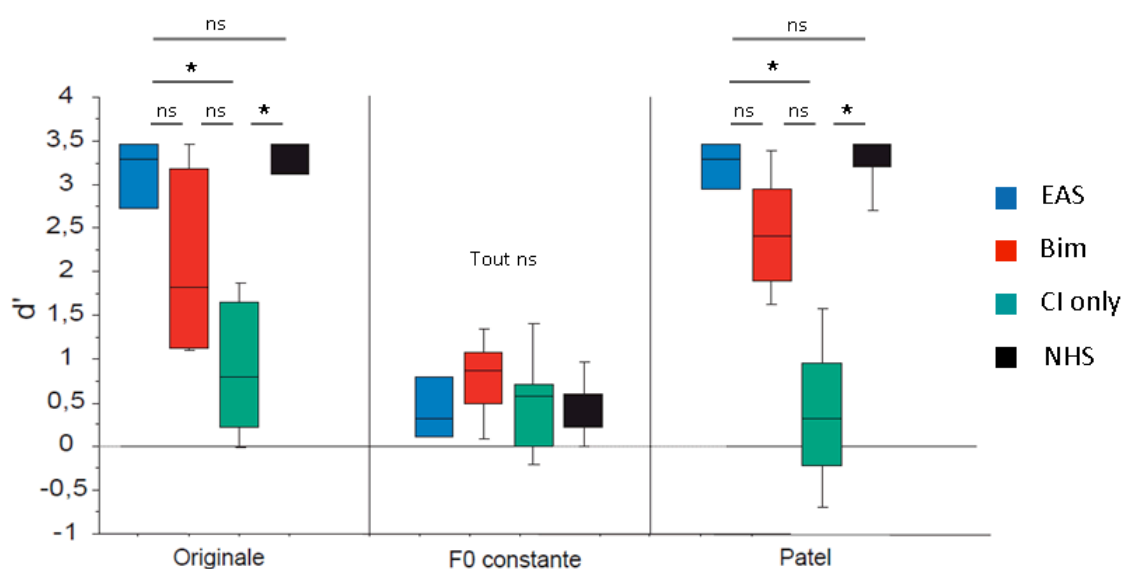


Figure V.5 : Synthèse des performances ( $d'$ ) obtenues dans chaque condition auditive pour chacun des groupes de sujets. On note que les performances des groupes EAS et NHS ne sont jamais différentes entre elles. En condition à F0 constante, aucune des comparaisons n'a retrouvé de différence significative entre les groupes. En condition Patel, les performances des groupes NHS, EAS et Bim sont supérieures à celles du groupe CI only.

### ***Sous – test 1 : condition auditive originale***

Cette condition nous a permis d'évaluer les capacités perceptives des sujets implantés cochléaires dans la vie réelle, en l'absence de toute manipulation acoustique des enregistrements. Globalement, les sujets implantés avec une audition résiduelle associée à la

stimulation par implant (groupes EAS et Bim) obtenaient de meilleures performances que les sujets implantés ne présentant pas de résidus auditifs, stimulés exclusivement par implant cochléaire (CI only).

En effet, les sujets des groupes EAS et Bim identifiaient correctement 94.8% (89.6-97.9) et 82.7% (74.4-92.3) des stimuli. Le score moyen de discrimination correcte dans le groupe CI only était de 65.4% (57.9-72.9), significativement supérieur au niveau de la chance évalué à 50%. Les performances du groupe contrôle NHS étaient logiquement élevées avec 97.9% (96.3-99) de réponses correctes.

Ces pourcentages ont été convertis en valeurs de  $d'$  et les analyses comparatives réalisées sur ces scores (voir figure V.5). Le  $d'$  moyen était de 3.1 (2.54-3.46) dans le groupe EAS, 2.07 (1.42-2.89) dans le groupe Bim, 0.91 (0.45-1.38) dans le groupe CI only et 3.3 (3.14-3.37) dans le groupe NHS. Dans tous les groupes, les valeurs moyennes de  $d'$  étaient situées au dessus du niveau de la chance ( significativement supérieures à 0). La performance du groupe EAS était comparable à celle du groupe contrôle NHS et meilleure que dans le groupe Bim. Le  $d'$  moyen du groupe Bim était inférieur à celui du groupe NHS. Les performances du groupe CI only étaient inférieures à celles des deux autres groupes de sujets implantés et du groupe NHS.

Dans les deux groupes des sujets implantés avec audition résiduelle, nous avons retrouvé une corrélation significative entre le délai post activation et les valeurs de  $d'$  en condition originale ( $r = 0.88$ ,  $p < 0.0001$ ). Cette relation n'était significative dans le groupe CI only ( $r = 0.1$ ,  $p = 0.79$ ). L'expérience avec l'implant pourrait donc permettre aux sujets présentant une audition résiduelle d'améliorer leurs performances pour la perception de la prosodie.

L'impact des modifications du signal au sein de chaque groupe de sujets est illustré dans la figure V.6.

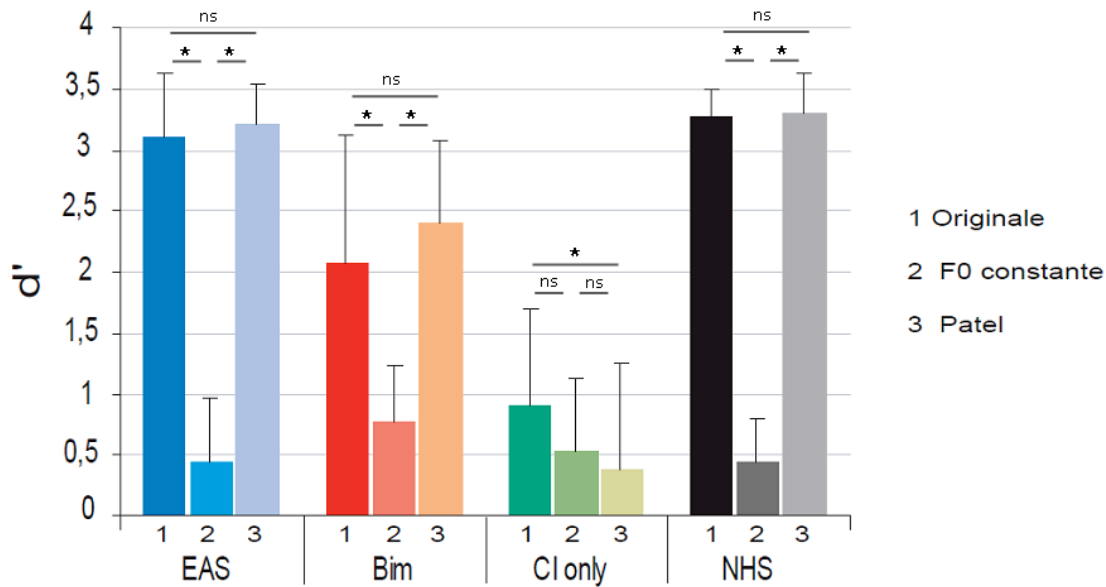


Figure V.6. Synthèse des performances ( $d'$ ) obtenues dans chaque groupe. On note un impact significatif lorsque les indices de variations de F0 sont enlevés dans le groupe NHS et dans les deux groupes de sujets implantés avec audition résiduelle. Dans le groupe CI only, il n'y a pas d'effet significatif de cette manipulation. En revanche, l'égalisation en durée et en amplitude pour la dernière syllabe (condition Patel) conduit à une chute de leurs performances, au niveau de la chance.

#### ***Sous – test 2 : condition auditive modifiée : F0 constante***

Cette condition nous a permis d'évaluer l'importance des variations de F0 dans chacun des 4 groupes de sujets étudiés. Les groupes de sujets implantés présentaient tous une diminution des performances moyennes par rapport à la condition originale.

En effet, le pourcentage de réponses correctes dans cette condition était de 57.3% (52.1-66.7) dans le groupe EAS, de 64.3% (57.7-69) dans le groupe Bim, de 58.7% (53.3-63.4) dans le groupe CI only, et de 58.1% (55.2-60.7). Les valeurs de  $d'$  correspondantes étaient de 0.45 (0.11-0.98) pour le groupe EAS, 0,78 (0.4-1.06) pour le groupe Bim et 0.53 (0.2-0.9) pour le groupe CI only. Dans le groupe NHS, le  $d'$  moyen était de 0.44 (0.27-0.61). Il n'y avait ainsi pas de différence significative entre les groupes dans cette condition et les performances restaient supérieures au niveau de la chance.



Les comparaisons entre les performances obtenues (en pourcentages de réponses correctes et  $d'$ ) en condition originale et après que les variations de F0 aient été supprimées retrouvaient une chute significative dans les groupes EAS, Bim et NHS :

- pour le groupe EAS, le  $d'$  moyen chutait ainsi de 3.1 (2.54-3.46) à 0.45 (0.11-0.98)
- pour le groupe Bim, le  $d'$  moyen chutait de 2.07 (1.42-2.89) à 0,78 (0.4-1.06)
- pour le groupe NHS, le  $d'$  moyen chutait de 3.3 (3.14-3.37) à 0.44 (0.27-0.61)

A l'inverse, la diminution des performances moyennes n'était pas significative dans le groupe CI only avec une valeur de  $d'$  moyenne de 0.91 (0.45-1.38) passant à 0.53 (0.2-0.9).

### ***Sous – test 3 : condition auditive modifiée : condition Patel***

Dans cette condition, les deux versions interrogative et affirmative d'une même phrase étaient similaires sur le plan acoustique jusqu'à la dernière syllabe. L'amplitude et la durée de cette dernière syllabe étaient égalisées d'une version à l'autre pour ne laisser comme indice saillant que les variations de F0.

Ces variations de F0 constituant l'indice prosodique principal chez le sujet normal, il apparaissait vraisemblable que les performances du groupe contrôle NHS ne seraient pas influencées par cette modification de l'enregistrement original.

Ainsi que nous l'attendions, les scores moyens du groupe NHS étaient élevés (98.2%, 95.7-99.5;  $d'$ :3.3, 3.05-3.41), comparables à leurs performances obtenues en condition originale. De la même manière, la modification acoustique des stimuli n'avait pas d'impact sur les performances moyennes des groupes EAS et Bim.

Pour le groupe EAS, les performances étaient comparables à celles obtenues en condition originale (pourcentage de réponses correctes : 96.8%, 92.7-100 ;  $d'$  :3.2, 2.85-3.46). Pour le groupe Bim, les performances n'étaient pas significativement différentes de la condition originale (pourcentage de réponses correctes : 86.9%, 80.9-93.4 ;  $d'$  :2.4, 2-2.94).

En revanche, les performances du groupe CI only diminuaient de manière significative et chutaient au niveau de la chance dans cette condition (pourcentage de réponses correctes : 55.4%, 47-64.6 ;  $d'$  :0.34, -0.14-0.89).

L'analyse comparative inter groupes (voir figure V.6) a retrouvé les différences qui étaient également constatées en condition originale avec des performances meilleures dans le groupe EAS par rapport au groupe Bim et au groupe CI only. Il n'y avait pas de différence entre le groupe EAS et le groupe contrôle NHS.

La suppression des indices d'amplitude et de durée n'avait donc aucune influence sur les performances des sujets NHS et des sujets implantés avec audition résiduelle (groupes EAS ou Bim). En revanche, les sujets implantés cochléaires sans audition résiduelle semblaient s'appuyer sur ces indices, dans une certaine mesure, puisque leurs capacités perceptives pour la prosodie linguistique diminuaient significativement lorsqu'ils étaient absents.

### 3.3 Discrimination de F0

Le F0DL a été calculé pour plusieurs fréquences (110Hz-220Hz-400Hz-500Hz-750Hz) et un F0DL moyen a été calculé à partir des résultats obtenus pour chacune de ces fréquences.

Les résultats obtenus par fréquence sont présentés dans le tableau IV.3. Le test de Kruskal-Wallis réalisé a mis en évidence un effet lié au groupe dans chacune des fréquences considérées. Cet effet était essentiellement lié au groupe CI only, qui obtenait des F0DL significativement et systématiquement plus élevés que les groupes EAS et NHS, quelle que soit la fréquence considérée (voir figure V.7). Les F0DL des groupes EAS et NHS n'étaient jamais significativement différents sauf pour la fréquence 750 Hz, pour laquelle le F0DL était meilleur dans le groupe des NHS. Si l'on compare les deux groupes de sujets implantés avec audition résiduelle, les sujets du groupe EAS présentaient de meilleurs F0DL que les Bim pour les fréquences 220 Hz et 500 Hz, ainsi qu'un meilleur F0DL moyen. Les sujets du groupe Bim présentaient quant à eux un meilleur F0DL que les CI only pour la fréquence 110 Hz. La comparaison inter groupe pour le F0DL moyen est illustrée figure V.8.

	F0DL 110 Hz	F0DL 220 Hz	F0DL 400 Hz	F0DL 500 Hz	F0DL 750 Hz	Mean F0DL
EAS	8,3% (5.3-12.3)	5,3% (3.5-7)	5,8% (3.7-8.9)	3,1% (1-4.3)	4,7% (2.5-7.6)	5,2% (3.8-5.9)
Bim	21,5% (9-20.5)	11,3% (8-58.4)	7,2% (3.3-10)	7,7% (4.7-11.4)	11,1% (2.5-9)	11,3% (6.7-17.2)
CI only	40,9% (27.9-80.3)	15,1% (14.3-92.7)	6,2% (7.1-140)	9% (8.6-146.7)	12,1% (8.9-32)	15,8% (16.2-91.6)
NHS	3,7% (1.9-8.2)	2,7% (1.7-4.1)	1,9% (1.7-4.1)	2,6% (1.4-4.2)	1,6% (1-2.4)	2,4% (1.5-3.4)

Tableau V.3. Valeurs moyennes du F0DL mesurées pour chaque fréquence standard avec l'intervalle de confiance à 95% indiqué entre parenthèses.

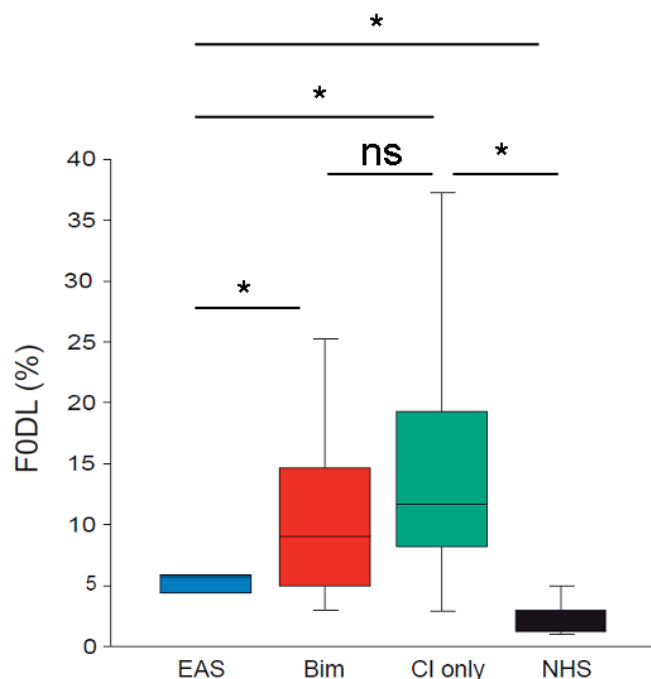


Figure V.8. Box plot illustrant les différences entre les valeurs du FODL moyen des 4 groupes. Les groupes EAS et NHS présentent un FODL plus bas, donc meilleur, que le groupe CI only.

A travers ces premiers résultats, nous avons montré que la perception de la prosodie linguistique et la discrimination de F0 étaient meilleures chez les sujets présentant une audition résiduelle. Chez les sujets implantés sans audition résiduelle (groupe CI only), nous avons observé un important déficit de perception de la prosodie linguistique et des niveaux de FODL variables. La deuxième partie de nos résultats s'intéresse aux relations entre le niveau d'audition résiduelle, les capacités de discrimination de F0 et les performances dans la tâche de discrimination question affirmation en condition originale.

### 3.4 Impact du niveau d'audition résiduelle

Nous avons évalué la relation entre les performances obtenues pour la discrimination question/affirmation en condition originale et le niveau d'audition résiduelle d'une part, et la relation entre le FODL et le niveau d'audition résiduelle d'autre part. Cette relation a été étudiée dans les deux groupes de sujets implantés cochléaires qui présentaient une audition résiduelle (groupe EAS et groupe Bim). Un sujet du groupe Bim a été exclu des analyses de corrélation en raison d'un FODL moyen mesuré au-delà de l'intervalle de confiance à 95% dans ce groupe après ré échantillonnage.

Nous avons ainsi considéré la perte auditive moyenne sur les fréquences 125, 250 et 500 Hz sur la meilleure oreille des sujets des groupes EAS et Bim. Nous avons sélectionné ces fréquences car elles rendent compte de la plupart des variations de la fréquence fondamentale de la voix humaine et c'est donc essentiellement dans cette région fréquentielle que l'intérêt d'une audition résiduelle peut s'exercer.

Le niveau d'audition résiduelle (125-500 Hz) et les valeurs du  $d'$  pour la perception de la prosodie linguistique en condition originale étaient fortement corrélés ( $r = -0.86$  ;  $p = 0.0006$ ), tout comme le niveau d'audition résiduelle et le FODL moyen ( $r = 0.85$  ;  $p = 0.0009$ , après exclusion des deux sujets outliers). Cette relation est illustrée dans la figure V.9 a et b.

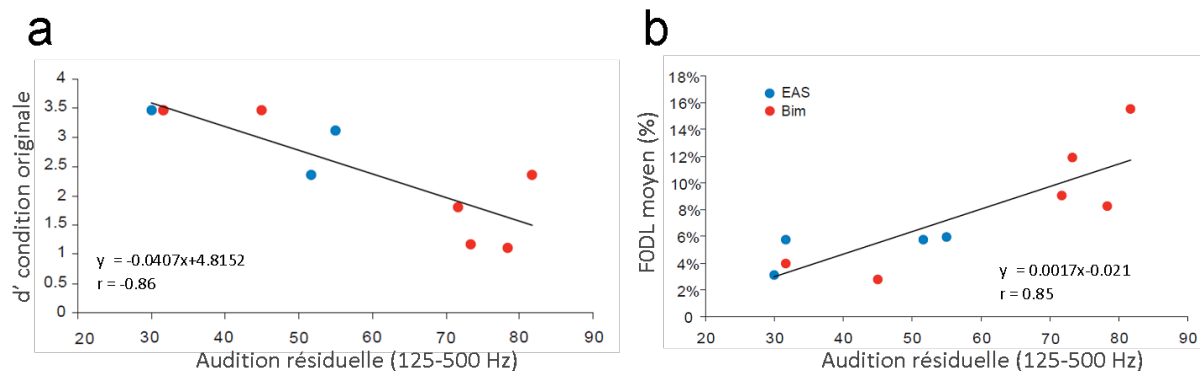


Figure V.9 a et b. a) Corrélation négative significative entre le niveau d'audition résiduelle sur les fréquences 125-500 Hz et les performances pour la prosodie linguistique en condition originale. Deux sujets EAS et Bim sont représentés par un point pour un  $d' = 3.46$  et un niveau d'audition résiduelle de 30 dB. b) Corrélation significative positive entre le niveau d'audition résiduelle sur les fréquences 125-500 Hz et le FODL moyen.

Il nous est donc apparu que le niveau d'audition résiduelle importait plus que son caractère uni ou bilatéral. La différence que nous avons observée entre les groupes EAS et Bim pour la perception de la prosodie linguistique en condition originale pourrait donc être rapportée à une différence de perte auditive. La figure IV.9a nous a permis d'identifier 6 sujets (les 4 sujets du groupe EAS et 2 sujets du groupe Bim) présentant des valeurs de  $d'$  élevées ( $> 2$ ) avec des résidus auditifs meilleurs que 60 à 70 dB sur les fréquences graves. Une analyse complémentaire des performances obtenues par ces six sujets comparées aux performances des cinq sujets du groupe Bim restant a permis de souligner l'importance du niveau de l'audition résiduelle permettant une perception de la prosodie linguistique au moins subnormale. En effet, ces 6 sujets avec des seuils auditifs meilleurs que 60-70 dB sur les

fréquences 125 à 500 Hz obtenaient une valeur moyenne de  $d'$  de 3.2 contre 1.5 pour les cinq sujets avec des seuils moins bons que 60 dB sur les fréquences 125 à 500 Hz. De la même manière, leur F0DL était meilleur comparé aux cinq autres sujets (4.6% versus 13.8%).

### **3.5 Relation entre le F0DL et la perception de la prosodie linguistique**

Nous avons fait l'hypothèse que cette relation pourrait être significative dans les groupes de sujets implantés avec une audition résiduelle puisque leurs performances pour la perception de la prosodie linguistique chutaient lorsque les variations de F0 étaient supprimées. En revanche, les performances des sujets du groupe CI only n'étaient pas significativement impactées par la suppression de l'indice F0 et il apparaissait plausible qu'il n'y aurait pas de relation entre le F0DL et les capacités pour la prosodie linguistique dans ce groupe.

En effet, dans les groupes EAS et Bim, nous avons retrouvé une corrélation négative significative entre F0DL et les valeurs de  $d'$  en condition originale ( $r = -0.67$  ;  $p = 0.03$ ). Dans le groupe CI only, il n'existait de corrélation significative ( $r = 0.4$  ;  $p = 0.22$ ).

Ces analyses complémentaires renforcent nos premières conclusions quant aux indices acoustiques utilisés par les sujets implantés cochléaires. Nous suggérons ainsi que des seuils auditifs résiduels meilleurs que 60 dB sur les fréquences graves permettent d'obtenir des performances subnormales pour la discrimination question/affirmation. Les sujets implantés cochléaires sans audition résiduelle pourraient adopter des stratégies compensatoires basées sur les variations d'amplitude et de durée pour réaliser cette tâche. Cette conclusion reste prudente dans la mesure où ces sujets montrent un déficit important de perception de la prosodie linguistique, faisant douter même de l'existence de leurs capacités perceptives dans ce domaine.

## **4. DISCUSSION**

### **4.1 Perception de la prosodie linguistique et audition résiduelle**

Nos résultats soulignent l'intérêt de l'audition résiduelle dans une tâche de perception de la prosodie linguistique. Nous avons montré que le niveau des résidus auditifs facilitait la

discrimination entre question et affirmation, et jouait probablement un rôle plus important que leur caractère uni ou bilatéral.

Certains sujets implantés avec des seuils auditifs meilleurs que 60-70 dB sur les fréquences graves étaient en effet susceptibles d'atteindre des performances comparables à celles des sujets normo entendant dans cette tâche. Cette conclusion est renforcée par l'existence d'une forte corrélation entre le niveau de perte auditive sur les fréquences graves et les valeurs du  $d'$  en condition originale. L'implication clinique directe de cette information concerne les patients présentant une surdité sévère à profonde bilatérale, déjà implantés d'un côté ou non, pour lesquels se pose la question d'une implantation cochléaire bilatérale. Le niveau de perte sur les fréquences 125 à 500 Hz pourrait être considéré comme un indicateur du bénéfice retiré des résidus auditifs et guider les équipes soignantes dans la difficulté de cette décision. L'évolution de la perception de la prosodie linguistique après implantation cochléaire mérite également une attention particulière. La relation entre le délai post activation et les performances en condition originale était forte et significative pour les sujets implantés des groupes présentant une audition résiduelle, ce qui n'était pas le cas pour les sujets CI only (voir figure V.10).

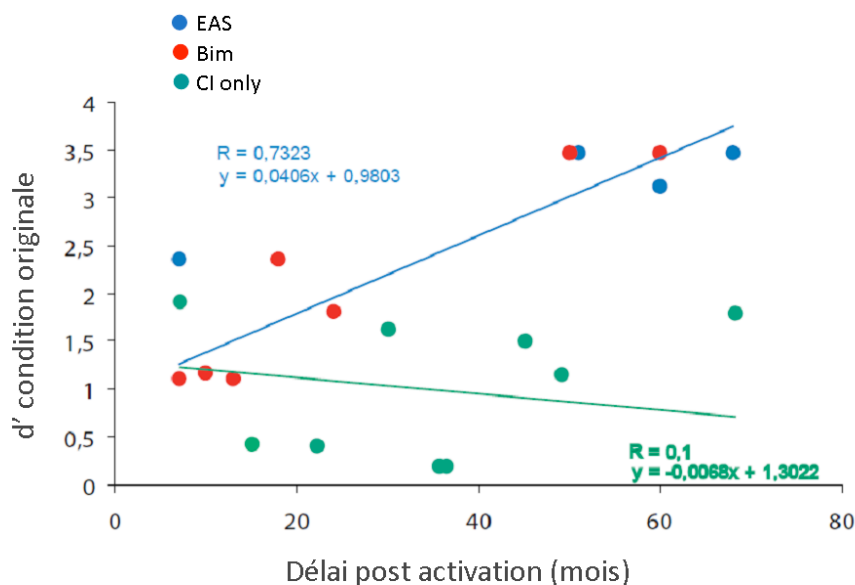


Figure V.10. Etude de la relation entre les valeurs du  $d'$  en condition originale et le délai post activation. Cette corrélation est fortement significative pour les groupes de sujets implantés avec audition résiduelle. En revanche, il n'y avait pas cette relation dans le groupe de sujets implantés cochléaires sans audition résiduelle (CI only).

L'hypothèse d'un effet favorable de l'expérience de la stimulation bimodale sur les performances auditives avait été évoquée dans une publication de Tyler et al. [Tyler et al., 2002]. Le bénéfice lié à la combinaison de l'information apportée par l'audition résiduelle et de la stimulation électrique reposerait sur les capacités de fusionner les deux types d'information. Ces capacités pourraient s'acquérir avec le temps, à la faveur de la réorganisation du système auditif central après implantation. La même équipe avait à nouveau évalué cette hypothèse trois ans plus tard en s'intéressant plus particulièrement aux effets binauraux pour la discrimination de la parole dans le bruit et la localisation spatiale de la source sonore. Leurs premières conclusions se voyaient nuancées puisque parmi les douze sujets bimodaux évalués, les deux présentant le gain le plus significatif en stimulation bimodale étaient implantés depuis seulement 3 mois. Le niveau d'audition résiduelle des sujets évalués n'était cependant pas précisé et l'on peut supposer que la mise en place de stratégies d'écoute en stimulation bimodale peut s'étendre dans le temps, à l'inverse des sujets implantés sans audition résiduelle, qui « apprennent » à n'entendre que la stimulation électrique et dont les principaux progrès sont réalisés pendant les premiers mois après l'implantation [Rouger et al., 2007, Lenarz et al., 2012].

L'audition résiduelle sur les fréquences graves permet aux sujets bénéficiant d'une stimulation électro acoustique ou bimodale d'avoir accès aux variations de F0, qui constitue l'indice prosodique principal. Cet accès à F0 a été mis en évidence dans plusieurs études portant sur différents domaines de la perception auditive chez le sujet implanté bénéficiant d'une stimulation bimodale. La perception de F0 permet ainsi au sujet de cibler le locuteur principal dans un environnement bruyant [Nootheboom and Terken, 1982, Assmann and Summerfield, 1990] ou de percevoir la dimension mélodique d'extraits musicaux [Gfeller et al., 2002b, Gantz et al., 2006].

Mais la combinaison de l'audition résiduelle à la stimulation électrique par l'implant transmet d'autres informations que F0 sur les fréquences graves, telles que les premières harmoniques pour les instruments ou les premiers formants pour la parole. Li et Loizou ont ainsi développé le concept de « glimpse » [Li and Loizou, 2008], ou coup d'œil, apporté par l'audition résiduelle. Les fréquences graves permettent de détecter et d'intégrer des informations paralinguistiques (caractéristiques du locuteur) à des informations purement linguistiques, comme le premier formant F1, qui contribue à la perception de certaines voyelles et du temps de voisement.

Si certains sujets implantés cochléaires présentant une audition résiduelle satisfaisante sur les fréquences graves peuvent percevoir la prosodie linguistique en s'appuyant sur les indices

apportés par les variations de F0, tout comme les sujets normo entendant, on observe un déficit important chez les sujets implantés sans audition résiduelle. En condition originale, les performances du groupe CI only demeuraient au-dessus du niveau de la chance mais ces résultats méritent discussion.

## **4.2 Perception de la prosodie linguistique par le groupe CI only**

Nous étions parvenus à des conclusions prudentes dans la description et les premières interprétations de nos résultats concernant ce groupe. Nous avons en effet souligné l'importance de leur déficit en condition originale, avec une valeur moyenne de  $d'$  correspondant à moins de 30% des capacités des sujets normo entendant ou des sujets implantés cochléaires avec des seuils sur les fréquences graves meilleurs que 60 dB. L'importance de ce déficit renvoie à la dissociation constatée chez les sujets implantés cochléaires dans nos études précédentes [Massida et al., 2011, Massida et al., 2013] entre la restauration satisfaisante de la discrimination de la parole et le déficit persistant pour la perception des attributs de la voix.

Les résultats obtenus dans les deux conditions auditives modifiées suggèrent que la discrimination entre question et affirmation dans ce groupe CI only repose sur d'autres indices que les variations de F0. En effet, la diminution des performances entre la condition originale et la condition à F0 constante ( $d'$  moyen diminuant de 0.91 à 0.53) n'était pas significative. En revanche, cette chute devenait significative dans la condition Patel où la durée et l'amplitude étaient égalisées ( $d'$  moyen diminuant de 0.91 à 0.34) avec un score qui n'était pas significativement différent du niveau de chance.

Ces observations, considérées dans le même temps que les difficultés de compréhension de la tâche demandée par certains sujets du groupe CI only, posent la question de l'existence de capacités perceptives pour la prosodie linguistique dans cette population. Pour certains sujets, la consigne a dû en effet être répétée et explicitée à plusieurs reprises avant que le test puisse être réalisé. Les principales limites résidaient dans la présentation de phrases neutres sur le plan syntactique, c'est-à-dire pouvant être perçues comme une question tout comme une affirmation. Pour certains sujets CI only, une phrase interrogative comporte nécessairement des indices syntaxiques, débute par « Est-ce que » ou contient un agencement verbe-sujet caractéristique (« Joue-t-il .. »). C'est donc la représentation même des variations prosodiques linguistiques qui est interrogée.



L'étude suivante nous a permis d'évaluer plus précisément cette hypothèse, en s'intéressant à d'autres indices prosodiques qui jouent un rôle démontré chez les sujets normo entendant : les indices prosodiques visuels. Les sujets implantés cochléaires sont en effet réputés pour développer des capacités visuelles supra normales, destinées à compenser le déficit présenté dans la modalité auditive. S'ils ne reconnaissent pas ou peu les indices prosodiques visuels, alors il est vraisemblable que la représentation de la prosodie linguistique soit déficitaire dans cette population.

## **5. ETUDE PRELIMINAIRE DE LA PERCEPTION DES INDICES VISUELS DE LA PROSODIE LINGUISTIQUE**

Nous présentons ici les résultats préliminaires d'une étude en cours de réalisation portant sur la perception des indices visuels de la prosodie linguistique par les sujets implantés cochléaires. Dans cette étude, nous nous sommes intéressés spécifiquement aux sujets implantés cochléaires sans audition résiduelle (groupe CI only), population qui présentait un déficit majeur de perception de la prosodie linguistique en condition auditive originale.

Cinq sujets implantés supplémentaires ont été inclus après avoir signé le consentement écrit prévu par le Comité de Protection des Personnes.

Les quinze sujets normo entendant inclus dans l'étude précédente ont également participé à cette étude préliminaire.

### **5.1 Méthodes**

#### *Evaluation de la perception de la prosodie linguistique*

La tâche de discrimination question/affirmation a également été utilisée pour cette étude en ajoutant une condition visuelle seule. Nous avons ainsi extrait l'enregistrement vidéo dissocié de l'enregistrement audio des 24 phrases répétées sous une forme affirmative ou sous une forme interrogative par les deux acteurs professionnels sollicités. Les indices prosodiques visuels étaient plus marqués pour la forme interrogative d'une phrase avec des mouvements d'anteflexion de la tête et des mouvements de sourcils en fin d'énoncé (voir figure V.11). Comme pour les autres sous – tests, la phrase était précédée d'une présentation sous forme

écrite, avant la présentation visuelle, pour éviter tout effort attentionnel dirigé vers la compréhension des mots par lecture labiale.

Les quatorze sujets ont également réalisé les sous – tests correspondant aux conditions auditives originale et modifiées (condition à F0 constante et condition Patel).



Figure V.11 : extraits du clip vidéo pour la phrase « il joue de la flûte » répété sous forme interrogative (haut) et sous forme affirmative (bas) par l'actrice. Les mouvements des sourcils, des paupières et l'avancée de la tête étaient plus marqués pour la forme interrogative.

### *Evaluation de la discrimination de la parole et de la lecture labiale*

Les performances des sujets pour la discrimination des mots dissyllabiques de la liste de Fournier, présentée à 65 dB, en modalité auditive seule, en modalité visuelle seule et en modalité audiovisuelle ont également été étudiées.

### *Analyse des résultats*

Le pourcentage de réponses correctes a été mesuré puis sa conversion en  $d'$  a été réalisée au moyen de l'équation de Creelman. Nous avons ensuite comparé les performances du groupe des sujets implantés cochléaires CI only aux performances du groupe contrôle des sujets normo entendant (NHS) en utilisant un ré échantillonnage par bootstrapping permettant le calcul et la comparaison des intervalles de confiance à 95% de distribution des valeurs obtenues dans chacun des groupes. Les comparaisons pour les performances intra – groupes (discrimination de la parole en modalité auditive seule, visuelle seule, et audiovisuelle pour

les sujets implantés ; d' en condition originale et en condition visuelle pour les deux groupes CI only et NHS)

Nos analyses complémentaires se sont ensuite portées sur l'existence d'éventuelles relations entre les capacités de lecture labiale, d'intégration visuo-auditive et la perception des indices visuels de la prosodie. Nous avons calculé le coefficient de corrélation  $r$  et son niveau de significativité (test de  $r$  en  $z$ ) pour ces différentes variables.

## 5.2 Résultats

### *Discrimination de la parole et lecture labiale*

Le groupe CI only était composé de 14 sujets âgés de 44 à 73 ans (médiane : 62 ans ; intervalle de confiance à 95% : 53.7 – 65.57). Huit sujets étaient implantés à droite et six sujets à gauche. On comptait neuf hommes pour cinq femmes.

Les résultats obtenus pour la discrimination de la parole sont illustrés dans la figure V.12. La discrimination moyenne des mots dissyllabiques en modalité auditive seule était de 79.6% (68.5-84). La reconnaissance moyenne des mots en lecture labiale était de 17.5% (9.8-26.9). Le gain en modalité audiovisuelle était significatif puisque la discrimination moyenne des mots dissyllabiques était de 96.4% (89,6-99,7).

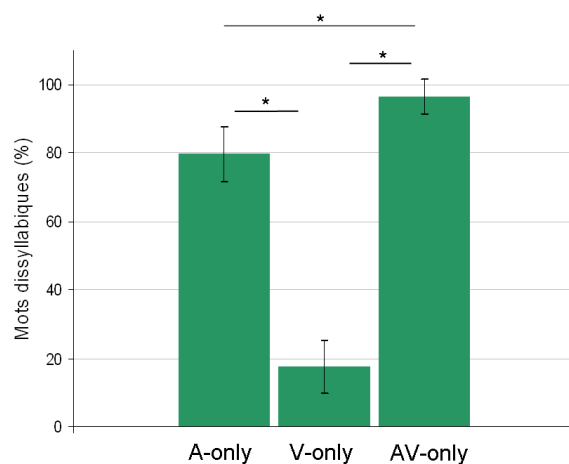


Figure V.12. Pourcentages de discrimination correcte des mots dissyllabiques selon la modalité de présentation. Le gain en condition audiovisuelle est significatif par rapport à la modalité auditive seule.

### *Perception de la prosodie linguistique*

Dans le groupe des sujets implantés cochléaires, le pourcentage moyen de réponses correctes était de 67.25% (57.9-72.9) en condition auditive originale et de 66% (55.5-72.7) en condition visuelle seule. Une fois convertis en  $d'$ , ces scores moyens étaient respectivement de 1 (0.51-1.4) et de 0.94 (0.36-1.28).

Dans le groupe contrôle des sujets normo entendant, le pourcentage moyen de réponses correctes en condition auditive originale était de 97.9% (96.4-99.7). En condition visuelle seule, ce pourcentage moyen était de 76.9% (+/-7.7). Les valeurs de  $d'$  correspondantes étaient 3.3 (+/-0.23) pour la condition auditive originale et 1.62 (+/-0.48) pour la condition visuelle.

Le calcul des intervalles de confiance à 95% des valeurs de  $d'$  après ré échantillonnage par bootstrapping dans les deux groupes montrait de meilleures performances obtenues dans le groupe NHS (intervalle de confiance à 95% pour le groupe NHS : 1.3-1.8 ; intervalle de confiance à 95% pour le groupe CI only : 0.58-1.25). Les résultats comparatifs sont illustrés dans la figure V.13.

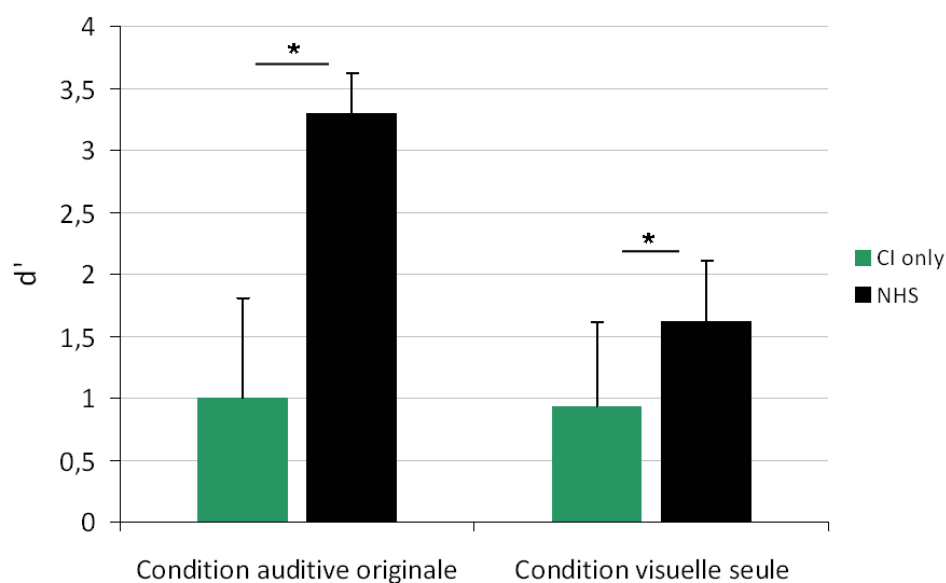


Figure V.13. Comparaison des valeurs de  $d'$  pour la perception de la prosodie linguistique en condition auditive originale et condition visuelle seule. Le groupe NHS présente de meilleures performances par rapport aux sujets implantés cochléaires.

Les sujets implantés cochléaires présentaient donc un déficit de perception des indices visuels de la prosodie linguistique par rapport aux sujets normo entendant.

### *Relation entre les capacités de lecture labiale et la perception des indices visuels de la prosodie linguistique*

Nous avons recherché l'existence d'une éventuelle relation entre les capacités d'intégration audiovisuelle pour la parole et les valeurs de  $d'$  pour la perception des indices visuels de la prosodie linguistique. Nos résultats font état d'une absence de corrélation entre ces variables, qu'elle concerne les performances en lecture labiale seule ( $r = 0.06$  ;  $p = 0.84$ ) ou le gain en pourcentage de points pour la modalité audiovisuelle ( $r = 0.06$  ;  $p = 0.83$ ) (voir figure V.14a et b).

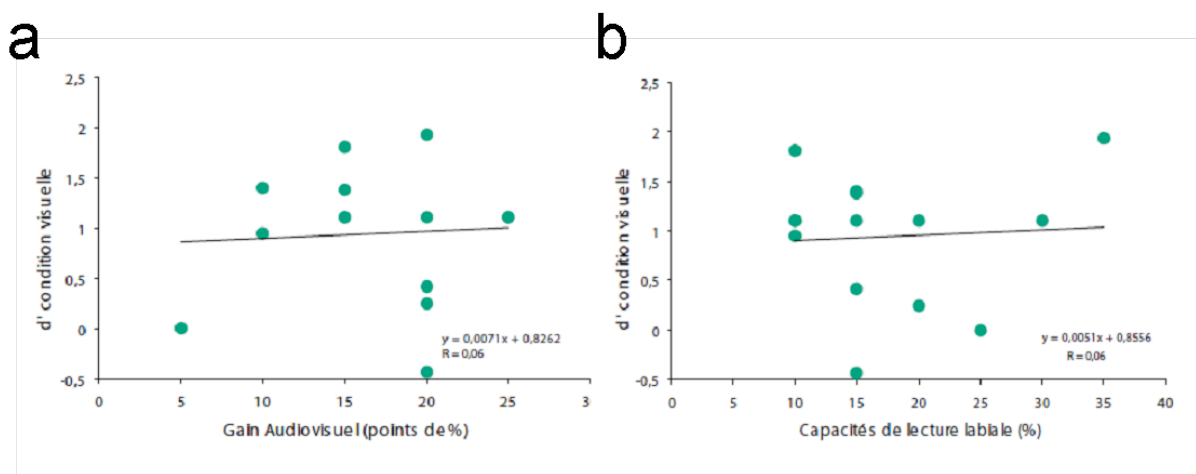


Figure V.14a et b. a) Etude de la corrélation entre les valeurs du  $d'$  en condition visuelle et le gain audiovisuel retrouvant une absence de corrélation. Deux sujets implantés cochléaires présentent les mêmes valeurs de  $d'$  pour un gain AV à 15 et 25 points. b) Etude de la relation entre les valeurs du  $d'$  en condition visuelle et les capacités de lecture labiale retrouvant une absence de corrélation. Deux sujets implantés cochléaires présentent les mêmes valeurs de  $d'$  pour une lecture labiale à 10% et 15%.

Cette absence de relation renforce l'hypothèse d'un déficit de représentation globale de la prosodie linguistique. Le déficit de perception est non seulement dissocié des performances en lecture labiale mais se manifeste en plus dans une modalité dans laquelle les sujets implantés cochléaires sont réputés performants.

### 5.3 Discussion préliminaire

#### *Un déficit global de perception de la prosodie chez les sujets implantés cochléaires*

Nos résultats font état d'un déficit de perception de la prosodie linguistique chez les sujets implantés cochléaires sans audition résiduelle. Ce déficit est majeur lorsque les sujets sont évalués en condition auditive, et persiste pour les indices prosodiques visuels, à moindre niveau si l'on compare leurs performances à celles des sujets normo entendant.

L'autre information notable est le caractère global, auditif et visuel, de la dissociation entre la restauration de la discrimination de la parole et la persistance du déficit de perception de la prosodie linguistique. L'existence de cette dissociation dans la modalité visuelle est surprenante quand on connaît l'importance de la vision chez les sujets sourds pour la reconnaissance de la parole. Cette thématique a fait l'objet de plusieurs travaux impliquant le CerCo et le service d'ORL, CHU Purpan. Ces études comportementales avaient porté sur les capacités de discrimination de la parole en modalité auditive seule, visuelle seule et audiovisuelle pour souligner l'importance de la vision dans la restauration de cette fonction. Les premières comparaisons entre sujets implantés cochléaires et sujets normo entendant pour la reconnaissance visuelle des phonèmes n'avaient pourtant pas retrouvé de différence significative entre les deux populations, avec des performances de l'ordre de 25% de reconnaissance correcte. C'est en analysant la reconnaissance visuelle des mots dissyllabiques que d'importantes différences entre sujets implantés et sujets normo entendant avaient été mises en évidence. En effet, la reconnaissance visuelle des mots dissyllabiques était bien meilleure chez les sujets implantés (moyenne : 35%) que chez les sujets normo entendant (moyenne : 9%). Il avait été conclu, en accord avec d'autres auteurs [Garstecki and O'Neill, 1980, Ronnberg et al., 1996], que la vision jouait un rôle facilitateur pour la compréhension de la parole par des processus cognitifs haut niveau, de type top-down, basés sur l'analyse du contexte sémantique, syntactique et lexical.

Le déficit visuel est d'autant plus surprenant quand on considère les résultats des études s'étant intéressées à la direction du regard chez les sujets sourds en situation de communication orale. Muir LJ et Richardson (2005) avaient ainsi constaté que les sujets sourds signant tendaient à fixer une région faciale inférieure, utilisant la vision périphérique pour intégrer les mouvements des mains et de la partie supérieure du visage. Watanabe K

(2011) et al. ont comparé les régions du visage sur lesquelles le regard était dirigé et la durée pendant laquelle ils étaient fixés chez des sujets sourds et des sujets normo entendant et ont quant à eux retrouvé une tendance plus importante à fixer les yeux chez les sujets sourds. Bien que les yeux soient ainsi fixés plus longtemps, les sujets sourds implantés de notre étude n'auraient donc pas traité les mouvements de sourcil comme des indices prosodiques pertinents. Pour Schyns et al. (Schyns, PG, 2002), ces propriétés de fixation pourraient être liées au processus d'identification du locuteur plutôt qu'à l'analyse de la prosodie linguistique ou affective.

Par ailleurs, les études menées par J. Rouger avaient insisté sur le maintien de ces performances visuelles et d'intégration audiovisuelle à long terme après l'implantation. La perception de la prosodie linguistique par les sujets implantés demeure quant à elle déficitaire et nous avons montré l'absence d'impact de l'expérience avec l'implant chez les sujets CI only sur leurs capacités auditives. Il en va de même pour les indices prosodiques visuels, dont la perception n'est nullement corrélée aux délais post activation. Cependant, les limites actuelles, et provisoires, de notre étude ne nous permettent pas d'affirmer l'absence de gain apporté par la vision pour la perception de la prosodie linguistique.

### ***Limites actuelles et perspectives de développement***

La principale limite actuellement présente dans nos résultats renvoie précisément à la notion de gain audiovisuel, qui n'a pas été évalué jusque là. Pour déterminer ce gain, il convient d'ajouter une condition audiovisuelle à l'ensemble des sous – tests. Cette condition nous permettra ainsi d'estimer l'évolution des performances par rapport aux conditions auditive originale et visuelle seule et de comparer l'éventuel gain obtenu à celui connu pour la discrimination de la parole.

La deuxième limite concerne la condition visuelle seule. Un déficit a été montré chez les sujets implantés cochléaires mais une inconnue demeure quant à l'objet de leur attention visuelle. Notre hypothèse est que les sujets implantés focalisent leur attention sur la reconnaissance de la parole. Ainsi, et bien que la phrase soit présentée sous forme écrite avant la présentation visuelle, les sujets implantés auraient naturellement tendance à regarder les mouvements articulatoires plutôt que les indices prosodiques visuels (mouvements des sourcils, des paupières, de l'ensemble de la tête). Cette hypothèse est en cours d'évaluation au moyen d'un eye tracker, qui permettra de documenter la direction prise par le regard des

sujets implantés et de mettre en évidence une éventuelle direction préférentielle vers les mouvements articulatoires.





## Chapitre VI

### Etude 4 : Etude EEG de l'intégration audiovisuelle du phénomène prosodique d'accentuation

## 1. INTRODUCTION

L'importance des indices visuels et de l'intégration audiovisuelle pour la reconnaissance de la parole par les sujets implantés cochléaires et son hypothétique déficit pour la perception de la prosodie ont été soulignés dans l'étude précédente. Chez le sujet normo entendant, l'analyse conjointe des indices acoustiques (variations de F0, d'amplitude et de durée) disponibles et des mouvements de l'extrémité céphalique (sourcils, menton, hochements de tête) accompagnant les phénomènes prosodiques permet d'optimiser la compréhension globale du message oral.

La démonstration de l'existence de cette dissociation chez le sujet implanté cochléaire revêt d'autant plus d'importance que l'intégration audiovisuelle des indices prosodiques par le sujet normo entendant s'est révélée efficace y compris en condition auditive dégradée. Il a ainsi été montré qu'un signal de parole dégradé était mieux compris en présence d'indices prosodiques visuels congruents [Munhall et al., 2004] et que la détection des frontières prosodiques était améliorée en condition audiovisuelle par rapport à chacune des conditions uni modales isolée [Barkhuysen et al., 2008]. Une étude réalisée par Jessica Foxton, post-doctorante au CerCo, a de plus mis en évidence une facilitation cross modale visuo-auditive spécifique à la prosodie, caractérisée par une amélioration du seuil de détection des changements de F0 et d'amplitude en présence d'indices visuels congruents.

Certains mouvements sont plus régulièrement associés aux phénomènes prosodiques, comme le mouvement du menton [Scarborough et al., 2009] ou la protrusion labiale [Dohen and Loevenbruck, 2009], qui ont été identifiés comme étant les indices visuels les plus marqués dans le phénomène d'accentuation. Ces mouvements précèdent généralement les variations acoustiques significatives de quelques millisecondes [Hadar et al., 1984] et pourraient donc être interprétés comme des signes annonçant les changements acoustiques à venir. Au niveau neuronal, ces indices visuels pourraient donc faciliter la réponse auditive.

Cette hypothèse a été évaluée par une exploration électro physiologique de l'activité corticale en réponse à un énoncé audiovisuel contenant mot accentué. Nous avons utilisé un paradigme oddball pour lequel les stimuli standards et discordants avaient le même contenu acoustique et ne différaient que par une accentuation visuelle portant sur un des mots de l'énoncé. Notre objectif était de montrer que l'accentuation visuelle du mot est accompagnée d'une sensation auditive également accentuée et serait ainsi analysée par le système auditif central. Cette

réponse neuronale spécifique au stimulus discordant se manifesterait par l'enregistrement d'une négativité de discordance (ou MMN pour mismatch negativity) distincte de la réponse au stimulus standard.

## **2. MATERIELS ET METHODES**

Cette étude a été réalisée sous la responsabilité de Jessica Foxton et de Pascal Barone. Jessica Foxton, Kuzma Strelnikov et moi-même avons participé à l'analyse des résultats.

### **2.1 Population**

Dix sujets normo entendant (âge moyen : 24 ; rang : 19-29) ont participé à cette étude. Tous les sujets étaient droitiers et on comptait sept femmes parmi les 10 participants. Aucun antécédent neurologique, psychiatrique ou ophtalmologique n'était rapporté. Tous les sujets ont été inclus après avoir reçu une information orale et écrite concernant l'étude, et signé un consentement écrit, ainsi que le prévoyait l'autorisation de réalisation de l'étude délivrée par le Comité de Protection des Personnes du Sud-ouest et d'Outre-mer (avis n°2-03-34).

### **2.2 Stimuli**

La conception des stimuli a été réalisée à partir de l'enregistrement audiovisuel d'une actrice professionnelle, qui devait répéter la séquence de mots « un-deux-trois » de deux manières différentes : une version sans accentuation et une version avec accentuation. Chacune des deux versions enregistrées durait 2.52 secondes. La durée de chaque mot et chaque silence entre les mots était similaire d'une version à l'autre. Le signal acoustique enregistré dans la version avec accentuation a été remplacé, à l'aide du logiciel Adobe Premiere Pro 7.0, par le signal acoustique enregistré dans la version sans accentuation. Ainsi, les deux versions avaient un contenu acoustique strictement identique et ne pouvaient être distinguées que par l'information visuelle relative à l'accentuation. Cette accentuation visuelle était essentiellement portée par des mouvements des sourcils, réalisés alors que l'actrice commençait à énoncer le mot « deux ». Ces mouvements étaient absents dans la version sans accentuation.

Les mouvements des sourcils précédaient le début du son « deux » d'environ 500 ms dans la version accentuée originale et dans la version hybride créée après remplacement du signal acoustique.

Nous nous sommes assurés que l'énoncé neutre « un-deux-trois » ne contenait pas d'indice prosodique auditif en demandant à 17 autres sujets normo entendant naïfs d'identifier le mot accentué dans l'énoncé. Les réponses de ces sujets se sont révélées aléatoires et permettaient donc de garantir la neutralité acoustique de l'énoncé présenté.

De plus, une fois le test terminé, nous avons demandé aux dix sujets participant à l'étude s'ils avaient perçu un mot accentué sur le plan auditif dans les stimuli présentés. Tous les sujets rapportaient avoir perçu le son correspondant au mot « deux » comme étant accentué.

### **2.3 Procédure**

Ainsi que le paradigme oddball le prévoit, nous avons présenté aux sujets le stimulus standard sans accentuation de manière répétée, et le stimulus discordant plus rarement, avec un délai constant de 400 ms entre chaque présentation.

Les clips vidéo étaient présentés aux sujets avec le son pour la moitié des séquences (condition audiovisuelle AV) et sans le son pour l'autre moitié (condition visuelle seule V). La condition visuelle seule avait été prévue au titre de contrôle, pour s'assurer que la MMN n'était pas de nature visuelle pure et correspondait bien à une facilitation visuo-auditive. Une tâche simple, sollicitant le maintien de l'attention, devait être réalisée par les sujets et consistait en la détection d'un élément visuel occasionnel. Ils devaient ainsi repérer l'apparition d'une croix blanche devant les yeux de l'actrice. Cette croix pouvait apparaître à quatre moments spécifiques dans le clip vidéo (1.0, 1.2, 1.4, ou 1.6 secondes). La tâche permettait ainsi de focaliser l'attention visuelle des sujets sur le tiers supérieur du visage de l'actrice et ce, plus particulièrement autour du moment où elle prononçait le mot « deux ».

La procédure comptait 6 séquences en condition AV et 6 séquences en condition V, l'ordre des conditions V et AV étant randomisés d'un sujet à l'autre. Dans chaque séquence, il y avait 100 présentations du stimulus standard, 25 présentations du stimulus discordant et 15 clips contenant la croix blanche devant les yeux de l'actrice. Il y avait ainsi 17.9% de stimuli discordants, 10.7% de clips cibles, et dans chaque condition, on dénombrait 150 stimuli discordants. L'ordre de présentation des stimuli (standard et discordant) était pseudo-randomisé car les dix premiers clips présentés correspondaient toujours au stimulus standard et deux stimuli discordants devaient être séparés d'au moins trois stimuli standard.

## 2.4 Enregistrement EEG et analyses

Les enregistrements ont été réalisés à partir d'électrodes de recueil fixées sur un bonnet (Oxford Instruments, UK). L'électrode de référence était placée sur le nez du sujet. Deux électrodes complémentaires ont été placées en regard des mastoïdes droite et gauche. L'électro-oculogramme (EOG) était enregistré au moyen de deux électrodes placées à proximité des yeux. Au total, 32 électrodes étaient donc utilisées : FP1-FP2-F3-F4-C3-C4-P3-P4-F7-F8-T3-T4-T5-T6-CB2-CZ-Fz-Pz-T5'-T6'-O1-O2-O1'-O2'-P3'-P4'-Pz'-Oz-CB1-M1-M2-VEOG. Les impédances étaient contrôlées pour rester inférieures à 10 kOhms. L'électro-encéphalogramme (EEG) et l'EOG étaient enregistrés à l'aide d'un amplificateur SynAmps (NeuroScan, El Paso, TX). Les signaux étaient échantillonnés à un taux de 1000 Hz et filtrés par un filtre passe-bas à 200 Hz.

L'analyse des enregistrements EEG a été réalisée à l'aide du logiciel Neuroscan. Le signal a été filtré (2-30 Hz) et l'analyse s'est concentrée sur la période allant de 700 ms à 1500 ms depuis le début du clip vidéo. Dans la suite du texte, le début de cette période est identifié comme 0 ms. Cette période a été sélectionnée car elle permet de concentrer les analyses sur les réponses neuronales au mot « deux ». C'est en effet à partir de 700 ms que vont s'initier les mouvements des sourcils précédant l'articulation du mot « deux ». Les périodes comprenant trop d'artefacts ont été rejetées, de même que les périodes enregistrées après la présentation d'un stimulus discordant ou d'une croix blanche. Après avoir rejeté ces périodes, nous disposons d'environ 510 enregistrements par condition et par sujet. Les signaux ainsi traités ont été moyennés séparément pour les stimuli standards et les stimuli discordants.

Les analyses statistiques ont cherché à déterminer s'il existait une différence significative entre les réponses électriques aux stimuli standards et les réponses aux stimuli discordants et ont donc été concentrées sur la période entourant le mot « deux ». Des ondes moyennées ont été calculées pour chaque sujet et les ondes obtenues au niveau de l'électrode d'intérêt Fz pour le stimulus standard et pour le stimulus discordant ont été comparées par un test t. Nous avons réalisé une analyse en densité de source du courant pour déterminer la topographie des réponses aux stimuli standards et discordants. La carte des activations différentielles a ainsi pu être déterminée.

Nous avons utilisé comme méthode d'analyse statistique des mesures répétées ANOVA 2x2 incluant comme facteurs la modalité (condition AV, condition V) et le type de stimulus (standard, discordant) et des comparaisons après ré échantillonnage par bootstrapping.

### 3. RESULTATS

#### 3.1 Tâche comportementale de détection

Les performances des sujets pour la détection de la croix blanche variaient entre 88% et 100% de réponses correctes (médiane : 99%) en condition audiovisuelle AV, entre 91% et 100% (médiane : 99%) pour la condition visuelle seule V. Le nombre de fausses alarmes variait entre 0 et 2 pour les conditions. Ces résultats témoignent du maintien de l'attention des sujets le long de l'enregistrement EEG, quelque soit le délai d'apparition de la croix blanche devant les yeux de l'actrice. L'accentuation visuelle du mot « deux » a été perçue par les sujets évalués.

#### 3.2 Résultats EEG

Comme il est classique dans les études recherchant une MMN (voir la revue de Naatanen et al. [Naatanen et al., 2007]), l'analyse des différences des réponses EEG pour le stimulus standard (sans accentuation) et le stimulus discordant (avec accentuation visuelle) a été réalisée pour l'électrode d'intérêt Fz. L'analyse des signaux obtenus en condition AV a révélé une négativité significative 360-400 ms par rapport au début de la période d'intérêt, ce qui correspond à 1060-1100 ms par rapport au début du clip vidéo ( $p < 0.05$ , 10000 permutations). La taille de cet effet a été évaluée à l'aide du test de Cohen [Cohen, 1992], appliqué à la fenêtre temporelle 360-400 ms en l'électrode Fz. Ce test a retrouvé une valeur  $d$  de 1.46, signant un large effet. Il n'a pas été retrouvé de négativité significative après le début du signal acoustique pour le mot « deux » (voir figure VI.1).

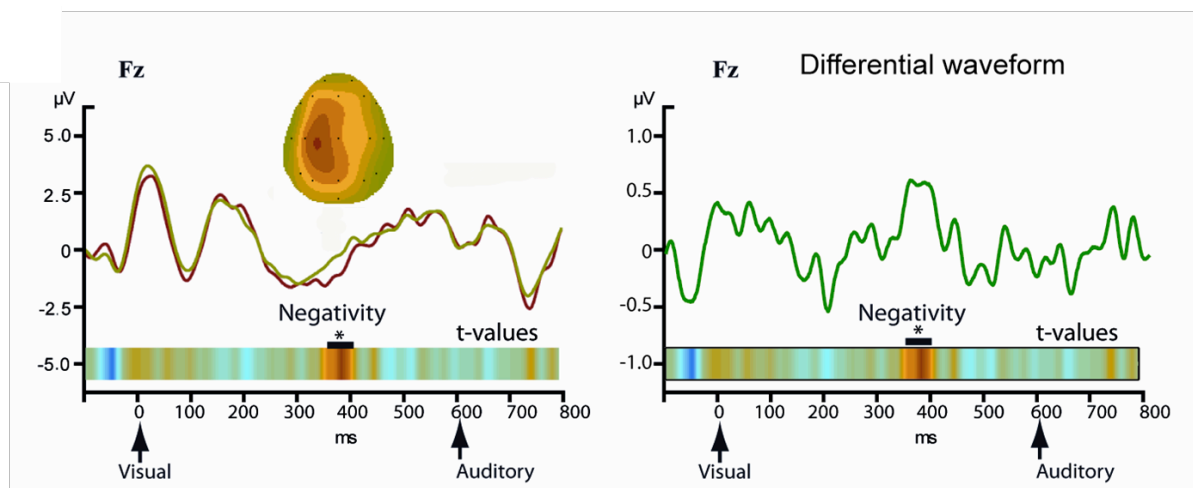


Figure VI.1. A gauche, potentiels évoqués moyennés en réponse au stimulus avec accentuation visuelle sur le mot « deux » (rouge) et au stimulus sans accentuation visuelle (vert) en condition audiovisuelle. A droite est présentée l'onde différentielle avec un pic de négativité significatif mesuré entre 360 et 400 ms.

La MMN à 360-400 ms en Fz a été retrouvée chez tous les sujets sauf un. Elle était en revanche observée pour tous les sujets en Cz, qui est l'électrode recueillant généralement les MMN auditives, y compris dans les études évaluant un signal de parole en condition audiovisuelle (voir figure VI.2).

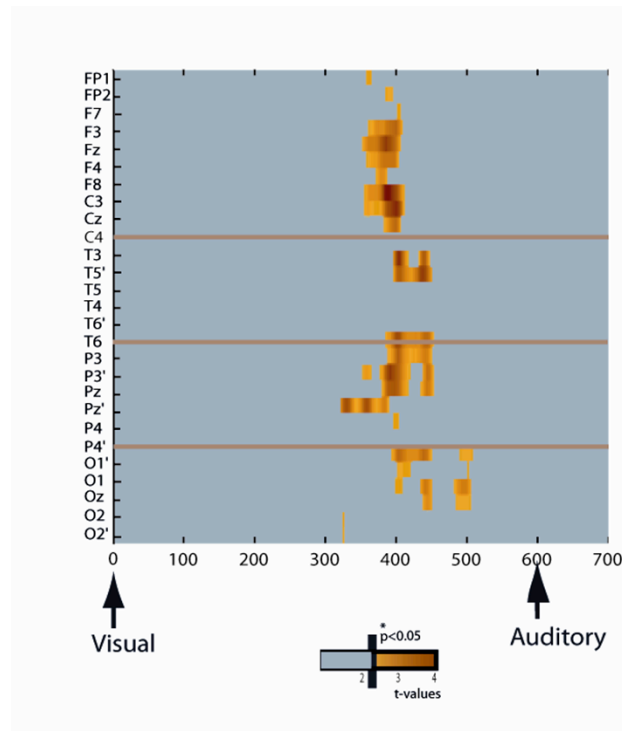


Figure VI.2. Analyse de la distribution spatiale et temporelle des valeurs de t obtenues au pic de la négativité. Le temps 0 correspond au début du mot « deux », soit 700 ms après le début du clip vidéo. La valeur de t correspondant au seuil de significativité ( $p = 0.05$ ) est 2.3 et les valeurs qui lui sont supérieures sont représentées par un code couleur.

Cette négativité était donc observée 200 ms avant l'apparition du signal acoustique correspondant au début du mot « deux », soit probablement en réponse aux indices visuels d'accentuation. Pour renforcer cette conclusion, nous avons réalisé une analyse image par image des différences de pixels entre le clip vidéo pour le stimulus standard (sans accentuation) et le clip vidéo pour le stimulus discordant (avec accentuation). Les différences entre la version sans accentuation et la version avec accentuation visuelle concernaient ainsi l'ouverture de la bouche (moyenne des différences :  $1.5 \pm 0.8$  pixels) et le mouvement des sourcils (moyenne des différences :  $2.7 \pm 1.5$  pixels). Ce mouvement était initié 500 ms avant le début du signal acoustique pour le mot « deux » et la négativité était observée, pour le stimulus discordant, juste après l'initiation de ce mouvement (voir figure VI.3). Cette



séquence temporelle suggère fortement que la réponse MMN soit liée aux indices prosodiques visuels précédant l'accentuation acoustique attendue.

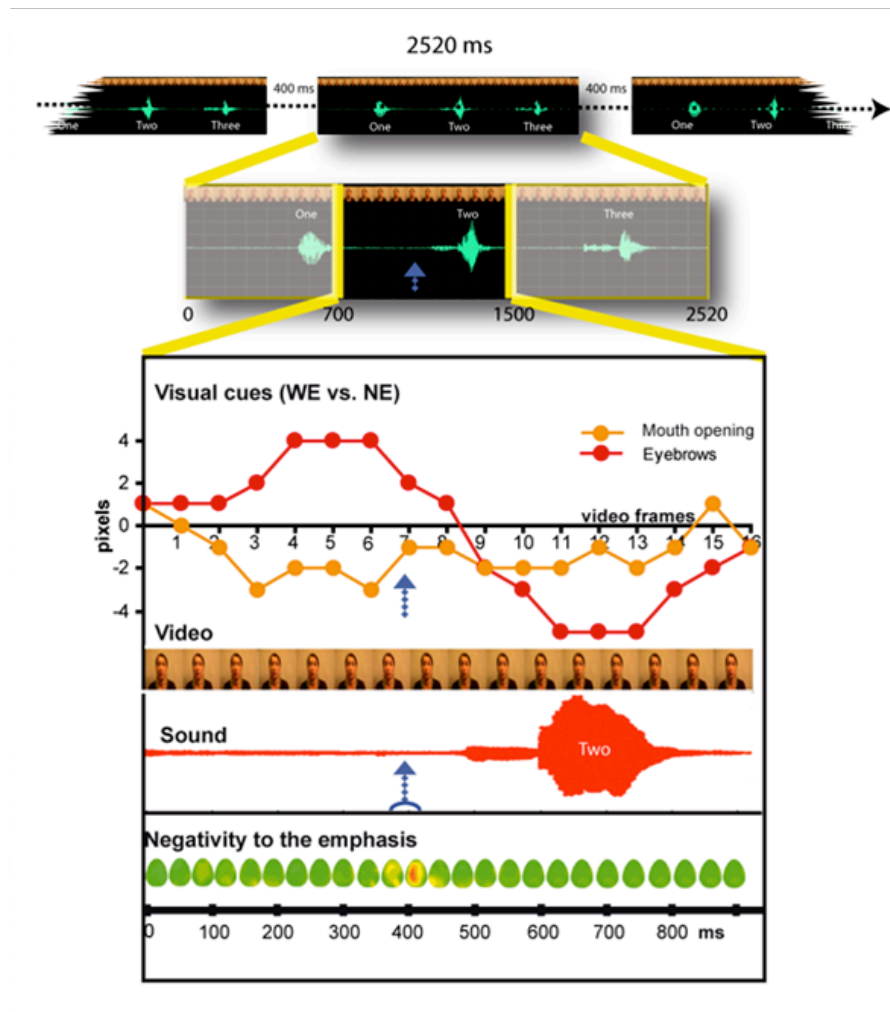


Figure VI.3. Analyse de la négativité mesurée en fonction des informations visuelles présentes dans clip vidéo. Les mouvements des sourcils (rouge) et de la bouche (protrusion labiale, orange) ont été analysés par pixel sur chaque image (en haut de la figure). La négativité mesurée en condition audiovisuelle (bas de la figure) se produit 150-200 ms après le début des mouvements, et environ 100 ms avant le voisement du son « deux ».

Cette réponse de type MMN pourrait également en rapport avec une MMN visuelle, puisqu'elle se produit dans un intervalle temporel pour lequel des négativités de discordance de N2 (250-450 ms) ont déjà été montrées en modalité visuelle seule (voir Pazo-Alvarez pour review, [Pazo-Alvarez et al., 2003]). Dans une étude cherchant à caractériser la MMN spécifique à l'effet McGurk, Saint-Amour et al. [Saint-Amour et al., 2007] ont démontré la spécificité AV de cette MMN en analysant d'une part l'effet de la modalité de présentation (audiovisuelle ou visuelle seule) et en analysant d'autre part la différence négativité obtenue entre les deux modalités de présentation (AV-V). Nous avons donc dans un premier temps

réalisé des mesures répétées ANOVA 2x2 qui ont révélé une interaction significative entre la condition de stimulation (V ou AV) et le type de stimulus (standard ou discordant) :  $F(1 ; 9) = 7.2, p < 0.05$ . Il n'y avait pas de négativité détectée dans la condition V en Fz ni Cz, quelque soit le moment de la période considéré.

Nous avons ensuite comparé par bootstrap les réponses des sujets pour le stimulus standard et le stimulus discordant au niveau de la fenêtre temporelle pour laquelle une réponse de type MMN avait été observée en condition AV (360-400 ms). Il n'y avait pas de négativité significative pour les électrodes centrales, frontales et temporales, en condition V. A 440ms, on retrouvait une légère différence avant correction ( $p > 0.05$ ) entre les réponses au stimulus discordant et les réponses au stimulus standard au niveau du pôle occipital (Oz). L'analyse de localisation de source CSD a mis en évidence une distribution occipito-pariétale de cette négativité. (voir figure VI.4).

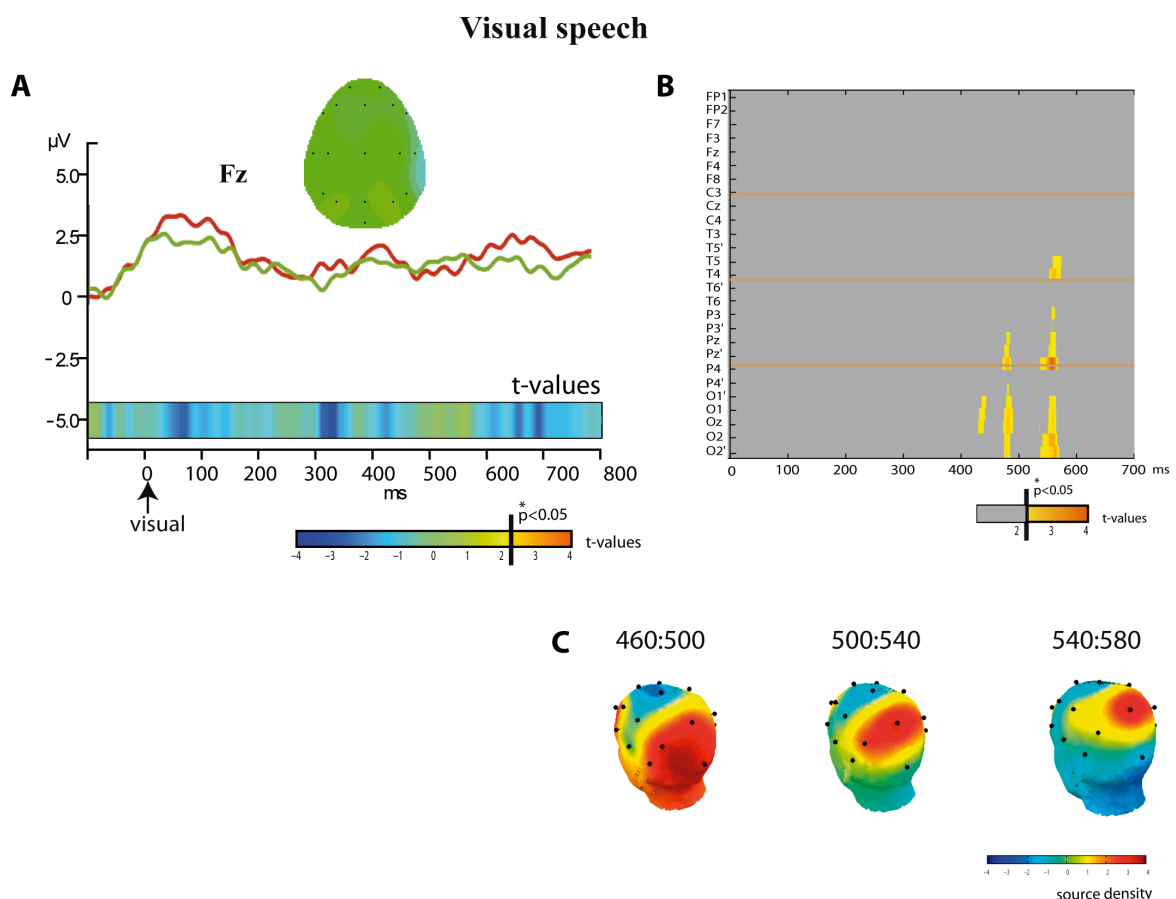


Figure VI.4. A) Au niveau de l'électrode Fz en condition V, aucune négativité de type MMN n'était détectée de manière significative en réponse au stimulus accentué sur le plan visuel (rouge) par rapport au stimulus non accentué (vert). B) L'analyse par cluster a cependant révélé une négative significative plus tardive, débutant à 440 ms, C) prédominant au niveau du pôle occipital et s'étendant en pariétal.

Dans un deuxième temps, pour chaque sujet, nous avons soustrait les potentiels évoqués par le stimulus standard et le stimulus discordant en condition V à ces potentiels évoqués en condition AV (AV-V, voir Saint Amour et al, 2007). Les ondes résultantes AV-V correspondant au stimulus standard ont été comparées aux ondes résultantes AV-V correspondant au stimulus discordant. Cette analyse comparative a été réalisée au moyen d'un bootstrap sur une durée comprenant l'intervalle d'intérêt 360-400 ms. Cette analyse a retrouvé une négativité significative en Fz entre 360 et 400 ms, confirmant sa spécificité pour la condition audiovisuelle (figure VI.5).

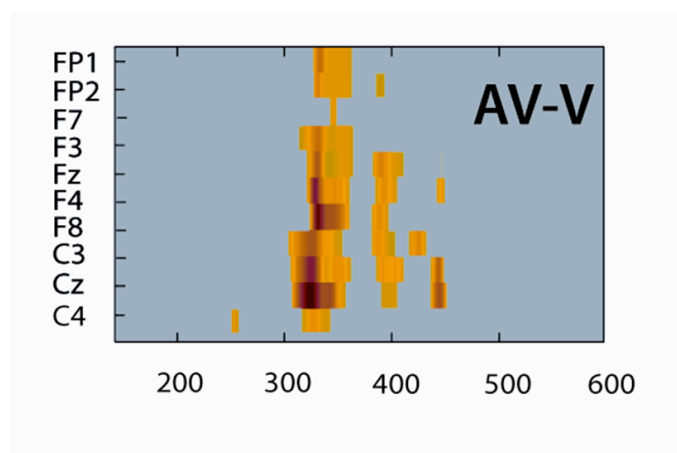


Figure VI.5. AV-V : Les différences obtenues en condition V entre les potentiels évoqués par le stimulus discordant avec accentuation visuelle et les potentiels évoqués par le stimulus standard sans accentuation ont été soustraites à ces mêmes différences obtenues en condition AV. La négativité est significative en Cz et Fz.

Les propriétés spatio-temporelles de la propagation de cette négativité sont illustrées dans la figure VI.6. Elle est ainsi initiée entre 320 et 360 ms au niveau du cortex occipito-pariétal postérieur ( $p < 0.05$ , correction par cluster dans le domaine temporel et spatial après 10000 permutations). Plus tard, pour l'intervalle entre 360 et 400 ms correspondant au pic de cette négativité, la réponse de type MMN s'étend vers l'avant aux régions auditives temporo-frontales. Elle se propage pour finir vers l'arrière en direction des régions occipitales entre 400 et 440 ms ( $p < 0.05$ , correction par cluster dans le domaine temporel et spatial après 10000 permutations). L'analyse de localisation des sources de courant CSD confirme l'origine occipitale de la négativité. Pour le pic de négativité obtenue entre 360 et 400 ms, l'analyse CSD localise les sources au niveau de la région pariétale gauche, de la région temporale postérieure gauche et de la région frontale inférieure. Les premières analyses par ANOVA 2x2 n'ont pas retrouvé de latéralisation significative mais la localisation de source a révélé une tendance plus marquée à la latéralisation gauche. Les régions impliquées dans cette

réponse de type MMN et la chronologie de leur activation constitue un argument supplémentaire en faveur de la nature visuo-auditive de cette négativité.

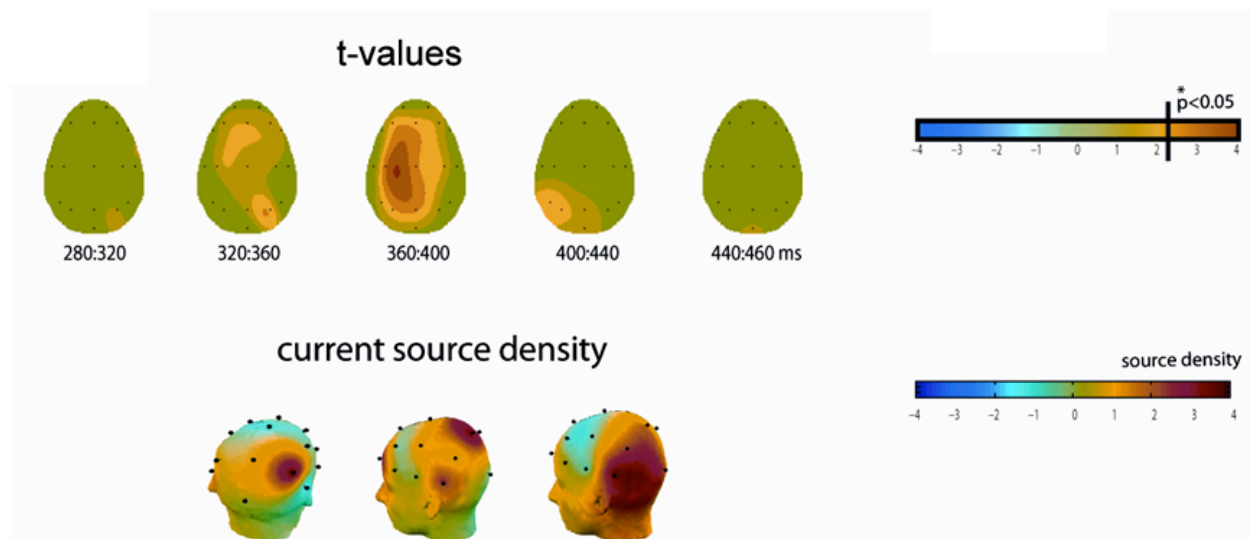


Figure VI.6. Caractéristiques spatio-temporelles de la négativité de type MMN enregistrée en condition AV. Sa propagation décrit une boucle avec une origine pariéto-occipitale, puis une extension en avant vers les régions fronto-temporales avant de se terminer dans la région occipitale. L'analyse de localisation de courant CSD retrouve une implication des régions frontale gauche et temporale postérieure gauche.

## 4. DISCUSSION

Nos analyses ont donc retrouvé une négativité spécifique à la condition audiovisuelle déclenchée par la présence d'indices visuels d'accentuation dans le stimulus discordant. Le fait qu'elle précède le début du signal acoustique est surprenant mais peut être interprété de la manière suivante : les indices prosodiques visuels d'accentuation annoncent des variations acoustiques attendues qui sont associées au phénomène d'accentuation. En d'autres termes, ils faciliteraient le codage prédictif de l'activité neuronale auditive.

### 4.1 Nature de la négativité mesurée

Une réponse de type MMN caractérise l'activité cérébrale automatique en réponse à un changement occasionnel détecté de manière pré-attentive au sein de stimuli auditifs identiques et répétés (voir Naatanen et al pour review, 2007). Cette réponse est classiquement évoquée par des changements concernant les paramètres acoustiques fondamentaux (fréquence, amplitude ou durée) mais elle déjà été mise en évidence pour l'intégration visuo-auditive de la parole [Saint-Amour et al., 2007, Ponton et al., 2009] et de la prosodie. De Gelder et al. [de

Gelder et al., 1999] ont ainsi souligné la précocité de l'intégration visage-voix pour la perception de la prosodie affective, en retrouvant une MMN en réponse à un stimulus audiovisuel occasionnel non congruent (visage exprimant la tristesse, voix exprimant la colère) au sein de stimuli audiovisuels répétés congruents (visage et voix exprimant la colère). Par ailleurs, les relations entre les indices visuels et auditifs sont démontrées pour plusieurs phénomènes prosodiques. L'accentuation dans le message oral est ainsi caractérisée sur le plan acoustique par une augmentation de l'amplitude et de F0 et sur le plan visuel par une augmentation de l'amplitude des mouvements de la tête [Hadar et al., 1984]. Les mouvements accentués des sourcils sont de même associés à une élévation de F0 [Cavé et al., 1996]. Cette association peut prendre une autre forme, avec l'effet de facilitation cross modale-visuo auditive mis en évidence par Jessica Foxton [Foxton et al., 2010].

Il semble donc probable que la négativité observée en condition audiovisuelle dans notre étude soit le témoin d'une activité auditive en rapport avec l'accentuation visuelle observée. L'absence de cette négativité en condition visuelle seule (au niveau de Cz et Fz) et sa persistance après soustraction des réponses évoquées en condition visuelle seule (AV-V) constituent des arguments supplémentaires en faveur de son rôle spécifique à l'intégration visuo-auditive de l'accentuation. De même, l'analyse de localisation des sources de courant CSD a retrouvé l'implication dans sa génération des régions temporales postérieures, connues pour intervenir dans l'intégration visuo-auditive. Cette région est en effet impliquée dans des tâches de reconnaissance audiovisuelle de phonèmes dégradés sur le plan auditif [Kilian-Hutten et al., 2011]. La négativité mesurée au niveau occipital, à 480 ms environ, dans la condition visuelle seule pourrait être rapprochée de l'onde N400, évoquée en réponse à des stimuli visuels non congruents [Proverbio and Riva, 2009]. La discordance visuelle présentée aux sujets peut en effet être interprétée comme tel.

Le principal problème posé par la négativité que nous avons mesurée réside dans ses caractéristiques temporelles. Elle précède en effet le début du signal acoustique correspondant au mot « deux », ce qui n'a pas été jusque là retrouvé dans la littérature. Les études portant sur l'effet McGurk [Sams et al., 1991, Colin et al., 2002, Saint-Amour et al., 2007, Ponton et al., 2009] montraient ainsi l'existence d'une MMN certes de courte latence (environ 175 ms), mais qui jamais ne précédait le stimulus auditif. Dans ces travaux, la courte durée de l'intervalle de temps séparant le début du geste articulatoire du début du son émis pourrait expliquer un chevauchement des réponses aux stimuli visuel et auditif. Le temps séparant le début des mouvements de sourcils du début du signal acoustique dans notre étude dépassait 400 ms, et la réponse à la discordance visuelle pourrait ainsi avoir été mesurée.

Une autre hypothèse, que nous soutenons plus volontiers, est que la réponse de type MMN mesurée rend compte d'une activité de facilitation visuo-auditive.

#### **4.2 Les indices prosodiques visuels ont-ils un rôle prédictif ?**

Plusieurs travaux comportementaux ont démontré que l'information visuelle permettait d'anticiper le signal de parole [Schwartz et al., 2004, Fort et al., 2012]. Fort et al (2012) ont par exemple étudié l'effet d'amorces visuelles, auditives et audiovisuelles sur la reconnaissance d'un mot présenté immédiatement après. L'amorce visuelle, c'est-à-dire la visualisation du début du geste articulatoire, permettait d'améliorer et d'accélérer la reconnaissance du mot. Les interactions entre les informations visuelles et l'activité neuronale auditive ont été mises en évidence pour la lecture labiale [Calvert et al., 1997](Calvert et al, 2010). Pour la reconnaissance audiovisuelle de syllabes [van Wassenhove et al., 2005], elles se manifestent sur le plan électro physiologique par un raccourcissement des ondes N1/P2 en présence d'indices visuels saillants.

Cette facilitation visuo-auditive pourrait également s'exercer pour la perception de la prosodie linguistique, dans le cadre d'une activité cérébrale visant à prédire le signal à venir. Plusieurs auteurs ont récemment développé ce concept de codage prédictif [Friston and Kiebel, 2009, Strelnikov et al., 2010], par lequel des attentes sensorielles sont codées par le cerveau avant d'être comparées au stimulus réel. Pour le traitement audiovisuel de la parole, l'information visuelle apportée par les mouvements labiaux facilite le traitement auditif en impliquant un réseau neuronal reliant les aires visuelles et auditives [Arnal et al., 2009, Arnal et al., 2011]. En proposant une tâche de lecture labiale sans signal auditif, Arnal et ses collaborateurs (2009) ont suggéré qu'une voie cortico-corticale directe transmet l'information relative aux mouvements articulatoires au cortex auditif et qu'une voie rétrograde indirecte transmet les erreurs entre les prédictions visuelles et le signal auditif. Ce réseau décrivait donc une boucle occipito-temporo-frontale et la comparaison du signal attendu versus le signal perçu siègeait dans le sulcus temporal supérieur [Ghazanfar et al., 2008, Arnal et al., 2009]. La réponse de type MMN que nous avons mesurée est générée et se propage selon des caractéristiques spatio-temporelles compatibles avec cette théorie. Son origine était en effet occipito-pariétale et son extension se faisait d'abord en avant vers les régions fronto-temporales avant de revenir en arrière vers le pôle occipital.

Notre étude souligne donc le rôle de l'intégration audiovisuelle dans la perception de la prosodie chez le sujet normo entendant en suggérant la facilitation de l'activité neuronale auditive par les indices prosodiques visuels. La réalisation de ce protocole chez des sujets implantés cochléaires pourrait permettre de confirmer sur le plan neuronal le déficit de perception des indices visuels suggéré par nos résultats comportementaux. Cette application aux sujets implantés cochléaires pose toutefois le problème des explorations électro physiologiques dans cette population, réalisables [Sandmann et al., 2010, Timm et al., 2012, Zhang et al., 2013] mais techniquement difficiles en raison de l'artefact provoqué par l'implant.

## Chapitre VII

### Etude 5 : Perception de la musique par le sujet implanté cochléaire Parallèles avec l'amusie congénitale



## 1. INTRODUCTION

La cérémonie d'ouverture du dernier congrès mondial d'Oto-Rhino-Laryngologie (IFOS 2013, Seoul) a marqué les esprits par la participation d'un orchestre exclusivement composé d'enfants et de jeunes implantés cochléaires. Il n'est ainsi pas rare de rencontrer des enfants implantés sourds pré linguaux ayant développé d'étonnantes facultés musicales et l'on compte parmi les jeunes implantés toulousains d'excellents pianistes, chanteurs, et même un accordéoniste. Si la dégradation spectro temporelle du signal acoustique par le processeur de l'implant est généralement accompagnée d'un déficit pour la perception musicale [Stabej et al., 2012, See et al., 2013], elle n'empêche pas pourtant certains enfants d'apprécier ni de pratiquer la musique comme les enfants normo entendant. Il n'en va pas de même pour les sujets sourds implantés post linguaux, qui rencontrent d'importantes difficultés à l'écoute de la musique, même chez ceux obtenant une excellente restauration de la parole. Leurs déficits se manifestent pour la détection du timbre, pour la perception de l'intonation, pour la reconnaissance de mélodies familières et surtout pour l'appréciation même de la musique. La perception du rythme pourrait quant à elle être relativement préservée, par la restitution partielle des indices temporels par l'implant.

Une population de sujets normo entendant présente ce type de déficit, centré sur la perception de la mélodie, limitant considérablement la reconnaissance de mélodies instrumentales et l'appréciation de la musique. Ces sujets, indemnes de toute pathologie auditive ou neurologique, souffrent donc d'une atteinte isolée de la perception musicale, appelée amusicie congénitale.

Initialement décrite en 1878 par Allen et également connue sous l'appellation « tone deafness » ou « note deafness » [Allen, 1878], l'amusicie congénitale toucherait près de 4% de la population générale. La symptomatologie reste discrète et de nombreux sujets atteints n'ont pas connaissance de leur condition car n'expriment rien d'autre qu'un profond désintérêt vis-à-vis de la musique. De nombreux travaux réalisés et/ou dirigés par I. Peretz depuis la fin des années 1990 ont permis de mieux connaître cette pathologie, tant au niveau étiologique que pour ses manifestations cliniques. Elle est ainsi définie comme un défaut de perception de la musique « qui ne peut être expliqué par une perte auditive, une lésion cérébrale, une déficience intellectuelle ou un manque d'exposition à la musique pendant l'enfance » [Hyde et al., 2011]. On sait également que ce déficit congénital est d'origine génétique [Peretz et al., 2007] et qu'il est caractérisé au niveau neuronal par des anomalies des réseaux fronto-temporaux [Hyde et al., 2007, Hyde et al., 2008, Tillmann et al., 2011a].

Si l'on considère isolément le déficit de perception musicale, les sujets implantés cochléaires pourraient a priori être rapprochés des sujets amusiques. C'est pourquoi nous avons évalué les capacités de perception musicale dans une population de sujets implantés cochléaires au moyen d'un test initialement développé afin de dépister les sujets amusiques. Notre étude sur la perception de la prosodie linguistique a mis évidence l'intérêt de l'audition résiduelle pour la détection de F0 et de ses variations dans la phrase et nous avons donc évalué son impact sur les capacités perceptives musicales des sujets implantés.

Outre l'éventuel bénéfice lié à la présence de restes auditifs, notre discussion s'attachera à déterminer les similitudes potentielles entre les sujets implantés cochléaires et les sujets amusiques, au niveau périphérique et neurofonctionnel.

## **2. MATERIELS ET METHODES**

### **2.1 Population**

Vingt sujets porteurs d'un implant cochléaire unilatéral ont été inclus dans cette étude. Cette population comprenait 12 femmes pour 8 hommes (sex ratio : 1.5), 11 implants cochléaires droits pour 9 implants cochléaires gauches. L'ensemble des données démographiques, cliniques et techniques relatives à l'implant utilisé est illustré dans le tableau VII.1. Tous les sujets ont été inclus après avoir reçu une information orale et écrite concernant l'étude, et signé un consentement écrit, ainsi que le prévoyait l'autorisation de réalisation de l'étude délivrée par le Comité de Protection des personnes du Sud-ouest et d'Outre-mer (n° : 08 161 03).

Nous avons déterminé les critères d'inclusions suivants :

- âge supérieur à 18 ans
- Acceptation du protocole et signature du consentement
- Durée de l'activation de l'implant cochléaire supérieure à 18 mois
- Discrimination correcte des mots dissyllabiques > 60%

Les deux derniers critères ont été établis afin de limiter le nombre de variables susceptibles d'intervenir sur les capacités perceptives musicales. Nous avons ainsi sélectionné des sujets présentant une discrimination de la parole satisfaisante afin de concentrer nos analyses sur la perception de la musique, et non sur les performances auditives globales. L'expérience avec l'implant est un facteur pronostique reconnu des résultats auditifs de l'implant et nous avons

limité son impact en sélectionnant des sujets implantés avec une durée d'activation supérieure à 18 mois.

Pour évaluer l'hypothèse d'un bénéfice lié à l'audition résiduelle, deux groupes ont été constitués en fonction de la présence ou de l'absence d'une audition acoustique résiduelle, utilisée en complément de la stimulation par l'implant. Nous avons ainsi comparé les performances obtenues dans un groupe de sujets implantés présentant une audition résiduelle (groupe dit bimodal, Bim) aux performances obtenues par un groupe de sujets implantés sans audition résiduelle (groupe dit implant cochléaire seul, CI only). Tous les sujets du groupe bimodal étaient porteurs d'une audioprothèse controlatérale à l'implant, dont le fonctionnement avait été vérifié avant la réalisation des tests. Leur niveau d'audition résiduelle était variable et les seuils auditifs mesurés en audiométrie tonale sont représentés dans le tableau VII.2.

Sujet	Âge	Cause de la surdité	Genre	Durée d'implantation (mois)	Côté IC	Marque Implant	Stratégie de codage
Bim 1	52	Inconnue	F	75	G	<i>Cochlear Hybrid L</i>	ACE 900 Hz
Bim 2	70	Lobstein	F	83	G	<i>Advanced Bionics</i>	HiRes S
Bim 3	63	Otospongiose	F	23	D	<i>Cochlear</i>	ACE 720 Hz
Bim 4	70	Inconnue	M	68	G	<i>Advanced Bionics</i>	HiRes S
Bim 5	64	Inconnue	F	37	G	<i>Advanced Bionics</i>	HiRes 120
Bim 6	47	Congénitale	F	35	D	<i>Cochlear CI 24</i>	ACE 720
Bim 7	47	Inconnue	M	130	D	<i>Cochlear CI 24</i>	ACE 900 Hz
Bim 8	56	Inconnue	F	62	D	<i>Cochlear CI 24</i>	ACE 900 Hz
Bim 9	78	Déhiscence	F	84	D	<i>Cochlear CI 24</i>	ACE 900 Hz
Bim 10	52	Inconnue	F	53	D	<i>Cochlear</i>	ACE 900 Hz
Bim 11	79	Inconnue	M	66	D	<i>Advanced Bionics</i>	HiRes 120
CI only 1	58	Otospongiose	M	37	G	<i>Advanced Bionics</i>	HiRes 120
CI only 2	46	Congénitale	M	114	D	<i>Cochlear CI 24</i>	ACE 720 Hz
CI only 3	50	Lobstein	F	36	D	<i>Med El Concerto</i>	FSP
CI only 4	45	Congénitale	F	84	D	<i>Cochlear CI 24</i>	ACE 900 Hz
CI only 5	71	Otospongiose	M	50	D	<i>Cochlear CI 24</i>	ACE 720 Hz
CI only 6	45	Inconnue	M	80	D	<i>Cochlear CI 24</i>	ACE 900 Hz
CI only 7	61	Otite chronique	F	17	G	<i>Cochlear CI 24</i>	ACE 900 Hz
CI only 8	76	Otite chronique	M	110	G	<i>Cochlear CI 24</i>	ACE 900 Hz
CI only 9	60	Inconnue	F	50	G	<i>Cochlear CI 24</i>	ACE 720 Hz

Tableau VII.1 : données démographiques et techniques relatives aux 20 sujets implantés cochléaires inclus et à l'implant utilisé

Sujet	125 Hz	250 Hz	500 Hz	750 Hz	1 kHz	2 kHz	4 kHz	8 kHz
Bim 1	40	50	65	75	85	x	x	x
Bim 2	na	90	85	90	95	85	95	x
Bim 3	90	90	90	90	90	95	110	x
Bim 4	30	40	50	60	x	x	x	x
Bim 5	50	55	70	80	90	120	x	x
Bim 6	80	80	90	120	120	x	x	x
Bim 7	40	55	100	110	115	x	x	x
Bim 8	20	20	40	75	95	x	x	x
Bim 9	70	80	80	80	80	80	70	75
Bim 10	25	25	45	50	60	80	x	x
Bim 11	90	85	75	85	90	95	115	x

Tableau VII.2 : seuils auditifs mesurés en audiométrie tonale oreilles nues, pour la meilleure oreille, chez les sujets implantés inclus dans le groupe de sujets avec audition résiduelle (groupe bimodal). Na : non renseigné. X : absence de réponse du sujet.

## 2.2 Appareillage et stimuli

### *Test de discrimination de la fréquence fondamentale*

Ce test a pour objectif de mesurer la distance minimale perceptible entre deux sons complexes de fréquences fondamentales différentes, comparables par ailleurs sur le plan acoustique. La littérature psycho acoustique utilise le terme de F0 difference limen (F0DL) pour faire référence à ce seuil de discrimination et nous emploierons cette appellation dans la suite de notre travail. Nous avons utilisé un test disponible on line, développé par Grassi et Soranzo [Grassi and Soranzo, 2009], qui permet de déterminer le F0DL pour plusieurs régions fréquentielles. Le F0DL a ainsi été mesuré pour les fréquences suivantes : 110-220-400-500-750 Hz.

Ce test utilise une méthode adaptative non paramétrique basée sur la maximum likelihood procedure (mlp), et est présenté sous Matlab® (the Mathworks, Inc., Natick, MA, USA). Les stimuli étaient des sons complexes à 5 harmoniques (1 F0 et 4 harmoniques), durant 250 ms chacun. Le test repose sur une procédure de choix forcé entre 3 stimuli, présentés 30 fois. Parmi ces trois stimuli, deux ont la même F0 et sont appelés F0 standards. Le troisième a une F0 plus élevée et est appelé F0 cible. Si le sujet identifie correctement le F0 cible deux fois de suite, la distance entre F0 standard et F0 cible est réduite. Si le sujet se trompe, la distance est augmentée. La procédure adaptative permet de déterminer le F0DL au terme des 30 présentations.

### *Test de dépistage de l'amusie*

Ce test a été développé par I. Peretz et son équipe (BRAMS, international laboratory for Brain, Music and Sound Research) et a été initialement conçu pour dépister les individus présentant une amusie congénitale [Peretz et al., 2008]. Il a fait l'objet d'une validation sur une cohorte de près de 300 sujets, après comparaison à l'évaluation par la Montreal Battery of Evaluation of Amusia [Peretz et al., 2003], qui représente le test de référence permettant d'étudier la perception musicale par différentes dimensions.

L'évaluation par MBEA permet ainsi d'illustrer la globalité des capacités musicales d'un sujet, qu'il soit normo entendant, amusique ou implanté cochléaire. Cette échelle comporte initialement six paramètres de mesure (échelle, contour, rythme, cadence, mémoire de travail) [Peretz et al., 2003]. Le test de dépistage de l'amusie est dérivé de l'échelle MBEA et repose sur la présentation de mélodies non familières, interdisant ainsi l'intervention d'un effet d'exposition antérieure. Il comprend 78 mélodies dérivées de 12 stimuli issus de la MBEA. Les mélodies ont été construites à partir des règles de musique occidentale, sur un mode majeur, avec une complexité suffisante pour acquérir une structure ayant un sens plutôt qu'une simple séquence de notes. Le nombre de notes au sein d'une mélodie variait entre 7 et 21, et le tempo variait entre 100 et 200 pulsations/minute. Les mélodies étaient jouées selon 10 timbres différents (piano, flûte, clarinette, cordes..).

Le test a été diffusé on line ( <http://www.brams.umontreal.ca/amusia-general/> ) et utilise l'encodage internet conventionnel (html, php). Les stimuli sont présentés au format MP3 afin d'être compatibles avec tous les équipements informatiques et d'occuper une mémoire réduite. Les réponses des sujets sont automatiquement enregistrées et classées par Microsoft Excel et Datafinder (Digimed systems Inc, 2006).

Ce test comprend trois sous-tests différents, chacun étant destiné à évaluer une dimension spécifique de la perception musicale.

#### *Sous - test 1 : Comparaison de deux mélodies*

Dans ce sous – test, les sujets devaient réaliser une tâche de discrimination de type même – différent. Deux mélodies étaient successivement présentées au sujet qui devait déterminer si elles identiques ou différentes. Trente paires de mélodies étaient ainsi présentées. Quinze d'entre elles étaient différentes et quinze étaient identiques (voir figure VII.1). La modification au sein d'une paire portait sur une seule note en modifiant sa fréquence tonale

(de 2 à 6 demi-tons) sans altérer le contour global de la mélodie : un contour ascendant (notes allant des graves aux aiguës) restait ascendant après modification, idem pour un contour descendant. Ce sous – test permet de rendre compte des capacités de perception de l’intervalle séparant deux notes. Dans la publication validant le test de dépistage on line, I. Peretz et ses collaborateurs rapportaient un pourcentage moyen de discrimination correcte de 90%.

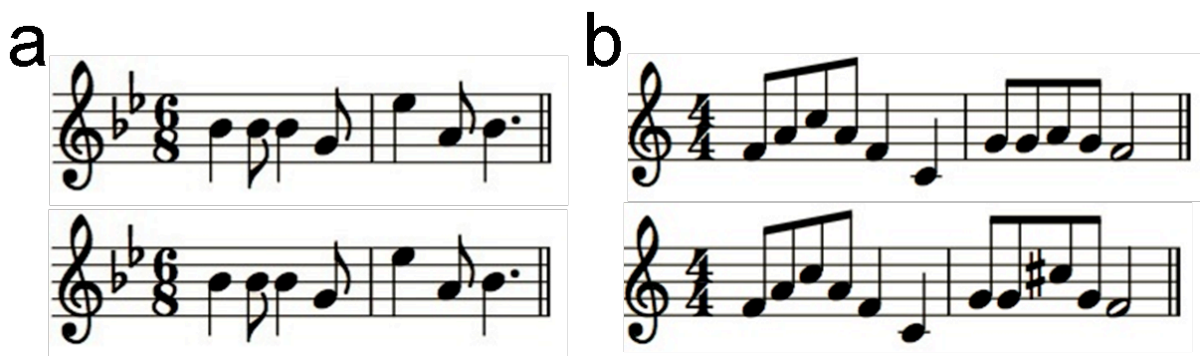


Figure VII.1. Exemples de paires de mélodies présentées au sujet, qui devait déterminer si les deux mélodies étaient identiques (a) ou différentes (b). La différence dans la deuxième mélodie ne portait que sur une note, séparée de la note dans la première mélodie par un intervalle variable allant de 2 à 6 demi-tons.

#### *Sous – test 2 : Rythme congruent / non congruent*

Ce sous – test était initialement réalisé à titre de contrôle dans le dépistage des sujets amusiques. En effet, ce déficit concerne uniquement le traitement des variations de F0 présentées dans une phrase musicale sans que la perception du rythme ou du timbre soit altérée.

Dans ce sous – test, les sujets devaient détecter un rythme non congruent dans une mélodie. Vingt quatre mélodies étaient successivement présentées aux sujets. Douze d’entre elles comprenaient une rupture de rythme, caractérisant le rythme non congruent, et les douze autres suivaient un rythme congruent. L’altération du rythme consistait en l’ajout d’un silence de 357 ms au sein de la mélodie (voir figure VII.2). Dans l’explication de la consigne au sujet, le rythme non congruent était défini comme une note « arrivant trop tôt » ou « trop tard » dans la mélodie. Chez le sujet normo entendant, ce sous – test peut poser problème aux amateurs de musique comportant naturellement des ruptures de rythme, comme dans certains morceaux de jazz ou dans certaines pièces de musique contemporaine. Leurs performances moyennes demeurent toutefois satisfaisantes avec plus de 80% de réponses correctes.



Figure VII.2. Exemple d'une mélodie suivant un rythme non congruent, comprenant un soupir réalisant une rupture dans le rythme global de l'extrait.

### *Sous – test 3 : Mélodie congruente / non congruente*

Dans ce sous – test, les sujets devaient détecter si la mélodie contenait une fausse note. Vingt quatre mélodies étaient successivement présentées aux sujets. Douze d'entre elles comprenaient une note n'appartenant pas à la tonalité de la mélodie et les douze autres ne contenaient pas de fausses notes. L'altération de la note consistait en son remplacement par une note un demi - ton en dessous ou en dessus, soit en dehors de la clé, rendant cette note dissonante au sein de la mélodie. Dans l'explication de la consigne au sujet, la mélodie dissonante était définie comme une mélodie « contenant une note incorrecte ». De la même manière que pour la perception du rythme, cette tâche peut dérouter chez les sujets normo entendant amateurs de musique contemporaine, à plus forte raison les amateurs de musique atonale. Mais les sujets normo entendant obtiennent tout de même un pourcentage moyen de reconnaissance correcte supérieur à 80%.

### *Performances attendues chez le sujet non amusique*

L'étape de validation initiale du test par I. Peretz et ses collaborateurs a donc nécessité l'inclusion d'une large cohorte (près de 300 sujets) pour étudier la distribution de leurs pourcentages de réponses correctes pour chacun des sous – tests et définir ainsi un niveau de performance normale. Ce niveau a été déterminé en fonction de l'âge des sujets (supérieur ou inférieur à 40 ans) et les auteurs ont estimé que les scores obtenus étaient anormalement bas, donc suspects d'amusie dans le contexte initial, s'ils étaient inférieurs à 2 écart-types en dessous de la moyenne. Les scores attendus en condition normale pour chacun des sous – tests sont représentés dans le tableau VII.3. Compte tenu de l'âge des sujets implantés cochléaires inclus dans l'étude, nous avons considéré le niveau de performance obtenu par les sujets de plus de 40 ans comme le niveau de référence.

Generation	Off-beat	Mistuned	Out-of-key	Global Score	Cut-off Score
Young (N= 158)	20.2 (2.1)	22.6 (1.8)	20.7 (2.7)	88.1% (7.2)	73.7%
Old (N= 65)	18.4 (2.2)	21.7 (2.0)	20.4 (3.1)	84.1% (7.0)	70.1%

Tableau VII.3. Scores moyens attendus présentés avec leurs écart-types pour chacun des sous – tests intégrant le test de dépistage on line. Il faut noter que les sujets âgés de plus de 40 ans présentent des scores légèrement inférieurs par rapport aux sujets plus jeunes, notamment pour la perception du rythme (off-beat).

### 2.3 Procédure et analyse des résultats

Tous les sujets implantés cochléaires ont été évalués dans la même cabine insonorisée du service d'ORL, CHU Purpan. Les performances des sujets pour la discrimination de la parole (mots monosyllabiques, mots dissyllabiques de la liste de Fournier présentés à 65 dB et discrimination des phrases dans le bruit à un rapport signal/bruit de +10 dB) étaient évaluées avant la passation du test de perception musicale.

Les informations et consignes relatives au test de dépistage de l'amusie ont été données avant chaque test et chaque sous test. Les stimuli acoustiques étaient transmis via un casque stéréophonique SENNHEISER HD 280 pro® à une intensité de 65 dB SPL, qui correspond au niveau d'intensité standard pour l'évaluation de la discrimination de la parole des patients après implantation cochléaire. Les sujets étaient tous évalués dans leurs meilleures conditions auditives, c'est-à-dire avec les réglages utilisés quotidiennement au niveau du processeur vocal et éventuellement avec leur audioprothèse controlatérale pour les sujets du groupe bimodal. Pour chaque tâche, le sujet donnait sa réponse oralement et l'expérimentateur validait sa réponse dans le logiciel. Aucune contrainte de temps n'était imposée. La séance durait environ 1 heure. La consigne était donnée avant chaque test et sous – test, et l'expérimentateur s'assurait que le sujet avait bien compris.

Les analyses statistiques descriptives et comparatives ont été réalisées à l'aide des logiciels Excel® et Statview®. L'étude comparative pour le F0DL a porté sur le rapport  $\Delta F/F$  standard exprimé en pourcentage et a été conduite sur l'ensemble des 5 régions fréquentielles testées et sur un F0DL moyen, correspondant à la moyenne des différents F0DL obtenus pour chacune des F0 standards évaluées. Pour l'analyse comparative des performances obtenues aux différents sous – tests du test de dépistage de l'amusie, nous avons réalisé des comparaisons portant sur le pourcentage de réponses correctes, puis sur le  $d'$ , indice de sensibilité calculé à l'aide de l'équation de Creelman.



La variabilité des valeurs obtenues au sein des deux groupes de sujets implantés limitait la pertinence d'analyses paramétriques et notre étude comparative des performances pour chaque sous – test entre le groupe bimodal et le groupe s'est appuyée une comparaison des intervalles de confiance à 95% des valeurs obtenues dans chaque groupe, après ré échantillonnage par bootstrapping. La recherche de corrélation entre les performances obtenues et le niveau d'audition résiduelle a été réalisée dans le groupe bimodal en calculant le coefficient de corrélation  $r$  et son niveau de significativité par un test de  $r$  en  $z$ . La recherche d'une relation entre les performances obtenues et la durée d'expérience avec l'implant a été réalisée dans les deux groupes en calculant le coefficient de corrélation  $r$  et son niveau de significativité. De la même manière, la relation potentielle entre le niveau de discrimination de la parole et les capacités musicales présentées a été évaluée par un test de  $r$  en  $z$ .

Des analyses complémentaires après ré échantillonnage ont été réalisées sur les principales caractéristiques des deux groupes de sujets implantés afin de s'assurer de leur comparabilité. La durée de surdité profonde, le délai post activation et les performances pour la discrimination de la parole (discrimination des mots monosyllabiques et dissyllabiques dans le silence, discrimination des phrases dans le bruit avec un rapport signal/bruit de +10 dB) ont ainsi été analysées.

### **3. RESULTATS**

#### **3.1 Description des groupes**

Les analyses descriptives des deux groupes de sujets implantés avaient pour premier objectif de s'assurer de leur comparabilité pour les principaux facteurs pronostiques reconnus de l'implantation cochléaire. Comme pour l'étude portant sur la prosodie linguistique et pour les mêmes raisons, la durée moyenne de surdité profonde n'a pas été intégrée à nos analyses. Il est en effet difficile de documenter précisément le moment où la perte auditive moyenne dépasse les 90 dB de perte et certains sujets du groupe bimodal présentaient une audition résiduelle controlatérale à l'implant substantielle, et ne pouvaient donc être considérés *stricto sensu* comme sourds profonds.

Les sujets implantés étaient comparables pour le délai post activation, témoin de l'expérience avec l'implant, avec une moyenne de 65 mois (intervalle de confiance à 95% : 50.6-84.1 mois) dans le groupe bimodal et une moyenne de 64.2 mois (intervalle de confiance à

95% :44.5-86) dans le groupe CI only. De la même manière, l'âge moyen et la distribution de ses valeurs étaient comparables entre les deux groupes (61.6ans, intervalle de confiance à 95% :50.3-68.4 dans le groupe bimodal ; 56.9 ans, intervalle de confiance à 95% : 50.5-64.5 dans le groupe CI only).

Les scores de discrimination de la parole étaient également comparables d'un groupe à l'autre, quel que soit le type de matériel utilisé. Les intervalles de confiance à 95% obtenus pour la distribution des scores obtenus pour la discrimination dans le silence des mots monosyllabiques et dissyllabiques ou la discrimination des phrases dans le bruit n'étaient ainsi pas différents entre les deux groupes, bien que les sujets du groupe bimodal tendent à obtenir de meilleures performances. Les résultats pour la discrimination de la parole sont illustrés dans la figure VII.3.

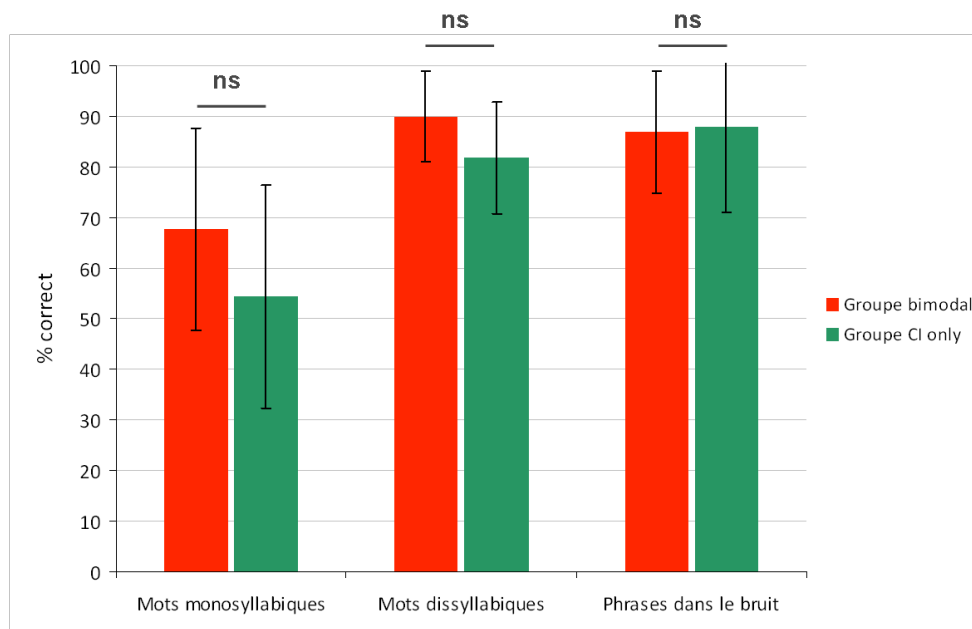


Figure VII.3. Pourcentage de discrimination correcte de la parole avec trois matériels de test différents. Il n'existe pas de différence significative entre le groupe bimodal et le groupe CI only, bien que les performances tendent à être meilleures dans le groupe bimodal.

### 3.2 Perception de la fréquence fondamentale : F0DL

Les capacités de discrimination de F0 ont été évaluées pour 5 fréquences standard différentes et comparées entre les groupes bimodal et CI only. Que ces analyses concernent les 5 F0DL ou le F0DL moyen obtenus, elles ne mettaient pas en évidence de différence significative entre les deux groupes de sujets implantés. Ces résultats ne contredisent pas nécessairement les résultats obtenus pour le même test lors de l'étude précédente sur la perception de la

prosodie linguistique. Nous rappelons en effet que l’audition résiduelle présentée par les sujets inclus dans le groupe bimodal ici était moins bonne que celle présentée par les sujets EAS et Bim de l’étude précédente.

Le tableau VII.4 récapitule les FODL mesurés pour chacune des fréquences standard testées ainsi que le FODL moyen. L’analyse comparative par bootstrap n’a mis en évidence aucune différence significative entre les deux groupes. Le F0 DL moyen était ainsi évalué à 11.3% (intervalle de confiance à 95% : 7.2-16.9) dans le groupe bimodal et à 16.5% (intervalle de confiance à 95% : 12.1-22.7) dans le groupe CI only.

	FODL 110Hz	FODL 220Hz	FODL 400Hz	FODL 500Hz	FODL 750Hz	F0 DL moyen
Bim	13.1% (8.6)	7.2% (5.3)	8.2% (6.2)	8.8% (7.1)	19.1% (27.7)	11.3% (8.3)
CI only	25.8% (13.2)	15% (12.2)	9.8% (7.1)	16.6% (12.8)	15.2% (15.8)	16.5% (8.4)

Tableau VII.4. Valeurs des FODL pour chacune des fréquences standards et du FODL moyen évaluées en fonction du groupe considéré. Il n’existait pas de différence significative entre les deux groupes quelle que soit la valeur considérée.

### 3.3 Perception de la musique par le test Amusia on line

Les résultats obtenus pour chacun des sous – tests par les sujets implantés ont été comparés aux résultats attendus d’une population sans déficit de perception musicale, tels qu’ils ont été définis lors de la validation de ce test. Ces résultats sont illustrés par la figure VII.4.

Les pourcentages de réponses correctes ont par la suite été convertis en d’ dans les deux groupes de sujets implantés (voir figure VII.5). Pour chacun des sous – tests, nous avons évalué la relation entre les performances obtenues et le délai post activation pour les deux groupes, entre les performances obtenues et le niveau d’audition résiduelle sur les fréquences 125 à 500 Hz pour le groupe bimodal.

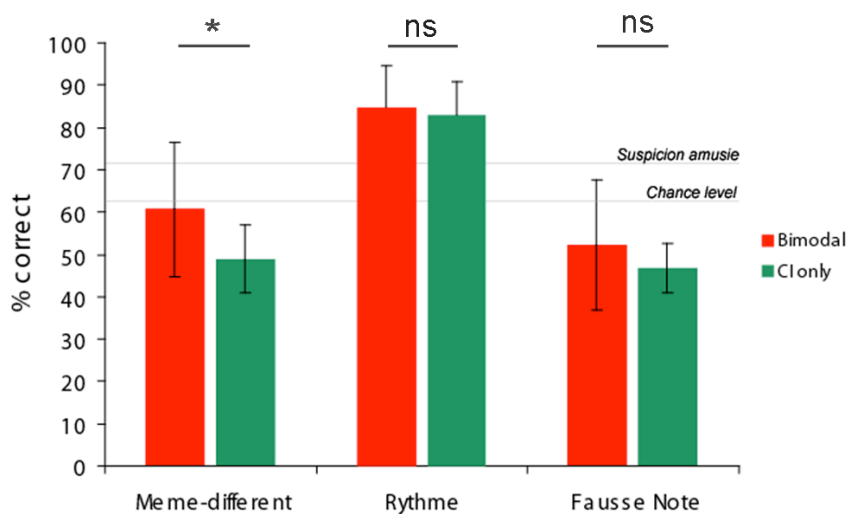


Figure VII.4. Pourcentage de reconnaissance correcte pour chacun des trois sous – tests réalisés avec représentation du niveau de chance et du seuil en deçà duquel les sujets normo entendant étaient considérés comme potentiellement amusiques dans la validation du test comme dépistage.

### ***Sous – test 1 : Comparaison de deux mélodies (même-différent)***

#### *Analyses comparatives entre les deux groupes de sujets implantés*

Les résultats obtenus par les deux groupes de sujets implantés pour ce sous – test font état d’un déficit de discrimination entre deux mélodies par rapport aux sujets normo entendant. Ce déficit est tel que leurs performances ne pouvaient être considérées comme différentes du niveau de chance. Dès lors, les analyses comparatives entre les deux groupes ont dû être interprétées avec précaution, en gardant à l’esprit l’importance du déficit quel que soit le groupe considéré.

La comparaison par bootstrap des pourcentages de réponses correctes lors de ce sous – test mettait en effet en évidence des performances légèrement supérieures dans le groupe bimodal par rapport au groupe CI only, avec une moyenne de 60.6% (intervalle de confiance à 95% : 54.3-72) contre 48.9% (intervalle de confiance à 95% : 44.3-53.2). Cette supériorité était également retrouvée lorsque les pourcentages étaient convertis en valeurs de  $d'$  pour s’intéresser à la sensibilité des sujets, en laissant de côté le biais lié à leur attitude pendant le test (voir figure VI.5). L’analyse par  $d'$  confirmait par ailleurs que les performances du groupe CI only ( $d'$  moyen : -0.25 ; intervalle de confiance à 95% : -0.86, 0.14) ne pouvaient être considérées comme statistiquement différentes du niveau de chance. Les valeurs de  $d'$

négatives chez certains sujets traduisaient l'existence de taux de fausses alarmes supérieurs aux taux de détection correcte chez les sujets concernés.

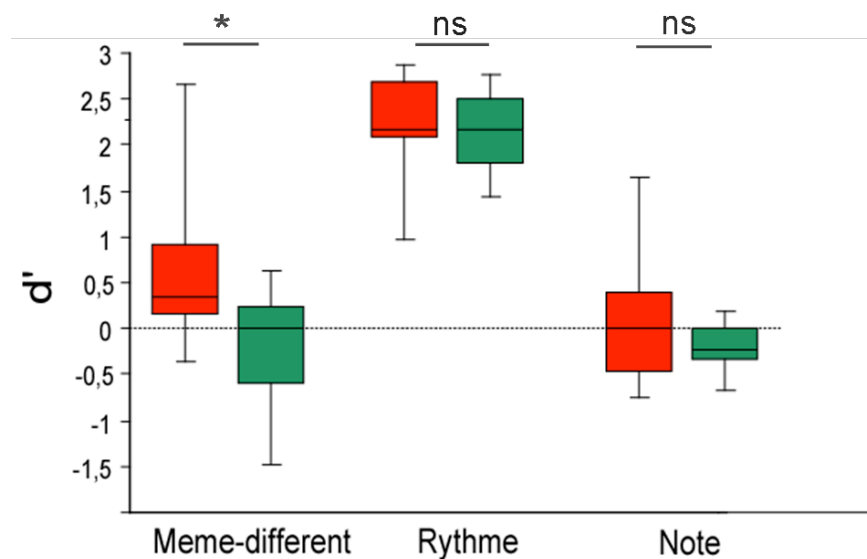


Figure VII.5. Valeurs de  $d'$  obtenues pour le groupe bimodal (rouge) et le groupe CI only (vert) pour chacun des sous – tests réalisés. Les performances du groupe bimodal étaient supérieures pour la comparaison de deux mélodies.

La valeur moyenne de  $d'$  dans le groupe bimodal (0.69) et son intervalle de confiance à 95% (0.22-1.5) étaient en revanche supérieurs au niveau de chance, définis par un  $d'$  égal à 0.

En nous appuyant sur les résultats obtenus lors de l'étude sur la perception de la prosodie linguistique, nous avons par la suite évalué la relation entre les performances ( $d'$ ) et le niveau d'audition résiduelle, ainsi que la durée d'expérience avec l'implant.

#### *Impact de l'audition résiduelle et de l'expérience avec l'implant*

Le niveau d'audition résiduelle moyen sur les basses fréquences (125-500 Hz) dans le groupe bimodal et les valeurs du  $d'$  pour la comparaison de deux mélodies étaient corrélées, ce dont témoignait le coefficient de corrélation  $r$ , calculé à -0.64 (test de  $r$  en  $z$  ;  $p=0.03$ ). Plus la perte auditive était importante, moins la comparaison de deux mélodies était donc performante (voir figure VII.6). Il n'existait en revanche pas de relation entre les valeurs de  $d'$  et l'expérience avec l'implant, évaluée par le délai post activation ( $r$  : 0.2 ; test de  $r$  en  $z$  ;  $p=0.3$  pour l'ensemble des sujets implantés), que ce soit dans le groupe bimodal ( $r = 0.14$  ;  $p=0.7$ ) ou dans le groupe CI only ( $r = 0.36$  ;  $p=0.35$ ).

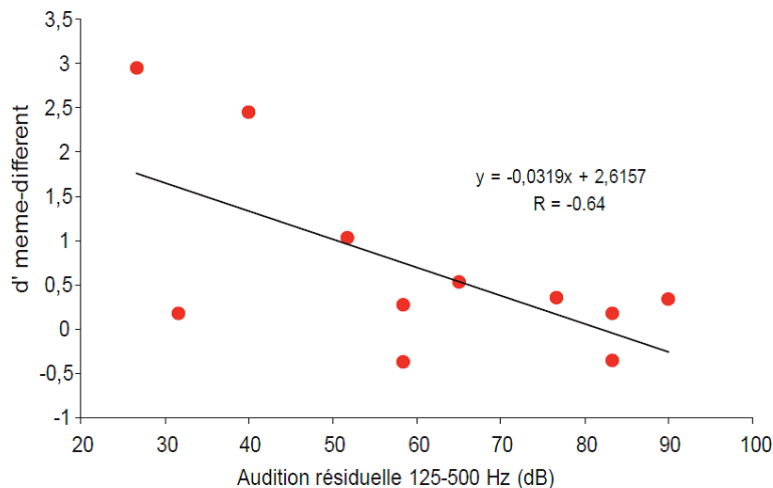


Figure VII.6. Etude de la relation entre le niveau d'audition résiduelle moyen calculé sur les fréquences 125 à 500 Hz dans le groupe bimodal et la performance pour la comparaison de deux mélodies représentée par la valeur du d'.

Nous avons ensuite cherché à déterminer, parmi les paires de mélodies différentes, si l'intervalle séparant la note modifiée entre la première et la deuxième mélodie avait un impact sur les performances des sujets implantés.

#### *Impact de l'intervalle pour la note modifiée entre les deux mélodies*

Parmi les 30 paires de mélodies présentées aux sujets, quinze d'entre elles contenaient deux mélodies identiques et les quinze autres contenaient deux mélodies différentes l'une de l'autre. Dans ce cas, la modification avait porté sur une seule note, en respectant le contour global (ascendant ou descendant) de la mélodie. Le nombre de demi-tons modifiant la note concernée variait d'une paire de mélodies dissemblables à l'autre et nous avons donc recherché une relation entre l'intensité de la modification de la note en demi-tons et le pourcentage de réponses correctes obtenues par les sujets. Le coefficient de corrélation était de 0.2 (test de r en z ; p=0.4) pour la population globale des sujets implantés, de -0.05 (p=0.8) si l'on considérait isolément le groupe bimodal, et enfin de 0.4 (p=0.1) si l'on considérait isolément le groupe CI only. Il n'y avait donc pas d'impact de l'intervalle pour la note modifiée entre les deux mélodies.

### *Sous – test 2 : Perception du rythme*

Cette tâche avait été intégrée au test Amusia on line au titre de tâche contrôle. Les sujets amusiques présentent en effet un déficit isolé du traitement des variations de F0 au sein de la mélodie, avec une perception des schémas rythmiques demeurant normale.

Dans notre étude, les sujets implantés cochléaires obtenaient également des performances comparables à celles des sujets normo entendant non amusiques dans cette tâche. Le pourcentage moyen de réponses correctes était ainsi de 84.5% (intervalle de confiance à 95% : 77.5-89.1) dans le groupe Bim et de 82.9% dans le groupe CI only (intervalle de confiance à 95% : 77.3-87.2). Une fois convertis en  $d'$ , ces scores demeuraient supérieurs aux scores obtenus dans les deux autres sous - tests ( $d'$  moyen : 2.18 pour le groupe bimodal et 2.15 pour le groupe CI only). Les deux groupes de sujets implantés présentaient donc des performances comparables à celles des normo entendant et comparables entre elles.

Ces résultats étaient attendus dans la mesure où la majorité des travaux sur la perception du rythme par les sujets implantés rapporte de bonnes performances pour sa détection [Schulz and Kerber, 1994, Gfeller et al., 1997, Gfeller et al., 2000] même avec les premières générations d'implants multi canaux [Gfeller and Lansing, 1991]. Cette préservation de la perception du rythme peut être expliquée par la qualité de la transmission des indices temporels liés à l'enveloppe du signal acoustique par l'implant. Ces indices sont essentiels à la discrimination de la parole, de la même manière qu'ils autorisent une perception normale de la structure rythmique présente dans les signaux de musique.

Pour ce sous – test, nous avons réalisé une analyse complémentaire visant à étudier la relation entre les valeurs du  $d'$  et leurs performances pour la discrimination de la parole. Cette analyse permettait d'évaluer l'hypothèse de représentations mentales distinctes de la structure rythmique pour la parole et pour la musique. Seules les performances du groupe CI only ont été incluses dans cette analyse, dans la mesure où les scores des sujets bimodaux pouvaient être influencés par l'audition résiduelle contralatérale à l'implant. Si l'on prenait en compte les performances obtenues à l'ensemble des tests de discrimination de la parole (mots monosyllabiques et dissyllabiques, phrases dans le bruit), on obtenait une relation avec les valeurs du  $d'$  pour le rythme qui tendait à être significative statistiquement (test de  $r$  en  $z$ ,  $r : 0.37$ ,  $p=0.051$ ). Une deuxième analyse de ce type n'a pas pris en compte la discrimination des phrases dans le bruit, test faisant largement intervenir les processus top down liés au contexte sémantique, et n'a pu établir de relation entre la discrimination de la parole (mots mono et dissyllabiques) et la perception du rythme ( $r : 0.29$ ,  $p=0.24$ ).

La structure rythmique d'un signal acoustique pourrait solliciter deux types de représentations mentales différentes selon qu'il s'agit d'un signal de musique ou d'un signal de parole.

***Sous – test 3 : Mélodie congruente / non congruente***

Les capacités de détection d'une fausse note apparaissaient particulièrement altérées dans les deux groupes de sujets implantés. Le pourcentage moyen de détection correcte était ainsi inférieur au niveau de la chance pour les deux groupes (52.3%, intervalle de confiance à 95% : 45-62.4 pour le groupe bimodal, 46.8% de détection correcte, intervalle de confiance à 95% : 41.6-49.6 pour le groupe CI only). Ces mauvaises performances étaient également illustrées par des valeurs de  $d'$  particulièrement faibles, atteignant des valeurs négatives pour 4 sujets dans le groupe bimodal et 7 des 9 sujets dans le groupe CI only. Ces valeurs négatives traduisaient encore une fois l'important taux de fausses alarmes dans les réponses des sujets implantés. Ainsi, le  $d'$  moyen était de 0.14 (intervalle de confiance à 95% : -0.27 ; 0.7) dans le groupe Bim et de -0.19 (intervalle de confiance à 95% : -0.43 ; -0.02) dans le groupe CI only. Les deux groupes de sujets implantés présentaient des performances particulièrement déficitaires pour la détection de fausses notes. Dans ce contexte, l'évaluation de l'impact de l'audition résiduelle dans le groupe bimodal était donc à interpréter avec précaution, dans la mesure où il est difficile de relier des capacités presque absentes à une autre variable. Deux sujets bimodaux présentaient toutefois des valeurs de  $d'$  (1.62) indiquant qu'ils avaient réussi, dans une certaine mesure, à réaliser la tâche. Nous avons donc recherché une corrélation entre les valeurs de  $d'$  dans ce sous – test et le niveau d'audition résiduelle moyen sur les fréquences graves (125-500 Hz). Cette relation était significative (voir figure VII.7) avec un coefficient de corrélation  $r$  de -0.79 ( $p=0.002$ ).



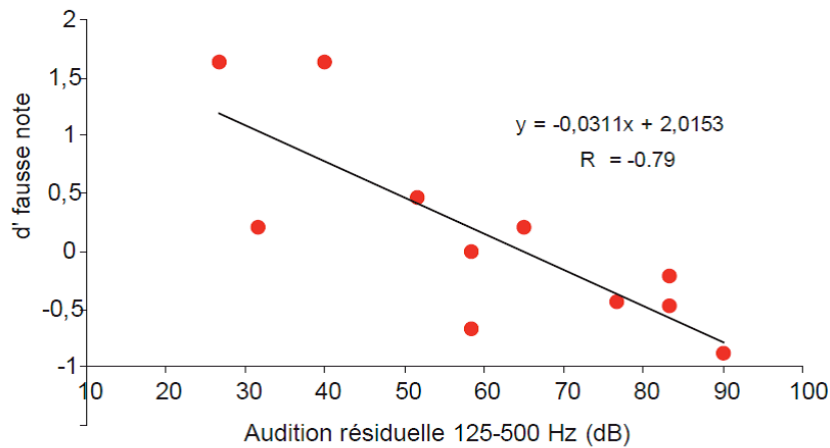


Figure VII.7. Etude de la relation entre le niveau d'audition résiduelle moyen calculé sur les fréquences 125 à 500 Hz dans le groupe bimodal et la performance pour la détection de fausses notes représentée par la valeur du d'.

Il n'y avait pas de corrélation significative entre le F0DL moyen et les performances obtenues dans cette tâche (pour la valeur du d' :  $r = -0.43$  ;  $p = 0.19$ ). Nous avons toutefois observé une forte corrélation entre les scores de d' obtenus dans cette tâche et ceux obtenus dans la tâche de comparaison de mélodies, qui implique le traitement d'une séquence de F0 différentes. Le coefficient de corrélation était alors  $r = 0.9$  ( $p < 0.0001$ ), traduisant une relation entre ces deux tâches faisant intervenir le traitement de F0.

Il est difficile d'apporter des conclusions solides à l'examen des résultats obtenus par les deux groupes de sujets implantés pour cette tâche, en dehors de l'important déficit qu'ils présentent. Deux sujets du groupe bimodal, présentant une audition résiduelle meilleure que 40 dB de perte sur les fréquences graves ont, dans une certaine mesure (pourcentage de réponses correctes : 79%), réussi la tâche. Les performances de ces deux sujets ont en effet dépassé le niveau de la chance mais également le seuil en deçà duquel une amusie est suspectée chez les sujets normo entendant.

## **4. DISCUSSION**

### **4.1 La perception de la musique chez le sujet implanté cochléaire**

Nos résultats apparaissent en accord avec les données de la littérature, constatant l'étendue des déficits présentés par les patients implantés cochléaires pour la perception de la musique.

Le principal facteur généralement incriminé dans ce déficit est la mauvaise restitution de F0 par l'implant cochléaire. Ce défaut de transmission semble en effet être l'origine des différents déficits présentés par les sujets implantés, depuis la perception jusqu'à la reconnaissance de mélodies familières ; la principale plainte des patients résidant dans l'absence de plaisir à l'écoute de la musique.

La reconnaissance de mélodies familières est un paradigme souvent utilisé pour mettre en évidence les difficultés des sujets implantés. Si les pourcentages de reconnaissance varient d'une étude à l'autre, de 15.6% pour Gfeller et al. [Gfeller et al., 2005] à 63.2% pour Kong et al. [Kong et al., 2004], tous témoignent de l'importance du déficit [Fujita and Ito, 1999, Gfeller et al., 2002a, Leal et al., 2003]. Ces études soulignent en règle générale l'impact naturellement positif des paroles sur la reconnaissance mais également l'effet du rythme, qui, s'il est modifié, accompagne une chute des performances [Kong et al., 2004]. Nos résultats montrent, dans cette tâche, une perception préservée du rythme qui constitue donc le principal indice sur lequel s'appuyer pour reconnaître les mélodies.

Le ressenti à l'écoute musicale a également fait l'objet de plusieurs travaux, comprenant généralement des échelles d'appréciation subjective. Ces travaux font généralement état de l'absence de plaisir à l'écoute de la musique avec près de 90% des patients qui rapportent avoir diminué leur écoute musicale après implantation et 38% qui affirment ne plus aimer la musique [Leal et al., 2003]. Mirza et al. [Mirza et al., 2003] ont évalué rétrospectivement le plaisir de 35 patients implantés à l'écoute de la musique sur une échelle visuelle analogique cotée de 0 à 10 et ont mis en évidence une chute du score moyen de 8.7 avant le début de la surdité à 2.6 après implantation. Le temps passé à écouter de la musique après implantation est logiquement corrélé au plaisir ressenti [Gfeller et al., 1998].

Le ressenti à l'écoute de la musique n'a en revanche pas de relation stricte avec les performances pour la reconnaissance d'instruments. Ainsi que pour la fréquence fondamentale, la perception du timbre est sévèrement altérée chez les sujets implantés avec

des performances généralement inférieures de moitié par rapport à celles des sujets contrôle normo entendant [Gfeller et al., 2002b], proches du niveau de la chance.

Si l'on considère nos résultats et les données concordantes issues de la littérature, il convient d'admettre que la perception musicale au travers de l'implant demeure un échec chez les sujets sourds profonds post-linguaux. Certains « star patients » affirment avoir retrouvé le goût de l'écoute musicale grâce à une auto-rééducation soutenue, basée sur l'écoute répétée de chansons appréciées avant la période de surdité profonde afin d'y identifier progressivement le signal correspondant à chaque instrument. Mais ces expériences restent rares et l'appui sur la mémoire auditive est une piste de rééducation à explorer sur le plan fondamental avant de le diffuser.

Les patients présentant une audition résiduelle, représentés par le groupe bimodal dans notre étude, pourraient bénéficier des informations acoustiques transmises par les fréquences graves. Les sujets bimodaux obtenaient ainsi de meilleurs scores que les sujets sans audition résiduelle, et des scores supérieurs au niveau de la chance si l'on considère le  $d'$ , pour une tâche faisant intervenir la perception de la mélodie. L'audition résiduelle transmet ainsi vraisemblablement l'information relative à F0 et aux premières harmoniques nécessaire à la sensation de hauteur tonale.

## **4.2 Intérêts de l'audition résiduelle**

La perception des changements de F0 est en effet facilitée par l'audition résiduelle controlatérale à l'implant ou par une audition résiduelle bilatérale. Notre étude sur la perception de la prosodie linguistique a démontré que les sujets implantés avec les restes auditifs les plus significatifs pouvaient atteindre des performances comparables à celles des normo entendant pour la perception de F0 et de l'intonation dans la phrase. La littérature portant sur la perception de la musique a également mis en évidence l'intérêt de l'audition résiduelle. Gfeller [Gfeller et al., 2007] et al ont ainsi retrouvé de meilleures performances pour la perception de l'intervalle séparant deux notes dans un groupe de 21 sujets bénéficiant d'une stimulation électro acoustique (audition résiduelle bilatérale) comparés à des sujets implantés sans audition résiduelle.

Certaines informations relatives au timbre peuvent également être transmises par les résidus auditifs sur les fréquences graves. Les premières harmoniques peuvent jouer un rôle important dans la reconnaissance des instruments et l'audition résiduelle sur les fréquences graves peut

permettre de les restituer au sujet implanté. Les patients EAS évalués par Gfeller et al. présentaient ainsi des scores supérieurs à 60% d'identification correcte, supérieurs à ceux de sujets implantés sans audition résiduelle dans la même tâche (environ 40%) et deux fois meilleurs que les taux d'identification correcte rapportés par Inverso et Limb [Inverso and Limb, 2010] dans leur étude sur la catégorisation (voir III. Catégorisation libre par le sujet implanté cochléaire). Ce taux d'identification était dépendant de la fréquence fondamentale à laquelle les mélodies étaient jouées, les instruments avec F0 basse (entre 131 et 262 Hz) étaient reconnus avec plus d'acuité que lorsqu'ils jouaient une mélodie avec une F0 plus élevée.

La question du seuil d'audition résiduelle utile pour la perception de la musique a été abordée dans une étude réalisée par notre équipe [El Fata et al., 2009], qui a comparé deux groupes de sujets implantés avec audition résiduelle en fonction de leurs seuils auditifs entre 125 et 1000 Hz (supérieur ou inférieur à 85 dB de perte auditive moyenne). La reconnaissance de mélodies familières purement instrumentales était supérieure dans le groupe d'implantés avec l'audition résiduelle meilleure que 85 dB et reposait presque exclusivement sur ces restes auditifs, dans la mesure où le taux de reconnaissance correcte n'était pas modifié si l'implant n'était pas utilisé et chutait en revanche si l'oreille partiellement fonctionnelle était bouchée (voir figure VII.8).

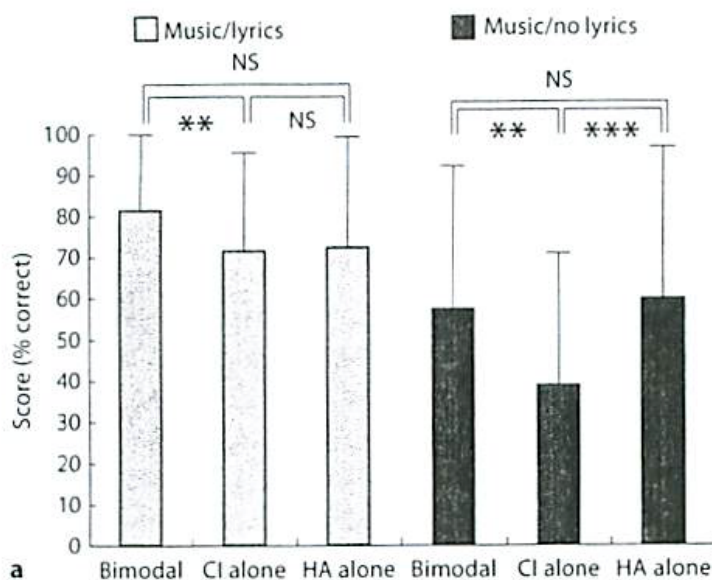


Figure VII.8. Pourcentage de reconnaissance de mélodies familières avec (gauche) et sans (droite) paroles chez 8 sujets implantés avec une audition résiduelle meilleure que 85 dB de perte. La reconnaissance des extraits sans paroles s'appuie exclusivement sur l'audition résiduelle et pas sur l'information transmise par l'implant. Tiré de ElFata et al. (2009)

Notre étude ne prévoyait pas de tâches de reconnaissance de mélodies familières mais évaluait des dimensions plus basiques de la perception musicale. Il apparaît difficile d'extraire de nos résultats une perte auditive seuil en deçà de laquelle l'audition résiduelle n'apporte pas réellement d'information complémentaire à l'implant pour la musique. Nous avons toutefois démontré que les performances pour les tâches impliquant la perception de F0 étaient dépendantes du niveau d'audition résiduelle.

Les sujets implantés cochléaires sans audition résiduelle présentaient quant à eux des résultats particulièrement déficitaires, ne dépassant pas le niveau de la chance dans les deux tâches impliquant la perception de F0 (comparaison de deux mélodies et détection de fausse note). Ce déficit, quand il est isolé chez les sujets normo entendant sans antécédents médicaux, peut conduire à suspecter une amusie congénitale. Ces deux populations, sujets amusiques et sourds implantés cochléaires, présentent en effet plusieurs similitudes à souligner.

### **4.3 Similitudes entre les sujets implantés cochléaires et sujets amusiques**

L'amusie se manifeste typiquement par l'incapacité de distinguer deux sons dont la F0 est séparée d'un demi-ton, ce qui correspond à deux notes adjacentes sur un clavier de piano, soit un FODL de 8.3% [Foxton et al., 2004]. Ce déficit est généralement isolé et centré sur l'écoute musicale, bien qu'il ait été rapporté un déficit limité de perception de la prosodie linguistique [Hutchins et al., 2010] ainsi qu'un défaut de sensibilité aux langues tonales [Tillmann et al., 2011b] chez plusieurs sujets amusiques.

Les corrélats neuronaux de l'amusie congénitale ont été étudiés sur le plan anatomique avec des anomalies mises en évidence dans des régions fronto-temporales droites, portant sur les concentrations relatives en substance blanche et en substance grise. Plusieurs travaux font ainsi état d'une concentration réduite de substance blanche avec une augmentation proportionnelle de la concentration en substance grise, dans les gyri frontal inférieur [Hyde et al., 2006] et temporal supérieur [Hyde et al., 2007, Albouy et al., 2013] droits. Ces anomalies ont été constatées alors que l'activité des cortex auditifs temporaux en réponse à des changements de F0, même minimes, s'est révélée comparable entre sujets amusiques et sujets contrôles dans plusieurs études utilisant l'IRMf [Hyde et al., 2011] ou la MMN en électro

physiologie [Moreau et al., 2009]. Les régions neuronales auditives bas niveau se voient donc transmettre l'information relative aux changements de F0 mais des centres de plus haut niveau ne peuvent l'intégrer. L'amusie congénitale pourrait dès lors être secondaire à un défaut de connectivité entre le gyrus frontal inférieur droit et le gyrus temporal supérieur droit, rendant impossible l'intervention de réseaux attentionnels dans le traitement des variations de F0. Cette hypothèse est appuyée par les travaux de Loui et al. [Loui et al., 2009] sur la connectivité fronto-temporale, diminuée chez les sujets amusiques par une altération fonctionnelle du faisceau arqué.

Sur le plan comportemental, les déficits présentés par les sujets amusiques et les sujets implantés cochléaires sont incontestablement différents. Le patient implanté est en difficulté pour percevoir la parole en présence de bruit concurrent, pour percevoir et catégoriser les bruits environnementaux, pour les informations prosodiques. Mais la cause commune aux déficits présentés par les deux populations pourrait être le déficit de perception de F0. Ce déficit demeure, il est vrai, essentiellement focalisé sur l'écoute musicale chez les sujets amusiques. Ce déficit est comparable entre les deux populations dans la mesure où il existe pour le traitement de séquences variations de F0, en plus du déficit simple de discrimination entre deux F0 différentes (voir Foxton, 2004 pour les amusiques ; Cousineau 2010 pour les sujets implantés). Les similitudes ne s'arrêtent pas à ce déficit commun. On peut en effet évoquer les capacités de perception du rythme, relativement préservées dans les deux populations. Il serait en effet inexact d'affirmer le caractère strictement normal de perception du rythme chez les sujets implantés, l'identification de subtiles modifications dans des schémas rythmiques diminuant leurs capacités d'identification [Kong et al., 2004]. Mais nos résultats, en accords avec d'autres travaux [Gfeller and Lansing, 1991, Schulz and Kerber, 1994, Gfeller et al., 1997] ne montrent pas de différence avec les sujets normo entendant. De plus, le déficit de traitement des variations de F0 chez les sujets amusiques pourrait influencer négativement leurs capacités perceptives pour le rythme et être à l'origine de discrets déficits [Foxton et al., 2006].

Il est difficile de statuer sur les similitudes entre amusie et implantation cochléaire sur le plan neuronal, du fait du manque d'études sur les corrélats neuronaux de la perception musicale chez le sujet implanté et de leurs limites techniques. Le défaut de traitement central des variations de F0 au sein d'une mélodie serait toutefois différent entre les deux populations. On sait que ce déficit se produit pour de hauts niveaux de traitement auditif chez les sujets amusiques, dans la mesure où des changements minimes de F0 produisent une réponse neuronale pré attentive de type MMN [Moreau et al., 2009, Peretz et al., 2009]. Il n'existe en

revanche pas de réponse neuronale pour les niveaux conscients supérieurs, ainsi qu'en témoigne l'absence d'onde P3 pour les changements de F0 inférieurs à un demi-ton dans l'étude de Moreau (2009). Chez les sujets implantés, ce déficit interviendrait à un niveau inférieur, pré attentif. Dans les travaux de Sandmann et al. [Sandmann et al., 2010] (2010), la réponse neuronale pré attentive de type MMN aux changements de F0 était en effet inférieure chez les sujets implantés par rapport aux sujets normo entendant en condition originale mais également en condition de simulation d'implant par vocoder (voir figure VII.9). Une dégradation plus importante du signal aurait cependant été nécessaire pour garantir la comparabilité des résultats avec les sujets implantés cochléaires, dont on sait qu'ils fonctionnent avec 8 à 10 canaux actifs.

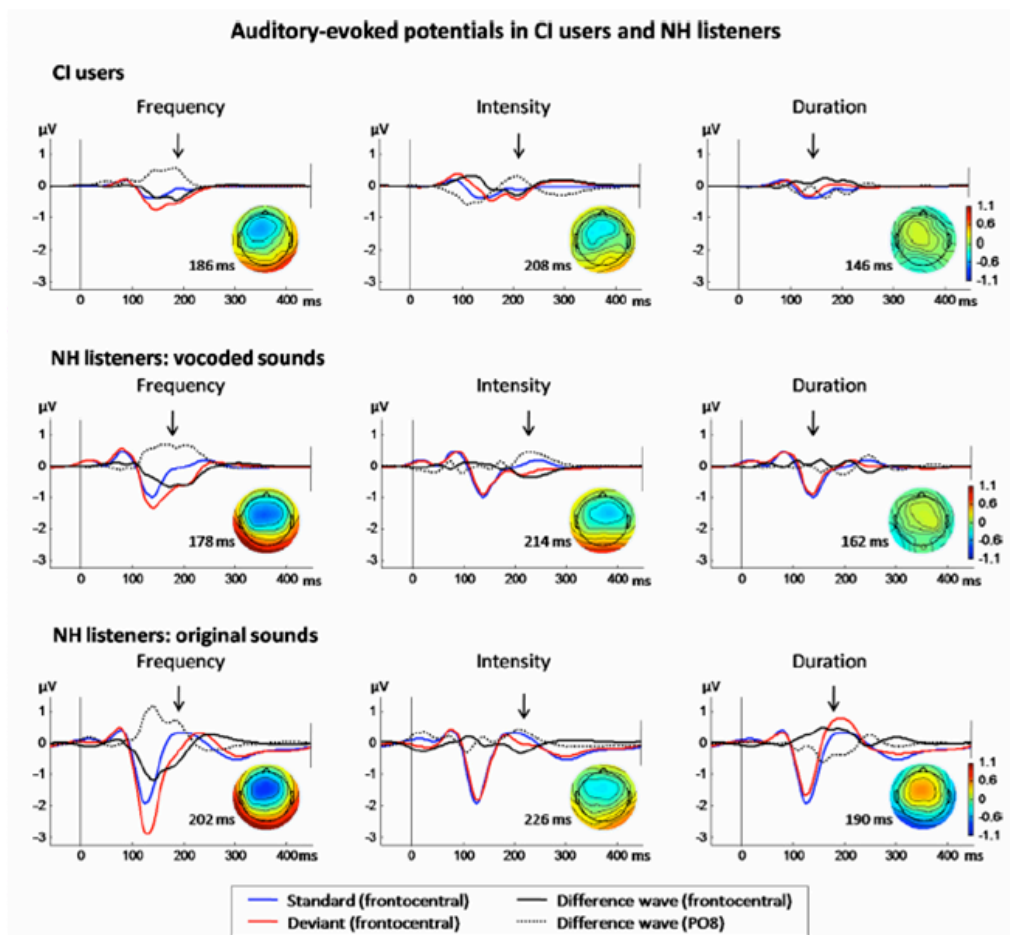


Figure VII.9. Réponse neuronale de type MMN à un changement de fréquence (gauche), d'intensité (milieu) ou de durée (droite), dans un paradigme oddball avec plusieurs stimuli déviants. Cette réponse est inférieure chez les sujets implantés comparés aux sujets normo entendant en condition originale et après vocoding.

La réorganisation du système auditif central secondaire à la surdité, puis secondaire à l'implantation, est nécessairement singulière. Mais on peut faire l'hypothèse que cette réorganisation est délétère pour la perception de la musique, de la même manière qu'elle l'est pour la perception de la voix humaine [Rouger et al., 2011]. L'ensemble de la réorganisation est en effet destiné à optimiser le traitement de la parole, en favorisant l'intégration audiovisuelle par le recrutement des aires de la voix dans le STS droit antérieur et le traitement phonologique bas niveau dans les régions postérieures des gyri temporaux supérieurs. La seule étude par PET scan des activations cérébrales en réponse à un signal de musique chez des sujets implantés cochléaires a effectivement retrouvé une hyper activation de régions temporales (gyri temporaux supérieurs et moyens) chez les implantés comparés aux sujets normo entendant, en réponse à un signal de parole. Cette hyper activation diminuait significativement en réponse à un signal de musique, qui, à la différence du signal de parole, ne pouvait donc hyper solliciter de régions neuronales pour favoriser son traitement.

#### **4.4 L'étude de la perception musicale chez le sujet sourd implanté cochléaire peut-elle être utile à la compréhension de la perception musicale chez le sujet normo entendant ?**

L'essentiel des recherches portant sur la perception musicale chez le sujet implanté cherche d'une part à évaluer l'étendue de leurs déficits et d'autre part à optimiser leurs capacités perceptives. Mais la contribution de ces recherches aux travaux menés chez le normo entendant demeure limitée. Certaines informations pourraient pourtant avoir un rôle important dans certaines controverses, notamment celle ayant trait à la spécificité du traitement cognitif de la musique.

Les doubles dissociations fonctionnelles entre perception de la musique et perception de la parole sont bien connues depuis la mise en évidence de lésions cérébrales atteignant spécifiquement la musique [Peretz et al., 1994, Piccirilli et al., 2000] et de lésions touchant spécifiquement la parole [Mendez, 2001]. Les études de cas après lésions cérébrales ont permis d'établir l'existence d'un système modulaire de reconnaissance musicale, indépendant du système traitant le signal de parole. Après une analyse acoustique bas niveau, ce système prévoit un traitement spécifique de l'information tonale contenue dans le signal musical, assurant une analyse du contour de la mélodie (ascendant, descendant), de l'intervalle séparant deux notes jouées simultanément ou de manière rapprochée (figure VII.10). Ces



étapes cognitives permettent de confronter une information tonale au répertoire musical dans lequel sont stockées les connaissances musicales. La spécificité « musique » de certains modules demeure encore incertaine, notamment pour le module d'analyse du contour, qui pourrait être « partagé » avec le langage pour le traitement de la prosodie linguistique. De même, le module d'analyse du rythme pourrait être commun aux traitements de la parole et de la musique.

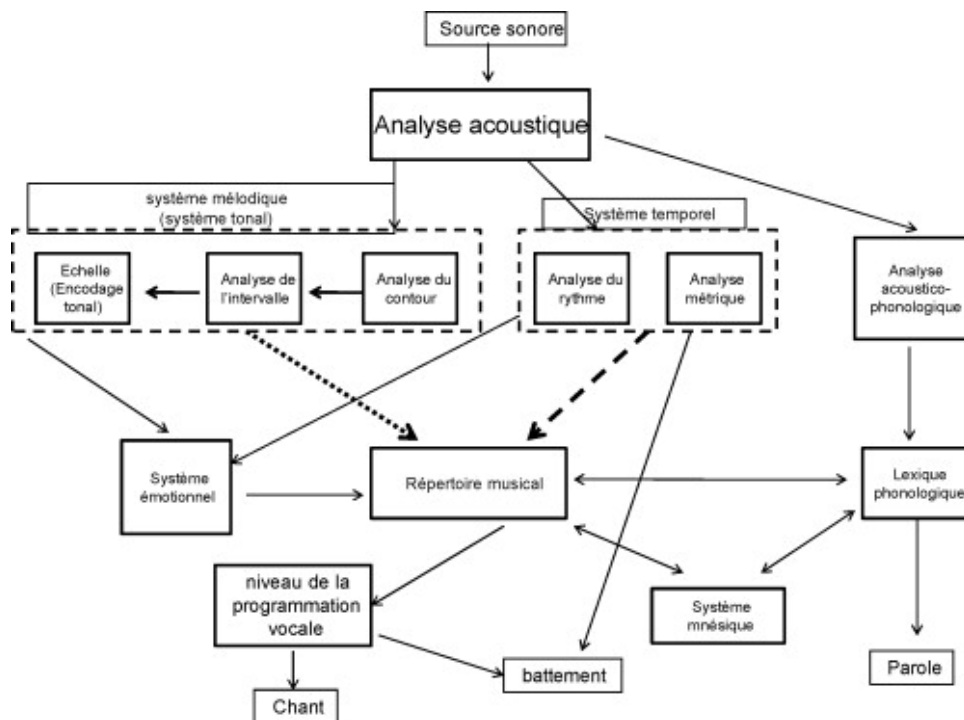


Figure VII.10. Système pluri modulaire de reconnaissance de la musique. Certains modules, tels que ceux réalisant l'analyse du contour et du rythme, pourraient être communs au traitement de la parole et de la musique. Adapté depuis Peretz et Coltheart (2003).

Les travaux sur les sujets implantés cochléaires pourraient contribuer à améliorer la connaissance de ce système modulaire de perception de la musique, notamment pour les modules faisant débat. Trehub et ses collaborateurs défendent la thèse d'un module unique d'analyse acoustique du rythme et ont insisté sur les similitudes entre les structures rythmiques présentes dans une langue donnée et la musique, même instrumentale pure, lui correspondant [Trehub and Hannon, 2006]. Les différences de durée des syllabes entre le français et l'anglais sont ainsi à l'origine de structures rythmiques spécifiques à chaque langue retrouvées dans la musique leur correspondant [Patel and Daniele, 2003]. Nos résultats sont à interpréter prudemment puisqu'ils portent sur un effectif réduit de sujets implantés et n'incluent qu'un test de perception du rythme. Ils pourraient toutefois suggérer que les

informations temporelles relatives au rythme, généralement comprises entre 2 et 20 Hz donc codées par l'implant, sont traitées différemment selon qu'il s'agit de musique ou de parole. En effet, les sujets implantés cochléaires présentaient des performances comparables aux sujets normo entendant pour la perception du rythme et surtout indépendantes vis-à-vis de leur niveau de discrimination de la parole. Nous avons ainsi montré qu'il n'existait pas de relation entre la discrimination des mots et leurs scores pour la tâche de perception du rythme. Des études comprenant plusieurs évaluations de la perception du rythme, de la cadence et du tempo mais également du contour de la mélodie, sur des effectifs plus importants, pourraient permettre de progresser dans la connaissance du système de reconnaissance de la musique et des éléments qui le composent.



## Chapitre VIII

### Discussion générale et conclusion

Les études comportementales que nous avons menées ont mis en évidence un déficit de la perception auditive para et extra linguistique chez le sujet sourd post lingual implanté cochléaire. Ce déficit est variable en intensité mais toujours majeur si l'on considère l'excellence des résultats obtenus pour la discrimination de la parole dans le silence. Le constat étant fait, l'heure est à l'analyse des mécanismes qui sous – tendent ces déficits et à l'élaboration de perspectives destinées à les améliorer.

Pour ce qui est des mécanismes à la base de ces déficits, deux facteurs ont été impliqués dans chacune de nos études comme responsables potentiels : la dégradation du signal acoustique par l'implant d'une part, et la réorganisation fonctionnelle corticale secondaire à la surdité d'autre part. Les mauvaises langues diront qu'il s'agit là de coupables idéaux ; seulement, au regard de nos résultats et à l'examen de la littérature, il apparaît probable que ces deux mécanismes rendent compte de la majeure partie des déficits présentés pour la perception des informations para et extra linguistiques.

Notre discussion générale sera donc axée autour de ces deux facteurs inter connectés, qui ne doivent pas être considérés séparément l'un de l'autre. Améliorer la technique dans le codage du signal n'a aucun sens si le cerveau n'est pas en mesure de traiter le message délivré. De la même manière, la réhabilitation suivant l'implantation doit permettre au cerveau « d'aller chercher » les indices pertinents transmis par l'implant.

## **1. CODAGE ET RESTITUTION DU SIGNAL**

Les principales stratégies de codage développées jusque là assurent une restitution relativement efficace des fluctuations à basse fréquence de l'enveloppe du signal, ce que Moore et bien d'autres identifient simplement comme l'enveloppe. Cette information temporelle est incontestablement l'indice le mieux restitué par l'implant, quelle que soit la stratégie de codage considérée.

Hors, un rôle essentiel dans la perception des informations para et extra linguistiques est joué par la fréquence fondamentale F0. Nos études « prosodie linguistique » et « musique » ont permis d'évaluer la perception des variations de F0 et ont souligné un déficit déjà rapporté par ailleurs [Vandali et al., 2005, Rogers et al., 2006, Cousineau et al., 2010]. Ces études ont présenté l'intérêt double d'inclure des sujets implantés utilisant leur audition résiduelle en complément de l'implant. Nous parlons ici d'intérêt double parce qu'il plaît au clinicien de constater les bénéfices retirés de la conservation ou de la stimulation de restes auditifs, comme au chercheur d'expliquer les mauvaises performances voire l'absence de performances des sujets implantés sans audition résiduelle. Les sujets implantés CI only

réalisent à peine la discrimination question-affirmation et ne peuvent percevoir la dimension mélodique, là où les sujets bimodaux présentant les meilleurs restes auditifs sont capables d'atteindre des performances presque normales. La différence majeure entre ces deux groupes de sujets tient au fait que les premiers ne peuvent pas percevoir F0 et les premiers harmoniques alors que les seconds peuvent traiter cette information.

La perception de F0 est donc susceptible d'apporter aux patients implantés un bénéfice substantiel et son codage par l'implant doit donc être optimisé. Faute de quoi certaines informations liées à l'intonation peuvent perdre jusqu'à leur sens, comme en témoignent les difficultés de certains sujets implantés à comprendre la tâche de discrimination entre question et affirmation dans l'étude « prosodie linguistique ». Pour ces sujets, une question devait nécessairement contenir des indices syntaxiques permettant de l'identifier comme telle. L'accès à sa représentation par l'information prosodique se révélait dans ces cas-là tout simplement impossible et les performances obtenues par ces sujets n'étaient pas différentes du niveau de la chance.

L'amélioration du codage de F0 fait actuellement l'objet de nombreux travaux. Le développement de la technologie dite « des canaux virtuels », par laquelle deux électrodes stimulées simultanément évoqueront une sensation de hauteur intermédiaire, n'a pour l'instant pas démontré de manière robuste une efficacité supérieure à celle des stratégies classiques. De la même manière, il n'existe à notre connaissance pas d'étude montrant d'importants effets des tentatives de restitution de la structure temporelle fine. Cependant, le regain d'intérêt pour la décomposition de la structure temporelle du signal s'est vu conforté par les études récentes portant sur les oscillations corticales. Il a été montré que l'activité neuronale de type high gamma [Kubanek et al., 2013] pouvait augmenter dans les régions corticales auditives et se synchroniser aux fluctuations à basse fréquence de l'enveloppe d'un signal de parole. Il est donc envisageable que la structure temporelle fine du signal puisse faire l'objet d'un traitement similaire et transmettre ainsi les informations relatives à F0 qu'elle porte. L'amélioration de la restitution de la structure temporelle fine pourrait toutefois être limitée par les capacités intrinsèques du nerf auditif qui demeurent incertaines après de longues périodes de surdité.

Plusieurs pistes visant à améliorer la spécificité de stimulation au sein de la cochlée sont par ailleurs à l'étude. L'évaluation d'une stimulation localisée des neurones cochléaires par lumière infra rouge est actuellement en cours [Matic et al., 2011, Xia et al., 2013] et semble limiter le problème de la diffusion de stimulation liée à l'utilisation de courant électrique. La forme des impulsions électriques stimulant le nerf auditif est également explorée et la

diffusion intra cochléaire du courant serait mieux contrôlée par des impulsions asymétriques [Undurraga et al., 2012].

## **2. INFLUENCES DE LA REORGANISATION CORTICALE SECONDAIRE A LA SURDITE**

La prosodie, la musique et même les bruits environnementaux font l'objet d'un traitement neuronal principalement latéralisé dans l'hémisphère droit, au niveau des régions temporales. Les phénomènes de plasticité secondaires à la surdité se révèlent particulièrement actifs dans ces régions et ont pour but de faciliter la perception de la parole. Une plasticité cross modale audition-vision touche ainsi le sulcus temporal supérieur droit, où siègent notamment les aires impliquées dans le traitement de la voix. Le recrutement de cette région pour des activités visuelles de type lecture labiale pourrait ainsi rendre compte d'une partie des mauvaises performances des patients implantés dans des tâches de détection de la voix [Massida et al., 2011] comme dans notre étude portant sur la perception du genre. Le changement de fonction des régions postérieures du gyrus temporal supérieur par des phénomènes de plasticité méta modale pourrait quant à lui altérer le traitement de la prosodie affective et des sons environnementaux, en faveur d'un traitement phonologique de bas niveau, encore destiné à optimiser la transmission de l'information de parole.

Les phénomènes de plasticité doivent également être considérés après restauration de l'audition par l'implant cochléaire. Une restauration satisfaisante de la discrimination de la parole par implant se traduit en effet par un processus de normalisation du traitement auditif, observé plus particulièrement au niveau des réseaux fronto temporaux gauches, qui pourrait témoigner d'une réactivation, au moins partielle, de la boucle audio phonatoire. Cette réactivation progressive des réseaux fronto temporaux gauches impliqués dans le traitement de la parole pourrait également être observée pour leurs alter ego droits, à la faveur d'une rééducation ciblant les informations para et extra linguistiques. La musique pourrait être le support idéal d'une telle rééducation car, bien que les faibles performances pour la perception de la musique après implantation soient la règle, notre étude sur la catégorisation auditive a révélé des capacités de la reconnaître. Cette étude a ouvert de grandes perspectives d'une manière générale, et notamment vis-à-vis de la perception musicale. L'inclusion d'un plus grand nombre de sujets d'une part, et des recherches de corrélations entre les analyses MDS et de nouveaux paramètres acoustiques d'autre part, permettront probablement d'identifier les

corrélats acoustiques des différentes dimensions en MDS. Il est possible d'imaginer des programmes de rééducation orthophonique ciblés sur la perception de ces indices. En association avec un réentraînement à l'écoute musicale, ces programmes pourraient faciliter la perception de la musique et d'autres informations para/extra linguistiques par des processus top down, dirigés par l'attention sélective. La plasticité corticale éventuellement induite par de tels programmes pourrait être monitorée par des études électro physiologiques ou en neuro imagerie fonctionnelle, en transposant des paradigmes déjà employés chez le sujet normo entendant aux sujets implantés cochléaires. L'exploration électro physiologique permettrait de déterminer, tout du moins de mieux cerner, les catégories de signaux acoustiques susceptibles d'être traitées sur le plan neuronal et d'orienter ainsi la rééducation orthophonique.

Nos travaux ont permis de souligner la contribution de F0 dans la perception auditive des informations para et extra linguistiques. Cet indice est accessible à certains patients implantés qui bénéficient d'une stimulation bimodale ou électro acoustique, grâce à l'audition résiduelle qu'ils présentent sur les basses fréquences. Les stratégies de codage doivent être améliorées afin d'assurer une meilleure transmission de F0 aux patients implantés ne présentant pas de restes auditifs. La réorganisation corticale secondaire à la surdité privilégie le traitement neuronal de la parole par l'optimisation de son intégration audiovisuelle, et du traitement phonologique chez certains sujets. Ces phénomènes sont particulièrement actifs dans l'hémisphère droit, et impliquent des réseaux fronto – temporaux qui interviennent normalement dans le traitement des informations para et extra linguistiques.

Nous envisageons deux principales directions à donner à ces travaux. La première concerne l'analyse de scène auditive, par laquelle un environnement acoustique naturellement complexe est perçu, et qui est encore insuffisamment documentée chez le sujet implanté cochléaire. Le deuxième que nous souhaitons développer est l'exploration électro physiologique dans cette population, difficile sur le plan technique, mais qui pourrait permettre de mieux comprendre les processus de perception de la prosodie et de la musique dans cette population.





## Références

- Agrawal, D., Timm, L., Viola, F. C., Debener, S., Buchner, A., Dengler, R., and Wittfoth, M. (2012). "ERP evidence for the recognition of emotional prosody through simulated cochlear implant strategies," *BMC neuroscience* 13, 113.
- Alain, C., Woods, D. L., and Covarrubias, D. (1997). "Activation of duration-sensitive auditory cortical fields in humans," *Electroencephalography and clinical neurophysiology* 104, 531-539.
- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., and Tillmann, B. (2013). "Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex," *Brain* 136, 1639-1661.
- Allen, G. (1878). "Note-Deafness," *Mind* 10, 157-167
- Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A. L. (2009). "Dual neural routing of visual facilitation in speech processing," *J Neurosci* 29, 13445-13453.
- Arnal, L. H., Wyart, V., and Giraud, A. L. (2011). "Transitions in neural oscillations reflect prediction errors generated in audiovisual speech," *Nature neuroscience* 14, 797-801.
- Assal, G., Zander, E., Kremin, H., and Buttet, J. (1976). "[Voice discrimination in patients with cerebral cortical lesions]," *Schweizer Archiv fur Neurologie, Neurochirurgie und Psychiatrie = Archives suisses de neurologie, neurochirurgie et de psychiatrie* 119, 307-315.
- Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies," *The Journal of the Acoustical Society of America* 88, 680-697.
- Bacciu, A., Pasanisi, E., Vincenti, V., Guida, M., Barbot, A., Berghenti, M., Forli, F., Berrettini, S., and Bacciu, S. (2004). "Comparison of speech perception performance between the Nucleus 24 and Nucleus 24 Contour cochlear implant systems," *Acta Otolaryngol* 124, 1155-1158.
- Bach, Y. R. P., Collins, C. C., Scadden, L. A., Holmlund, G. W., and Hart, B. K. (1970). "Display techniques in a tactile vision-substitution system," *Medical & biological illustration* 20, 6-12.
- Barkhuysen, P., Krahmer, E., and Swerts, M. (2008). "The interplay between the auditory and visual modality for end-of-utterance detection," *The Journal of the Acoustical Society of America* 123, 354-365.
- Bassim, M. K., Buss, E., Clark, M. S., Kolln, K. A., Pillsbury, C. H., Pillsbury, H. C., 3rd, and Buchman, C. A. (2005). "MED-EL Combi40+ cochlear implantation in adults," *The Laryngoscope* 115, 1568-1573.
- Baumann, U., and Nobbe, A. (2004). "Pulse rate discrimination with deeply inserted electrode arrays," *Hearing research* 196, 49-57.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., and Martin, A. (2004). "Unraveling multisensory integration: patchy organization within human STS multisensory cortex," *Nature neuroscience* 7, 1190-1192.
- Belin, P., Fecteau, S., and Bedard, C. (2004). "Thinking the voice: neural correlates of voice perception," *Trends in cognitive sciences* 8, 129-135.

- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). "Voice-selective areas in human auditory cortex," *Nature* 403, 309-312.
- Bendor, D., and Wang, X. (2005). "The neuronal representation of pitch in primate auditory cortex," *Nature* 436, 1161-1165.
- Bennett, S. (1981). "Vowel formant frequency characteristics of preadolescent males and females," *The Journal of the Acoustical Society of America* 69, 231-238.
- Bernstein, J. G., and Oxenham, A. J. (2006). "The relationship between frequency selectivity and pitch discrimination: effects of stimulus level," *The Journal of the Acoustical Society of America* 120, 3916-3928.
- Blamey, P., Arndt, P., Bergeron, F., Bredberg, G., Brimacombe, J., Facer, G., Larky, J., Lindstrom, B., Nedzelski, J., Peterson, A., Shipp, D., Staller, S., and Whitford, L. (1996). "Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants," *Audiology & neuro-otology* 1, 293-306.
- Blamey, P., Artieres, F., Baskent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallego, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., Rousset, A., Schauwers, K., Skarzynski, H., Skarzynski, P. H., Sterkers, O., Terranti, A., Truy, E., Van de Heyning, P., Venail, F., Vincent, C., and Lazard, D. S. (2012). "Factors Affecting Auditory Performance of Postlinguistically Deaf Adults Using Cochlear Implants: An Update with 2251 Patients," *Audiology & neuro-otology* 18, 36-47.
- Boersma, P., and D. W. (2009). "Praat: doing phonetics by computer".
- Bradvik, B., Dravins, C., Holtas, S., Rosen, I., Ryding, E., and Ingvar, D. H. (1991). "Disturbances of speech prosody following right hemisphere infarcts," *Acta neurologica Scandinavica* 84, 114-126.
- Bregman, A. (1990). *Auditory Scene Analysis: the perceptual Organization of Sound* (Mass.: Bradford Books, MIT Press, Cambridge).
- Bregman, A. S., Levitan, R., and Liao, C. (1990). "Fusion of auditory components: effects of the frequency of amplitude modulation," *Perception & psychophysics* 47, 68-73.
- Bross, M. (1979). "Response bias in deaf and hearing subjects as a function of motivational factors," *Perceptual and motor skills* 49, 779-782.
- Brozinsky, C. J., and Bavelier, D. (2004). "Motion velocity thresholds in deaf signers: changes in lateralization but not in overall sensitivity," *Brain research* 21, 1-10.
- Bruce, V., and Young, A. (1986). "Understanding face recognition," *Br J Psychol* 77 ( Pt 3), 305-327.
- Bryan, K. (1989). "Language prosody and the right hemisphere," *Aphasiology* 3, 285-299.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., and Jancke, L. (2005). "Scanning silence: mental imagery of complex sounds," *NeuroImage* 26, 1119-1127.
- Busby, P. A., and Plant, G. L. (1995). "Formant frequency values of vowels produced by preadolescent boys and girls," *The Journal of the Acoustical Society of America* 97, 2603-2606.
- Calmels, M. N., Saliba, I., Wanna, G., Cochard, N., Fillaux, J., Deguine, O., and Fraysse, B. (2004). "Speech perception and speech intelligibility in children after cochlear implantation," *International journal of pediatric otorhinolaryngology* 68, 347-351.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. (1997). "Activation of auditory cortex during silent lipreading," *Science (New York, N.Y)* 276, 593-596.
- Campanella, S., and Belin, P. (2007). "Integrating face and voice in person perception," *Trends in cognitive sciences* 11, 535-543.

- Capek, C. M., Waters, D., Woll, B., MacSweeney, M., Brammer, M. J., McGuire, P. K., David, A. S., and Campbell, R. (2008). "Hand and mouth: cortical correlates of lexical processing in British Sign Language and speechreading English," *Journal of cognitive neuroscience* 20, 1220-1234.
- Carlyon, R. P. (1994). "Further evidence against an across-frequency mechanism specific to the detection of frequency modulation (FM) incoherence between resolved frequency components," *The Journal of the Acoustical Society of America* 95, 949-961.
- Carlyon, R. P., Long, C. J., and Micheyl, C. (2012). "Across-channel timing differences as a potential code for the frequency of pure tones," *J Assoc Res Otolaryngol* 13, 159-171.
- Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., and Espesser, R. (1996). "About the relationship between eyebrow movements and F0 variations," in *ICSLP* (Philadelphia).
- Cedolin, L., and Delgutte, B. (2010). "Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve," *J Neurosci* 30, 12712-12724.
- Cermak, G. W., and Cornillon, P. C. (1976). "Multidimensional analyses of judgements about traffic noise," *The Journal of the Acoustical Society of America* 59, 1412-1420.
- Chatterjee, M., Sarampalis, A., and Oba, S. I. (2006). "Auditory stream segregation with cochlear implants: A preliminary report," *Hearing research* 222, 100-107.
- Chouard, C. H., and MacLeod, P. (1976). "Implantation of multiple intracochlear electrodes for rehabilitation of total deafness: preliminary report," *The Laryngoscope* 86, 1743-1751.
- Cohen, J. (1992). "A power primer," *Psychological bulletin* 112, 155-159.
- Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., and Deltenre, P. (2002). "Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory," *Clin Neurophysiol* 113, 495-506.
- Cooper, H. R., and Roberts, B. (2007). "Auditory stream segregation of tone sequences in cochlear implant listeners," *Hearing research* 225, 11-24.
- Cousineau, M., Demany, L., Meyer, B., and Pressnitzer, D. (2010). "What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant," *Hearing research* 269, 34-41.
- Cousineau, M., Demany, L., Meyer, B., and Pressnitzer, D. (2011). "What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant," *Hearing research* 269, 34-41.
- Cullington, H. E., and Zeng, F. G. (2011). "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification," *Ear and hearing* 32, 16-30.
- de Gelder, B., Bocker, K. B., Tuomainen, J., Hensen, M., and Vroomen, J. (1999). "The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses," *Neuroscience letters* 260, 133-136.
- Demany, L., and Semal, C. (2002). "Learning to perceive pitch differences," *The Journal of the Acoustical Society of America* 111, 1377-1388.
- Dietrich, S., Ackermann, H., Szameitat, D. P., and Alter, K. (2006). "Psychoacoustic studies on the processing of vocal interjections: how to disentangle lexical and prosodic information?," *Progress in brain research* 156, 295-302.
- Dohen, M., and Loevenbruck, H. (2009). "Interaction of audition and vision for the perception of prosodic contrastive focus," *Language and speech* 52, 177-206.
- Donaldson, G. S., Kreft, H. A., and Litvak, L. (2005). "Place-pitch discrimination of single-versus dual-electrode stimuli by cochlear implant users (L)," *The Journal of the Acoustical Society of America* 118, 623-626.

- Donnelly, P. J., Guo, B. Z., and Limb, C. J. (2009). "Perceptual fusion of polyphonic pitch in cochlear implant users," *The Journal of the Acoustical Society of America* 126, EL128-133.
- Dorman, M. F., and Loizou, P. C. (1997). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," *The American journal of otology* 18, S113-114.
- Dorman, M. F., Loizou, P. C., and Fitzke, J. (1998a). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," *Ear and hearing* 19, 481-484.
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (1998b). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6-20 channels," *The Journal of the Acoustical Society of America* 104, 3583-3585.
- Dufor, O., Serniclaes, W., Sprenger-Charolles, L., and Demonet, J. F. (2009). "Left premotor cortex and allophonic speech perception in dyslexia: a PET study," *NeuroImage* 46, 241-248.
- Eddins, D. A., Hall, J. W., 3rd, and Grose, J. H. (1992). "The detection of temporal gaps as a function of frequency region and absolute noise bandwidth," *The Journal of the Acoustical Society of America* 91, 1069-1077.
- Edgerton, B. J., Prietto, A., and Danhauer, J. L. (1983). "Cochlear implant patient performance on the MAC battery," *Otolaryngologic clinics of North America* 16, 267-280.
- Eggermont, J. J. (2007). "Correlated neural activity as the driving force for functional changes in auditory cortex," *Hearing research* 229, 69-80.
- Eisenberg, L. S., Berliner, K. I., House, W. F., and Edgerton, B. J. (1983). "Status of the adults' and children's cochlear implant programs at the House Ear Institute," *Annals of the New York Academy of Sciences* 405, 323-331.
- El Fata, F., James, C. J., Laborde, M. L., and Fraysse, B. (2009). "How much residual hearing is 'useful' for music perception with cochlear implants?," *Audiology & neuro-otology* 14 Suppl 1, 14-21.
- Faure, A. (2000). "Des mots aux sons. Comment parle-t-on du timbre musical?," (EHESS, Paris).
- Finney, E. M., and Dobkins, K. R. (2001). "Visual contrast sensitivity in deaf versus hearing populations: exploring the perceptual consequences of auditory deprivation and experience with a visual language," *Brain research* 11, 171-183.
- Firszt, J. B., Koch, D. B., Downing, M., and Litvak, L. (2007). "Current steering creates additional pitch percepts in adult cochlear implant recipients," *Otol Neurotol* 28, 629-636.
- Fishman, K. E., Shannon, R. V., and 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, 1201-1215.
- Fishman, Y. I., Micheyl, C., and Steinschneider, M. (2013). "Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey," *J Neurosci* 33, 10312-10323.
- Fitzgerald, M. B., Shapiro, W. H., McDonald, P. D., Neuburger, H. S., Ashburn-Reed, S., Immerman, S., Jethanamest, D., Roland, J. T., and Svirsky, M. A. (2007). "The effect of perimodiolar placement on speech perception and frequency discrimination by cochlear implant users," *Acta oto-laryngologica* 127, 378-383.
- Fonagy, I. (1958). "[Phonetics & psychoanalysis; psychological remarks on the stress accent]," *Psyche* 12, 63-72.

- Fonagy, I. (1966). "Electrophysiological and acoustic correlates of stress and stress perception," *Journal of speech and hearing research* 9, 231-244.
- Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., and Spinelli, E. (2012). "Visual speech triggers lexical access," *Language and Cognitive Processes* iFirst 1-17.
- Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., and Griffiths, T. D. (2004). "Characterization of deficits in pitch perception underlying 'tone deafness'," *Brain* 127, 801-810.
- Foxton, J. M., Nandy, R. K., and Griffiths, T. D. (2006). "Rhythm deficits in 'tone deafness'," *Brain and cognition* 62, 24-29.
- Foxton, J. M., Riviere, L. D., and Barone, P. (2010). "Cross-modal facilitation in speech prosody," *Cognition* 115, 71-78.
- Fraysse, B., Macias, A. R., Sterkers, O., Burdo, S., Ramsden, R., Deguine, O., Klenzner, T., Lenarz, T., Rodriguez, M. M., Von Wallenberg, E., and James, C. (2006). "Residual hearing conservation and electroacoustic stimulation with the nucleus 24 contour advance cochlear implant," *Otol Neurotol* 27, 624-633.
- French, N., and Steinberg, J. C. (1947). "Factors governing the intelligibility of speech sounds," *The Journal of the Acoustical Society of America* 19, 90-119.
- Friesen, L. M., Shannon, R. V., Baskent, D., and 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, 1150-1163.
- Frijns, J. H., de Snoo, S. L., and ten Kate, J. H. (1996). "Spatial selectivity in a rotationally symmetric model of the electrically stimulated cochlea," *Hearing research* 95, 33-48.
- Friston, K., and Kiebel, S. (2009). "Predictive coding under the free-energy principle," *Philosophical transactions of the Royal Society of London* 364, 1211-1221.
- Fu, Q. J., and Galvin, J. J., 3rd (2008). "Maximizing cochlear implant patients' performance with advanced speech training procedures," *Hearing research* 242, 198-208.
- Fujiki, N., Naito, Y., Hirano, S., Kojima, H., Shiomi, Y., Nishizawa, S., Konishi, J., and Honjo, I. (1999). "Correlation between rCBF and speech perception in cochlear implant users," *Auris, nasus, larynx* 26, 229-236.
- Fujiki, N., Naito, Y., Hirano, S., Kojima, H., Shiomi, Y., Nishizawa, S., Konishi, J., and Honjo, I. (2000). "Cortical activity and speech perception performance in cochlear implant users," *Advances in oto-rhino-laryngology* 57, 32-35.
- Fujita, S., and Ito, J. (1999). "Ability of nucleus cochlear implantees to recognize music," *The Annals of otology, rhinology, and laryngology* 108, 634-640.
- Gaillard, P. (ed). (2009). *Laissez-nous trier! TCL-LabX et les tâches de catégorisation libre de sons*. (L'Harmattan, Paris).
- Gantz, B. J., Turner, C., and Gfeller, K. E. (2006). "Acoustic plus electric speech processing: preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant," *Audiology & neuro-otology* 11 Suppl 1, 63-68.
- Garnham, C., O'Driscoll, M., Ramsden, R., and Saeed, S. (2002). "Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets," *Ear and hearing* 23, 540-552.
- Garstecki, D. C., and O'Neill, J. J. (1980). "Situational cue and strategy influence on speechreading," *Scandinavian audiology* 9, 147-151.
- Gaudrain, E., Grimault, N., Healy, E. W., and Bera, J. C. (2008). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," *The Journal of the Acoustical Society of America* 124, 3076-3087.
- Geier, L. L., and Norton, S. J. (1992). "The effects of limiting the number of Nucleus 22 cochlear implant electrodes programmed on speech perception," *Ear and hearing* 13, 340-348.

- Gelfer, M. P., and Mikos, V. A. (2005). "The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels," *J Voice* 19, 544-554.
- Geurts, L., and Wouters, J. (2001). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *The Journal of the Acoustical Society of America* 109, 713-726.
- Geurts, L., and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *The Journal of the Acoustical Society of America* 115, 844-852.
- Gfeller, K., Christ, A., Knutson, J. F., Witt, S., Murray, K. T., and Tyler, R. S. (2000). "Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients," *Journal of the American Academy of Audiology* 11, 390-406.
- Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S., and DeBus, B. (1998). "Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults," *Journal of the American Academy of Audiology* 9, 1-19.
- Gfeller, K., and Lansing, C. R. (1991). "Melodic, rhythmic, and timbral perception of adult cochlear implant users," *Journal of speech and hearing research* 34, 916-920.
- Gfeller, K., Olszewski, C., Rychener, M., Sena, K., Knutson, J. F., Witt, S., and Macpherson, B. (2005). "Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults," *Ear and hearing* 26, 237-250.
- Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (2002a). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," *Cochlear implants international* 3, 29-53.
- Gfeller, K., Turner, C., Oleson, J., Zhang, X., Gantz, B., Froman, R., and Olszewski, C. (2007). "Accuracy of cochlear implant recipients on pitch perception, melody recognition, and speech reception in noise," *Ear and hearing* 28, 412-423.
- Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J. (2002b). "Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal," *The Annals of otology, rhinology, and laryngology* 111, 349-356.
- Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., and Knutson, J. F. (1997). "Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users," *Ear and hearing* 18, 252-260.
- Ghazanfar, A. A., Chandrasekaran, C., and Logothetis, N. K. (2008). "Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys," *J Neurosci* 28, 4457-4469.
- Ghitza, O., Giraud, A. L., and Poeppel, D. (2012). "Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence," *Frontiers in human neuroscience* 6, 340.
- Giraud, A. L., and Poeppel, D. (2012). "Cortical oscillations and speech processing: emerging computational principles and operations," *Nature neuroscience* 15, 511-517.
- Giraud, A. L., Truy, E., Frackowiak, R. S., Gregoire, M. C., Pujol, J. F., and Collet, L. (2000). "Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients," *Brain* 123 ( Pt 7), 1391-1402.
- Giraudet, F., and Avan, P. (2012). "Auditory neuropathies: understanding their pathogenesis to illuminate intervention strategies," *Current opinion in neurology* 25, 50-56.
- Goldstein, J. L. (1973). "An optimum processor theory for the central formation of the pitch of complex tones," *The Journal of the Acoustical Society of America* 54, 1496-1516.
- Gordin, A., Papsin, B., James, A., and Gordon, K. (2009). "Evolution of cochlear implant arrays result in changes in behavioral and physiological responses in children," *Otol Neurotol* 30, 908-915.

- Grassi, M., and Soranzo, A. (2009). "MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimation," *Behavior research methods* 41, 20-28.
- Green, K. M., Julyan, P. J., Hastings, D. L., and Ramsden, R. T. (2005). "Auditory cortical activation and speech perception in cochlear implant users: effects of implant experience and duration of deafness," *Hearing research* 205, 184-192.
- Grey, J. M. (1977). "Multidimensional perceptual scaling of musical timbres," *The Journal of the Acoustical Society of America* 61, 1270-1277.
- Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (2001). "Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing," *British journal of audiology* 35, 173-182.
- Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (2002). "Evidence for two pitch encoding mechanisms using a selective auditory training paradigm," *Perception & psychophysics* 64, 189-197.
- Grose, J. H., Hall, J. W., 3rd, and Buss, E. (2001a). "Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and marker duration, frequency separation, and mode of presentation," *J Assoc Res Otolaryngol* 2, 388-398.
- Grose, J. H., Hall, J. W., 3rd, Buss, E., and Hatch, D. (2001b). "Gap detection for similar and dissimilar gap markers," *The Journal of the Acoustical Society of America* 109, 1587-1595.
- Gygi, B., Kidd, G. R., and Watson, C. S. (2004). "Spectral-temporal factors in the identification of environmental sounds," *The Journal of the Acoustical Society of America* 115, 1252-1265.
- Hadar, U., Steiner, T. J., and Rose, F. C. (1984). "Involvement of head movement in speech production and its implications for language pathology," *Advances in neurology* 42, 247-261.
- Hall, D. A., and Plack, C. J. (2007). "The human 'pitch center' responds differently to iterated noise and Huggins pitch," *Neuroreport* 18, 323-327.
- Hall, D. A., and Plack, C. J. (2009). "Pitch processing sites in the human auditory brain," *Cereb Cortex* 19, 576-585.
- Hannon, E. E. (2009). "Perceiving speech rhythm in music: listeners classify instrumental songs according to language of origin," *Cognition* 111, 404-410.
- HAS (2007). "TRAITEMENT DE LA SURDITE PAR POSE D'IMPLANTS COCHLEAIRES OU D'IMPLANTS DU TRONC CEREBRAL," (Haute Autorité de Santé).
- Herzog, H., Lamprecht, A., Kuhn, A., Roden, W., Vosteen, K. H., and Feinendegen, L. E. (1991). "Cortical activation in profoundly deaf patients during cochlear implant stimulation demonstrated by H2(15)O PET," *Journal of computer assisted tomography* 15, 369-375.
- Hickok, G., and Poeppel, D. (2004). "Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language," *Cognition* 92, 67-99.
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *The Journal of the Acoustical Society of America* 97, 3099-3111.
- Hochmair-Desoyer, I. J., Hochmair, E. S., Fischer, R. E., and Burian, K. (1980). "Cochlear prostheses in use: recent speech comprehension results," *Archives of oto-rhino-laryngology* 229, 81-98.
- House, D. (1994). "Perception and production of mood in speech by cochlear implant users," in *Proceedings of the International Conference on Spoken Language Processing* (Online archive of International Speech Communication Association proceedings available at: <http://www.isca-speech.org>), pp. 2051-2054.



- House, W. F. (1978). "The clinical value of single electrode systems in auditory prostheses," *Otolaryngologic clinics of North America* 11, 201-208.
- Hutchins, S., Gosselin, N., and Peretz, I. (2010). "Identification of Changes along a Continuum of Speech Intonation is Impaired in Congenital Amusia," *Frontiers in psychology* 1, 236.
- Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (2007). "Cortical thickness in congenital amusia: when less is better than more," *J Neurosci* 27, 13028-13032.
- Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). "Evidence for the role of the right auditory cortex in fine pitch resolution," *Neuropsychologia* 46, 632-639.
- Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., and Peretz, I. (2006). "Morphometry of the amusic brain: a two-site study," *Brain* 129, 2562-2570.
- Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). "Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia," *Cereb Cortex* 21, 292-299.
- Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., Kato, T., Nakamura, A., Hatano, K., Kojima, S., and Nakamura, K. (1997). "Vocal identification of speaker and emotion activates different brain regions," *Neuroreport* 8, 2809-2812.
- Inverso, Y., and Limb, C. J. (2010). "Cochlear implant-mediated perception of nonlinguistic sounds," *Ear and hearing* 31, 505-514.
- Iseli, M., Shue, Y. L., and Alwan, A. (2007). "Age, sex, and vowel dependencies of acoustic measures related to the voice source," *The Journal of the Acoustical Society of America* 121, 2283-2295.
- Ito, J., Sakakibara, J., Honjo, I., Iwasaki, Y., and Yonekura, Y. (1990). "Positron emission tomographic study of auditory sensation in a patient with a cochlear implant," *Archives of otolaryngology--head & neck surgery* 116, 1437-1439.
- James, C., Albegger, K., Battmer, R., Burdo, S., Deggouj, N., Deguine, O., Dillier, N., Gersdorff, M., Laszig, R., Lenarz, T., Rodriguez, M. M., Mondain, M., Offeciers, E., Macias, A. R., Ramsden, R., Sterkers, O., Von Wallenberg, E., Weber, B., and Fraysse, B. (2005). "Preservation of residual hearing with cochlear implantation: how and why," *Acta oto-laryngologica* 125, 481-491.
- Johnson, W. F., Emde, R. N., Scherer, K. R., and Klinnert, M. D. (1986). "Recognition of emotion from vocal cues," *Archives of general psychiatry* 43, 280-283.
- Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (2000). "Functional specificity in the right human auditory cortex for perceiving pitch direction," *Brain* 123 ( Pt 1), 155-163.
- Kawahara H, M.-K. I., de Cheveigné A. (1999). "Restructuring speech representations using pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds," *Speech Commun* 27, 187-207.
- Kent, R. D., and Forner, L. L. (1979). "Developmental study of vowel formant frequencies in an imitation task," *The Journal of the Acoustical Society of America* 65, 208-217.
- Kiefer, J., Muller, J., Pfenningdorff, T., Schon, F., Helms, J., von Ilberg, C., Baumgartner, W. D., Gstoettner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., and Baur, S. (1997). "Speech understanding in quiet and in noise with the CIS speech-coding strategy (MED EL Combi-40) compared to the MPEAK and SPEAK strategies (Nucleus)," *Advances in oto-rhino-laryngology* 52, 286-290.
- Kileny, P. R., Zwolan, T. A., Telian, S. A., and Boerst, A. (1998). "Performance with the 20 + 2L lateral wall cochlear implant," *The American journal of otology* 19, 313-319.

- Kilian-Hutten, N., Vroomen, J., and Formisano, E. (2011). "Brain activation during audiovisual exposure anticipates future perception of ambiguous speech," *NeuroImage* 57, 1601-1607.
- Klatt, D. H., and Klatt, L. C. (1990). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," *The Journal of the Acoustical Society of America* 87, 820-857.
- Koch, D. B., Downing, M., Osberger, M. J., and Litvak, L. (2007). "Using current steering to increase spectral resolution in CII and HiRes 90K users," *Ear and hearing* 28, 38S-41S.
- Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (2004). "Music perception with temporal cues in acoustic and electric hearing," *Ear and hearing* 25, 173-185.
- Kotz, S. A., and Paulmann, S. (2007). "When emotional prosody and semantics dance cheek to cheek: ERP evidence," *Brain research* 1151, 107-118.
- Kovacic, D., and Balaban, E. (2009). "Voice gender perception by cochlear implantees," *The Journal of the Acoustical Society of America* 126, 762-775.
- Kriegstein, K. V., and Giraud, A. L. (2004). "Distinct functional substrates along the right superior temporal sulcus for the processing of voices," *NeuroImage* 22, 948-955.
- Kruskal, W. H. (1978). "Committee on national statistics," *Science (New York, N.Y)* 200, 380.
- Kubaneck, J., Brunner, P., Gunduz, A., Poeppel, D., and Schalk, G. (2013). "The tracking of speech envelope in the human cortex," *PloS one* 8, e53398.
- Lazard, D. S., Lee, H. J., Gaebler, M., Kell, C. A., Truy, E., and Giraud, A. L. (2010). "Phonological processing in post-lingual deafness and cochlear implant outcome," *NeuroImage* 49, 3443-3451.
- Leal, M. C., Shin, Y. J., Laborde, M. L., Calmels, M. N., Verges, S., Lugardon, S., Andrieu, S., Deguine, O., and Fraysse, B. (2003). "Music perception in adult cochlear implant recipients," *Acta oto-laryngologica* 123, 826-835.
- Lenarz, M., Sonmez, H., Joseph, G., Buchner, A., and Lenarz, T. (2012). "Long-term performance of cochlear implants in postlingually deafened adults," *Otolaryngol Head Neck Surg* 147, 112-118.
- Li, N., and Loizou, P. C. (2008). "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," *The Journal of the Acoustical Society of America* 123, 2287-2294.
- Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., and Chauvel, P. (1998). "Contribution of different cortical areas in the temporal lobes to music processing," *Brain* 121 ( Pt 10), 1853-1867.
- Limb, C. J., Molloy, A. T., Jiradejvong, P., and Braun, A. R. (2010). "Auditory cortical activity during cochlear implant-mediated perception of spoken language, melody, and rhythm," *J Assoc Res Otolaryngol* 11, 133-143.
- Linville, S. E. (1996). "The sound of senescence," *J Voice* 10, 190-200.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proceedings of the National Academy of Sciences of the United States of America* 103, 18866-18869.
- Loui, P., Alsop, D., and Schlaug, G. (2009). "Tone deafness: a new disconnection syndrome?," *J Neurosci* 29, 10215-10220.
- Ludlow, A., Heaton, P., Rosset, D., Hills, P., and Deruelle, C. (2010). "Emotion recognition in children with profound and severe deafness: do they have a deficit in perceptual processing?," *Journal of clinical and experimental neuropsychology* 32, 923-928.

- Luo, H., and Poeppel, D. (2012). "Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex," *Frontiers in psychology* 3, 170.
- Luo, X., and Fu, Q. J. (2007). "Frequency modulation detection with simultaneous amplitude modulation by cochlear implant users," *The Journal of the Acoustical Society of America* 122, 1046-1054.
- Marcell, M. M., Borella, D., Greene, M., Kerr, E., and Rogers, S. (2000). "Confrontation naming of environmental sounds," *Journal of clinical and experimental neuropsychology* 22, 830-864.
- Marozeau, J. (2004). "Effet de la fréquence fondamentale sur le timbre," (Université René Descartes, Paris).
- Marozeau, J., and de Cheveigne, A. (2007). "The effect of fundamental frequency on the brightness dimension of timbre," *The Journal of the Acoustical Society of America* 121, 383-387.
- Marozeau, J., de Cheveigne, A., McAdams, S., and Winsberg, S. (2003). "The dependency of timbre on fundamental frequency," *The Journal of the Acoustical Society of America* 114, 2946-2957.
- Marrinan, M. S., Roland, J. T., Jr., Reitzen, S. D., Waltzman, S. B., Cohen, L. T., and Cohen, N. L. (2004). "Degree of modiolar coiling, electrical thresholds, and speech perception after cochlear implantation," *Otol Neurotol* 25, 290-294.
- Marx, M., James, C., Elhadi, T., Martin-Dupont, N., Laborde, M. L., Deguine, O., and Fraysse, B. (2011). "Probability of improvement as a function of pre operative residual hearing," *Audiology & neuro-otology* 16, 3-25.
- Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., and Deguine, O. (2011). "Voice discrimination in cochlear-implanted deaf subjects," *Hearing research* 275, 120-129.
- Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., and Deguine, O. (2013). "Gender categorization in cochlear-implanted deaf subjects," *Journal of speech and hearing research*, Accepted.
- Matic, A. I., Walsh, J. T., Jr., and Richter, C. P. (2011). "Spatial extent of cochlear infrared neural stimulation determined by tone-on-light masking," *Journal of biomedical optics* 16, 118002.
- McAdams, S., Winsberg, S., Donnadiou, S., De Soete, G., and Krimphoff, J. (1995). "Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes," *Psychological research* 58, 177-192.
- McMillan, N. A., and Creelman, C. D. (eds). (1991). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
- Mendez, M. F. (2001). "Generalized auditory agnosia with spared music recognition in a left-hander. Analysis of a case with a right temporal stroke," *Cortex; a journal devoted to the study of the nervous system and behavior* 37, 139-150.
- Merzenich, M. M., White, M. W., Leake, P. A., Schindler, R. A., and Michelson, R. P. (1977). "Further progress in the development of multichannel cochlear implants," *Transactions. Section on Otolaryngology* 84, 181-182.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., and von Cramon, D. Y. (2002). "fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences," *Human brain mapping* 17, 73-88.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., and von Cramon, D. Y. (2004). "Brain activity varies with modulation of dynamic pitch variance in sentence melody," *Brain and language* 89, 277-289.

- Mirza, S., Douglas, S. A., Lindsey, P., Hildreth, T., and Hawthorne, M. (2003). "Appreciation of music in adult patients with cochlear implants: a patient questionnaire," *Cochlear implants international* 4, 85-95.
- Moore, B. C. (1973). "Frequency difference limens for short-duration tones," *The Journal of the Acoustical Society of America* 54, 610-619.
- Moore, B. C. (1989). *An introduction to the psychology of hearing* (Academic Press Limited, London).
- Moore, B. C. (2008). "The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people," *J Assoc Res Otolaryngol* 9, 399-406.
- Moore, B. C., Alcantara, J. I., and Dau, T. (1998). "Masking patterns for sinusoidal and narrow-band noise maskers," *The Journal of the Acoustical Society of America* 104, 1023-1038.
- Moore, B. C., Peters, R. W., and Glasberg, B. R. (1990). "Auditory filter shapes at low center frequencies," *The Journal of the Acoustical Society of America* 88, 132-140.
- Moore, B. C., Tyler, L. K., and Marslen-Wilson, W. (2008). "Introduction. The perception of speech: from sound to meaning," *Philosophical transactions of the Royal Society of London* 363, 917-921.
- Moreau, P., Jolicoeur, P., and Peretz, I. (2009). "Automatic brain responses to pitch changes in congenital amusia," *Annals of the New York Academy of Sciences* 1169, 191-194.
- Mortensen, M. V., Mirz, F., and Gjedde, A. (2006). "Restored speech comprehension linked to activity in left inferior prefrontal and right temporal cortices in postlingual deafness," *NeuroImage* 31, 842-852.
- Most, T., Gaon-Sivan, G., Shpak, T., and Luntz, M. (2012). "Contribution of a contralateral hearing aid to perception of consonant voicing, intonation, and emotional state in adult cochlear implantees," *Journal of deaf studies and deaf education* 17, 244-258.
- Muchnik, C., Efrati, M., Nemeth, E., Malin, M., and Hildesheimer, M. (1991). "Central auditory skills in blind and sighted subjects," *Scandinavian audiology* 20, 19-23.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., and Vatikiotis-Bateson, E. (2004). "Visual prosody and speech intelligibility: head movement improves auditory speech perception," *Psychological science* 15, 133-137.
- Murray, I. R., and Arnott, J. L. (1993). "Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion," *The Journal of the Acoustical Society of America* 93, 1097-1108.
- Naatanen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). "The mismatch negativity (MMN) in basic research of central auditory processing: a review," *Clin Neurophysiol* 118, 2544-2590.
- Nagel, H. N., Shapiro, L. P., and Nawy, R. (1994). "Prosody and the processing of filler-gap sentences," *Journal of psycholinguistic research* 23, 473-485.
- Naito, Y., Hirano, S., Okazawa, H., Takahashi, H., Ishizu, K., Fujiki, N., Shiomi, Y., Kawano, M., Yonekura, Y., and Honjo, I. (1997). "Central auditory processing of speech in cochlear implant users demonstrated by positron emission tomography," *Advances in oto-rhino-laryngology* 52, 19-23.
- Naito, Y., Okazawa, H., Honjo, I., Hirano, S., Takahashi, H., Shiomi, Y., Hoji, W., Kawano, M., Ishizu, K., and Yonekura, Y. (1995). "Cortical activation with sound stimulation in cochlear implant users demonstrated by positron emission tomography," *Brain research* 2, 207-214.
- Naito, Y., Tateya, I., Fujiki, N., Hirano, S., Ishizu, K., Nagahama, Y., Fukuyama, H., and Kojima, H. (2000). "Increased cortical activation during hearing of speech in cochlear implant users," *Hearing research* 143, 139-146.

- Nakata, T., Trehub, S. E., and Kanda, Y. (2012). "Effect of cochlear implants on children's perception and production of speech prosody," *The Journal of the Acoustical Society of America* 131, 1307-1314.
- Narne, V. K. (2013). "Temporal processing and speech perception in noise by listeners with auditory neuropathy," *PloS one* 8, e55995.
- Neville, H. J., and Lawson, D. (1987). "Attention to central and peripheral visual space in a movement detection task: an event-related potential and behavioral study. I. Normal hearing adults," *Brain Res* 405, 253-267.
- NIH (1995). "National Institute of Health consensus conference. Cochlear implants in adults and children.," *JAMA* 274, 1955-1961.
- NIHCDC (1988). "Cochlear implants," in *National Institute of Health Consensus Development Conference*.
- Nooteboom, S. G., and Terken, J. M. (1982). "What makes speakers omit pitch accents? An experiment," *Phonetica* 39, 317-336.
- Oxenham, A. J., Bernstein, J. G., and Penagos, H. (2004). "Correct tonotopic representation is necessary for complex pitch perception," *Proceedings of the National Academy of Sciences of the United States of America* 101, 1421-1425.
- Patel, A. D., and Daniele, J. R. (2003). "An empirical comparison of rhythm in language and music," *Cognition* 87, B35-45.
- Patel, A. D., Peretz, I., Tramo, M., and Labrecque, R. (1998). "Processing prosodic and musical patterns: a neuropsychological investigation," *Brain and language* 61, 123-144.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). "The processing of temporal pitch and melody information in auditory cortex," *Neuron* 36, 767-776.
- Pazo-Alvarez, P., Cadaveira, F., and Amenedo, E. (2003). "MMN in the visual modality: a review," *Biological psychology* 63, 199-236.
- Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). "A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging," *J Neurosci* 24, 6810-6815.
- Pereira, C. (2000). "The perception of vocal affect by cochlear implantees," in *Cochlear implants*, edited by S. Waltzman, and N. Cohen (Thieme medical, New York), pp. 343-345.
- Peretz, I., Champod, A. S., and Hyde, K. (2003). "Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia," *Annals of the New York Academy of Sciences* 999, 58-75.
- Peretz, I., and Coltheart, M. (2003). "Modularity of music processing," *Nature neuroscience* 6, 688-691.
- Peretz, I., Cummings, S., and Dube, M. P. (2007). "The genetics of congenital amusia (tone deafness): a family-aggregation study," *American journal of human genetics* 81, 582-588.
- Peretz, I., Gosselin, N., Belin, P., Zatorre, R. J., Plailly, J., and Tillmann, B. (2009). "Music lexical networks: the cortical organization of music recognition," *Annals of the New York Academy of Sciences* 1169, 256-265.
- Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L., Gagnon, B., Trimmer, C., Paquette, S., and Bouchard, B. (2008). "On-line identification of congenital amusia," *Music perception* 25, 331-343.
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (1994). "Functional dissociations following bilateral lesions of auditory cortex," *Brain* 117 ( Pt 6), 1283-1301.

- Perkins, J. M., Baran, J., and Gandour, J. (1996). "Hemispheric specialization in processing intonation contours," *Aphasiology* 10, 343-356.
- Pernet, C. R., and Belin, P. (2012). "The role of pitch and timbre in voice gender categorization," *Frontiers in psychology* 3, 23.
- Peterson, C. C., and Siegal, M. (1995). "Deafness, conversation and theory of mind," *Journal of child psychology and psychiatry, and allied disciplines* 36, 459-474.
- Petrushin, V. (2000). "Emotion recognition in speech signal: experimental study, development, and application," in *Proceedings of the International Conference on Spoken Language Processing* (Online archive of International Speech Communication Association proceedings available at: <http://www.isca-speech.org>).
- Piccirilli, M., Sciarra, T., and Luzzi, S. (2000). "Modularity of music: evidence from a case of pure amusia," *Journal of neurology, neurosurgery, and psychiatry* 69, 541-545.
- Picton, T. W., Goodman, W. S., and Bryce, D. P. (1970). "Amplitude of evoked responses to tones of high intensity," *Acta oto-laryngologica* 70, 77-82.
- Poitevineau, J. (ed). (2009). *Prendre ses distances : de l'usage raisonné (raisonnable) de quelques statistiques pour les arbres additifs* (L'Harmattan, Paris).
- Pols, L. C., van der Kamp, L. J., and Plomp, R. (1969). "Perceptual and physical space of vowel sounds," *The Journal of the Acoustical Society of America* 46, 458-467.
- Ponton, C. W., Bernstein, L. E., and Auer, E. T., Jr. (2009). "Mismatch negativity with visual-only and audiovisual speech," *Brain topography* 21, 207-215.
- Proksch, J., and Bavelier, D. (2002). "Changes in the spatial distribution of visual attention after early deafness," *Journal of cognitive neuroscience* 14, 687-701.
- Proops, D. W., Donaldson, I., Cooper, H. R., Thomas, J., Burrell, S. P., Stoddart, R. L., Moore, A., and Cheshire, I. M. (1999). "Outcomes from adult implantation, the first 100 patients," *The Journal of laryngology and otology* 24, 5-13.
- Proverbio, A. M., and Riva, F. (2009). "RP and N400 ERP components reflect semantic violations in visual processing of human actions," *Neuroscience letters* 459, 142-146.
- Puschmann, S., Uppenkamp, S., Kollmeier, B., and Thiel, C. M. (2010). "Dichotic pitch activates pitch processing centre in Heschl's gyrus," *NeuroImage* 49, 1641-1649.
- Rauschecker, J. P. (2001). "Cortical plasticity and music," *Annals of the New York Academy of Sciences* 930, 330-336.
- Rauschecker, J. P., and Korte, M. (1993). "Auditory compensation for early blindness in cat cerebral cortex," *J Neurosci* 13, 4538-4548.
- Rauschecker, J. P., Tian, B., Korte, M., and Egert, U. (1992). "Crossmodal changes in the somatosensory vibrissa/barrel system of visually deprived animals," *Proceedings of the National Academy of Sciences of the United States of America* 89, 5063-5067.
- Roby, B. (2012). "La perception musicale chez le sujet sourd implanté cochléaire. Intérêts de l'audition résiduelle," in *Faculté de médecine* (Université Paul Sabatier, Toulouse).
- Röder, B., and Rosler, F. (eds). (2004). *Compensatory plasticity as a consequence of sensory loss* (MIT Press, Cambridge, MA).
- Rogers, C. F., Healy, E. W., and Montgomery, A. A. (2006). "Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions," *The Journal of the Acoustical Society of America* 119, 2276-2287.
- Ronnberg, J., Samuelsson, S., Lyxell, B., and Arlinger, S. (1996). "Lipreading with auditory low-frequency information. Contextual constraints," *Scandinavian audiology* 25, 127-132.
- Rosch, E. (1973). "Natural Categories," *Cognitive Psychology* 4, 328-350.
- Rosen, S. (1992). "Temporal information in speech: acoustic, auditory and linguistic aspects," *Philosophical transactions of the Royal Society of London* 336, 367-373.

- Rouger, J., Fraysse, B., Deguine, O., and Barone, P. (2008). "McGurk effects in cochlear-implanted deaf subjects," *Brain Res* 1188, 87-99.
- Rouger, J., Lagleyre, S., Demonet, J. F., Fraysse, B., Deguine, O., and Barone, P. (2011). "Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients," *Human brain mapping*.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (2007). "Evidence that cochlear-implanted deaf patients are better multisensory integrators," *Proceedings of the National Academy of Sciences of the United States of America* 104, 7295-7300.
- Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., and Foxe, J. J. (2007). "Seeing voices: High-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion," *Neuropsychologia* 45, 587-597.
- Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., and Simola, J. (1991). "Seeing speech: visual information from lip movements modifies activity in the human auditory cortex," *Neuroscience letters* 127, 141-145.
- Sandler, W. (2012). "The Phonological Organization of Sign Languages," *Language and linguistics compass* 6, 162-182.
- Sandler, W., and Lillo-Martin, D. (eds). (2006). *Sign language and linguistics universals* (Cambridge University Press).
- Sandmann, P., Kegel, A., Eichele, T., Dillier, N., Lai, W., Bendixen, A., Debener, S., Jancke, L., and Meyer, M. (2010). "Neurophysiological evidence of impaired musical sound perception in cochlear-implant users," *Clin Neurophysiol* 121, 2070-2082.
- Sattah, S., and Tversky, A. (1977). "Additive similarity trees," *Psychometrica* 42, 319-345.
- Saur, D., Kreher, B. W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M. S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). "Ventral and dorsal pathways for language," *Proceedings of the National Academy of Sciences of the United States of America* 105, 18035-18040.
- Scarborough, R., Keating, P., Mattys, S. L., Cho, T., and Alwan, A. (2009). "Optical phonetics and visual perception of lexical and phrasal stress in English," *Language and speech* 52, 135-175.
- Scherer, K. (2003). "Vocal communication of emotion: A review of research paradigms," *Speech Communication* 40, 227-256.
- Schindler, R. A., and Kessler, D. K. (1987). "The UCSF/Storz cochlear implant: patient performance," *The American journal of otology* 8, 247-255.
- Schooneveldt, G. P., and Moore, B. C. (1987). "Comodulation masking release (CMR): effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," *The Journal of the Acoustical Society of America* 82, 1944-1956.
- Schulz, E., and Kerber, M. (1994). *Music perception with the MED-EL implants* (Datenkonvertierung, Reproduktion und Druck, Vienna).
- Schwartz, J. L., Berthommier, F., and Savariaux, C. (2004). "Seeing to hear better: evidence for early audio-visual interactions in speech identification," *Cognition* 93, B69-78.
- See, R. L., Driscoll, V. D., Gfeller, K., Kliethermes, S., and Oleson, J. (2013). "Speech intonation and melodic contour recognition in children with cochlear implants and with normal hearing," *Otol Neurotol* 34, 490-498.
- Shackleton, T. M., and Carlyon, R. P. (1994). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *The Journal of the Acoustical Society of America* 95, 3529-3540.

- Shafiro, V., Gygi, B., Cheng, M. Y., Vachhani, J., and Mulvey, M. (2011). "Perception of environmental sounds by experienced cochlear implant patients," *Ear and hearing* 32, 511-523.
- Shafiro, V., Sheft, S., Gygi, B., and Ho, K. T. (2012). "The influence of environmental sound training on the perception of spectrally degraded speech and environmental sounds," *Trends in amplification* 16, 83-101.
- Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). "Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians," *J Neurosci* 23, 5545-5552.
- Shahin, A. J., Picton, T. W., and Miller, L. M. (2009). "Brain oscillations during semantic evaluation of speech," *Brain and cognition* 70, 259-266.
- Shamma, S., and Klein, D. (2000). "The case of the missing pitch templates: how harmonic templates emerge in the early auditory system," *The Journal of the Acoustical Society of America* 107, 2631-2644.
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," *Hearing research* 11, 157-189.
- Shannon, R. V., Fu, Q. J., and Galvin, J., 3rd (2004). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta oto-laryngologica*, 50-54.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science (New York, N.Y)* 270, 303-304.
- Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proceedings of the National Academy of Sciences of the United States of America* 99, 3318-3323.
- Skinner, M. W., Fourakis, M. S., Holden, T. A., Holden, L. K., and 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 and hearing* 20, 443-460.
- Speer, S. R., Kjølgaard, M. M., and Dobroth, K. M. (1996). "The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities," *Journal of psycholinguistic research* 25, 249-271.
- Stabej, K. K., Smid, L., Gros, A., Zargi, M., Kosir, A., and Vatovec, J. (2012). "The music perception abilities of prelingually deaf children with cochlear implants," *International journal of pediatric otorhinolaryngology* 76, 1392-1400.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., and Formisano, E. (2009). "Sound categories are represented as distributed patterns in the human auditory cortex," *Curr Biol* 19, 498-502.
- Starr, A., Picton, T. W., Sininger, Y., Hood, L. J., and Berlin, C. I. (1996). "Auditory neuropathy," *Brain* 119 ( Pt 3), 741-753.
- Straatman, L. V., Rietveld, A. C., Beijen, J., Mylanus, E. A., and Mens, L. H. (2010). "Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid," *The Journal of the Acoustical Society of America* 128, 1884-1895.
- Strelnikov, K., Rouger, J., Demonet, J. F., Lagleyre, S., Fraysse, B., Deguine, O., and Barone, P. (2010). "Does brain activity at rest reflect adaptive strategies? Evidence from speech processing after cochlear implantation," *Cereb Cortex* 20, 1217-1222.
- Strelnikov, K. N., Vorobyev, V. A., Chernigovskaya, T. V., and Medvedev, S. V. (2006). "Prosodic clues to syntactic processing--a PET and ERP study," *NeuroImage* 29, 1127-1134.



- Terhardt, E. (1974). "Pitch, consonance, and harmony," *The Journal of the Acoustical Society of America* 55, 1061-1069.
- Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., and Peretz, I. (2011a). "Congenital Amusia (or Tone-Deafness) Interferes with Pitch Processing in Tone Languages," *Frontiers in psychology* 2, 120.
- Tillmann, B., Rusconi, E., Traube, C., Butterworth, B., Umiltà, C., and Peretz, I. (2011b). "Fine-grained pitch processing of music and speech in congenital amusia," *The Journal of the Acoustical Society of America* 130, 4089-4096.
- Timm, L., Agrawal, D., F, C. V., Sandmann, P., Debener, S., Buchner, A., Dengler, R., and Wittfoth, M. (2012). "Temporal feature perception in cochlear implant users," *PLoS one* 7, e45375.
- Trehub, S. E., and Hannon, E. E. (2006). "Infant music perception: domain-general or domain-specific mechanisms?," *Cognition* 100, 73-99.
- Tyler, R. S., Moore, B. C., and Kuk, F. K. (1989). "Performance of some of the better cochlear-implant patients," *Journal of speech and hearing research* 32, 887-911.
- Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (2002). "Patients utilizing a hearing aid and a cochlear implant: speech perception and localization," *Ear and hearing* 23, 98-105.
- UKCISG (2004). "Criteria of candidacy for unilateral cochlear implantation in postlingually deafened adults III: prospective evaluation of an actuarial approach to defining a criterion," *Ear and hearing* 25, 361-374.
- Undurraga, J. A., Carlyon, R. P., Macherey, O., Wouters, J., and van Wieringen, A. (2012). "Spread of excitation varies for different electrical pulse shapes and stimulation modes in cochlear implants," *Hearing research* 290, 21-36.
- Van Lancker, D., and Kreiman, J. (1987). "Voice discrimination and recognition are separate abilities," *Neuropsychologia* 25, 829-834.
- Van Lancker, D. R., and Canter, G. J. (1982). "Impairment of voice and face recognition in patients with hemispheric damage," *Brain and cognition* 1, 185-195.
- Van Lancker, D. R., Cummings, J. L., Kreiman, J., and Dobkin, B. H. (1988). "Phonagnosia: a dissociation between familiar and unfamiliar voices," *Cortex; a journal devoted to the study of the nervous system and behavior* 24, 195-209.
- van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). "Visual speech speeds up the neural processing of auditory speech," *Proceedings of the National Academy of Sciences of the United States of America* 102, 1181-1186.
- Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., and McDermott, H. J. (2005). "Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies," *The Journal of the Acoustical Society of America* 117, 3126-3138.
- Vanderveer, N. J. (1980). "Ecological acoustics: human perception of environmental sounds," *Dissertation Abstracts International* 40, 4543.
- Venail, F., Sicard, M., Piron, J. P., Levi, A., Artieres, F., Uziel, A., and Mondain, M. (2008). "Reliability and complications of 500 consecutive cochlear implantations," *Archives of otolaryngology--head & neck surgery* 134, 1276-1281.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., and Giraud, A. L. (2003). "Modulation of neural responses to speech by directing attention to voices or verbal content," *Brain research* 17, 48-55.
- Vongphoe, M., and Zeng, F. G. (2005). "Speaker recognition with temporal cues in acoustic and electric hearing," *The Journal of the Acoustical Society of America* 118, 1055-1061.

- Wackym, P. A., Firszt, J. B., Gaggl, W., Runge-Samuels, C. L., Reeder, R. M., and Raulie, J. C. (2004). "Electrophysiologic effects of placing cochlear implant electrodes in a perimodiolar position in young children," *The Laryngoscope* 114, 71-76.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). "Analyzing pitch chroma and pitch height in the human brain," *Annals of the New York Academy of Sciences* 999, 212-214.
- Warrier, C. M., and Zatorre, R. J. (2004). "Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task," *Brain* 127, 1616-1625.
- Weintraub, S., Mesulam, M. M., and Kramer, L. (1981). "Disturbances in prosody. A right-hemisphere contribution to language," *Archives of neurology* 38, 742-744.
- Weisel, A. (1985). "Deafness and perception of nonverbal expression of emotion," *Perceptual and motor skills* 61, 515-522.
- Wiefferink, C. H., Rieffe, C., Ketelaar, L., De Raeve, L., and Frijns, J. H. (2013). "Emotion understanding in deaf children with a cochlear implant," *Journal of deaf studies and deaf education* 18, 175-186.
- Williams, C. E., and Stevens, K. N. (1972). "Emotions and speech: some acoustical correlates," *The Journal of the Acoustical Society of America* 52, 1238-1250.
- Wilson, B. S., and Dorman, M. F. (2008). "Cochlear implants: a remarkable past and a brilliant future," *Hearing research* 242, 3-21.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature* 352, 236-238.
- Wingfield, A. (1975). "Acoustic redundancy and the perception of time-compressed speech," *Journal of speech and hearing research* 18, 96-104.
- Wong, P. C., Nusbaum, H. C., and Small, S. L. (2004). "Neural bases of talker normalization," *Journal of cognitive neuroscience* 16, 1173-1184.
- Xia, N., Wu, X. Y., Wang, X., Mou, Z. X., Wang, M. Q., Gu, X., Zheng, X. L., and Hou, W. S. (2013). "Pulsed 808-nm infrared laser stimulation of the auditory nerve in guinea pig cochlea," *Lasers in medical science*.
- Yildirim, S., Bulut, M., and Lee, M. (2004). "An acoustic study of emotions expressed in speech," in *Proceedings of the International Conference on Spoken Language Processing* (Online archive of International Speech Communication Association proceedings available at: <http://www.isca-speech.org>), pp. 2193-2196.
- Zeng, F. G. (2002). "Temporal pitch in electric hearing," *Hearing research* 174, 101-106.
- Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargava, A., Wei, C., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proceedings of the National Academy of Sciences of the United States of America* 102, 2293-2298.
- Zhang, F., Benson, C., and Cahn, S. J. (2013). "Cortical encoding of timbre changes in cochlear implant users," *Journal of the American Academy of Audiology* 24, 46-58.



# Annexes

## **1. Publications en rapport avec la thèse**

### **Gender recognition in cochlear implanted subjects.**

Z. Massida, M. Marx, P. Belin, C. James, B. Fraysse, P. Barone and O. Deguine. (contribution égale des deux premiers auteurs). *Journal of Speech Language and Hearing Research*. Accepté

### **Cortical prediction of auditory emphasis by facial expressions during audiovisual**

**continuous speech.** K. Strelnikov, J. M. Foxtan, M. Marx and P. Barone. *Brain Topography*.  
En revision

### **Speech prosody perception in cochlear implant users with and without residual hearing**

M. Marx, C. James, JM Foxtan, A. Capber, B. Fraysse, P. Barone and O. Deguine. *Ear and hearing*. En revision

## **2. Communications orales en rapport avec la thèse (orateur)**

### ***20th IFOS World Congress, Juin 2013, Seoul***

Prosodic cues in cochlear implant users

### ***Congrès National SFORL, 2012 :***

Perception de la musique par le patient implanté cochléaire.

### ***Journées scientifiques GEMO, Janvier 2012***

Perception de la hauteur tonale et de la prosodie linguistique par le patient implanté cochléaire. Intérêts de l'audition résiduelle.

### ***Congrès National SFORL, 2011 :***

Perception de la hauteur tonale et de la prosodie linguistique par le patient implanté cochléaire. Intérêts de l'audition résiduelle.

*Journées scientifiques GEMO, Janvier 2010*

Perception de la voix et de la prosodie par le patient implanté cochléaire : projet d'étude.

**Prix GEMO 2010.**

**3. Communications affichées en rapport avec la thèse**

*Congrès international Cochlear implants Stockholm, juin 2010 :*

Voice detection through a cochlear implant. Z. Massida, M. Marx

Voice features perception with a cochlear implant. M. Marx, Z. Massida

**4. Travaux encadrés pendant la période de la thèse**

*Perception de la hauteur tonale et de la prosodie linguistique par le patient implanté cochléaire.* A. Capber. Master 2 Neuropsychologie 2009

*Perception musicale par le patient implanté cochléaire.* B. Roby. Thèse de Médecine, 2012.

**5. Tableau récapitulatif des sujets implantés cochléaires inclus**

Sujets	Catégorisation	Prosodie affective	Prosodie linguistique	Musique	Age	Cause de la surdité	Genre	Délai post activation (mois)	Modèle d'implant	Côté implanté	Audition résiduelle
AL..R	x				72	Inconnue	F	71	Cochlear CI 24	droit	non
AN..J	x			x	53	Inconnue	F	68	Cochlear CI 24	droit	non
AN..Ch	x			x	48	Génétique	F	53	Cochlear CI 24	droit	oui
AU..A	x				54	Otospongiose	M	36	Cochlear CI512	bilat	non
BE..C	x			x	56	Inconnue	F	77	Cochlear Hybrid L	droit	oui
CH..J	x			x	70	Lobstein	F	84	Advanced Bionics HiRes	gauche	oui
DA..C	x			x	52	Lobstein	F	55	MedEl sonata	droit	non
ES..P	x			x	48	Inconnue	M	134	Cochlear CI 24	droit	non
GE..J	x				52	Inconnue	M	90	Cochlear CI 24	gauche	non
GI..G	x				64	Otospongiose	F	42	Cochlear CI 24	droit	non
LA..C	x			x	62	Otite chronique	F	33	Cochlear CI512	droit	oui
LE..M	x			x	47	Inconnue	M	129	Cochlear CI22	droit	non
MA..S	x	x		x	66	Inconnue	F	36	Advanced Bionics HiRes	gauche	oui
MA..A	x				76	Inconnue	F	23	Cochlear CI512	droit	non
RA..C	x		x	x	70	Inconnue	M	68	Advanced Bionics HiRes	droit	oui
SE..M	x			x	45	Inconnue	F	83	Cochlear CI 24	droit	non
FA..P		x			38	Ototoxicité	M	2	Cochlear MRA	gauche	oui
CA..C		x			56	Labyrinthite	F	85	Cochlear CI 24	droit	non
LL..S		x			73	Inconnue	M	85	Cochlear CI 24	droit	non
MA..B		x	x		61	Otospongiose	F	12	Cochlear CI 24	droit	oui
GE..M		x			53	Inconnue	F	24	Cochlear CI 24	droit	non
GI..F		x			72	Ototoxicité	M	42	Cochlear CI 24	gauche	non

Catégorisation	Prosodie affective	Prosodie linguistique	Musique	Age	Cause de la surdité	Genre	Délai post activation (mois)	Modèle d'implant	Côté implanté	Audition résiduelle
FO..R	x		x	76	Inconnue	M	36	Advanced Bionics HiRes	gauche	oui
JA..J	x	x		70	Otospongiose	M	10	Advanced Bionics HiRes	droit	oui
AV..C	x			72	Inconnue	M	12	Advanced Bionics HiRes	droit	oui
SU..M	x			66	Inconnue	F	6	Cochlear CI 24	droit	oui
HE..J	x			85	Inconnue	F	10	MedEl sonata	droit	non
AR..J	x			52	Traumatisme	M	109	Cochlear CI 22	droit	non
LI..D	x	x		70	Inconnue	F	84	Cochlear CI 24	gauche	non
SI..C		x		43	Inconnue	F	68	Cochlear CI 24	droit	oui
DA..V		x	x	51	Inconnue	F	60	Cochlear Hybrid L	gauche	oui
CA..N		x		47	Inconnue	F	51	Cochlear Hybrid L	droit	oui
PA..B		x		70	Inconnue	M	7	Cochlear MRA	gauche	oui
SK..J		x		62	Inconnue	M	7	Cochlear CI 24	gauche	oui
BR..D		x		42	Inconnue	M	18	Cochlear CI 24	gauche	oui
RO..N		x		43	Inconnue	F	50	Cochlear CI 24	droit	oui
LE..E		x		21	Inconnue	F	10	Cochlear CI 24	gauche	oui
GA..J		x		68	Inconnue	M	13	Advanced Bionics HiRes	gauche	oui
HO..C		x		76	Traumatisme	M	15	Advanced Bionics HiRes	gauche	non



Catégorisation	Prosodie		Musique	Age	Cause de la surdité	Genre	Délai post activation (mois)	Modèle d'implant	Côté implanté	Audition résiduelle
	affektive	linguistique								
GO..L	x			49	Inconnue	F	24	MedEl sonata	droit	oui
CA..P	x			67	Inconnue	M	109	Cochlear CI 24	gauche	non
LE..P		x		54	Inconnue	M	44	Advanced Bionics HiRes	droit	non
PR..M		x		73	Meniere	F	7	Advanced Bionics HiRes	droit	non
RI..G		x		76	Meniere	M	26	Advanced Bionics HiRes	droit	non
GA..J		x	x	44	Inconnue	M	60	Cochlear CI 24	gauche	non
RI..R		x	x	69	Otospongiose	M	36	Cochlear CI 24	droit	non
AL..JJ		x		69	Inconnue	M	55	Cochlear CI 24	gauche	non
GI..G			x	62	Otospongiose	F	23	Cochlear CI 24	droit	non
CO..P			x	78	Déhiscence	F	84	Cochlear CI 24	droit	oui
BE..J			x	58	Otospongiose	M	108	Advanced Bionics HiRes	gauche	non
DU..J			x	76	Otite chronique	M	110	Cochlear CI 24	gauche	non
FO..M			x	60	Inconnue	F	50	Cochlear CI 24	gauche	non

## **6. Autres publications orientées implantation cochléaire, réalisées pendant la thèse**

**Cochlear implants in elderly: performance outcomes in the long term. L'implant cochléaire chez le sujet âgé : évolution des performances à long terme.**

M. Marx, D. Gutierrez, B. Lepage, E. Koury, ML Laborde, O. Deguine, B. Fraysse. *Rev laryngol Otol Rhinol (Bord)*.2011, 132 (4-5) :187-91.

**L'implant cochléaire : indications actuelles et perspectives.** Mathieu Marx, Olivier Deguine, Nadine Cochard, Chris James, Marie-Laurence Laborde et Bernard Fraysse. Monographie Amplifon 2011 « Recherche fondamentale en ORL »

**Probability of improvement after cochlear implantation as a function of pre operative residual hearing.**

M. Marx, C. James, T. ElHadi, N. Martin-Dupont, ML Laborde, O. Deguine, B. Fraysse. *Audiology Neurotology*, 2011 (16) Suppl 1.

**Development and evaluation of the modiolar research array--multi-centre collaborative study in human temporal bones.**

Briggs RJ, Tykocinski M, Lazsig R, Aschendorff A, Lenarz T, Stöver T, Fraysse B, Marx M, Roland JT Jr, Roland PS, Wright CG, Gantz BJ, Patrick JF, Risi F. *Cochlear Implants Int*. 2011 Aug;12(3):129-39.

**Benefits from upgrade to the CP810™ sound processor for Nucleus® 24 cochlear implant recipients.** Mosnier I, Marx M, Venail F, Loundon N, Roux-Vaillard S, Sterkers O. *Eur Arch Otorhinolaryngol*. 2013 Feb 14. [Epub ahead of print]

**Reliability of cone beam computed tomography in scalar localization of the electrode array: a radio histological study.** Marx M, Risi F, Escudé B, Durmo I, James C, Deguine O, Fraysse B. *Eur Arch Otorhinolaryngol*. 2013, Mar 28 [Epub ahead of print]

**Cochlear implant feasibility in rhesus macaque monkey: anatomical and radiological results.** Marx M, Girard P, Ecludé B, Barone P, Fraysse B, Deguine O. (contribution égale des deux premiers auteurs). *Otol. Neurotol*. Accepté

**European multi-centre study of the Nucleus Hybrid L24 cochlear implant.**

T. Lenarz, C. James, D. Cuda, A. Fitzgerald-O'Connor, B. Frachet, J. Frijns, T. Klenzner, R. Laszig, M. Manrique, M. Marx, P. Merkus, E. Mylanus, E. Officiers, J. Pesch, A. Ramos-Macias, A. Robier, O. Sterkers and A. Uziel. *Int. J Audiol.* *Accepté*



## Gender categorization in cochlear-implanted deaf subjects

Journal:	<i>Journal of Speech, Language, and Hearing Research</i>
Manuscript ID:	JSLHR-H-12-0132.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Massida, Zoé; CNRS UMR 5549, Cerveau & Cognition Marx, Mathieu; CHU Purpan, Oto-Rhino-Laryngologie & Oto-Neurologie; CNRS UMR 5549, Cerveau & Cognition Belin, Pascal; University of Glasgow, Voice Neurocognition Laboratory James, Chris; Cochlear, Cochlear France Fraysse, Bernard; CHU Purpan, Oto-Rhino-Laryngologie & Oto-Neurologie BARONE, Pascal; CNRS UMR 5549, Cerveau & Cognition Deguine, Olivier; CHU Purpan, Oto-Rhino-Laryngologie & Oto-Neurologie
Keywords:	Cochlear Implants, Voice, Speech perception

**Gender categorization in cochlear-implanted deaf subjects.**

Massida Z.<sup>1,2,¶</sup>, M. Marx<sup>1,2,5,¶</sup>, P. Belin<sup>3</sup>, C. James<sup>4,5</sup>, B. Fraysse<sup>5</sup>, P. Barone<sup>1,2\*</sup> and O. Deguine<sup>1,2,5</sup>

**Shortened title: Gender categorization in CI**

- 1., Université Toulouse, CerCo, Université Paul Sabatier.
2. CNRS, UMR 5549, 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

Authors contribution. ¶ ZM and MM contributed equally to this work.

\* Corresponding author

Pascal Barone

e-mail: [pascal.barone@cerco.ups-tlse.fr](mailto:pascal.barone@cerco.ups-tlse.fr)

CNRS-Universite Paul Sabatier Toulouse 3, Centre de Recherche Cerveau & Cognition

UMR 5549, Pavillon Baudot CHU Purpan, 31062 Toulouse CEDEX9, France.

Phone: +33 (0)5 62 17 37 79

Fax: +33 (0)5 62 17 28 09

Number of words:

Abstract: 198

Introduction: 971

Discussion: 1675

**Abstract.**

*Purpose.* The present study examined the ability of cochlear implanted (CI) deaf subjects to discriminate voice gender and how this ability evolved as a function of CI experience.

*Method.* We presented a continuum of voice samples created by voice morphing with 9 intermediate acoustic parameter steps between typical male and typical female: this method allowed us to evaluate gender categorization when acoustical features were specific to a gender but also for more ambiguous cases, when fundamental frequency or formant distribution were located between typical values.

*Results.* Results showed a global, though variable, deficit for voice gender categorization in CI recipients compared to normal hearing subjects (NHS). This deficit was stronger for ambiguous stimuli in the voice continuum: average performance for CI users were 58% lower than average scores for NHS in cases of ambiguous stimuli and 19% points lower for typical male/female voices. We found no significant improvement in voice gender categorization with CI experience.

*Conclusions.* These results emphasize the dissociation between recovery of speech recognition and voice feature perception after cochlear implantation. This large and durable deficit may be related to spectral and temporal degradation induced either by CI sound coding or to central voice processing deficit.

**Key words.** Voice processing. Deafness. Cochlear implant. Gender.

## Introduction

A cochlear implant (CI) is an effective neuroprosthesis that allows deafened subjects to recover auditory abilities, especially speech recognition. Indeed speech recognition scores in CI subjects often exceed 80% for sentences or words in quiet (Rouger et al., 2007; UKCISG, 2004; Wilson & Dorman, 2008). Efficient encoding of temporal envelope information in speech and intense speech therapy following implantation are probably responsible for these remarkable outcomes. However other aspects of speech perception remain difficult for CI users, including speech understanding in noise (Fu, Shannon, & Wang, 1998; Munson & Nelson, 2005), prosody recognition and music perception (El Fata, James, Laborde, & Fraysse, 2009; Gfeller et al., 2002; Kong, Cruz, Jones, & Zeng, 2004; Leal et al., 2003). These difficulties can be related to the limitations of current cochlear implants, providing significantly fewer effective frequency channels than normal hearing (see (McKay, 2005) for review). Cochlear implant users usually experience difficulties in recognizing paralinguistic information due to degradation of the acoustic signal after acoustic-to-electric coding. Indeed, sound coding results in significant spectral degradation as the original sound spectrum is necessarily divided into a limited number of frequency bands that correspond to the number of activated cochlear electrodes, alternately stimulated depending on the sound processor. Further the auditory information is not only spectrally degraded but also fine temporal structure is not represented which is important for certain aspects of speech recognition (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Zeng et al., 2005).

By manipulating spectral and temporal components of speech, it has been shown that linguistic information is not supported by the same spectro-temporal modulation as for paralinguistic information (Elliott & Theunissen, 2009). Indeed, several studies have revealed a dissociation between speech recognition recovery and discrimination of human voice features, such as identity or emotion of voice perception (Cleary & Pisoni, 2002; Cleary, Pisoni, & Kirk, 2005; Vongphoe & Zeng, 2005). In addition, we have shown that CI recipients present a large deficit in distinguishing the human voice from environmental

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
noise when relevant speech information is not available; such as the longer-term temporal envelope (Massida et al., 2011). It is important to take into account such a deficit in voice and voice feature perception because it has a large negative impact on the social communication of the CI users.

13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
In the present study we focused on the ability of CI users to discriminate a specific feature of the human voice: the gender. In a hierarchical model of voice perception, as proposed by Belin et al. (Belin, Fecteau, & Bedard, 2004), recognition of voice attributes, such as identity cues, can be performed after a first low level acoustic analysis that distinguishes human voice from environmental sounds. Voice gender perception relies on several acoustical cues among which the main ones are: voice fundamental frequency (F0) which determines voice pitch, and the formant distribution that underlies voice timbre. F0 constitutes an important cue for voice gender recognition as the F0 values for a female voice typically differ from a male voice by one octave or more (see (Titze, 1989)). Fundamental frequency reflects vocal fold vibration frequency and therefore mainly relates their size and tension. Voice timbre is often referred to as "voice quality" and also provides useful information regarding voice gender. Formants are composed of harmonics of F0 and result from resonances in the vocal tract; mostly the supra laryngeal cavities: Vocal tract size, length and shape strongly differ between genders (in adults) and so formants frequencies differ between males and females. According to Lavner et al. (Lavner Y, 2000), the third and fourth formants would be particularly useful to convey identity cues to the listener.

42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
Until now, previous studies have used a two-alternative forced choice (2AFC) between one typical male and one typical female voice to test the subjects' ability to discriminate gender. Using such a paradigm, CI subjects present a performance level of correct gender identification that ranges from 70% to 95% (Fu, Chinchilla, & Galvin, 2004). In addition, the use of a noise band vocoder that mimics the processing of a sound by a cochlear implant allowed Fu et al. to perform a direct comparison between CI subjects and normal hearing subjects. Gender recognition performance in CI users was comparable to normal hearing subjects with 4 to 8 channels, which highlighted the role of



1  
2  
3  
4  
5  
6  
7 both spectral and fine temporal cues in voice gender discrimination (Fu et al., 2004; Fu,  
8 Chinchilla, Nogaki, & Galvin, 2005).

9  
10 However the voice F0 distribution among the general population is highly variable  
11 (Andrews & Schmidt, 1997) and in CI users gender categorization appeared to be  
12 impaired in cases of ambiguous F0 values within the region of overlap of distributions for  
13 males/females (Kovacic & Balaban, 2009). In order to evaluate precisely the impairment  
14 of voice gender perception in CI users when voice attributes are ambiguous, we used a  
15 task involving stimuli created from morphing from male voice to female voice, which  
16 included several acoustical intermediates on a continuum between one typical male voice  
17 and one typical female voice (Pernet & Belin, 2012). Such a paradigm allowed us to  
18 quantify the ability of CI users to perform voice gender categorization in a tightly  
19 controlled fashion especially when acoustical features presented varying degrees of  
20 ambiguity. The present study had two main goals: first to precisely quantify the deficit in  
21 gender categorization of CI users especially when acoustic cues of the voice (pitch and  
22 timber) were ambiguous. We hypothesized that CI subjects performance would be  
23 particularly affected when the voice features were ambiguous. Our second objective was  
24 to analyze the development of CI users performance in such a categorization task as a  
25 function of experience with CI. Further, most studies on auditory recovery in CI patients  
26 reveal a large variability in performance levels (Rouger et al., 2007; Wilson & Dorman,  
27 2008). The origin of such variability is diverse, from the etiology of deafness, the severity  
28 of deafness (residual hearing) and others sources including the CI sound coding strategy.  
29 In order to rule out such sources of variability we performed a longitudinal analysis on a  
30 selected group of patients from the time of implantation up to more than 18 months of  
31 experience. Based on our previous findings concerning human voice discrimination, we  
32 expected that gender discrimination would not be influenced by time after implantation.  
33 This hypothesis is underlied by the differences between acoustic cues required for  
34 linguistic and paralinguistic information perception and by the hierarchical organization of  
35 voice processing in the brain (Campanella & Belin, 2007).  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## **2. Materials and Methods**

### **2.1 Participants.**

Fourteen *normally hearing* (NHS) native French speakers (7 males; age: mean  $\pm$ s.d.: 24.6  $\pm$  2.9 years) were included in this study. Subjects' histories were collected through a questionnaire and none of the subjects participating in the study had a history of auditory, neurological or psychiatric disorders.

A total of 42 CI subjects participated in the study. They were divided into two groups (Table 1). *The transversal group*: 32 unilateral cochlear implanted (CI) deaf adults (15 males, 17 females; age: 54.5  $\pm$  15 years; 15 right CI, 17 left CI). *The Follow-up group*: 10 cochlear implanted deaf adults (5 males, 5 females; age: 51.9  $\pm$  16; 5 right CI, 5 left CI).

The CI subjects were on average older than the normal hearing control subjects but their age ranged from 20 to 80 years. Hearing can be affected by age (Arehart, Souza, Muralimanohar, & Miller, ; Gratton & Vazquez, 2003); for example when a gender discrimination task is performed through a cochlear implant simulation (Schvartz & Chatterjee, 2012; Souza, Arehart, Miller, & Muralimanohar, 2011). However we found no correlation between age and performance level in our discrimination tasks, albeit for our relatively limited group of patients. In consequence we adopted the strategy of comparing the patient performance in voice discrimination with that obtained from a homogeneous group of young normally hearing subjects with no evidence of hearing loss. Voice discrimination was tested during regular visits to the ENT department following the standard rehabilitation program. Post-activation delay varied from 1 day to 131 months and subjects were divided into 4 different sub-groups according to duration of implant use. All subjects had postlingually acquired profound bilateral deafness of diverse etiologies (meningitis, chronic otitis, otosclerosis, etc.) and durations. Only 2 subjects had presented with sudden deafness which occurred 2 and 3 years before CI,

1  
2  
3  
4  
5  
6  
7 respectively. In all the other subjects, the deafness was progressive. Duration of hearing  
8 loss for each patient is shown in Table 1. Because of this progressive hearing  
9 impairment, the duration of deafness could not be reliably defined and consequently we  
10 did not attempt to correlate this measure with any of the performance levels presented  
11 by these subjects. Further, 16 subjects of the transversal group and 4 of the follow-up  
12 had a hearing aid in the non-implanted ear and used it in daily life. These subjects were  
13 always tested with the implant alone. We have analyzed separately the performances of  
14 the subjects (slope values of the psychometric categorization function) who have a daily  
15 use of a hearing aid, but we did not observe a difference compared to the patients using  
16 the CI alone.  
17

18  
19 Information concerning the subjects is provided in Table 1 for the transversal and for the  
20 follow-up group. All participants gave written informed consent prior to their inclusion in  
21 the study.  
22  
23

## 24 25 26 27 28 29 30 31 **2.2 Stimulus material.**

32 All stimuli used in our experiment were developed at the Voice Neurocognition Laboratory  
33 (Univ. Glasgow). We used in this experiment a subtest of the Voice Perception  
34 Assessment (VPA) battery. This subtest was used in a previous published work by Pernet  
35 and Belin (Pernet & Belin, 2012). The task requires subjects to categorize by gender  
36 voice stimuli from a morphing-generated voice continuum between a male and a female  
37 voice speaking the same syllable "had". The two extreme voices each correspond to an  
38 average voice with a spectrum made of 16 voices of the same gender. Morphing was  
39 performed using STRAIGHT (Hideki Kawahara, University of Wakayama) (Kawahara H,  
40 1999) in Matlab (the Mathworks, Inc., Natick, MA, USA). STRAIGHT performs an  
41 instantaneous pitch-adaptive spectral smoothing to separate the contributions of the  
42 glottal source (including F0) from the supra-laryngeal filtering (distribution of spectral  
43 peaks, including the first formant F1) to the voice signal. Voice stimuli are decomposed  
44 by STRAIGHT into five parameters: fundamental frequency (F0), formant frequencies,  
45 duration, spectro-temporal density and aperiodicity; each parameter can be  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

1  
2  
3  
4  
5  
6 independently manipulated. Anchor points, i.e time-frequency landmarks were  
7 determined in both extreme voices based on easily recognizable features of the  
8 spectrograms. The temporal landmarks were defined as the onset, the offset and initial  
9 burst of the sound. Spectro-temporal anchors were the first and second formant at onset  
10 of phonation, onset of formant transition and end of phonation. Using the temporal  
11 anchors, elements of the continuum were equalized in duration (39.2 ms long, i.e, 17289  
12 data points at 44100 Hz). Morphed stimuli were then generated by re-synthesis based on  
13 a logarithmic interpolation of female and male anchor templates and spectrograms in  
14 steps of 10%. We thus obtained a continuum of 11 voices ranging from 100% female  
15 (re-synthesized female stimulus) to 100% male (re-synthesized male stimulus) with 9  
16 gender-interpolated voices in 10% steps (90% female-10% male, 80% female-20%  
17 male,..., 10% female-90% male) (see fig. 1A and B). Values of F0, F1, F2, F3 and F4 for  
18 each position in the continuum are presented in Table 2.

19  
20  
21 In addition, voice/non-voice discrimination was also tested in eleven patients as in a  
22 previous study we performed in CI patients (Massida et al., 2011). This test assesses CI  
23 subjects' abilities to distinguish human voice sounds from environmental sounds. Two  
24 sets of 500 ms long stimuli were created: the first set contained 55 different human  
25 voice stimuli, including 29 speech stimuli (phonemes presented in a/h/-vowel/d context,  
26 words in different languages, or non semantic syllables) and 26 non speech stimuli (e.g  
27 laughs, coughs). The second set contained 55 non voice stimuli consisting of a wide  
28 variety of environmental sounds, including sounds from cars, telephone, bells, streaming  
29 water. Neither group contained animal vocalizations.

30  
31 All subjects were tested on open set recognition for French dissyllabic words obtained  
32 from the standard speech therapist's Fournier lists presented through loudspeakers.

### 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 **2.3 Stimulus presentation and procedure.**

50  
51 Subjects were tested in a sound-attenuated chamber with volume adjusted to 72 dB SPL.  
52 For NHS the intensity was measured at the ear, while for CI subjects, intensity was  
53 measured at the distance of the patient from the loudspeakers. NHS were tested at the  
54  
55

1  
2  
3  
4  
5  
6 CerCo Laboratory and CI subjects in the ENT Department. Stimuli (16-bits, mono, 22050  
7 Hz sampling rate) were presented binaurally to the control group via Sennheiser EH250  
8 headphones, as in the regular speech recognition assessments. The stimuli were  
9 presented to the CI users in sound field through loudspeakers (KINYO, model PS-240).

10  
11 The 11 voices of the continuum were each presented 10 times in a pseudo-random order  
12 (110 stimuli presentations). The task for NHS and CI subjects was a 2-alternative forced  
13 choice (2AFC) categorization: male vs. female. NHS were tested with a 1-s inter-trial  
14 delay and were instructed to respond as quickly and accurately as possible using the left  
15 or right control buttons of the computer keyboard corresponding to their answer (male or  
16 female). The response keys were counterbalanced across subjects. CI subjects were  
17 tested with a 1.5 s inter-trial delay and were instructed to answer as accurately as  
18 possible, with no reference to reaction time. Despite the longer inter-trial delay, the short  
19 duration of the stimuli (39.2 ms) makes the task difficult for the CI subjects, and in few  
20 trials some subjects did not provide a response. Such cases were more often present in  
21 subjects tested on the first day of activation of the implant in 8% of trials (mean=7.8  
22  $\pm$ 11.1). In experienced subjects, such behavior occurred in less than 2% of trials  
23 (mean=2.2 $\pm$ 3.6). In consequence, these trials were excluded from the analysis and only  
24 definite responses were retained. For each voice on the continuum we measured the  
25 percent of "female" responses. The test lasted about 5 minutes for NHS and from 7 to 10  
26 minutes for CI subjects.

27  
28 For the follow-up group, CI subjects were tested on the first day of implant activation,  
29 and then at all sessions with the hospital speech therapist: at 1 month, 3 months, 6  
30 months, 9 months, 12 months and 18 months after implant activation. Each time the  
31 procedure was the same.

32  
33 For the voice/non-voice discrimination task, eleven CI subjects were also tested in the  
34 same acoustic conditions as for the gender discrimination task (loudspeakers at 72 dB  
35 SPL), using a 2-alternative forced choice (2AFC) categorization: voice vs non voice. CI  
36 subjects were tested with a 1.5 s inter-trial delay and were instructed to answer as  
37 accurately as possible, using the left or right control buttons of the computer keyboard

1  
2  
3  
4  
5  
6  
7 corresponding to their answer (Voice or Non-Voice). The response keys were  
8 counterbalanced across subjects.  
9

### 10 11 **2.3 Analysis**

12  
13 As a primary measure we used the rate of responses "female" for each of the 11 voices,  
14 from the male voice to the female voice. A Boltzmann sigmoidal function was fitted to  
15 these points using a non-linear least-squares procedure (Levenburg-Marquart, Origin  
16 v6.1). Categorization was studied among several criteria: Firstly, we measured on the  
17 curve the abscissa for which the subjects responded 'female' on 50% of the trials (C50  
18 threshold). Secondly, we estimated the slope of the psychometric function, which is  
19 maximal at the centre of symmetry of the curve, that is the point for subjective equality.  
20

21  
22 The slope was calculated using an original Matlab function and corresponded to the value  
23 of the derivative function of the sigmoid psychometric curve at the point for subjective  
24 equality. We used the slope value as the main criterion of voice gender categorization  
25 ability, particularly sensitive when the stimulus is ambiguous. In a normal categorization  
26 task, the slope reflects an abrupt shift from male to female categorization (see Figure  
27 1C). In case of deficit in gender discrimination, the subject responses will be more  
28 variable for a given stimulus, especially for the ambiguous voices. Such behavior would  
29 be expressed as a flatter psychometric function and consequently lower slope value.  
30

31  
32 Thirdly, we analyzed the percent correct of gender recognition at for typical male and  
33 female voices at the ends of the continuum.  
34

35  
36 To analyze the global effect of the implantation delay on C50 threshold, slope and correct  
37 recognition measures, we used an ANOVA for independent measures. Bonferroni/Dunn  
38 test was used for post-hoc analysis. A t-test was used to measure differences between CI  
39 subjects and NHS and between recognition of male and female voices.  
40

41  
42 Concerning voice/non-voice discrimination performance, we measured  $d'$ , which is a  
43 criterion of perception sensitivity independent from decision bias, relying on hit rate and  
44 false alarm rate according to Signal Detection Theory (Swets, Harris, McElroy, & Rudloe,  
45 1966; Tanner & Swets, 1954).  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

### **3. Results**

#### **3.1 Normally hearing subjects.**

As explained in the Methods section, the advantage of our gender categorization task is that several criteria can be analyzed to evaluate subjects' performance. Figure 1C illustrates the performance of the NHS group during the categorization task. Subjects categorized correctly the voice at the extremities (mean 99.6%  $\pm$ 1.9) which corresponded to re-synthesized typical, unambiguous voices (100% male or 100% female). When the voice stimuli were morphed and corresponded to ambiguous voices (50% on the continuum), subjects show categorization performance at chance level. In between, the performance of the subjects is shifted progressively toward the major feature (predominantly male or female) of the morphed voice. Globally, the psychometric curves of the subjects can be fitted with a sigmoid function from which two supplementary criteria can be computed for comparison with subjects' performance: the C50 threshold correspond to the stimulus of the continuum which elicit a chance level categorization; the slope of the sigmoid at the center of symmetry of the curve. For the NHS control group, the mean threshold C50 value was 53.8% ( $\pm$ 6.6) and thus it corresponded closely to the most ambiguous stimulus. The mean slope value was 4,3 ( $\pm$ 1,9).

#### **3.2 Cochlear implanted deaf subjects: transversal group.**

We have been able to analyze the performance of 32 CI subjects who were distributed in 4 sub groups according to duration of CI experience (post-activation delay). Globally, our test revealed a strong and durable deficit in voice gender categorization for CI subjects.

1  
2  
3  
4  
5  
6  
7 *Gender categorization.* Performance of CI subjects varied according to duration of CI  
8 experience, and most of the CI subjects presented a strong impairment when they were  
9 tested during the first month of CI use. Fig 1D illustrates representative responses in CI  
10 group, characterized by a high variability compared to those for the NHS group (dashed  
11 line). Some subjects managed to perform the task as accurately as NHS (black line)  
12 while others failed and presented responses systematically biased toward female voices  
13 (see #patient 36) for most stimuli of the continuum. Such behavior corresponded to  
14 recently implanted subjects (less than a month) and in 4 of these subjects (3 at the first  
15 day of activation and 1 after one month of activation of the implant) the slope and C50  
16 threshold could not be computed because of a lack of a good fit of the psychometric  
17 curve to a sigmoidal function ( $R^2 = 0.0179$ ; Chi squared: 170.4).

18  
19  
20  
21  
22  
23  
24  
25 For each post-activation delay, the performance of CI subjects in gender categorization  
26 remained significantly lower to that observed in NHS. This could be directly assessed by a  
27 simple inspection of the psychometric functions: The CI subjects evaluated at the first  
28 day of activation presented poor abilities to categorize gender (Fig 2.A) but their  
29 responses were not at chance level as their proportion of "Female" responses varied  
30 along the continuum. Gender categorization didn't improve with CI experience, since  
31 mean slope values ranged from 1,8 ( $\pm 0,9$ ) for CI subjects tested on first day of  
32 activation to 2,3 ( $\pm 9$ ) for CI subjects evaluated between 6 and 18 months post activation  
33 and 2,0 ( $\pm 0,9$ ) for subjects evaluated more than 18 months after implantation (see Fig.  
34 2B). Besides, correlation between mean slope and CI subjects' age did not reveal any  
35 significant influence of age in gender categorization performance ( $R^2=0.0131$ ).

36  
37  
38  
39  
40  
41  
42  
43  
44 The performance of CI subjects remained globally inferior to that that observed in the  
45 NHS control group (mean slope 4,3 $\pm$ 1,9). In all CI subject groups, whatever the post-  
46 activation delay (Fig 2.B), the slope of the psychometric curve remained significantly  
47 lower than that of the NHS group (t-test:  $t = [3.52; -3.006]$ ,  $p < 0.05$ ).

48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
The C50 threshold value characterizes position on the continuum for which subjects  
respond at chance level. Compared to NHS (mean threshold 53.8%  $\pm$ 6.6), CI subjects  
seemed to categorize female voice earlier on in the continuum while the morphed voice is



1  
2  
3  
4  
5  
6  
7 composed of a slightly higher proportion of male voice features. Such shift could either  
8 be due to the implant insertion depth or to the specific parameters if the individual's CI  
9 processor. However, it is not possible to conclude on this specific variability because this  
10 apparent difference with NHS was only observed in two sub-groups of patients while  
11 there was no significant difference in thresholds between CI subject groups (ANOVA:  $F(3,$   
12  $24)=0.918$ ,  $p>0.05$ ,  $\text{power}=0.216$ ). The difference was only significant in newly  
13 implanted subjects (First day of activation; t-test:  $t=-2.265$ ,  $p=0,034$ ) and for subjects  
14 with 6 to 18 months of experience of the implant (t-test:  $t=-2.827$ ,  $p=0,011$ ). Intra-  
15 group threshold analyses for the CI subjects did not show any difference according to  
16 post-activation delay. At the first day of activation the mean threshold was 43.1%  
17 ( $\pm 10.2$ ) and increased to 46.9% ( $\pm 16.3$ ) for subjects evaluated at or after 18 months CI  
18 experience.  
19

20  
21  
22  
23  
24  
25  
26  
27  
28 *Gender Recognition.* Gender recognition can be assessed by analyzing the subjects'  
29 responses for the two extreme voices of the continuum, 100% male voice and 100%  
30 female voice. Compared to NHS, CI subjects at activation had a significant deficit in  
31 recognizing both male (t-test:  $t=-2.818$ ,  $p=0,01$ ) and female voices (t-test:  $t=-3.202$ ,  
32  $p=0,004$ , see Fig 3). However, this deficit was lower than the deficit quantified from the  
33 slope analysis: the newly implanted CI subjects scored only 19% points lower (male and  
34 female voices combined) than NHS (i.e. for typical male/female voices). Conversely when  
35 considering the slope values, corresponding to the categorization of the most ambiguous  
36 voices, the CI subjects' slope values at activation was reduced by 58% compared to NHS.  
37 Gender recognition tended to improve with increasing the experience of the implant, but  
38 this trend did not reach significance level (ANOVA: Male voice:  $F(3,28)=2.170$ ,  $p>0.05$ ,  
39  $\text{power}=0.485$ ; Female voice:  $F(3,28)=1.923$ ,  $p>0.05$ ,  $\text{power}=0.434$ ). In addition, only  
40 the intermediate group of CI patients with post-activation delay of 6-18 months shows  
41 performance similar to the NHS group (t-test:  $t=2.786$ ,  $p=0.01$ ). As for gender  
42 categorization, there was no significant relationship between CI subjects' age and their  
43 recognition scores (for male voice,  $R^2=0.124$ ; for female voice,  $R^2=0.013$ ).  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Lastly, it is important to mention that there was no difference between performance for male and female voice recognition at any period post-implantation (t-test:  $t = [-0.820:0.430]$ ,  $p > 0.05$ ). Thus, characteristic gender voices are equally identified by CI subjects.

*Voice Discrimination.* A limited number of subjects ( $n=11$ ) were also tested for their ability to discriminate the human voice from environmental sounds (Massida et al., 2011). Our previous study showed a deficit in such a task in early stages after implantation with an improvement in performance after several months of experience. In the current limited set of patients we confirm the weak performance of CI subjects in such a task even after more than 18 months of experience with CI (mean  $d'$   $1.5 \pm 0.7$ ). Such values were significantly lower than those observed in normal hearing subjects (mean  $d'$ :  $4.3 \pm 0.5$  in Massida et al 2011). A statistical analysis did not reveal a correlation between performance in discriminating voices from non-voices ( $d'$ ) and gender categorization (threshold, slope and correct recognition of the gender, all cases  $p > 0.05$ ).

*Speech recognition.* As previously reported, CI users presented a significant improvement of word recognition scores following implantation (ANOVA;  $F(3,28)=52.626$ ,  $p < 0.0001$ , power=1). They reached 75% ( $\pm 8.7$ ) disyllabic word recognition after 6-18 months post-implantation (Fig 4). However CI subjects' performance remained below that of NHS, even after more than 18 months of experience with CI (t-test:  $t = [-16.5:10.9]$ ,  $p < 0.0001$ ).

We looked for a correlation between performance levels in speech recognition and gender categorization. In the full set of CI subjects ( $n = 32$ ) that encompassed novices and experienced users, analysis revealed no significant correlation (Fisher test,  $r^2 = [0.01:0.03]$ ,  $p > 0.05$ ) between word recognition scores (in %) and any performance criteria for gender categorization (slope, C50 threshold, male and female voice recognition). The relationship between speech recognition scores and gender

1  
2  
3  
4  
5  
6 categorization performance was further examined using an ANOVA analysis within  
7 subjects which revealed the independency between the two variables whatever the delay  
8 post implantation ( $F(3,28)=52,48$ ,  $p<0,0001$ ). This result adds further evidence that the  
9 recovery of speech recognition is independent of the recovery level in voice gender  
10 discrimination.  
11  
12  
13  
14  
15  
16  
17  
18  
19

### 20 **3.3 Cochlear implanted deaf subjects: follow-up group.**

21 We performed a longitudinal follow-up of 10 subjects, from the first day of activation of  
22 the cochlear implant to 18 months after activation. Unfortunately, the number of subjects  
23 was limited and performance was variable across the different post-implantation delays  
24 to allow reliable paired statistical analysis: The analysis of this follow-up group tended to  
25 confirm the results obtained in the transversal group. Indeed, performance in gender  
26 categorization did not seem to improve with CI experience whereas mean speech  
27 recognition score improved from 14% up to 70% correct (Fig 5). At any delay post-  
28 implantation, the mean slope value for this CI subjects group was always lower than for  
29 normal hearing controls and were similar to that observed in the transversal group. As  
30 for the transversal group, there was no significant relation between post activation delay  
31 and gender categorization slope (Fisher test,  $r^2 = 0.0008$ ,  $p > 0.05$ ). The C50 threshold  
32 showed a high variability with mean values ranging between 38 and 56%, but these  
33 values did not differ significantly from that for NHS. Gender recognition, evaluated for  
34 typical male/female voices, tended to improve from the first day of activation but, as  
35 obtained in the transversal group, they never reached NHS performance levels.  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

47 The transversal and longitudinal group analyses lead to very similar results, showing a  
48 deficit in gender categorization that did not improve significantly with CI experience.  
49 Because of differences in the post-activation delay, direct statistical comparisons were  
50 precluded. However, by simple observation at time of implantation the performance level  
51 (mean slope) was similar in the transversal and follow-up groups (1,8 vs 1,5.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 respectively). Similarly, after over 18 months of experience of implantation, the two  
8 groups present very comparable slope values (2,0 and 1,7). In spite of a regular  
9 repetition of the gender discrimination test in the longitudinal group, no improvement  
10 was observed, a result which rules out the implication of learning mechanisms.  
11  
12

#### 13 14 15 **4. Discussion** 16

17  
18 The voice gender recognition test we have used is part of the Voice Perception  
19 Assessment (VPA) test. It consists of a speaker discrimination task between sounds of a  
20 voice continuum obtained by morphing between a male and a female voice. This test  
21 allowed us to reveal a clear handicap of cochlear implanted recipients in voice gender  
22 categorization independent of duration of CI experience. The test was highly sensitive  
23 and therefore it can be used to evaluate pitch and timbre-related paralinguistic aspects of  
24 auditory recovery in CI subjects.  
25  
26  
27  
28  
29

30  
31  
32 ***Voice processing in CI recipients.*** Cochlear implantation re-activates the  
33 auditory system of profoundly deaf subjects by electrical stimulation through 12-22  
34 electrodes implanted into the cochlea. In spite of the fact that such a device provides  
35 relatively poor quality spectral information (Shannon, Zeng, Kamath, Wygonski, & Ekelid,  
36 1995), CI recipients benefit from an excellent recovery of speech recognition  
37 performance (UKCISG, 2004) albeit remaining below that of normally hearing subjects.  
38 Further paralinguistic aspects of speech perception remain deficient, especially  
39 concerning the processing of voice and voice features. Indeed, voice gender identification  
40 is known to be impaired in CI recipients (Fu et al., 2004; Fu et al., 2005; Kovacic &  
41 Balaban, 2009). However, in such tasks at a population level, the deficit of CI subjects  
42 appears to be satisfactory; usually 70-90% correct identification according to the study  
43 and the type of test. Such performance levels are not considerably lower than those for  
44 speech recognition which typically reach about 80% correct (i.e. for disyllabic words in  
45 experienced CI recipients) (Rouger et al., 2007). But, CI subjects present a large  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

1  
2  
3  
4  
5  
6  
7 variability in voice gender identification which is related to their hearing history (Kovacic  
8 & Balaban, 2010). Here using a continuum of morphed voices we have been able to show  
9 that the deficit in gender discrimination of the subjects is much more substantial than  
10 previously shown when compared to normally hearing subjects. The VPA allows us to  
11 assess the patient's performance using male and female voices with varying pitch and  
12 formant characteristics. In such a test, categorization scores within the continuum,  
13 expressed by the slope of the psychometric function, were more than 50% lower than  
14 that observed in normal hearing subjects even after a long period of CI experience. In  
15 addition, such a deficit remained nearly constant, independently from the duration of CI  
16 use, as performance for new and experienced users did not differ significantly.

17  
18  
19  
20  
21  
22  
23  
24 These results should be compared to our previous analysis of voice/non-voice  
25 discrimination in deaf CI subjects (Massida et al., 2011). In agreement with previous  
26 work (Proops et al., 1999; Tye-Murray, Tyler, Woodworth, & Gantz, 1992) we have  
27 shown that CI subjects are highly impaired in discriminating environmental sounds from  
28 voice stimuli. As in the present results for gender categorization, the performance level of  
29 experienced CI subjects in voice discrimination is weak, being more than 63% points  
30 lower than that observed in control subjects. Again, the ability of CI subjects to perform  
31 such voice/non-voice discrimination task does not improve with the delay post-  
32 implantation. Similarly the other features which can be extracted from a human voice  
33 such as the talker identity, the emotion, the familiarity, are also only weakly recognized  
34 by CI recipients (Cleary et al., 2005; Fu et al., 2004; Vongphoe & Zeng, 2005). In  
35 contrast, some CI patients have developed specific auditory strategies to recognize some  
36 aspects of paralinguistic information, such as speech prosody. Although voice pitch  
37 stands as the main acoustic cue to perceive this category of stimuli, some CI listeners  
38 may also rely on intensity properties to perceive speech prosody. For instance, intensity  
39 variation may contribute to their auditory analysis when CI subjects have to distinguish  
40 a question from a statement (Peng, Lu, & Chatterjee, 2009). Further, integration of  
41 auditory and visual information (eyebrows movements, head noddings) might help them  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 for speech and affective prosody (Foxton, Riviere, & Barone, 2010; Munhall, Jones,  
8 Callan, Kuratate, & Vatikiotis-Bateson, 2004).

9  
10 In conclusion, in spite of the possibility for CI subjects to develop adaptive strategies to  
11 increase discrimination of voice cues (see (Li & Fu, 2011)), including the use of other  
12 sensory modalities (Barone & Deguine, 2011), most of the auditory processes that are  
13 required for voice perception (Belin et al., 2004) are severely impaired in CI recipients  
14 very likely as a consequence of the relatively crude auditory information delivered by the  
15 implant.  
16  
17  
18  
19

20  
21  
22 **Voice gender discrimination in CI recipients.** Voice gender recognition relies on  
23 several acoustic cues among which are the fundamental frequency (F0) and the formant  
24 structure which both differ largely between male and female voices (Klatt & Klatt, 1990).  
25 Consequently, gender discrimination can be considered as a pitch-related task: A  
26 temporal pitch corresponding to F0, and a spectral pitch corresponding to the position of  
27 formants; these both work in the same sense, lower for male voice, higher for female  
28 voice. But the signal processing necessary for the CI removes fine structure information  
29 which would normally support temporal pitch extraction and the limited number of  
30 implanted electrodes would constitute a major limitation on the performance of CI  
31 subjects for spectral pitch perception. Indeed, using complex sounds that differ in F0, CI  
32 subjects present much higher thresholds in pitch discrimination than normal-hearing  
33 subjects (Cousineau, Demany, Meyer, & Pressnitzer, 2011; Laneau, Wouters, & Moonen,  
34 2006; Rogers, Healy, & Montgomery, 2006; Vandali et al., 2005; Zeng, 2002). Even for  
35 the best frequency range, CI subjects cannot achieve difference limens lower than 5-10%  
36 of F0 (Geurts & Wouters, 2001). Such results are compatible with the results obtained in  
37 the present study when considering gender discrimination performance for unambiguous  
38 male or female voices at the extremities of the continuum. CI subjects obtained on  
39 average 90% correct depending on the duration of CI experience, which is compatible  
40 with the F0 distance that separated the most distant male and female voices (126 vs 212  
41 Hz).  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

However, the performance of CI subjects decreased when the voice became more ambiguous after reducing the differences in both pitch and formant structure. When approaching the androgynous voice, responses of the subjects became more variable leading to a slope value of the psychometric function much lower than that for normal hearing control subjects. Kovacic et al (Kovacic & Balaban, 2009) used a different method, involving an adaptive procedure to evaluate voice gender perception as a function of voice F0 differences. In a sample of twenty individual talkers from each gender, some F0 values were contained in an overlapping frequency range (between 138 and 163 Hz) that closely matched the central region of our continuum. Recognition scores for these voices were significantly affected, dropping to chance level even in good CI performers. Thus, F0 is undoubtedly a prominent cue in voice gender identification but other individual acoustic characteristics also play a role.

Indeed, both spectral and temporal cues are involved in gender discrimination especially when the F0 distance is reduced as shown by the performance of NHS stimulated with a processor that simulate a cochlear implant processor (Fu et al., 2004; Fu et al., 2005; Gonzalez & Oliver, 2005). Correct gender discrimination among subjects ranged from 75 to 95%, which was similar to that of NHS listening to between 4 and 8 spectral channels. In NHS listening to vocoded stimuli, both speech detection and gender recognition improved as the number of spectral channels used was increased. Speech detection significantly improved when spectral channels increased from 4 to 8. Voice gender recognition improvement was more particularly observed for the switch from 8 to 16 spectral channels. Concerning the manipulation of the temporal information in each spectral condition, speech detection was not influenced by envelope filters: performance remained stable when envelope cut off frequency was increased from 40 Hz to 80Hz, up to 320 Hz. In contrast, gender recognition performance was significantly better for higher envelope cut-off frequencies, which suggests that providing more temporal cues to cochlear implanted subjects enhances their voice features perception. However, timbre perception, using formants structure, and temporal modulation processing abilities convey relevant cues regarding voice gender, especially when there is a small F0

1  
2  
3  
4  
5  
6  
7 difference as in our morphed voice in the middle part of the continuum. For large F0  
8 distances, reducing temporal cues did not alter voice gender discrimination even when  
9 stimuli were degraded through a 16 channel processor (Fu et al., 2005).

10  
11 While CI patients display a strong deficit in gender categorization (present study,  
12 (Kovacic & Balaban, 2009)) their performance is above chance level once the stimulus is  
13 progressively moving away from the androgynous voice on the continuum. This implies  
14 that they have developed adaptive strategies to compensate from the low spectral  
15 resolution provided by the implant. Our preliminary results in a pitch and timbre  
16 covarying condition indicate that pitch, though its perception is poor in CI users, remains  
17 the main feature for voice gender discrimination (Massida PhD dissertation 2010). Timbre  
18 manipulations did not affect gender recognition. More work is needed to understand how  
19 pitch cues are used by CI patients to perform in pitch related tasks, from paralinguistic to  
20 musical domains.  
21  
22  
23  
24  
25  
26  
27

28 To our knowledge, the impact of device experience on voice perception has never  
29 been specifically studied while it has been shown that several factors of the hearing  
30 history of the subjects are important such as the age onset of deafness as well as the  
31 age of implantation (Kovacic & Balaban, 2010). Such results were observed both in the  
32 transversal group and in the follow-up group where individual subjects were tested  
33 regularly for more than 18 months. Both gender categorization, and gender recognition  
34 tended to improve but these trends did not reach statistical significance. These results  
35 suggest that in spite of a strong adaptation to the relatively poor information delivered  
36 by the implant, CI recipients have not been able to develop specific strategies, including  
37 a more effective use of temporal cues or use of spectral cues, to categorize voices which  
38 are less typically male or female.  
39  
40  
41  
42  
43  
44  
45  
46  
47

48  
49 ***Gender discrimination and speech recognition in CI recipients.*** Our results showed  
50 a clear dissociation between the recovery of speech recognition and the capacity of CI  
51 subjects to discriminate voice gender. While functional speech recognition is acquired  
52 during the first year post-implantation, gender discrimination remains very poor even  
53  
54  
55



1  
2  
3  
4  
5  
6  
7 after several year of CI experience. Similar results were observed for voice/non voice  
8 recognition (Massida et al., 2011). However, while a dissociation between speech  
9 recognition and perception of paralinguistic features is usually shown in cochlear  
10 implanted subjects (Dorman, Gifford, Spahr, & McKarns, 2008; Vongphoe & Zeng, 2005)  
11 some authors have observed a correlation between phoneme perception and gender  
12 discrimination in CI recipients. Significant positive correlations were found for phonemes  
13 recognition in noise and voice gender recognition scores when inter-gender voice  $F_0 < 50$   
14 Hz (Li & Fu, 2011). Besides functional spectral resolution in CI users, this finding  
15 underlines the prominent role of temporal modulation detection in the 150-200 Hz range  
16 and would also be supported by individual differences in high-level central processes (Li  
17 & Fu, 2011).  
18  
19  
20  
21  
22  
23  
24  
25  
26

27 **Conclusion.** The Voice Perception Assessment (VPA) is an original protocol that  
28 allows in particular precise evaluation of the ability of CI recipients to discriminate voice  
29 gender across a frequency range that covers most of the continuum between a typical  
30 male and female voice. The performance of the CI recipient on the VPA revealed a strong  
31 impairment in the discrimination of more ambiguous voices, while the recognition of  
32 stereotype male and female voices approached that for normal hearing subjects. Such  
33 impairment is permanent and does not disappear after several months or year of  
34 experience of the implant. The deficit found for CI subjects reflects the degradation of  
35 both spectral and temporal cues by CI sound coders, in agreement with the weak ability  
36 of CI users to perform pitch discrimination. Further, our results are complementary to  
37 previous work that shows a durable impairment in CI users in processing features of the  
38 voice, from voice discrimination to perception of gender, familiarity and emotional  
39 content of the voice. Besides a deficit due to the limitation of the implant processor and  
40 as for speech perception, functional integrity of the cortical areas devoted to voice  
41 processing (Belin et al., 2004), is probably affected by the several years of auditory  
42 deprivation (Rouger et al., 2011).  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 **Acknowledgments.** We thank the cochlear-implanted and normally hearing subjects for  
8 their participation in this study, ML Laborde for help in collecting the data, C. Marlot for  
9 help in bibliography, J. Rouger for help in analysis of the data. This work was supported  
10 by a Cifre Convention to ZM (Cochlear France SAS-ANRT N°979/2006), the ANR  
11 Plasmody (ANR-06-Neuro-021-04 to BP and OD), The DRCI Toulouse (Direction de la  
12 Recherche Clinique et de l'Innovation to MM) and the recurrent funding of the CNRS (to  
13 OD and BP), and by BBSRC grant BB/E003958/1 to PBe.  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**References.**

- Andrews, M. L., & Schmidt, C. P. (1997). Gender presentation: perceptual and acoustical analyses of voice. *J Voice, 11*(3), 307-313.
- Arehart, K. H., Souza, P. E., Muralimanohar, R. K., & Miller, C. W. Effects of age on concurrent vowel perception in acoustic and simulated electroacoustic hearing. *J Speech Lang Hear Res, 54*(1), 190-210.
- Barone, P., & Deguine, O. (2011). *Multisensory processing in cochlear implant listener*. New York: Springer Science + Business Media.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci, 8*(3), 129-135.
- Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn Sci, 11*(12), 535-543.
- Cleary, M., & Pisoni, D. B. (2002). Talker discrimination by prelingually deaf children with cochlear implants: preliminary results. *Ann Otol Rhinol Laryngol Suppl, 189*, 113-118.
- 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.
- Cousineau, M., Demany, L., Meyer, B., & Pressnitzer, D. (2011). What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant. *Hear Res, 269*(1-2), 34-41.
- 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.
- El Fata, F., James, C. J., Laborde, M. L., & Fraysse, B. (2009). How much residual hearing is 'useful' for music perception with cochlear implants? *Audiol Neurootol, 14 Suppl 1*, 14-21.
- Elliott, T. M., & Theunissen, F. E. (2009). The modulation transfer function for speech intelligibility. *PLoS Comput Biol, 5*(3), e1000302.
- Foxton, J. M., Riviere, L. D., & Barone, P. (2010). Cross-modal facilitation in speech prosody. *Cognition, 115*(1), 71-78.
- Fu, Q. J., Shannon, R. V., & Wang, X. (1998). Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. *J Acoust Soc Am, 104*(6), 3586-3596.
- 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.
- Geurts, L., & Wouters, J. (2001). Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. *J Acoust Soc Am, 109*(2), 713-726.
- 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.
- Gonzalez, J., & Oliver, J. C. (2005). Gender and speaker identification as a function of the number of channels in spectrally reduced speech. *J Acoust Soc Am, 118*(1), 461-470.
- Gratton, M. A., & Vazquez, A. E. (2003). Age-related hearing loss: current research. *Curr Opin Otolaryngol Head Neck Surg, 11*(5), 367-371.
- Kawahara H, M.-K. I., de Cheveigné A. (1999). Restructuring speech representations using pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun, 27*, 187-207.

- 1  
2  
3  
4  
5  
6  
7 Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality  
8 variations among female and male talkers. *J Acoust Soc Am*, *87*(2), 820-857.
- 9 Kong, Y. Y., Cruz, R., Jones, J. A., & Zeng, F. G. (2004). Music perception with temporal  
10 cues in acoustic and electric hearing. *Ear Hear*, *25*(2), 173-185.
- 11 Kovacic, D., & Balaban, E. (2009). Voice gender perception by cochlear implantees. *J Acoust*  
12 *Soc Am*, *126*(2), 762-775.
- 13 Kovacic, D., & Balaban, E. (2010). Hearing history influences voice gender perceptual  
14 performance in cochlear implant users. *Ear Hear*, *31*(6), 806-814.
- 15 Laneau, J., Wouters, J., & Moonen, M. (2006). Improved music perception with explicit pitch  
16 coding in cochlear implants. *Audiol Neurootol*, *11*(1), 38-52.
- 17 Lavner Y, G. I., Rosenhouse J. (2000). The effects of acoustic modifications on the  
18 identification of familiar voices speaking isolated vowels. *Speech Commun*, *30*, 9-26.
- 19 Leal, M. C., Shin, Y. J., Laborde, M. L., Calmels, M. N., Verges, S., Lugaardon, S., et al.  
20 (2003). Music perception in adult cochlear implant recipients. *Acta Otolaryngol*, *123*(7), 826-  
21 835.
- 22 Li, T., & Fu, Q. J. (2011). Voice gender discrimination provides a measure of more than  
23 pitch-related perception in cochlear implant users. *Int J Audiol*, *50*(8), 498-502.
- 24 Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception  
25 problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl*  
26 *Acad Sci U S A*, *103*(49), 18866-18869.
- 27 Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., et al. (2011). Voice  
28 discrimination in cochlear-implanted deaf subjects. *Hear Res*, *275*(1-2), 120-129.
- 29 McKay, C. M. (2005). Spectral processing in cochlear implants. *Int Rev Neurobiol*, *70*, 473-  
30 509.
- 31 Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004).  
32 Visual prosody and speech intelligibility: head movement improves auditory speech  
33 perception. *Psychol Sci*, *15*(2), 133-137.
- 34 Munson, B., & Nelson, P. B. (2005). Phonetic identification in quiet and in noise by listeners  
35 with cochlear implants. *J Acoust Soc Am*, *118*(4), 2607-2617.
- 36 Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on  
37 speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiol*  
38 *Neurootol*, *14*(5), 327-337.
- 39 Pernet, C. R., & Belin, P. (2012). The role of pitch and timbre in voice gender categorization.  
40 *Front Psychol*, *3*, 23.
- 41 Proops, D. W., Donaldson, I., Cooper, H. R., Thomas, J., Burrell, S. P., Stoddart, R. L., et al.  
42 (1999). Outcomes from adult implantation, the first 100 patients. *J Laryngol Otol Suppl*, *24*,  
43 5-13.
- 44 Rogers, C. F., Healy, E. W., & Montgomery, A. A. (2006). Sensitivity to isolated and  
45 concurrent intensity and fundamental frequency increments by cochlear implant users under  
46 natural listening conditions. *J Acoust Soc Am*, *119*(4), 2276-2287.
- 47 Rouger, J., Lagleyre, S., Demonet, J. F., Fraysse, B., Deguine, O., & Barone, P. (2011).  
48 Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients.  
49 *Hum Brain Mapp*.
- 50 Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., & Barone, P. (2007). Evidence  
51 that cochlear-implanted deaf patients are better multisensory integrators. *Proc Natl Acad Sci*  
52 *U S A*, *104*(17), 7295-7300.
- 53 Schvartz, K. C., & Chatterjee, M. (2012). Gender identification in younger and older adults:  
54 use of spectral and temporal cues in noise-vocoded speech. *Ear Hear*, *33*(3), 411-420.
- 55 Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech  
56 recognition with primarily temporal cues. *Science*, *270*(5234), 303-304.

- 1  
2  
3  
4  
5  
6  
7 Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2011). Effects of age on F0  
8 discrimination and intonation perception in simulated electric and electroacoustic hearing.  
9 *Ear Hear*, 32(1), 75-83.
- 10 Swets, J. A., Harris, J. R., McElroy, L. S., & Rudloe, H. (1966). Computer-aided instruction  
11 in perceptual identification. *Behav Sci*, 11(2), 98-104.
- 12 Tanner, W. P., Jr., & Swets, J. A. (1954). A decision-making theory of visual detection.  
13 *Psychol Rev*, 61(6), 401-409.
- 14 Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *J*  
15 *Acoust Soc Am*, 85(4), 1699-1707.
- 16 Tye-Murray, N., Tyler, R. S., Woodworth, G. G., & Gantz, B. J. (1992). Performance over  
17 time with a nucleus or Ineraid cochlear implant. *Ear Hear*, 13(3), 200-209.
- 18 UKCISG. (2004). Criteria of candidacy for unilateral cochlear implantation in postlingually  
19 deafened adults III: prospective evaluation of an actuarial approach to defining a criterion.  
20 *Ear Hear*, 25(4), 361-374.
- 21 Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., & McDermott, H. J.  
22 (2005). Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing  
23 strategies. *J Acoust Soc Am*, 117(5), 3126-3138.
- 24 Vongphoe, M., & Zeng, F. G. (2005). Speaker recognition with temporal cues in acoustic and  
25 electric hearing. *J Acoust Soc Am*, 118(2), 1055-1061.
- 26 Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant  
27 future. *Hear Res*, 242(1-2), 3-21.
- 28 Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hear Res*, 174(1-2), 101-106.
- 29 Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., et al. (2005).  
30 Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci U S A*,  
31 102(7), 2293-2298.
- 32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 **Figure. legends**  
8

9 **Figure 1. A.** Spectrograms corresponding to the two extremities of the vocal continuum  
10 (100% female and 100% male). **B.** Spectral representation of the eleven elements of the  
11 vocal continuum. **C.** Psychometric curve of gender categorization in normal hearing  
12 subjects. **D.** Representative example of the performance of CI subjects with various  
13 delays post-implantation.  
14  
15  
16

17  
18  
19 **Figure 2. A.** Psychometric curves of CI subjects. Each curve represents the mean curve  
20 of a group of CI subjects at different delays post-implantation (see Table 1). Normal-  
21 hearing values are indicated by the dashed line curve. **B.** Slope values computed at the  
22 C50. Slope values were significantly higher in NHS than in any CI patients group (\*), no  
23 significant difference was found across CI patients groups (see text).  
24  
25  
26  
27

28  
29 **Figure 3.** Performance of CI subjects in voice gender recognition corresponding to the  
30 hit scores typical, unambiguous, male or female voices.  
31  
32

33  
34 **Figure 4.** Recovery of auditory speech recognition scores in CI subjects.  
35  
36

37  
38 **Figure 5.** Longitudinal follow-up of a second group of 10 CI subjects: performance in  
39 voice gender categorization and in speech recognition.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Subject number	Group	Age (year)	Hearing loss (year)	Model of implant	Processor	Strategy	Hearing aid
1	1st day	63	>20	<i>Cochlear</i>	Freedom	ACE	
2		46	>30	<i>Cochlear</i>	Freedom	ACE	
3		48	>40	<i>Advanced Bionic</i>	Auria	HiRes-S	
4		43	>30	<i>Cochlear</i>	Freedom	ACE	
5		21	>10	<i>Cochlear</i>	Freedom	ACE	
6		67	>10	<i>Cochlear</i>	Freedom	ACE	Y
7		48	>30	<i>Cochlear</i>	Freedom	ACE	
8		47	>40	<i>MedEl</i>	OPUS 2	FSP	Y
9		53	>40	<i>Cochlear</i>	Freedom	ACE	Y
10	1 - 6 months	67	>30	<i>Cochlear</i>	Freedom	ACE	
11		45	3	<i>Cochlear</i>	Freedom	ACE	Y
12		71	>40	<i>Advanced Bionic</i>	Harmony	HiRes-S	Y
13	6 to 18 months	35	>30	<i>Cochlear</i>	Freedom	ACE	Y
14		21	>10	<i>Cochlear</i>	Freedom	ACE	Y
15		51	>30	<i>Advanced Bionic</i>	Auria	HiRes-S	Y
16		40	>30	<i>Cochlear</i>	Freedom	ACE	Y
17		63	>20	<i>Cochlear</i>	Freedom	ACE	Y
18		53	< 5	<i>Advanced Bionic</i>	Auria	HiRes-S	Y
19		81	NA	<i>Advanced Bionic</i>	Auria	HiRes-S	Y
20		74	2	<i>Cochlear</i>	Freedom	ACE	Y
21		48	>30	<i>Cochlear</i>	Freedom	ACE	Y
22		71	NA	<i>Cochlear</i>	Freedom	ACE	
23	More than 18 months	48	>30	<i>Cochlear</i>	Freedom	ACE	Y
24		41	>30	<i>Cochlear</i>	ESPrit 3G	ACE	
25		38	>30	<i>Cochlear</i>	ESPrit 3G	ACE	
26		61	>40	<i>Advanced Bionic</i>	Auria	HiRes-S	
27		75	>10	<i>Cochlear</i>	ESPrit 3G	ACE	
28		59	>20	<i>Cochlear</i>	ESPrit 3G	ACE	
29		59	NA	<i>Cochlear</i>	Sprint	ACE	
30		58	>40	<i>Cochlear</i>	ESPrit 3G	SPEAK	
31		68	>10	<i>Cochlear</i>	Sprint	ACE	
32		75	>20	<i>Cochlear</i>	ESPrit 22	SPEAK	
Subject number	Group	Age (year)	Hearing loss (year)	Model of implant	Processor	Strategy	Hearing aid
33	Longitudinal	63	>20	<i>Cochlear</i>	Freedom	ACE	
34		46	>30	<i>Cochlear</i>	Freedom	ACE	
35		48	>40	<i>Advance Bionic</i>	Auria	HiRes-S	
36		21	>10	<i>Cochlear</i>	Freedom	ACE	
37		48	>30	<i>Cochlear</i>	Freedom	ACE	
38		47	>40	<i>MedEl</i>	OPUS 2	FSP	Y
39		53	>40	<i>Cochlear</i>	Freedom	ACE	Y
40		67	>30	<i>Cochlear</i>	Freedom	ACE	
41		45	3	<i>Cochlear</i>	Freedom	ACE	Y
42		81	NA	<i>Advance Bionic</i>	Auria	HiRes-S	Y

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

Formatted: English (U.S.)

**Table 1. Patients' summary.** In the individual columns, the age of CI subjects is in years. The estimation of the duration of hearing loss is given in years: not available (NA); The model of cochlear implant and the processor and strategies are provided. The last column indicates which patients used a hearing aid in daily life

	Mean F0 (Hz)	Mean F1 (Hz)	Mean F2 (Hz)	Mean F3 (Hz)	Mean F4 (Hz)
Position 1	125.86	752.77	1436.59	2598.76	3859.52
Position 2	133.08	777.39	1480.86	2663.10	3977.32
Position 3	140.49	791.59	1506.54	2688.29	4042.53
Position 4	148.20	810.47	1547.96	2738.83	4114.92
Position 5	155.83	830.00	1585.38	2780.41	4153.51
Position 6	163.99	856.36	1632.17	2815.70	4198.26
Position 7	173.10	877.23	1668.88	2846.01	4200.36
Position 8	182.24	904.72	1711.57	2881.42	4241.36
Position 9	191.73	922.13	1740.51	2906.60	4281.48
Position 10	202.10	941.89	1777.30	2946.08	4319.14
Position 11	212.71	967.21	1803.72	2963.55	4360.48

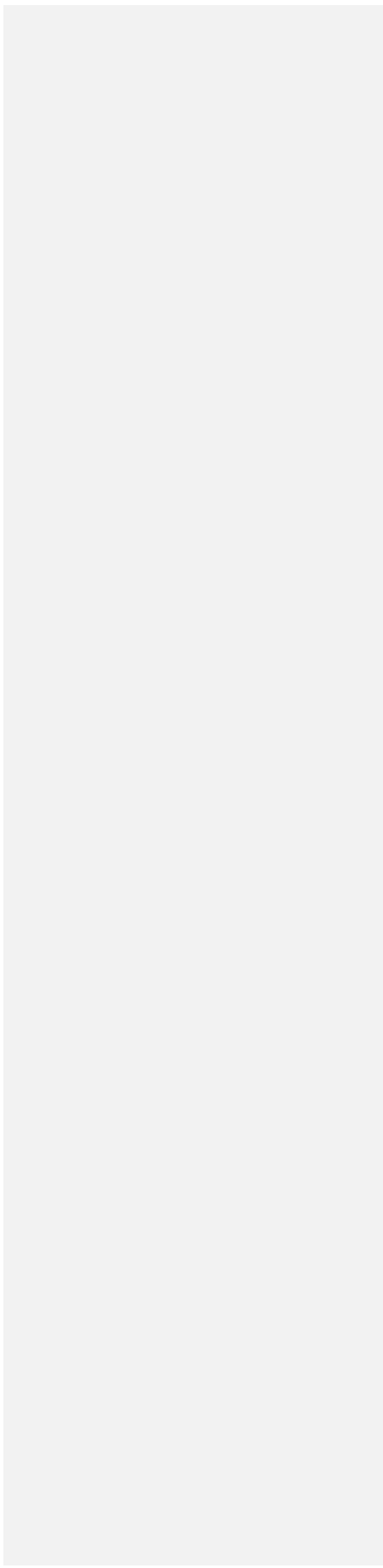
**Table 2.** Values of F0, F1, F2, F3 and F4 for each position of the voice stimuli across the male (position 1) to the female voice (position 11) in the continuum.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

|

For Peer Review



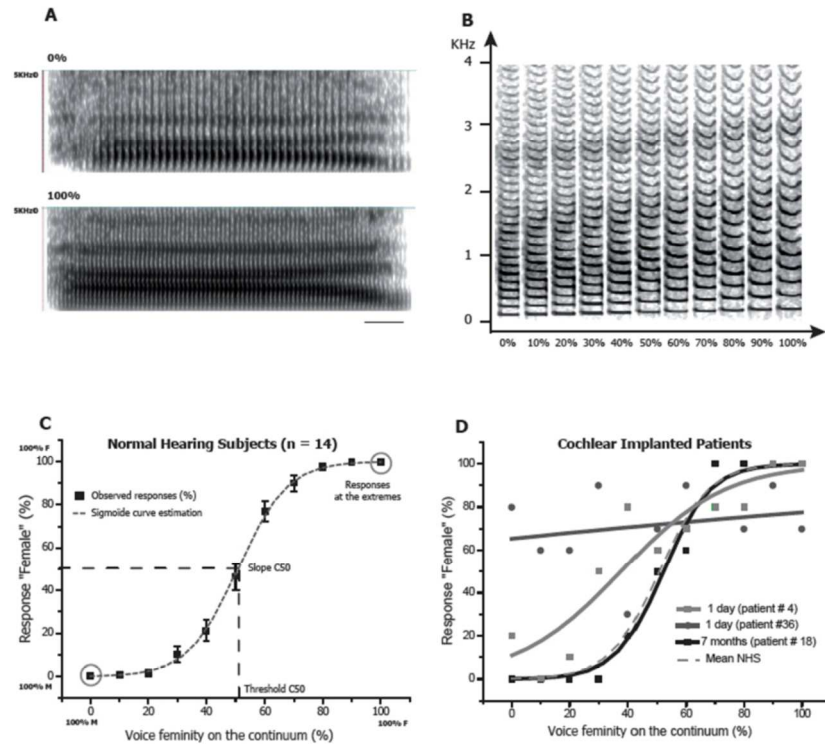


Figure 1. A. Spectrograms corresponding to the two extremities of the vocal continuum (100% female and 100% male). B. Spectral representation of the eleven elements of the vocal continuum. C. Psychometric curve of gender categorization in normal hearing subjects. D. Representative example of the performance of CI subjects with various delays post-implantation.

275x232mm (87 x 87 DPI)

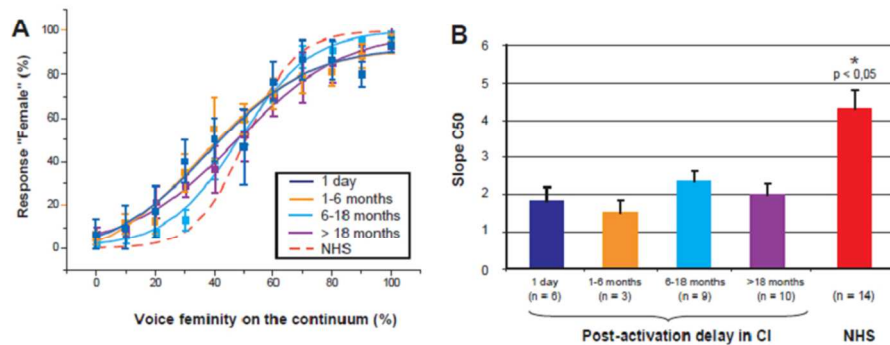


Figure 2. A. Psychometric curves of CI subjects. Each curve represents the mean curve of a group of CI subjects at different delays post-implantation (see Table 1). Normal-hearing values are indicated by the dashed line curve. B. Slope values computed at the C50. Slope values were significantly higher in NHS than in any CI patients group (\*), no significant difference was found across CI patients groups (see text). 254x190mm (96 x 96 DPI)

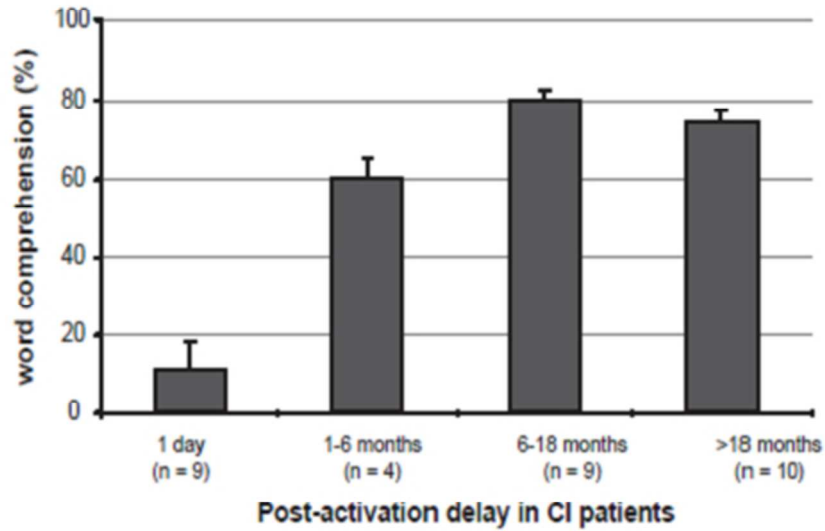


Figure 4. Recovery of auditory speech recognition scores in CI subjects.  
138x100mm (87 x 87 DPI)

Review

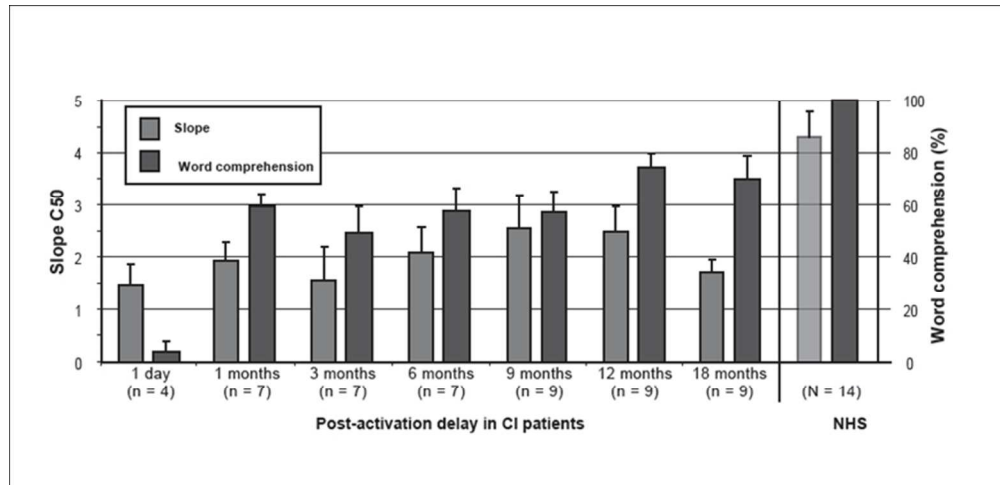


Figure 5. Longitudinal follow-up of a second group of 10 CI subjects: performance in voice gender categorization and in speech recognition.  
248x117mm (89 x 90 DPI)

Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Subject number	Group	Age (year)	Hearing loss (year)	Model of implant	Processor	Strategy	Hearing aid
1	1st day	63	>20	<b>Cochlear</b>	Freedom	ACE	
2		46	>30	<b>Cochlear</b>	Freedom	ACE	
3		48	>40	<b>Advanced Bionic</b>	Auria	HiRes-S	
4		43	>30	<b>Cochlear</b>	Freedom	ACE	
5		21	>10	<b>Cochlear</b>	Freedom	ACE	
6		67	>10	<b>Cochlear</b>	Freedom	ACE	Y
7		48	>30	<b>Cochlear</b>	Freedom	ACE	
8		47	>40	<b>MedEI</b>	OPUS 2	FSP	Y
9		53	>40	<b>Cochlear</b>	Freedom	ACE	Y
10	1 - 6 months	67	>30	<b>Cochlear</b>	Freedom	ACE	
11		45	3	<b>Cochlear</b>	Freedom	ACE	Y
12		71	>40	<b>Advanced Bionic</b>	Harmony	HiRes-S	Y
13		35	>30	<b>Cochlear</b>	Freedom	ACE	Y
14	6 to 18 months	21	>10	<b>Cochlear</b>	Freedom	ACE	Y
15		51	>30	<b>Advanced Bionic</b>	Auria	HiRes-S	Y
16		40	>30	<b>Cochlear</b>	Freedom	ACE	Y
17		63	>20	<b>Cochlear</b>	Freedom	ACE	Y
18		53	< 5	<b>Advanced Bionic</b>	Auria	HiRes-S	Y
19		81	NA	<b>Advanced Bionic</b>	Auria	HiRes-S	Y
20		74	2	<b>Cochlear</b>	Freedom	ACE	Y
21		48	>30	<b>Cochlear</b>	Freedom	ACE	Y
22		71	NA	<b>Cochlear</b>	Freedom	ACE	
23	More than 18 months	48	>30	<b>Cochlear</b>	Freedom	ACE	Y
24		41	>30	<b>Cochlear</b>	ESPrIt 3G	ACE	
25		38	>30	<b>Cochlear</b>	ESPrIt 3G	ACE	
26		61	>40	<b>Advanced Bionic</b>	Auria	HiRes-S	
27		75	>10	<b>Cochlear</b>	ESPrIt 3G	ACE	
28		59	>20	<b>Cochlear</b>	ESPrIt 3G	ACE	
29		59	NA	<b>Cochlear</b>	Sprint	ACE	
30		58	>40	<b>Cochlear</b>	ESPrIt 3G	SPEAK	
31		68	>10	<b>Cochlear</b>	Sprint	ACE	
32		75	>20	<b>Cochlear</b>	ESPrIt 22	SPEAK	
Subject number	Group	Age (year)	Hearing loss (year)	Model of implant	Processor	Strategy	Hearing aid
33	Longitudinal	63	>20	<b>Cochlear</b>	Freedom	ACE	
34		46	>30	<b>Cochlear</b>	Freedom	ACE	
35		48	>40	<b>Advanced Bionic</b>	Auria	HiRes-S	
36		21	>10	<b>Cochlear</b>	Freedom	ACE	
37		48	>30	<b>Cochlear</b>	Freedom	ACE	
38		47	>40	<b>MedEI</b>	OPUS 2	FSP	Y
39		53	>40	<b>Cochlear</b>	Freedom	ACE	Y
40		67	>30	<b>Cochlear</b>	Freedom	ACE	
41		45	3	<b>Cochlear</b>	Freedom	ACE	Y
42		81	NA	<b>Advance Bionic</b>	Auria	HiRes-S	Y

**Table 1. Patients' summary.** In the individual columns, the age of the patient is provided in years. The estimation of the duration of hearing loss is given in years: non available (NA); The model of cochlear implant and the processor and strategies are provided. The last column indicates which patients use a hearing aid in daily life

	Mean F0 (Hz)	Mean F1 (Hz)	Mean F2 (Hz)	Mean F3 (Hz)	Mean F4 (Hz)
Position 1	125.86	752.77	1436.59	2598.76	3859.52
Position 2	133.08	777.39	1480.86	2663.10	3977.32
Position 3	140.49	791.59	1506.54	2688.29	4042.53
Position 4	148.20	810.47	1547.96	2738.83	4114.92
Position 5	155.83	830.00	1585.38	2780.41	4153.51
Position 6	163.99	856.36	1632.17	2815.70	4198.26
Position 7	173.10	877.23	1668.88	2846.01	4200.36
Position 8	182.24	904.72	1711.57	2881.42	4241.36
Position 9	191.73	922.13	1740.51	2906.60	4281.48
Position 10	202.10	941.89	1777.30	2946.08	4319.14
Position 11	212.71	967.21	1803.72	2963.55	4360.48

**Table 2.** Values of F0, F1, F2, F3 and F4 for each position of the voice stimuli across the male (position 1) to the female voice (position 11) in the continuum.

## **Introduction**

Speech provides meaning through its semantic content but also via syntactic arrangement in the sentence and the way words are uttered. Indeed, the meaning of a sentence can be altered using intonation variation, syllable or word emphasis and pauses in discourse [Wingfield, 1975, Speer et al., 1996]. The acoustic correlates of these phenomena are variations in voice fundamental frequency (F0), amplitude, and duration [Nagel et al., 1994]. These specific characteristics of language are referred to as supra-segmental information and gathered under the term of prosody.

Affective prosody yields information regarding the emotional status of the speaker, which can be inferred based upon specific acoustic parameter changes. Primary emotions such as anger, fear or happiness have been the focus of many studies of prosody. Anger in speech is for example associated with a higher voice fundamental frequency (F0) and a higher amplitude level on emphasized syllables [Fonagy, 1966, Williams and Stevens, 1972]. Fonagy [Fonagy, 1958] reported that happiness also leads to an increase in F0 and F0 range with an increased speaking rate.

Prosody may also convey linguistic content, relying on the relative arrangement of lexical units and intonation variations in the oral discourse. The way the words are grouped together can facilitate the extraction of the meaning of the message. For instance, it helps distinguish a command from a simple observation. Linguistic prosody, or speech prosody, also exerts its influence in the recognition of the sentential mode, i.e the discrimination between a question and a statement or an exclamation. In French, as in English an interrogative sentence typically ends with an ascending intonation that reflects a rising F0, whereas intonation remains stable or decreases in a statement sentence [Patel et al, 2008].

### *Prosody perception in cochlear implanted deaf patients*

For more than 25 years, cochlear implants (CI) have been shown to restore excellent speech recognition [Rouger et al., 2007, Wilson and Dorman, 2008], however other aspects of auditory perception remain problematic for many patients. Music perception for CI recipients is generally limited to rhythmic cues [Nimmons et al, 2008] and to following speech in songs [Gfeller & Lancing, 1991, El Fata et al, 2009] due to limitations in sound processing and the means of stimulation. Instrumental passages are hardly perceived and recognized because the representation of melody and timbre rely on too many acoustical cues that do not appear to be efficiently transmitted through the implant [Gfeller et al., 2002b]. Similarly for prosody perception, many studies highlight poor recognition of speech intonation [Green et al.,



2002, Peng et al., 2008] and lexical tone [Peng et al., 2004] for both pre and post-lingually deafened CI recipients. Nakata et al. [Nakata et al., 2012] assessed perception of affective prosody by CI children and their normal hearing counterparts and showed a strong deficit in emotion recognition with CI. Scores were 35 to 60 percentage points lower in the CI group for “happy” or “sad” and for “angry” respectively.

The poor perception for music and affective speech prosody perception seen in CI recipients is often attributed to a deficit in the representation of F0 due to sound coding. Indeed, resolution for both place and temporal information for F0 is lacking in current CIs. Place F0 encoding is limited by the number of channels (from 12 to 22 electrodes current CIs) and possibly the relatively shallow insertion depth which does not allow stimulation of the locations tuned for low frequencies in the cochlea. F0 temporal encoding in current strategies is available for slowly-varying envelope (<50 Hz) and periodicity cues up to 300 Hz [Shannon, 1983] but does not appear to be always effective in this range [Milczynski et al., 2009], and certainly appears absent for higher modulation rates [Geurts and Wouters, 2001, Vandali et al., 2005]. Sound coding strategies using fixed rate carriers essentially discard temporal fine structure in the input signal except for low F0s which are represented by modulation. Speech recognition in quiet can be supported by the low-frequency information conveyed in the envelope but speech recognition in noise [Stickney et al., 2005] and other challenging situations such as prosody [Chatterjee and Peng, 2008] require the correct perception of much higher amplitude modulation rates.

F0 representation might not be impaired where acoustic residual hearing coexists with CI. The perception of sound via residual acoustic hearing is generally fused with that of electric stimulation to form a unique and integrated auditory percept [Tyler et al., 2002]. The residual acoustic hearing can be either contralateral to the CI as in “bimodal” stimulation, or both contralateral and ipsilateral to the CI as is the case for “hybrid” or electric acoustic stimulation (EAS). Bimodal and hybrid stimulation have shown advantages in auditory tasks that rely on F0 perception. Thus, CI patients with substantial residual hearing get better scores for speech recognition in noise [Armstrong et al., 1997, Adunka et al., 2010] and melody identification [Kong et al., 2004, El fata et al., 2009] than traditional CI subjects. Straatman et al. [Straatman et al., 2010] found that bimodal CI children discriminated more accurately questions from statement if they wore their hearing aid.

Using a statement/question discrimination paradigm, Peng et al. [Peng et al., 2009] pointed to the necessity of cooperating cues in the signal to improve CI patients’ performance. More specifically, CI subjects relied on the congruency of intensity and F0 cues to perform the task

and these findings suggested the role of intensity cues in intonation recognition, which should not be neglected.

We hypothesized that speech prosody perception would be superior in CI patients with bimodal stimulation or EAS because these patients have access to acoustic F0 cues through their residual hearing. Our corollary assumption was that better the residual hearing, the better prosody and F0 perception should be, as has been shown for song recognition [El Fata et al., 2009]. Our second main hypothesis was that CI patients without residual hearing would rely on other cues than F0 to discriminate questions from statements. Therefore the objectives of our study were to evaluate question/statement discrimination abilities in three different groups of cochlear implanted patients according to the presence/absence and level of residual hearing and to identify which acoustic cues were more important for any one group. We also tested pitch perception for simple harmonic tones in order to measure their overall capacity to perceive F0 without a speech context.

## **Subjects and methods**

### *Participants*

Twenty-one postlingually deafened adult cochlear implant (CI) users participated in the study. Four of them had post-operative bilateral residual hearing and were included in the group called “electric-acoustic stimulation” (EAS). Seven subjects had significant levels of residual hearing in the ear contralateral to the CI and were included in the “bimodal” group (Bim). The other ten subjects had no residual acoustic hearing (“CI only” group). Subjects’ characteristics are given in Table 1. For EAS and Bim groups, better ear pure-tone average (125-250-500 Hz PTA) thresholds measured before testing are represented in table 2. The three groups were comparable with respect to aetiology of deafness. All CI users were assessed in their daily listening mode, i.e using unilateral or bilateral hearing aids, in order to measure optimal performance and to avoid learning and/or experience effects. We intended to focus our study on speech prosody perception and thus included subjects with good speech perception scores in quiet (dissyllabic word recognition >70% correct). CI experience varied across subjects which allowed us assess its impact on performances.

Fifteen normal hearing subjects (NHS) served as controls – for example for the modified stimulus conditions. These fifteen subjects had no history of otologic pathology or neurological disorder. All subjects were required to give informed consent to participate to the study and an institutional ethics committee approval was obtained (n°AFSSAPS:B90183-20).

### *Speech prosody testing*

We evaluated speech prosody perception using a statement-question discrimination task. A sentence, containing no syntactic indices, was presented to the subject, who had to determine if this sentence was a question or a statement. Sentences were chosen from Patel's study [Patel et al., 1998] where sentences pairs were acoustically similar. In order to evaluate the relative contribution of different acoustic cues in this perception task, three different sub tests were performed.

In the first condition ("original" condition), new original and natural recordings were presented. A total of 30 sentences were recorded from two professional actors (one man, one woman) who uttered them once in an interrogative mode and once in a statement mode.

In a second condition (flat F0), we removed F0 variations from the original stimuli to evaluate the role of F0 variations in statement question discrimination. Using Praat® [Boersma and D, 2009], the mean F0 in the sentence was calculated from the original recording, and this mean F0 was then maintained constant through the whole sentence. To perform well under this condition, subjects would have to rely on time and amplitude variations as cues.

In a third condition, we employed the stimuli from Patel's study in a "synthesized" condition leaving F0 variations in the last syllable (e.g. last syllable or last monosyllabic word) as the only salient cue for discrimination with equal amplitude and duration.

For each of the three sub tests, the 24 sentences were presented in random order, first visually in large type on a laptop screen for 2 seconds, then acoustically, so that subjects focus on the discrimination task and not on speech understanding. Practise items were presented before each subtest.

### *F0 detection threshold*

We aimed to determine the smallest perceptible difference in F0 between two stimuli, which is often referred as the F0 difference limen (F0DL). We used the non parametric adaptive procedure described by Grassi and Soranzo [Grassi and Soranzo, 2009], available online, that estimates a F0DL for various standard F0s, using a method based on a maximum likelihood procedure (mlp). Stimuli were five-harmonic tone complexes including F0. The test was run using Matlab® code (the Mathworks, Inc., Natick, MA, USA). We measured several F0DLs with standard F0s being: 110 Hz, 220 Hz, 400 Hz, 500 Hz, and 750 Hz. The test used a 3

alternative forced choice procedure with 30 presentations of 3 stimuli. In each trial two of the sounds had the same F0 and were called F0 standards. The other one had a higher F0 and was called F0 target. Between those 3 stimuli, subjects had to determine the stimulus with the highest F0. If the subject correctly identified the sound with the highest F0, the distance between the sounds' F0 was adaptively reduced until the smallest discriminable distance in Hz between the frequencies was determined.

### *Speech recognition*

All subjects were tested on open set recognition of French dissyllabic words using the commonly-used Fournier word lists, presented through loudspeakers at 60 dB SPL.

MBAA sentence list scores were also collected in quiet and in noise. These lists are used as a standard CI evaluation for speech recognition in noise (Marx et al., 2012)

### *Test set-up*

All subjects were assessed sitting comfortably in a sound attenuated audiometric booth. All auditory stimuli were presented through Sennheiser® circumaural ear phones at an intensity level of 60 dB SPL in order to match with conventional speech recognition measurements. Speech prosody tests were performed using Presentation® software and F0 threshold was assessed using mlp function in Matlab (the Mathworks, Inc., Natick, MA, USA). The speech prosody sub tests and F0 threshold determination order was balanced across subjects in order to avoid training effects. No repetition of any test stimulus was permitted. The examiner was also inside the booth, recording the subjects' response and running the test. The whole procedure took one to one and a half hours including training phases.

### *Data analysis*

In each question-statement discrimination subtest, we used as a primary measure the hit rates for questions and statements converted to  $d'$  values (Creelman, 1962).  $d'$  is a criterion of perception sensitivity independent of decision bias, relying on hit rate and false alarm rate according to Signal Detection Theory [Tanner and Swets, 1954, Swets et al., 1966]. The F0 detection thresholds (FODL) were expressed as Weber fractions: smallest detected (F0 in Hz-F0 standard in Hz)/F0 standard, expressed as a percentage.

Analyses of variance (ANOVA) were performed on speech prosody scores and FODLs for each subtest/condition with group as factor. Fischer's tests were used to perform multiple comparisons. T tests were used to perform paired comparisons between the three subtests of

the speech prosody perception procedure for each group. We determined Pearson correlation coefficients for each correlation analysis. All statistical analyses were performed using Statview® software.

## **Results**

### *Groups' description*

Twenty one post linguually deafened adult cochlear implanted subjects were included in 3 different groups according to the absence (CI only, n=10) or presence of residual hearing (EAS group with bilateral residual hearing, n=4; bimodal group with unilateral residual hearing, contralateral to the CI, n=7).

Mean age was 52,7 years (43-70) in the EAS group, 52,4 years in the bimodal group (21-68) and 66,9 years (44-76) in the CI only group. Mean age was significantly higher in the CI only group compared to bimodal group (Fischer's test,  $p=0,02$ )

Mean delay post implantation was 46,5 months (+/-27,23) for the EAS group, 26 months (+/-20,76) for the bimodal group and 45,4 months (+/-26,15) for the CI only group. Comparative analysis revealed a shorter experience with CI for the bimodal group than for the CI only group (Fischer's test,  $p=0,038$ ). There was a trend for shorter experience for the EAS group versus the bimodal group (Fischer's test,  $p=0,082$ ). However, there were no significant differences between groups for speech recognition in quiet scores. Mean correct disyllabic words recognition was 83,7 % (+/- 6,29) for the EAS group, 87,1% for the bimodal group (+/-11,85) and 79,5% (+/-10,74) for the CI only group. Mean correct sentence recognition in quiet was 98,2% (+/- 2,06) for the EAS group, 90% (+/-22,19) for the bimodal group and 95,1% (+/-5,46) for the CI only group. Our objective to study comparable groups regarding speech perception abilities was then achieved.

Fifteen NHS were also included to determine a normal performance pattern. Mean age was 27,9 years old (median:24; range:18-57).

### *Speech prosody perception*

Results of speech prosody perception different subtests are illustrated in figure 1.

#### *Original condition*

This condition was tested in order to evaluate speech prosody perception for "naturally" produced sentences, when all acoustic cues are present before signal processing. Globally, subjects with residual hearing (EAS and bimodal groups) performed better than subjects without any residual hearing (CI only group). EAS subjects and bimodal subjects scored

94,8% (+/- 5,24) and 82,7% (+/-13,27) correct. In contrast, the CI only group only correctly recognized 65,4% (+/-12,88). Mean score for NHSs was 97,9% (+/- 2,63).

Raw scores were converted to  $d'$  scores (see figure 1), using Creelman's equation [Creelman, 1962]. Mean  $d'$  was 3,1 (+/-0,52) for the EAS group, 2,07 (+/-1,056) for the bimodal group, 0,91 (+/-0,786) for the CI only group and 3,27 (+/-0,234) for the NHS group. Performance of EAS was comparable to that of NHS ( $p=0,64$ ) and CI only subjects had significantly poorer results than the two other groups of CI users ( $p=0,001$  vs bimodal;  $p<0,0001$  vs EAS). The EAS group performed significantly better than the bimodal group ( $p=0,02$ ).

Experience with CI was significantly and positively related to  $d'$  scores in CI subjects with residual hearing (EAS and bimodal group ( $R=0,88$   $p<0,0001$ )). In CI only group, there was no relation between experience and prosody perception performance ( $R=0,1$ ,  $p=0,79$ ).

#### *Constant F0 condition*

This condition allowed us to evaluate the contribution of F0 variations in statement/question discrimination. When variations in F0 were removed from the sentences, CI subjects displayed a drop in performance. Mean score was 57,3% (+/-9,24) for the EAS group, 64,3% (+/-8,28) for the bimodal group and 58,7% (+/-10,10) for the CI only group. Corresponding  $d'$  values were 0,45 (+/-0,51) in EAS group, 0,78 (+/-0,46) in bimodal group and 0,53 (+/-0,37) in CI only group. Multiple comparisons using Fischer's test did not find any significant differences between CI groups. Paired comparisons for  $d'$  in EAS group showed a significant decrease in performance ( $t=5,1$ ;  $p=0,01$ ) between original and constant F0 conditions. The decrease was not significant for the bimodal ( $t=2,38$ ;  $p=0,55$ ) and CI only groups ( $t=2,38$ ;  $p=0,55$ ).

NHS' performance was significantly affected by the absence of F0 contour. Their score moved from near perfect in the original condition (% correct: 97,9% +/-2,63;  $d'$ : 3,3 +/-0,23) to near chance 58% (+/- 5,63), with  $d'=0,4$  (+/-0,35) – similar to that for CI groups.

#### *Synthesized condition*

In this condition, the last word or syllable of the sentence was equalized between the interrogative and affirmative mode for duration and amplitude. Therefore subjects could only rely on F0 variations to distinguish questions from statements. As the F0 contour is the prominent cue in speech prosody perception, we expected that NHS' scores would not be affected by this acoustic manipulation.

Indeed, NHS showed high scores (mean 98,2% $\pm$ 3,72;  $d'$ :3,3 $\pm$ 0,33). Similarly, performance for EAS and bimodal subjects was not affected by these synthetic stimuli. Paired comparisons for  $d'$  (3,2 $\pm$ 0,33 for EAS group; 2,4 $\pm$ 0,67 for bimodal group) did not reveal any significant difference between their performance in the original and the synthetic condition (respectively  $t=-0,34$ ,  $p=0,76$ ;  $t=-1,8$ ,  $p=0,12$ ). In contrast, CI only subjects showed a significant drop in scores for the synthetic condition versus the original condition ( $t=2,3$ ;  $p=0,047$ ) with  $d' = 0,91$  ( $\pm$ 0,786) for the original condition compared to 0,34 ( $\pm$ 0,89) for the synthetic condition.

### *F0 perception*

F0DLs were measured for several frequency ranges (110Hz-220Hz-400Hz-500Hz-750Hz) for each subject and an aggregate, mean F0DL was then calculated. Two subjects (Bim 3 and CI only 6) displayed particularly poor F0 discrimination with a mean F0DL exceeding two standard deviations from their group's mean (respectively 24% and 171%) and were thereby excluded from further analyses. An ANOVA was performed on F0DLs for the three CI patient groups, and NHS for all frequencies (see figure 2). There was a significant effect of group ( $F(3;175)=9,9$ ;  $p<0,0001$ ). Mean F0DL was significantly higher (poorer) for the CI only group compared to the EAS group (16% $\pm$ 13 vs 5,2% $\pm$ 1,4,  $p=0,028$ ) and compared to the NHS group (2,4% $\pm$ 1,9,  $p=0,0002$ ). When we analyzed frequency by frequency, the EAS group had significantly better/lower F0DLs than the CI only group when F0 standard was 110Hz and 220 Hz ( $p\leq 0,05$ ). Otherwise, no differences could be identified due to the wide variability in CI only subjects' performance.

These results showed that intonation perception and F0 discrimination were better in CI subjects with residual acoustic hearing. CI only subjects displayed a significant deficit in speech prosody perception and variable F0DLs. The relations between residual hearing level, F0 discrimination and speech prosody perception were then further investigated.

### *Impact of residual hearing level*

We evaluated the relation between speech prosody performance, F0DL perception and residual hearing level in the two groups of subjects with substantial residual hearing (EAS and bimodal groups) because residual hearing may convey information about F0. We considered hearing levels in non-operated ears, with the best hearing thresholds for each subject, and computed pure tone average low-frequency hearing loss using 125, 250, 500 and 750 Hz.

Mean residual hearing level and speech prosody  $d'$  scores in the original condition were highly correlated ( $R=-0,84$ ;  $p=0,001$ ); as were mean residual hearing level and mean F0DL ( $R=-0,85$ ;  $p=0,0009$ ) (see figure 3a and b). It appeared that the level of residual hearing may have a greater impact on intonation perception rather than whether a subject had significant levels of residual hearing in one or both ears. In other words, the differences shown so far between the EAS and bimodal groups may be attributed to the residual hearing level for subjects in each group rather than whether they were listening using unilateral or bilateral low-frequency acoustic hearing. From simple observation of figure 4, one can identify six subjects with high question-statement discrimination scores ( $d' > 2$ ) with a residual hearing level better than 60 dB and five subjects with poorer residual hearing and  $d'$  scores. The six best performing subjects were represented by all four subjects in EAS group and two subjects in the bimodal group with the best residual hearing. The speech prosody perception scores for these two bimodal subjects in the original condition were significantly higher than those of the five other subjects (mean  $d'=3,2 \pm 0,20$  vs mean  $d'=1,5 \pm 0,31$ ;  $t = 5,66$ ;  $p = 0,0003$ ). Similarly, their F0DLs were better than the six bimodal subjects with poorer residual hearing (mean F0DL =  $0,046 \pm 0,014$  vs  $0,138 \pm 0,065$ ;  $t = -3,44$ ;  $p = 0,007$ ).

#### *Relation between F0 perception and speech prosody performance*

We assumed that F0 cues were relevant for CI subjects with residual hearing since  $d'$  scores decreased significantly for the constant F0 condition. In contrast, performance in CI only subjects was not affected when F0 variations were removed. It was therefore plausible that this relation would not be significant in this group of subjects.

When we considered the 21 implanted subjects, we found a global significant negative correlation between F0DL and question/statement discrimination in original condition (Pearson's  $R=-0,47$ ;  $p=0,03$ ). When the correlation was performed only for subjects with residual hearing (EAS and bimodal groups) a higher correlation was obtained between the two variables (Pearson's  $R=-0,67$ ;  $p=0,03$ ). Within the CI only group, no significant correlation was found (Pearson's  $R=0,4$ ;  $p=0,22$ ).

## **Discussion**

### *Residual hearing advantages in F0 related tasks*

Our results highlight how residual hearing may help CI patients for prosody perception. Here CI subjects using bimodal stimulation or EAS achieved better performance than CI only subjects in discriminating questions from statements in the “real life”<sub>2</sub> original condition. In



our study, CI subjects with residual acoustic hearing better than 60 dB HL in the range 125-750 Hz achieved high  $d'$  scores, which were comparable to that of NHSs. Furthermore, we found strong correlations between residual hearing level and F0DL ( $R=-0,85$ ;  $p=0,0009$ ), and also between residual hearing level and recognition of speech prosody in terms of  $d'$  ( $R=-0,84$ ;  $p=0,001$ ).

Benefits for bimodal stimulation or electric acoustic hearing have been shown for speech recognition in noise [Armstrong et al., 1997, Fraysse et al., 2006, Adunka et al., 2010] or for music perception [Leal et al., 2003, El Fata et al., 2009]. For speech prosody, Straatman et al. [Straatman et al., 2010] showed a bimodal advantage in 17 CI children with residual hearing. They compared their question/statement discrimination performance when children were wearing or were not wearing their contralateral hearing aid. Using their residual hearing, correct recognition was improved between 6 to 11 percentage points. Cullington and Zeng [Cullington and Zeng, 2010] compared bimodal stimulation to bilateral CI in different tasks, including affective prosody [Ross et al., 1997]. Although no significant difference in this specific test was found between bilateral and bimodal CI users, performance tended to be better in bimodal patients, who displayed comparable scores to NHS controls, in contrast to bilateral CI subjects.

Though our series did not include bilateral cochlear implanted subjects, we studied CI patients with comparable speech in quiet recognition scores to focus on the hypothetical benefit of residual hearing in F0 and speech prosody perception. The positive impact of residual hearing was genuine since both EAS and bimodal groups showed better speech prosody perception than CI only subjects. Performance for the bimodal group should be interpreted with caution since delay post activation tended to be shorter than that of other CI groups, which might have lowered their performance. However, bimodal CI subjects displayed better  $d'$  scores than CI only subjects in original condition and two of them (bim1 and bim5) showed a particularly high  $d'$  score (3.46), comparable to NHSs. These two subjects had the best residual hearing for low frequencies within the bimodal group (see table 2). Combined with the results obtained by EAS subjects, we suggest that average low-frequency residual hearing better than 60 dB HL allows for excellent speech prosody perception. Indeed, our six best performers were the subjects with best acoustic residues and they achieved near normal scores (more than 95% correct recognition) compared to the CI only subjects (mean correct recognition score: 65,8%). Another factor that might have facilitated prosody perception among subjects with residual hearing is experience with CI. Indeed, we observed a significant correlation between delay post activation and  $d'$  scores in the original condition across EAS and bimodal subjects.

This effect might be related to the delay subjects need to fuse acoustic and electric hearing information [Tyler et al., 2002].

#### *Information provided by residual acoustic hearing*

First, residual hearing allows for acoustic transmission of F0 and thereby more accurate perception of intonation in the sentence. Indeed, mean F0DL was significantly better in the EAS subjects and the bimodal subjects with the best acoustic hearing. The natural efficacy of acoustic transmission of F0 may help CI subjects in music perception [Gfeller et al., 2002a, Gantz et al., 2006] or for recognizing speech in noise [Nooiteboom and Terken, 1982, Assmann and Summerfield, 1990]. But residual hearing may convey additional information, plausibly relevant for speech prosody perception. Kong and Carlyon [Kong and Carlyon, 2007] showed that the EAS advantage for speech in noise recognition persisted at low signal to noise ratios even when F0 cues were removed from the low-passed stimulus. Furthermore, this advantage disappeared at high SNR levels when F0 cues were preserved but low-frequency phonetic cues were eliminated. Besides F0, low-frequency hearing would then provide useful information to target the talker in background noise. Li and Loizou [Li and Loizou, 2008] suggested that low-frequency hearing was used to glimpse speech information in modulated background noise and to detect and integrate acoustic cues characterizing the speaker. Speech harmonics falling in the low frequencies are perceived and provide reliable F0 information. The better the low frequency hearing is, the better the resolution of harmonics. In addition, residual hearing may also contribute to the representation of the first formant (F1) frequency, which is critical for vowel and stop-consonant identification (voicing information), even at low signal to noise ratios. To get CI only subjects to perform as well as EAS or bimodal users in F0 related situations (speech in noise recognition, music perception, prosody perception, talker identification), F0 coding, and ideally the spectral resolution in the low frequency region, should therefore be improved.

#### *F0 encoding in CIs*

Our results regarding F0DL in CI only group are consistent with the literature, both quantitatively and qualitatively. Indeed, it is reported that best CI performers reach F0DL around 10% of a reference F0 and we found an average F0DL of 15,8% (+/-13). Moreover, substantial variability in F0 discrimination, as seen here, has consistently been reported in the literature: Pressnitzer et al [Pressnitzer et al., 2005] reported F0DLs between 2 and 7

semitones, that is between 16,5% and 58,5%. Geurts and Wouters [Geurts and Wouters, 2001] found F0DLs between 4% and 13% at 150 Hz for modulated pulse trains presented on single channels. In their study, at 250 Hz, one subject was not able to perform the task and in the three others F0DL varied between 5% and 12%.

Although most CI subjects display impaired F0 discrimination, they are still able to perform gross F0 related tasks, such as recognition of typical male and female voices. But finer F0 discrimination is required for more difficult F0 settings, such as melody perception or gender categorization of ambiguous voices. In another study by our group, we showed that CI users' recognition of typical male and female voices reached more than 90% of NHS's performance, but this level dropped by 63 percentage points in cases of more ambiguous voices. This is similar to Kovacic et al's [Kovacic and Balaban, 2009] findings where gender categorization was accurate for CI users outside of the overlapping male-female F0 region between 138 and 163 Hz.

Enhancing the temporal representation of F0 is one of the main approaches that has been adopted. Studies by Green et al. [Green et al., 2004, Green et al., 2005] assessed a 100% modulated sawtooth-like envelope distributed to all electrodes and found an improvement of intonation recognition, but this resulted in reduced vowel recognition. Vandali et al. [Vandali et al., 2005] evaluated different experimental strategies that enhanced temporal information. One strategy was intended to code fine temporal structure in each channel using positive temporal peaks in the band-pass filtered signal. The other strategies provided more F0 modulation cues, aligned across all channels. The experimental strategies provided better scores than the clinical strategy for sung vowels without affecting speech recognition. Another approach, aimed to improve F0 place coding, was devised by Geurts and Wouters [Geurts and Wouters, 2004]. Their strategy extracted the frequency of the first harmonic and steered the output current between two distinct electrodes in order to induce an intermediate F0 percept. Results showed improved F0 detection but speech recognition was not assessed. Peng et al. [Peng et al., 2009] demonstrated the facilitating role of intensity for intonation recognition when its variations are congruent with F0 contour information. In their study, the proportion of "question" recognition was higher when the stimulus ended up with an ascending F0 and an increment of intensity. This proportion decreased in cases of conflicting cues, i.e when an ascending F0 and a decreased intensity were present at the end of the stimulus. Our results obtained in CI only group also lead us to consider intensity as a prosodic cue. In the synthetic condition of the question-statement test in the current study, CI only subjects showed a significant decrease in recognition scores compared to their performance in

the original condition (mean  $d'$  dropped from 0,91 to 0,34; mean correct recognition scores from 65,4% to 55,4%, i.e chance level). Therefore, the absence of intensity cues significantly and negatively influenced CI only subject's performance in statement-question recognition. However it should be emphasised that overall their performance was very poor even for the original condition.

### *Limitations*

This study was conducted on a small number of subjects but the differences between groups were large and statistically significant. One of the main limitations is the difference between mean ages across groups, which might have affected CI only group's performance. Several studies demonstrated the effect of ageing on F0 related tasks [Souza et al., 2011, Schwartz and Chatterjee, 2012]. In our study however, we did not find any correlation between age and mean F0DL ( $R=0,24$ ;  $p=0,3$ ), nor between age and  $d'$  in the original condition ( $R=0,36$ ;  $p=0,1$ ).

### **Conclusion**

Residual hearing combined with CI provides F0 via low frequency harmonic information that is required for speech prosody perception. In our study, residual hearing level was highly correlated to both F0 detection and question-statement discrimination. Furthermore, mean acoustic residual hearing better than 60 dB HL for 125-750Hz provided near-normal performance in speech prosody recognition. In children these results raise the issue of bilateral implantation and the need for a proper pre-operative hearing thresholds evaluation. The failure of current CI strategies to represent F0 and the potentially high contribution of residual hearing both suggest caution in cases of doubtful low-frequency hearing thresholds and might support sequential bilateral cochlear implantation rather than simultaneous implantation after an attentive follow-up of the acoustic hearing in the non-implanted ear.

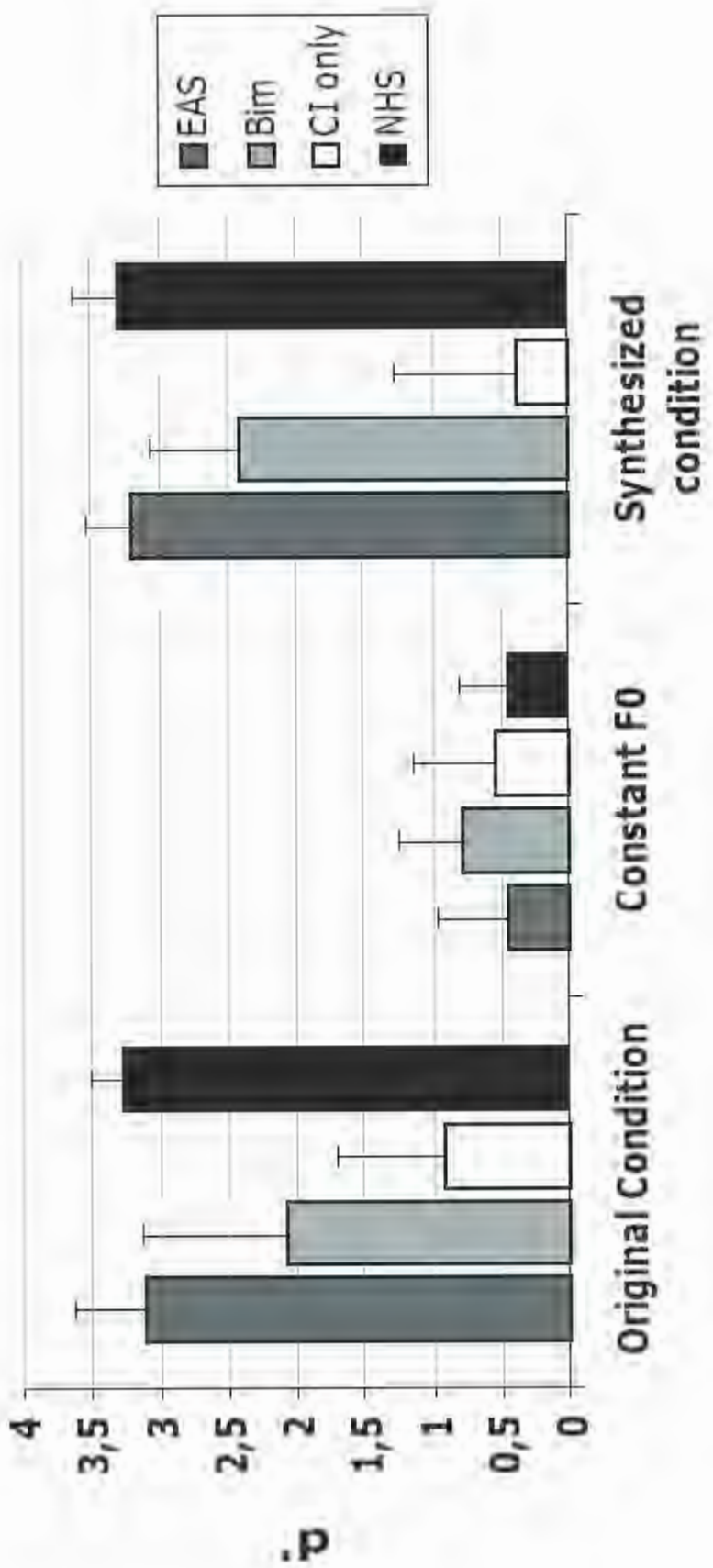
Acknowledgements: we thank cochlear implanted subjects and normal hearing subjects for their participation in this study, M-L Laborde and A. Capber for their help in collecting the data, C. Marlot for her help in bibliography. This work was supported by Direction de la Recherche Clinique et de l'Innovation, CHU Toulouse, the ANR Plasmody (PB) and CNRS (MM).

## References

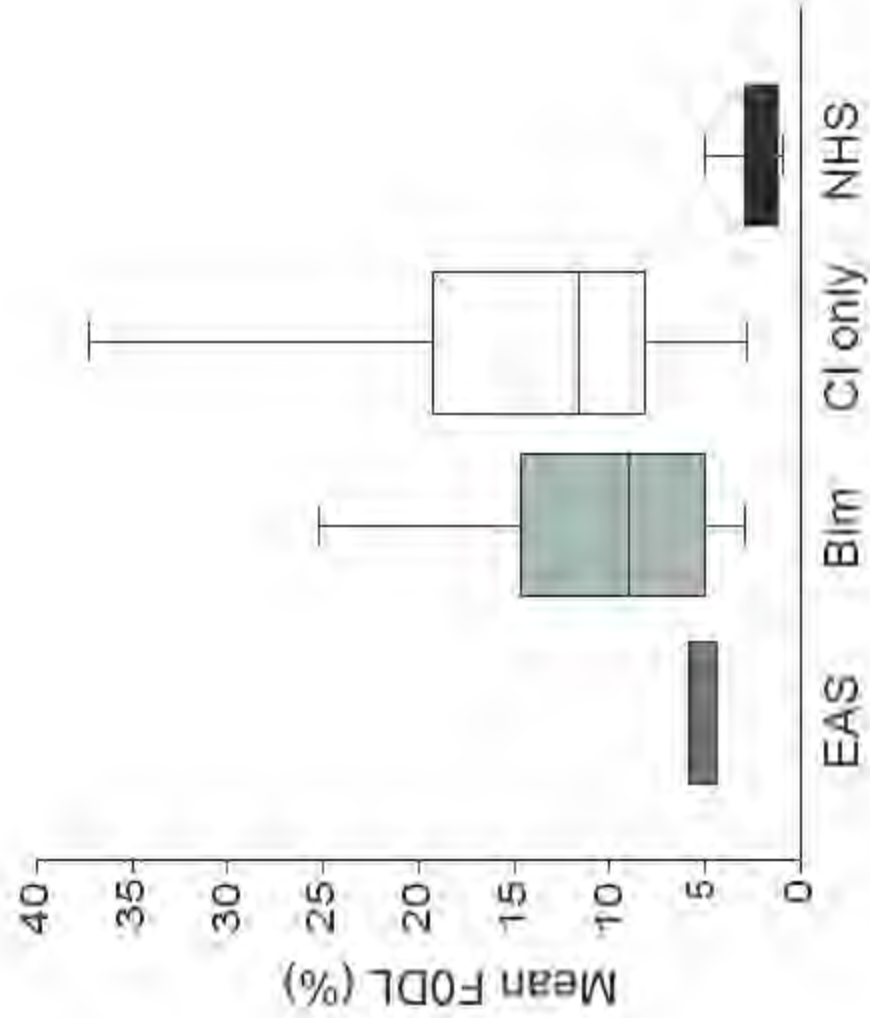
- Adunka, O. F., Pillsbury, H. C., Adunka, M. C., and Buchman, C. A. (2010). "Is electric acoustic stimulation better than conventional cochlear implantation for speech perception in quiet?," *Otol Neurotol* 31, 1049-1054.
- Armstrong, M., Pegg, P., James, C., and Blamey, P. (1997). "Speech perception in noise with implant and hearing aid," *The American journal of otology* 18, S140-141.
- Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: vowels with different fundamental frequencies," *The Journal of the Acoustical Society of America* 88, 680-697.
- Boersma, P., and D, W. (2009). "Praat: doing phonetics by computer".
- Chatterjee, M., and Peng, S. C. (2008). "Processing F0 with cochlear implants: Modulation frequency discrimination and speech intonation recognition," *Hearing research* 235, 143-156.
- Creelman, C. (1962). "Human discrimination of auditory duration," *The Journal of the Acoustical Society of America* 34, 582-593.
- Cullington, H. E., and Zeng, F. G. (2010). "Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear," *Ear and hearing* 31, 70-73.
- El Fata, F., James, C. J., Laborde, M. L., and Fraysse, B. (2009). "How much residual hearing is 'useful' for music perception with cochlear implants?," *Audiology & neuro-otology* 14 Suppl 1, 14-21.
- Fonagy, I. (1958). "[Phonetics & psychoanalysis; psychological remarks on the stress accent]," *Psyche* 12, 63-72.
- Fonagy, I. (1966). "Electrophysiological and acoustic correlates of stress and stress perception," *Journal of speech and hearing research* 9, 231-244.
- Fraysse, B., Macias, A. R., Sterkers, O., Burdo, S., Ramsden, R., Deguine, O., Klenzner, T., Lenarz, T., Rodriguez, M. M., Von Wallenberg, E., and James, C. (2006). "Residual hearing conservation and electroacoustic stimulation with the nucleus 24 contour advance cochlear implant," *Otol Neurotol* 27, 624-633.
- Gantz, B. J., Turner, C., and Gfeller, K. E. (2006). "Acoustic plus electric speech processing: preliminary results of a multicenter clinical trial of the Iowa/Nucleus Hybrid implant," *Audiology & neuro-otology* 11 Suppl 1, 63-68.
- Geurts, L., and Wouters, J. (2001). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," *The Journal of the Acoustical Society of America* 109, 713-726.
- Geurts, L., and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *The Journal of the Acoustical Society of America* 115, 844-852.
- Gfeller, K., Turner, C., Mehr, M., Woodworth, G., Fearn, R., Knutson, J. F., Witt, S., and Stordahl, J. (2002a). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," *Cochlear implants international* 3, 29-53.
- Gfeller, K., Witt, S., Woodworth, G., Mehr, M. A., and Knutson, J. (2002b). "Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal," *The Annals of otology, rhinology, and laryngology* 111, 349-356.
- Grassi, M., and Soranzo, A. (2009). "MLP: a MATLAB toolbox for rapid and reliable auditory threshold estimation," *Behavior research methods* 41, 20-28.

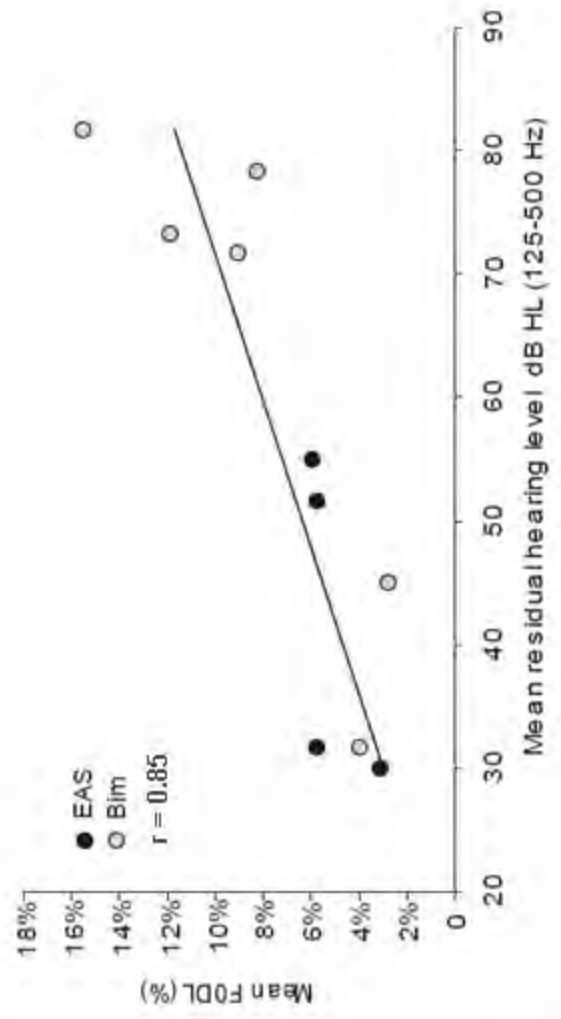
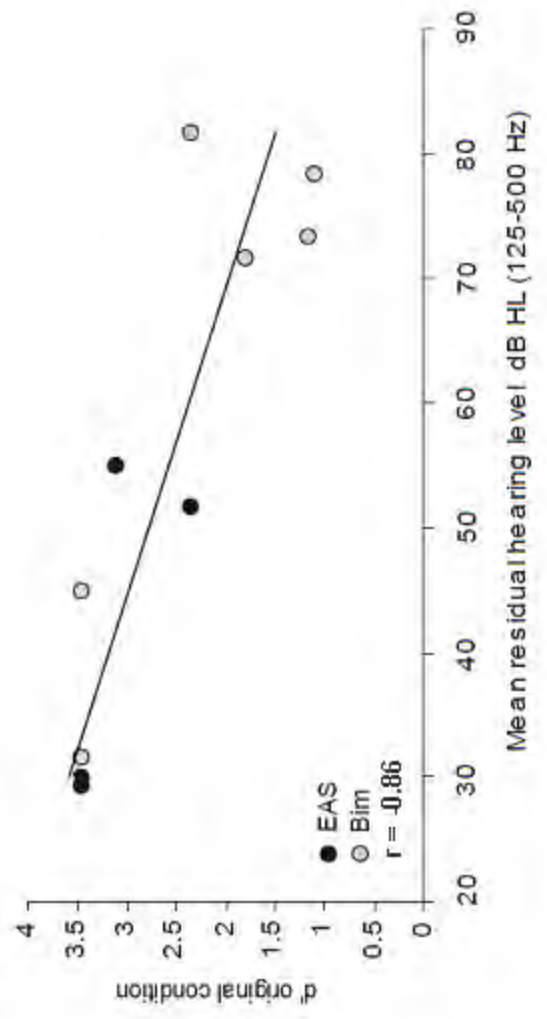
- Green, T., Faulkner, A., and Rosen, S. (2002). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," *The Journal of the Acoustical Society of America* 112, 2155-2164.
- Green, T., Faulkner, A., and Rosen, S. (2004). "Enhancing temporal cues to voice pitch in continuous interleaved sampling cochlear implants," *The Journal of the Acoustical Society of America* 116, 2298-2310.
- Green, T., Faulkner, A., Rosen, S., and Macherey, O. (2005). "Enhancement of temporal periodicity cues in cochlear implants: effects on prosodic perception and vowel identification," *The Journal of the Acoustical Society of America* 118, 375-385.
- Kong, Y. Y., and Carlyon, R. P. (2007). "Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation," *The Journal of the Acoustical Society of America* 121, 3717-3727.
- Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (2004). "Music perception with temporal cues in acoustic and electric hearing," *Ear and hearing* 25, 173-185.
- Kovacic, D., and Balaban, E. (2009). "Voice gender perception by cochlear implantees," *The Journal of the Acoustical Society of America* 126, 762-775.
- Leal, M. C., Shin, Y. J., Laborde, M. L., Calmels, M. N., Verges, S., Lugardon, S., Andrieu, S., Deguine, O., and Fraysse, B. (2003). "Music perception in adult cochlear implant recipients," *Acta oto-laryngologica* 123, 826-835.
- Li, N., and Loizou, P. C. (2008). "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," *The Journal of the Acoustical Society of America* 123, 2287-2294.
- Nagel, H. N., Shapiro, L. P., and Nawy, R. (1994). "Prosody and the processing of filler-gap sentences," *Journal of psycholinguistic research* 23, 473-485.
- Nakata, T., Trehub, S. E., and Kanda, Y. (2012). "Effect of cochlear implants on children's perception and production of speech prosody," *The Journal of the Acoustical Society of America* 131, 1307-1314.
- Nooteboom, S. G., and Terken, J. M. (1982). "What makes speakers omit pitch accents? An experiment," *Phonetica* 39, 317-336.
- Patel, A. D., Peretz, I., Tramo, M., and Labreque, R. (1998). "Processing prosodic and musical patterns: a neuropsychological investigation," *Brain and language* 61, 123-144.
- Peng, S. C., Lu, N., and Chatterjee, M. (2009). "Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners," *Audiology & neuro-otology* 14, 327-337.
- Peng, S. C., Tomblin, J. B., Cheung, H., Lin, Y. S., and Wang, L. S. (2004). "Perception and production of mandarin tones in prelingually deaf children with cochlear implants," *Ear and hearing* 25, 251-264.
- Peng, S. C., Tomblin, J. B., and Turner, C. W. (2008). "Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing," *Ear and hearing* 29, 336-351.
- Pressnitzer, D., Bestel, J., and Fraysse, B. (2005). "Music to electric ears: pitch and timbre perception by cochlear implant patients," *Annals of the New York Academy of Sciences* 1060, 343-345.
- Ross, E. D., Thompson, R. D., and Yenkosky, J. (1997). "Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions," *Brain and language* 56, 27-54.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., and Barone, P. (2007). "Evidence that cochlear-implanted deaf patients are better multisensory integrators," *Proceedings of the National Academy of Sciences of the United States of America* 104, 7295-7300.

- Schvartz, K. C., and Chatterjee, M. (2012). "Gender identification in younger and older adults: use of spectral and temporal cues in noise-vocoded speech," *Ear and hearing* 33, 411-420.
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," *Hearing research* 11, 157-189.
- Souza, P., Arehart, K., Miller, C. W., and Muralimanohar, R. K. (2011). "Effects of age on F0 discrimination and intonation perception in simulated electric and electroacoustic hearing," *Ear and hearing* 32, 75-83.
- Speer, S. R., Kjelgaard, M. M., and Dobroth, K. M. (1996). "The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities," *Journal of psycholinguistic research* 25, 249-271.
- Stickney, G. S., Nie, K., and Zeng, F. G. (2005). "Contribution of frequency modulation to speech recognition in noise," *The Journal of the Acoustical Society of America* 118, 2412-2420.
- Straatman, L. V., Rietveld, A. C., Beijen, J., Mylanus, E. A., and Mens, L. H. (2010). "Advantage of bimodal fitting in prosody perception for children using a cochlear implant and a hearing aid," *The Journal of the Acoustical Society of America* 128, 1884-1895.
- Swets, J. A., Harris, J. R., McElroy, L. S., and Rudloe, H. (1966). "Computer-aided instruction in perceptual identification," *Behavioral science* 11, 98-104.
- Tanner, W. P., Jr., and Swets, J. A. (1954). "A decision-making theory of visual detection," *Psychological review* 61, 401-409.
- Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (2002). "Patients utilizing a hearing aid and a cochlear implant: speech perception and localization," *Ear and hearing* 23, 98-105.
- Vandali, A. E., Sucher, C., Tsang, D. J., McKay, C. M., Chew, J. W., and McDermott, H. J. (2005). "Pitch ranking ability of cochlear implant recipients: a comparison of sound-processing strategies," *The Journal of the Acoustical Society of America* 117, 3126-3138.
- Williams, C. E., and Stevens, K. N. (1972). "Emotions and speech: some acoustical correlates," *The Journal of the Acoustical Society of America* 52, 1238-1250.
- Wilson, B. S., and Dorman, M. F. (2008). "Cochlear implants: a remarkable past and a brilliant future," *Hearing research* 242, 3-21.
- Wingfield, A. (1975). "Acoustic redundancy and the perception of time-compressed speech," *Journal of speech and hearing research* 18, 96-104.







**A****B**

# Brain Topography

## Brain prediction of auditory emphasis by facial expressions during audiovisual continuous speech --Manuscript Draft--

<b>Manuscript Number:</b>	BTOP-D-13-00019
<b>Full Title:</b>	Brain prediction of auditory emphasis by facial expressions during audiovisual continuous speech
<b>Article Type:</b>	SI: Auditory Cortex 2012
<b>Section/Category:</b>	Cognitive neuroscience
<b>Keywords:</b>	audio-visual speech; prosody; mismatch; predictive coding
<b>Corresponding Author:</b>	Kuzma Strelnikov FRANCE
<b>Corresponding Author Secondary Information:</b>	
<b>Corresponding Author's Institution:</b>	
<b>Corresponding Author's Secondary Institution:</b>	
<b>First Author:</b>	Kuzma Strelnikov
<b>First Author Secondary Information:</b>	
<b>Order of Authors:</b>	Kuzma Strelnikov Jessica Foxton Mathieu Marx Pascal Barone
<b>Order of Authors Secondary Information:</b>	
<b>Abstract:</b>	<p>The visual cues involved in auditory speech processing are not restricted to information from lip movements but also include head or chin gestures and facial expressions such as eyebrow movements. The fact that visual gestures precede the auditory signal implicates that visual information may influence the auditory activity. As visual stimuli are very close in time to the auditory information for audiovisual syllables, the cortical response to them usually overlaps with that for the auditory stimulation; the neural dynamics underlying the visual facilitation for continuous speech therefore remains unclear. We presented video clips with even continuous speech as the frequent stimuli and with one word visually emphasized by the speaker as the non-frequent stimuli. Negativity in the resulting ERPs was detected after the start of the emphasizing articulatory movements but before the auditory stimulus, a finding that was confirmed by the statistical comparisons of the audiovisual and visual stimulation. No such negativity was present in the control visual-only condition. The propagation of this negativity formed a loop between the visual and fronto-temporal regions. Thus, in continuous speech, the visual modality evokes predictive coding for the auditory speech, which is analysed by the cerebral cortex in the context of the phrase even before the arrival of the corresponding auditory signal.</p>
<b>Suggested Reviewers:</b>	Mireille Besson Mireille.Besson@incm.cnrs-mrs.fr Specialist in the auditory and audiovisual prosody processing.  Virginie van Wassenhove virginie.van-wassenhove@cea.fr Specialist in audiovisual speech processing.  Jean Vroomen J.Vroomen@uvt.nl

	Specialist in audiovisual interactions.
	Anne-Lise Giraud Anne-Lise.Giraud@unige.ch Specialist in audiovisual speech processing.
<b>Opposed Reviewers:</b>	

To the Editor.

Please find enclosed the manuscript entitled “Brain prediction of auditory emphasis by facial expressions during audiovisual continuous speech” which we would like to submit for publication in *Brain Topography*.

It is now widely accepted that the simultaneous integration of different sensory modalities is important for our perceptual experience of the world. This is even more important when it concerns social communication through speech processing because speech is multisensory by nature. The present work investigates at the neuronal level the cross-modal effects for speech intonation (prosody) during continuous speech. During EEG recording, we developed an original protocol based on the McGurk principle but using prosody-related visual information encompassing head and eyebrows movements. In an oddball paradigm, our results demonstrate that visual features of speech prosody induce a MMN-like response that precedes the auditory speech information. This demonstrates for the first time that visual gestures play an important role in ecological speech processing and can be used to predict certain features of the auditory utterance such as pitch or intensity of the emphasized word. In line with the evidence of visuo-auditory interactions at the level of the auditory system during speech processing, our work provides further cues to understand how the visual features of speech prosody can modulate the auditory activity.

A publication in *Brain Topography* would offer us the unique opportunity to reach a broad readership interested in the implications of our results. Firstly, our work is the first to address the neuronal mechanisms by which prosody-related visual features participate in auditory speech perception. Our results are likely to stimulate further research of prosody as a multimodal percept though it has traditionally been considered solely in terms of its auditory features. Secondly, our data highlight the importance of the predictive mode of brain functioning in ecological communicative situations and opens further perspectives for the studies of predictive cues and their impact on speech processing.

Altogether, we believe that our study will bring the readers of *Brain Topography* important new insights into human brain functioning during ecological audiovisual speech processing.

I recommend the following qualified referees:

Mireille Besson Institut de Neurosciences Cognitives de la Méditerranée CNRS  
Mireille.Besson@incm.cnrs-mrs.fr

Virginie van Wassenhove    INSERM-CEA Cognitive Neuroimaging Unit, France  
virginie.van-wassenhove@cea.fr

Jean Vroomen    Tilburg University, Tilburg, The Netherlands  
J.Vroomen@uvt.nl

Anne-Lise Giraud  
Département des Neurosciences Fondamentales, Université de Genève.  
Anne-Lise.Giraud@unige.ch

I declare that there is no conflict of interest in this research.

Sincerely yours,  
Pascal Barone

1  
2 Kuzma Strelnikov<sup>1,2\*α</sup>, Jessica Foxton<sup>1,2\*</sup>, Mathieu Marx<sup>1,2,3</sup> and Pascal  
3 Barone<sup>1,2</sup>

4 **Brain prediction of auditory emphasis by facial expressions during**  
5 **audiovisual continuous speech**  
6  
7

8  
9  
10  
11  
12 \*Joint first authors  
13

14  
15  
16  
17 1. Université Toulouse, CerCo, Université Paul Sabatier  
18

19 2. CNRS, UMR 5549. Faculté de Médecine de Rangueil, Toulouse France  
20

21 3. Service d'Oto–Rhino–Laryngologie, Hopital Purpan, Toulouse, France  
22  
23  
24

25  
26 α Corresponding author  
27

28 e-mail: kuzma@cerco.ups-tlse.fr  
29  
30

31 CNRS CERCO UMR 5549 Pavillon Baudot CHU Purpan BP 25202,  
32

33  
34 31052 Toulouse Cedex  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Abstract

1  
2  
3  
4  
5 The visual cues involved in auditory speech processing are not restricted to  
6  
7 information from lip movements but also include head or chin gestures and facial  
8  
9 expressions such as eyebrow movements. The fact that visual gestures precede the  
10  
11 auditory signal implicates that visual information may influence the auditory  
12  
13 activity. As visual stimuli are very close in time to the auditory information for  
14  
15 audiovisual syllables, the cortical response to them usually overlaps with that for  
16  
17 the auditory stimulation; the neural dynamics underlying the visual facilitation for  
18  
19 continuous speech therefore remains unclear. We presented video clips with even  
20  
21 continuous speech as the frequent stimuli and with one word visually emphasized  
22  
23 by the speaker as the non-frequent stimuli. Negativity in the resulting ERPs was  
24  
25 detected after the start of the emphasizing articulatory movements but before the  
26  
27 auditory stimulus, a finding that was confirmed by the statistical comparisons of  
28  
29 the audiovisual and visual stimulation. No such negativity was present in the  
30  
31 control visual-only condition. The propagation of this negativity formed a loop  
32  
33 between the visual and fronto-temporal regions. Thus, in continuous speech, the  
34  
35 visual modality evokes predictive coding for the auditory speech, which is  
36  
37 analysed by the cerebral cortex in the context of the phrase even before the arrival  
38  
39 of the corresponding auditory signal.  
40  
41  
42  
43  
44  
45  
46  
47  
48

49 *Key words:* audio-visual speech; prosody; mismatch; predictive coding  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



# Introduction

Ecological speech has a multisensory nature and is mostly continuous, involving whole words and phrases but not separate phonemes or syllables. Speech perception combines both auditory and visual features in face-to-face communication beginning from early childhood. The information provided by the auditory and visual channels is relatively redundant, and in natural situations most speech processing is supported solely by the auditory system. However, in degraded auditory conditions, congruent visual information from lip movements added to the auditory speech signal significantly increases the accuracy of speech comprehension (Sumbly and Pollack 1954), a phenomenon which is equivalent to an increase in the signal to noise ratio (Ross et al. 2007). An important characteristic of audio-visual natural speech is that the visual cues originating from the mouth opening precede the auditory information by 100-200ms and are strongly correlated with the auditory speech envelope (Chandrasekaran et al. 2009). This temporal dynamic implies that lip movements provide strong predictive cues for the auditory information. Such facilitatory audio-visual interactions can be observed at the behavioral level when analyzing speech intelligibility, but they are also present at the neuronal level, as expressed as a visual modulation of auditory speech processing in the human auditory cortex (Besle et al. 2004; Besle et al. 2008; Hertrich et al. 2007).

However, the visual cues involved in auditory speech processing are not restricted to information from the lip movements, but they also include head and chin gestures and facial expressions such as eyebrow movements (Munhall et al. 2004). These visual cues are especially important at the supra-segmental level of

1 verbal or emotional communication, corresponding to the speech prosody.  
2  
3 Locution prosody consists of modulations in the pitch, amplitude, and duration  
4  
5 patterns of the words in a phrase, and enables listeners to distinguish between  
6  
7 questions and statements, finished and unfinished phrases, or to recognize the  
8  
9 emotional state of the speaker. Visual, non-verbal gestures contribute to speech  
10  
11 prosody, and head and eyebrow movements have been shown to correlate with  
12  
13 modulations in the pitch and amplitude of the talker's voice (Munhall et al. 2004;  
14  
15 Hadar et al. 1984; Vatikiotis-Bateson and Yehia 2002). Again, based on the  
16  
17 facilitatory rules that govern multisensory interactions, visual prosody-related  
18  
19 information improves auditory speech comprehension in situations of degraded  
20  
21 auditory information (Munhall et al. 2004; Barkhuysen et al. 2008). Recently, we  
22  
23 have been able to show that prosodic visual cues can affect not only the linguistic  
24  
25 aspects of speech but also the more fundamental levels of auditory processing.  
26  
27 Indeed, we have demonstrated that the visual features of speech prosody can  
28  
29 induce a crossmodal facilitation in detecting the auditory features of prosody, as  
30  
31 expressed as a decrease in the threshold for detecting amplitude changes (Foxton  
32  
33 et al. 2010).

41 This study suggests that visual cues improve sensitivity to loudness  
42  
43 changes and consequently can improve speech-sound processing. A fundamental  
44  
45 question then emerges as to the underlying neuronal mechanisms of such an  
46  
47 audio-visual facilitation. The fact that these visual gestures precede the auditory  
48  
49 signal by several milliseconds (Hadar et al. 1984) means that the visual  
50  
51 information may play a predictive role, which could influence the neuronal  
52  
53 auditory responses. While it is known that visual cues from lip-reading can  
54  
55 modulate auditory cortical responses (Campbell 2008), no evidence for such an  
56  
57 interaction exists for prosody-related visual information. The purpose of the  
58  
59  
60  
61  
62  
63  
64  
65

1 present study was to investigate the temporo-spatial neural dynamics underlying  
2 this visually driven facilitation in continuous speech. In contrast to studies of  
3  
4 separate phonemes and syllables, this study focuses on the intonation emphasis of  
5  
6 a word in a phrase.  
7

8  
9 In order to examine the effect of the visual prosodic information on  
10 auditory responses during continuous speech, we used an oddball paradigm in  
11  
12 which the standard and deviant stimuli contained the same auditory information,  
13  
14 but where the deviant stimuli had visual emphasis on one word. Our hypothesis  
15  
16 was that the visual cues would give the impression that there is an increase in the  
17  
18 amplitude/pitch of this word, and that this would be detected as such by the  
19  
20 auditory system and expressed as a brain response similar to the auditory  
21  
22 mismatch negativity (MMN) response (Näätänen et al. 2007). This approach is  
23  
24 analogous to the one developed for the McGurk effect to study linguistic audio-  
25  
26 visual interactions, where the visual information alone can change the perceived  
27  
28 syllable (McGurk and MacDonald 1976). In these studies, a significant MMN has  
29  
30 been found, demonstrating a visual influence on speech processing in the auditory  
31  
32 system (Colin et al. 2002; Saint-Amour et al. 2007; Besle et al. 2005). In our  
33  
34 prosodic visual-auditory protocol, the auditory and visual information were not  
35  
36 congruent with respect to one word in a phrase: the auditory information had a  
37  
38 neutral prosodic pattern, while the visual information emphasized one of the  
39  
40 words in the phrase. This paradigm enabled us to explore the visual influence on  
41  
42 prosodic processing independently of any auditory modulation.  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

# Materials and methods

## Participants

Ten normally-hearing native French speakers (mean age 24, range 19-29) participated in the study. Seven of the participants were female and all were right-handed. All of the subjects reported no auditory, neurological, or psychiatric disease, and all had normal or corrected to normal vision. All of the participants gave their full-informed consent prior to their participation in this study, in accordance with the Declaration of Helsinki (1968). The study was approved by the local research ethics committee (Comité Consultatif de Protection des Personnes dans la Recherche Biomédicale Toulouse II Avis N°2-03-34/Avis N°2).

## Stimuli

An actress was asked to say the words “un-deux-trois” (translated into English as “one-two-three”) twice: once with *no emphasis* (NE), and again with *emphasis* (WE) on the word “two”. She was asked to try to produce the words with the same timing. Several video recordings were taken using a professional Sony digital camera (recording with 25 frames per second, with a resolution of 720 x 576 pixels, and a sound sample rate of 48000Hz) in a sound-attenuating chamber. Two video clips were chosen, one with no emphasis and the other with emphasis. Both videos were 2.52 seconds in duration. Using Adobe Premier Pro 7.0, the sound in the deviant video clip was replaced with the sound from the ‘no-emphasis’ video clip. In this way, the two video clips had exactly the same sound (always from the ‘no-emphasis’ clip), and differed only in terms of the visual signal. Thus the WE video can be considered as an incongruent audio-visual stimulus with respect to the prosodic information (see (Foxton et al. 2010). The

1 visual features of emphasis included raised eyebrows during the word “two” that  
2 were not present in the ‘no-emphasis’ video clip (Figure 1).  
3

4  
5 These visual gestures were clearly dissociated from the auditory stimulus  
6  
7 as they preceded the sound “two” by a period of about 500 ms, making them a  
8  
9 good predictor of the emphasised word in the intact with-emphasis video (before  
10  
11 sound replacement). In the dubbed video clip, we maintained this same delay  
12  
13 between the visual and auditory cues.  
14  
15

16  
17 To verify the neutral prosody of the phrase, we tested a supplementary set  
18  
19 of subjects (n = 17) asking what word was emphasized. The majority (80%) of  
20  
21 them did not hear any emphasis on the word “two” but indicated randomly other  
22  
23 words when tested with the auditory-only recording ( $p < 0.001$ ,  $\chi^2$ -test). In  
24  
25 addition, in the deviant audio-visual clip, the illusion of the auditory word “two”  
26  
27 as being emphasized was reported by every one of the 10 subjects who  
28  
29 participated in the EEG study (they were asked after the study), as well as in every  
30  
31 one of 8 additional controls.  
32  
33  
34  
35  
36  
37  
38  
39

## 40 **Procedure**

41  
42 For this study, we adopted an oddball design, whereby participants were  
43  
44 presented with repetitions of the “no-emphasis” video clip interspersed with  
45  
46 occasional presentations of the deviant “with-emphasis” video clip, with a  
47  
48 constant inter-stimulus onset asynchrony of 2920ms. As the deviant stimulation  
49  
50 was located in the middle of the phrase, it could not cause an expectation effect  
51  
52 for the whole phrase.  
53  
54

55  
56 For half of the runs, the video clips were presented with the sound (audio-  
57  
58 visual condition-AV); for the other half, the video clips were presented without  
59  
60 the sound (visual-alone condition-V). During the runs, participants were asked to  
61  
62  
63  
64  
65

1 detect rare white crosses that occasionally appeared over the eyes during the video  
2 clips. They had to respond by pressing a button. The crosses appeared after one of  
3  
4 four possible delays with respect to the start of the video (1.0, 1.2, 1.4, or 1.6  
5  
6 seconds). This task ensured that the participants focused their attention on the  
7  
8 videos, and especially around the time of the word “two”. By limiting the white  
9  
10 crosses to the eyes, we focused the participants’ attention on a facial area where  
11  
12 visual features of emphasis are prominent, given the presence of raised eyebrows  
13  
14 and widened eyes during the emphasised word.  
15  
16  
17  
18

19 There were six AV runs and six V runs, each lasting approximately 7  
20  
21 minutes. The order of the AV and V runs was randomised for each participant. In  
22  
23 each run, there were 100 standard videos, 25 deviant videos, and a further 15  
24  
25 standard videos that contained a target white cross. 17.9 % of the videos were  
26  
27 deviants and 10.7% of the videos were targets, and there were a total of 150  
28  
29 deviants for each condition. Within each run, the order of the videos was pseudo-  
30  
31 randomised, with the constraint that the first ten videos were always standards and  
32  
33 that there were at least three standard videos between the deviants.  
34  
35  
36  
37  
38  
39  
40  
41

## 42 **EEG recordings and analyses**

43  
44 EEG recordings were obtained using an elastic cap (Oxford Instruments,  
45  
46 UK) fitted with tin electrodes and with a reference electrode placed on the tip of  
47  
48 the nose. The cap electrode locations were in accordance with the 10-20 system  
49  
50 with additional electrodes from the 10-10 system, and were at locations: FP1-FP2-  
51  
52 F3-F4-C3-C4-P3-P4-F7-F8-T3-T4-T5-T6-CB2-CZ-Fz-Pz-T5'-T6'-O1-O2-O1'-  
53  
54 O2'-P3'-P4'-Pz'-Oz-CB1-M1-M2-VEOG (32 electrodes). Additional electrodes  
55  
56 were placed on the right and left mastoids, and the vertical electro-oculogram  
57  
58 (VEOG) was recorded between electrodes placed above and below the eyes. The  
59  
60  
61  
62  
63  
64  
65

1 ground electrode was placed along the midline in front of Fz, and electrode  
2 impedances were kept below 10 kOhms. EEG and EOG data were recorded using  
3 a SynAmps amplifier (NeuroScan, El Paso, TX), with a sampling rate of 1000Hz,  
4 and low-pass filtered at 200Hz.  
5  
6

7  
8  
9 A standardised EEG analysis procedure was followed using Neuroscan software.  
10  
11 The data was band-pass filtered at 2-30 Hz (slope 48 dB / octave, FIR) and  
12  
13 epoched from 700 to 1500 ms with respect to the beginning of the video clip. In  
14  
15 the text and figures, this corresponds to the “zero” of the epoch used for the  
16  
17 analysis. This timing was chosen to target the analysis towards the visual and  
18  
19 auditory parts of the word of interest: "two" (Fig 1). We chose this onset on the  
20  
21 basis of the timing of the visual gestures, and the “zero” corresponds closely to  
22  
23 the minimum difference between the NE and WE videos for both the eyebrow  
24  
25 movements and mouth opening. The epochs were baseline-corrected for a time  
26  
27 period of 100 ms before the epoch. This baseline corresponded to the interval  
28  
29 between the auditory “one” and “two”. Epochs containing excessive residual  
30  
31 artefacts were excluded (artefact rejection outside -40:40  $\mu$ V), as were epochs for  
32  
33 videos immediately following the deviants or targets to exclude the possible  
34  
35 contamination by the mismatch negativity to the standard stimulus following the  
36  
37 deviant one (Sams et al. 1984). After the rejection, for each condition we had the  
38  
39 average of 510 trials per subject. The recording sites plus an electrode placed on  
40  
41 the right mastoid were referenced online to the left mastoid electrode and digitally  
42  
43 re-referenced offline to the algebraic average of the left and right mastoids  
44  
45 (Naaatanen et al. 2007). Averages were calculated separately for the standard and  
46  
47 the deviant videos.  
48  
49  
50  
51  
52  
53  
54  
55  
56

57  
58 Statistical analyses were carried out in order to determine whether there  
59  
60 were significant differences between the waveforms for the standard and deviant  
61  
62  
63  
64  
65

1 videos. These focused on the time period during the word “two” where the visual  
2 differences related to emphasis were present. Average waveforms were calculated  
3  
4 for each participant, and then waveforms for standards and deviants were  
5  
6 compared with a paired t-test for the electrode of interest, Fz. To sharpen surface  
7  
8 topographies, we conducted the current source density (CSD) analysis separately  
9  
10 for standard and deviant stimuli and calculated the differential map using the CSD  
11  
12 toolbox  
13  
14 (<http://psychophysiology.cpmc.columbia.edu/software/CSDtoolbox/index.html>),  
15  
16 which computes scalp surface Laplacian or current source density (CSD)  
17  
18 estimates for surface potentials.  
19  
20  
21  
22  
23

24 The basic methods of statistical estimation were a 2x2 repeated measures  
25  
26 ANOVA including the modality (audio-visual, visual) and type of stimulation  
27  
28 (standard, deviant) factors, and a bootstrap test with bias-corrected and  
29  
30 accelerated confidence intervals (Carpenter and Bithell 2000). The family-wise  
31  
32 error rate was controlled when necessary using the cluster correction in the  
33  
34 temporal and spatial domains (Maris and Oostenveld 2007).  
35  
36  
37  
38  
39  
40  
41

## 42 **Results**

### 43 **Behavioural results**

44  
45 The participants obtained high scores on the detection task (a white cross),  
46  
47 which ranged from 88% to 100% correct detection (median 99%) for the audio-  
48  
49 visual condition, and from 91% to 100% detection (median 99%) for the visual-  
50  
51 alone condition. The number of false alarms ranged from 0 to 2 for both  
52  
53 conditions. These results demonstrate that the participants’ attention was focused  
54  
55 on the videos during the recordings. Because there was a variable delay in the  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1 presentation of the visual target, we were confident that participants were attentive  
2 to the visually emphasised word ‘two’.  
3  
4  
5  
6  
7

## 8 **EEG results**

9

10 To search for a MMN response, we performed a contrast analysis by first  
11 comparing the EEG signal for the standard (NE) and deviant (WE) audio-visual  
12 conditions at the Fz electrode of interest, as is normally done for MMN analyses  
13 (Naatanen et al. 2007). The SNR in Fz was 1.7, which permits to distinguish ERPs  
14 even in the continuous EEG without averaging (Quian Quiroga and Garcia 2003).  
15  
16  
17  
18  
19  
20  
21

22 As we had no *a priori* hypothesis as to the timing of the negativity, we  
23 performed a temporal cluster correction of the t-values obtained from paired t-  
24 tests of the whole waveform; the correction controlled for the family-wise error  
25 rate in the temporal domain. Using this procedure, significant negativity was  
26 detected at 360-400 ms relative to the beginning of the epoch, which corresponds  
27 to 1060-1100 ms with respect to the onset of the video ( $p_{\text{corr}} < 0.05$  (10 000  
28 permutations), see Figure 1).  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

39 We computed the size of the effect using the Cohen’s test (Cohen 1992).  
40 This method, applied at the temporal window of 360-400 ms at the electrode Fz,  
41 resulted in d values of 1.46, which classified the size of the observed effect as  
42 large.  
43  
44  
45  
46  
47

48 At the Fz electrode, we did not observe a significant negativity after the onset of  
49 the auditory word “two,” even at the uncorrected level of significance. As shown  
50 in Figure 2, the cluster with negativity at 360-400 ms covers a set of the frontal,  
51 central, and temporal electrodes, and the significance of this spatial distribution of  
52 negativity was confirmed using the cluster correction in the spatial domain  
53 ( $p_{\text{corr}} < 0.05$  (10 000 permutations)).  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Concerning the between-subject stability of the effect at the Fz electrode, the negativity at 360-400 ms was found for all but one of the subjects. However, it was present for each subject at the Cz electrode, a location that is usually found to elicit an auditory MMN, including previous audio-visual speech MMN studies (Kislyuk et al. 2008).

As the detected negativity occurred at about 200 ms before the beginning of the auditory “two,” it is probably related to the prosodic visual gestures. In an attempt to relate this MMN-like response to the visual cues, we performed a frame-by-frame analysis of the differences (in pixels, see Figure 1) between the NE standard and WE deviant video clips. This analysis showed differences in the size of the mouth opening (mean  $1.5 \pm 0.8$  pixels), a feature known to be related to the spectral structure of voice sounds (see (Chandrasekaran et al. 2009)). But the largest variations concerned the eyebrow movements, which were found to present a mean amplitude difference between the NE and WE conditions of 2.7 ( $\pm 1.5$  pixels) with bimodal temporal dynamics. It is important to note that these eyebrow movements began about 500 ms before the related sound “two.” As can be seen from the frame-by-frame timing in Figure 1, the MMN-like response occurred immediately after the initiation of a large eyebrow lift in the emphasized video clip. This timing strongly suggests a causal link between the MMN-like response and the facial prosodic gestures.

However, this MMN-like activity could be a visual MMN, as has been demonstrated in some McGurk studies (see (Saint-Amour et al. 2007)). To explore this possibility, we analyzed the ERPs during the visual-only runs in which the visual deviant and standard were the same as in the AV conditions. The 2x2 repeated measures ANOVA at 360-400 ms revealed a significant interaction between the modality of stimulation and the type of stimulation (standard and

1 deviant):  $F(1,9)=7.2$ ,  $p<0.05$ , Power of the effect 0.7. No significant negativity  
2 was detected for the emphasized relative to the even speech in the visual modality  
3  
4 at any time period for the Fz and Cz electrodes of interest, even at the uncorrected  
5  
6 level of significance (Figure 3A).  
7

8  
9 A bootstrap analysis showed no significant effect for the visual-only  
10 condition ( $p>0.05$ ) at central, frontal, and temporal groups of electrodes for the  
11  
12 360-400 ms time window, where negativity had been observed in the audio-visual  
13  
14 condition. However, at the occipital pole (electrode Oz), a slight difference was  
15  
16 observed at the uncorrected level ( $p>0.05$ ) between the deviant and standard ERP  
17  
18 signals at a later stage (440 ms, Figure 3B). The CSD analysis revealed the  
19  
20 occipito-parietal distribution of the sources for this late negativity (Figure 3C).  
21  
22  
23  
24

25  
26 As a supplementary part of the analysis, to further delineate the  
27  
28 audiovisual nature of the observed response, we subtracted the ERP signal of the  
29  
30 visual condition from the audiovisual one, prior to contrasting the standard and  
31  
32 deviant conditions for each subject (AV-V contrast, see (Saint-Amour et al.  
33  
34 2007)). A bootstrap test was used to compare the differences between the standard  
35  
36 and deviant waveforms in a window that covered the 360-400 ms time window of  
37  
38 interest (see Fig 2 B). This analysis revealed the presence of negativity at the Fz  
39  
40 electrode ( $p<0.05$ ), thus confirming that the MMN-like response was specific to  
41  
42 the audio-visual condition.  
43  
44  
45  
46  
47

48  
49 The spatio-temporal pattern of the MMN-like response in the audio-visual  
50  
51 condition is shown in the the right panel of Figure 2A. This statistical cluster plot  
52  
53 (shown at the uncorrected level of  $p<0.05$ ) reveals a clearly organized pattern of  
54  
55 statistically significant negativity across the electrodes, which was confirmed by a  
56  
57 cluster correction in the temporal and spatial domains. It can be seen that the  
58  
59  
60  
61  
62  
63  
64  
65

1 MMN-like response was more pronounced at the central and frontal electrode  
2 sites and had no significant left-right asymmetries ( $p>0.05$ , bootstrap test).  
3

4 The spatial dynamics of the negativity propagation are presented in Figure  
5 2C. The earliest part (320-360ms) provides evidence that the MMN-like response  
6 originates in the posterior parieto-occipital visual cortex ( $p<0.05$ , cluster  
7 correction in the temporal and spatial domains with 10 000 permutations). Later,  
8 between 360-400ms, corresponding to the peak of the MMN-like response, it  
9 propagates forwards to the auditory temporo-frontal regions, before ending in a  
10 backward propagation towards the occipital region at 400-440ms ( $p<0.05$ , cluster  
11 correction in the temporal and spatial domains with 10 000 permutations). In the  
12 CSD analysis, the sources in the occipital regions confirm the early occipital  
13 negativities, but for the main negativity effect at 360-400 ms, the large central  
14 negativity corresponds to the sources in the left parietal region, in the left  
15 posterior temporal region, and in the inferior frontal region (Figure 2C). Though  
16 the tendency to the left side lateralization was not significant in the classical  
17 analysis, it became more pronounced in the CSD analysis. This spatio-temporal  
18 source analysis confirms the multisensory visuo-auditory nature of the negativity,  
19 which was only observed in the audiovisual condition.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 It has recently been suggested that low-pass filtering below 40 Hz may  
44 induce a timing uncertainty in ERPs (Vanrullen 2011). However, another study  
45 indicates that filtering might be safe if the cut-off frequencies are far from the  
46 frequencies of the effects (Rousselet 2012). The negativity in our study has a  
47 frequency of about 11 Hz, which is within the chosen band-pass filter of 2-30 Hz.  
48  
49  
50  
51  
52  
53  
54

55 No gender effect on the negativity was detected ( $p>0.5$ ).  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Discussion

1  
2  
3  
4  
5           Speech processing represents what is probably the most striking example  
6  
7 of multisensory interactions, in which complementary information from lip  
8  
9 movements and sound signals are merged into coherent percepts for phonetic  
10  
11 perception. Our present results provide new evidence that facial movements play  
12  
13 an important role in continuous speech processing and can be used to predict  
14  
15 certain features of the auditory utterance. Using an oddball audio-visual paradigm  
16  
17 with the same sound across standards and deviants, we detected a significant  
18  
19 negativity after the start of the emphasizing visual movements but, surprisingly,  
20  
21 before the auditory stimulus. This negativity was confirmed by the statistical  
22  
23 comparisons of the audio-visual and visual stimulations. No such negativity was  
24  
25 present in the control visual-only condition, in spite of the same emphasizing  
26  
27 movements. Thus, the observed brain response is specific to the auditory  
28  
29 expectation in a audio-visual condition; in the multisensory context, the visual  
30  
31 modality evokes facilitating predictive coding for the auditory speech. These  
32  
33 results provide the first demonstration at the neuronal level of predictive  
34  
35 relationships between prosody-related facial movements, such as eyebrow  
36  
37 movements and speech acoustic signals.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

### **A MMN-like response induced by eyebrow movements**

48  
49           The MMN in oddball protocols corresponds to the automatic brain  
50  
51 responses evoked by changes in repeated auditory stimuli. This negativity has  
52  
53 been observed in response to changes across a large number of different acoustic  
54  
55 features, such as intensity, frequency, and duration (review in (Näätänen et al.  
56  
57 2007)). Several MMN studies have used incongruent situations such as the  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

McGurk effect to study audio-visual interactions at the level of the auditory cortex. MMN-like responses have been reported for natural ecological stimuli (such as a hammer hitting a nail(Ullsperger et al. 2006)), and for emotional face/voice interactions (de Gelder et al. 1999), but they are especially prominent for visuo-auditory speech sounds where the deviance is limited to the visual stimulus (Colin et al. 2002; Saint-Amour et al. 2007; Ponton et al. 2009; Sams et al. 1991). In this latter case it has been shown that the McGurk effect elicits a significant MMN-like waveform at Fz, despite the auditory stimulus in the deviants being the same as in the standards. Source analysis has located the generators of this response to the auditory cortex (Saint-Amour et al. 2007; Mottonen et al. 2002), confirming the modulating influence of visual information at the early stages of auditory processing, a concept that is now widely accepted for multisensory interactions (Cappe et al. 2010; Schroeder and Foxe 2005). Our present results support there being a visual influence on auditory speech processing, and they clarify the major features of these audio-visual interactions. In addition, whereas in all previous reports the MMN response occurred at around 150 to 250 ms after the auditory onset, the MMN-like response precedes the auditory stimulation in our study.

Prosodic visual gestures have been previously shown to be closely related to sound modulations during speech production (Guaitella et al. 2009; Hadar et al. 1984; Munhall et al. 2004), and several studies have revealed a correlation between the acoustic and visual features of speech (Scarborough et al. 2009; Vatikiotis-Bateson and Yehia 2002; Barker and Berthommier 1999; Jiang et al. 2002); for example, between the fundamental frequency of the voice and the speaker's eyebrow movements (Cavé et al. 1996). In addition, the visual features of speech prosody have been shown to exert a cross-modal facilitation on auditory

1 thresholds for intensity changes (Foxton et al. 2010). It is probable that the  
2 significant negativity we observed in the present study is involved in evoking the  
3 illusory percept of emphasis on the auditory word “two,” as was reported by the  
4 subjects. Importantly, we did not observe a MMN-like negativity at the Fz or Cz  
5 electrodes when contrasting the standard and deviant conditions in the visual-only  
6 condition. Only an ancillary difference was observed at later stages of visual  
7 processing. Furthermore, when the visual signals were subtracted from the audio-  
8 visual ones, the MMN-like response remained robust. These results rule out the  
9 possibility that the MMN-like response observed in the audio-visual condition  
10 was induced by the visual mismatch alone. Rather, it shows that the MMN  
11 resulted from the visual features of emphasis being used to predict a change in the  
12 auditory utterance, such as in the pitch or intensity of the stressed word.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

29 In previous MMN studies using the McGurk effect, audio-visual reactions  
30 were not found to precede the auditory stimulus (Colin et al. 2002; Saint-Amour  
31 et al. 2007; Ponton et al. 2009; Sams et al. 1991) as in the present study. In these  
32 studies, an auditory MMN was reported at short latencies (~ 175ms, see (Colin et  
33 al. 2002; Saint-Amour et al. 2007; Ponton et al. 2009; Sams et al. 1991)) after the  
34 acoustic phoneme, but never before. This absence of a preceding MMN signal  
35 may be explained by the fact that the visual stimuli, such as the mouth opening,  
36 were very close in time to the auditory information (Chandrasekaran et al. 2009)  
37 so that the brain reaction to them would overlap with the auditory stimulation. In  
38 our study, we have been able to show the brain response to the visual deviance,  
39 because the facial movements preceded the auditory information by more than  
40 400 ms. The absence of an illusory “auditory” MMN in our study could be  
41 explained by a limit in the sensitivity of the MMN approach. Our previous  
42 behavioural finding(Foxton et al. 2010) showed that visual cues increase the  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 auditory threshold to prosody only by a few dB. In classical oddball protocols  
2 based on auditory intensity changes(Naatanen et al. 2007), MMN can be observed  
3  
4 with intensity differences of more than 3 dB, corresponding to at least 10%  
5  
6 between the deviant and standard stimuli.  
7

8  
9 It is possible that the observed negativity in the audio-visual condition may  
10 also reflect an electrophysiological N2b response as part of the orienting  
11 complex(Halgren et al. 1995), overlapping with MMN-like activity. An  
12 interesting perspective for further studies of audio-visual continuous speech  
13 would be to distinguish between the pre-attentive MMN-like audio-visual  
14 mismatch and the orienting complex.  
15  
16  
17  
18  
19  
20  
21  
22  
23

24 Our results provide strong evidence for the involvement of audio-visual  
25 networks during the audio-visual condition. First, the observed negativity was not  
26 restricted to the occipital sites as in the studies using the visual oddball paradigm  
27 ((Campanella et al. 2002; e.g., Tales et al. 1999; see Kimura in press for the recent  
28 review)for the recent review), but rather had a fronto-central distribution as has  
29 been found in the auditory MMN studies (Naatanen et al. 2007). Secondly, the  
30 CSD analysis confirms the implication of the posterior temporal regions, which  
31 are known to be involved in audio-visual integration during speech processing.  
32 This cortical region is involved, for example, in semantic decision under cross-  
33 modal influence (Kang et al. 2006) and in visually-based deciphering of  
34 ambiguous auditory phonemes (Kilian-Hutten et al. 2011); Kilian-Hutten et al.  
35 (2011) proposed that the posterior temporal region is involved in predicting  
36 forthcoming auditory phonemes on the basis of the visual information, which  
37 corresponds closely to the present study. Lastly, the absence of negativity at the  
38 same latency for the visual-only deviance, further supports the involvement of  
39 audio-visual networks.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

The occipital negativity at about 480 ms in the visual-only condition may be related to the N400 component, which is known to be elicited in response to visual incongruence (Proverbio and Riva 2009), which is in our case an incongruence with respect to visual expectations.

### **A predictive role of visual prosodic information during speech processing**

Using the incongruent audio-visual prosodic features, we observed a pre-auditory MMN-like activity which is not present during the same visual stimuli in silence. We hypothesize that this response anticipates the auditory changes that follow the visual emphasis. This hypothesis is consistent with the general predictive role of visual information when processing auditory speech (van Wassenhove et al. 2005; Arnal et al. 2009), simply because there is a systematic delay between the visual gestures and the sound. This predictive and facilitative role has been demonstrated using behavioral tests, where information from the lip movement allows participants to distinguish between phonemes (e.g. [gy gu dy du] vs. [ty tu ky ku] (Schwartz et al. 2004)). At the neuronal level, electrophysiological studies have clearly demonstrated that the auditory cortex is sensitive to visual information, with activations found during speech reading (Calvert et al. 1997), and with visually-induced modifications of auditory responses (Besle et al. 2009; Davis et al. 2008; Reale et al. 2007). These multisensory interactions at early stages of auditory processing have an important facilitatory role in preparing the auditory system for optimal responses (Lakatos et al. 2007). For example, early audio-visual interactions for syllables are manifested as a latency shortening of the N1/P2 responses, which relates to the salience of the visual input(van Wassenhove et al. 2005). This suggests that the visual input

1 carries a specific predictive value for the auditory utterance. Furthermore, the  
2 visual facilitation of auditory responses is positively correlated with the predictive  
3 proficiency of the lip movements (van Wassenhove et al. 2005; Arnal et al. 2009).  
4 Altogether, this demonstrates the clear impact of visual speech information on  
5 neuronal auditory responses. Our results suggest that the visual prosodic cues give  
6 rise to similar effects on auditory brain activations, but on a larger time-scale.  
7

8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

During auditory perception, it has been demonstrated that brain reactions to prosodic violations depend on the expectations derived from the general context, including the pitch contour and linguistic expectations (Colombo et al. 2011). In our study, the MMN-like response is dependent on the multisensory context, as it is only present during the audio-visual blocks. We attribute this to the predictive role of the prosodic visual cues within the context of audio-visual speech. These results are consistent with context-dependent audio-visual interactions that have been observed in ERP studies using non-speech protocols (Stekelenburg and Vroomen 2007, 2009). Specifically, it has been found that if the preceding visual stimuli do not provide anticipatory information, the auditory N1 component is not affected.

Current theories claim that the brain generates predictions about the sensory environment, which are then compared to the actual incoming signal (Friston and Kiebel 2009; Strelnikov 2010, 2007). In the case of speech processing, lip movements provide predictive information which facilitates auditory processing, and this involves a specific network linking the visual and auditory areas (Arnal et al. 2009; Arnal et al. 2011). For syllable perception, there is functional connectivity between the visual motion and auditory areas, which relates to the degree of visual predictability. It has been suggested that a fast direct cortico-cortical pathway conveys visual motion parameters to the auditory cortex,

1 and that a slower indirect feedback pathway signals the error between the visual  
2 prediction and the auditory input (Arnal et al. 2009). There is evidence that the  
3 comparison of the signals involves the superior temporal sulcus ((Arnal et al.  
4 2009), see also (Ghazanfar et al. 2008)). Our analysis of the MMN-like  
5 component in continuous speech revealed a similar cortical network, with the  
6 negativity originating in the occipito-parietal electrodes, propagating forward to  
7 the fronto-temporal sites and then returning back to the occipital region. Similar  
8 occipito-temporo-frontal loops have previously been shown for lip-reading  
9 syllables in silence (Arnal et al. 2009). This indicates that the visual information  
10 gives rise to auditory predictions in a fronto-temporal network, which are then  
11 compared to the next incoming auditory information. One might consider that the  
12 absence of a purely auditory MMN in our study could be due to the absence of an  
13 auditory mismatch as detected by the temporal region (Arnal et al. 2009; Arnal et  
14 al. 2011).

## 37 **Conclusions**

42 Speech comprehension constitutes what is probably a unique process of  
43 multisensory integration. In addition to the redundant information carried by  
44 visual lip movements, which are crucial in degraded auditory situations, the visual  
45 cues play a predictive role for the acoustic signals. Here we furthered our  
46 understanding of the facilitative role of visual information during continuous  
47 audio-visual speech processing. The MMN-like activity we observed suggests that  
48 facial movements, which convey the prosodic stress of a word in continuous  
49 speech, act as a predictive representation for the forthcoming auditory inputs.

1 These mechanisms of audio-visual interactions may be particularly important in  
2 patients with auditory deficits preventing prosody perception. Indeed, cross-modal  
3 prosody processing can not only enhance the comprehension of speech in  
4 general(Munhall et al. 2004), but could also compensate for the reduced pitch  
5 perception as in cochlear implanted deaf patients (Chatterjee and Peng 2008;  
6 Donnelly et al. 2009).

#### 17 Acknowledgements

18 We thank E. Barbeau for help in the first pilot study and C. Marlot for the bibliography. This study  
19 was supported by the Human Frontier Science Program (to JMF), the DRCI Toulouse (Direction  
20 de la Recherche Clinique et de l'Innovation to KS and MM), the ANR (ANR *Plasmody* ANR-11-  
21 BSHS2-0008 (to BP), and the recurrent funding of the CNRS (to BP).

References

- 1  
2  
3  
4 Arnal LH, Morillon B, Kell CA, Giraud AL (2009) Dual neural routing of visual  
5  
6 facilitation in speech processing. *J Neurosci* 29 (43):13445-13453  
7  
8  
9 Arnal LH, Wyart V, Giraud AL (2011) Transitions in neural oscillations reflect  
10  
11 prediction errors generated in audiovisual speech. *Nat Neurosci* 14 (6):797-801  
12  
13  
14 Barker J, Berthommier F Evidence of correlation between acoustic and visual  
15  
16 features of speech. In: Ohala JJ, Hasegawa Y, Ohala M, Granville D, Bailey AC  
17  
18 (eds) 14th International Congress of Phonetic Sciences, San Francisco, USA,  
19  
20 1999. The Congress organizers at the Linguistics Department, University of  
21  
22 California, Berkeley, pp 199-202  
23  
24  
25  
26 Barkhuysen P, Kraemer E, Swerts M (2008) The interplay between the auditory  
27  
28 and visual modality for end-of-utterance detection. *J Acoust Soc Am* 123 (1):354-  
29  
30 365  
31  
32  
33 Besle J, Bertrand O, Giard MH (2009) Electrophysiological (EEG, sEEG, MEG)  
34  
35 evidence for multiple audiovisual interactions in the human auditory cortex. *Hear*  
36  
37 *Res* 258 (1-2):143-151  
38  
39  
40 Besle J, Fischer C, Bidet-Caulet A, Lecaigard F, Bertrand O, Giard MH (2008)  
41  
42 Visual activation and audiovisual interactions in the auditory cortex during speech  
43  
44 perception: intracranial recordings in humans. *J Neurosci* 28 (52):14301-14310  
45  
46  
47 Besle J, Fort A, Delpuech C, Giard MH (2004) Bimodal speech: early suppressive  
48  
49 visual effects in human auditory cortex. *Eur J Neurosci* 20 (8):2225-2234  
50  
51  
52 Besle J, Fort A, Giard MH (2005) Is the auditory sensory memory sensitive to  
53  
54 visual information? *Exp Brain Res* 166 (3-4):337-344  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK,  
2 Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during  
3 silent lipreading. *Science* 276 (5312):593-596  
4  
5  
6  
7 Campanella S, Gaspard C, Debatisse D, Bruyer R, Crommelinck M, Guerit JM  
8  
9 (2002) Discrimination of emotional facial expressions in a visual oddball task: an  
10 ERP study. *Biological psychology* 59 (3):171-186  
11  
12  
13  
14 Campbell R (2008) The processing of audio-visual speech: empirical and neural  
15 bases. *Philos Trans R Soc Lond B Biol Sci* 363 (1493):1001-1010  
16  
17  
18  
19 Cappe C, Thut G, Romei V, Murray MM (2010) Auditory-Visual Multisensory  
20 Interactions in Humans: Timing, Topography, Directionality, and Sources. *J*  
21  
22 *Neurosci* 30 (38):12572-12580  
23  
24  
25  
26  
27 Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what?  
28 A practical guide for medical statisticians. *Statistics in medicine* 19 (9):1141-1164  
29  
30  
31 Cavé C, Guaitella I, Bertrand R, Santi S, Harlay F, Espesser R About the  
32 relationship between eyebrow movements and Fo variations. In: ICSLP,  
33 Philadelphia, 1996. pp 2175-2178  
34  
35  
36  
37  
38  
39 Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009)  
40 The natural statistics of audiovisual speech. *PLoS computational biology* 5  
41 (7):e1000436  
42  
43  
44  
45  
46 Chatterjee M, Peng SC (2008) Processing F0 with cochlear implants: Modulation  
47 frequency discrimination and speech intonation recognition. *Hear Res* 235 (1-  
48 2):143-156  
49  
50  
51  
52  
53  
54 Cohen J (1992) A power primer. *Psychol Bull* 112 (1):155-159  
55  
56  
57  
58 Colin C, Radeau M, Soquet A, Demolin D, Colin F, Deltenre P (2002) Mismatch  
59 negativity evoked by the McGurk-MacDonald effect: a phonetic representation  
60 within short-term memory. *Clin Neurophysiol* 113 (4):495-506  
61  
62  
63  
64  
65

1 Colombo L, Deguchi C, Boureux M, Sarlo M, Besson M (2011) Detection of  
2 pitch violations depends upon the familiarity of intonational contour of sentences.  
3  
4 Cortex 47 (5):557-568  
5  
6  
7 Davis C, Kislyuk D, Kim J, Sams M (2008) The effect of viewing speech on  
8 auditory speech processing is different in the left and right hemispheres. Brain  
9  
10 Res 1242:151-161  
11  
12  
13 de Gelder B, Bocker KB, Tuomainen J, Hensen M, Vroomen J (1999) The  
14 combined perception of emotion from voice and face: early interaction revealed  
15 by human electric brain responses. Neurosci Lett 260 (2):133-136  
16  
17  
18  
19 Donnelly PJ, Guo BZ, Limb CJ (2009) Perceptual fusion of polyphonic pitch in  
20 cochlear implant users. J Acoust Soc Am 126 (5):EL128-133  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Friston K, Kiebel S (2009) Predictive coding under the free-energy principle.  
Philos Trans R Soc Lond B Biol Sci 364 (1521):1211-1221

Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between  
the superior temporal sulcus and auditory cortex mediate dynamic face/voice  
integration in rhesus monkeys. J Neurosci 28 (17):4457-4469

Guaitella I, Santi S, Lagrue B, Cave C (2009) Are eyebrow movements linked to  
voice variations and turn-taking in dialogue? An experimental investigation. Lang  
Speech 52 (Pt 2-3):207-222

Hadar U, Steiner TJ, Rose FC (1984) Involvement of head movement in speech  
production and its implications for language pathology. Adv Neurol 42:247-261

Halgren E, Baudena P, Clarke JM, Heit G, Marinkovic K, Devaux B, Vignal JP,  
Biraben A (1995) Intracerebral potentials to rare target and distractor auditory and

visual stimuli. II. Medial, lateral and posterior temporal lobe.

Electroencephalography and clinical neurophysiology 94 (4):229-250

Hertrich I, Mathiak K, Lutzenberger W, Menning H, Ackermann H (2007)

Sequential audiovisual interactions during speech perception: a whole-head MEG study. *Neuropsychologia* 45 (6):1342-1354

Jiang J, Alwan A, Keating PA, Auer ET, Bernstein LE (2002) On the relationship between face movements, tongue movements and speech acoustics. *EURASIP Journal on Applied Signal Processing* 11:1174–1188

Kang E, Lee DS, Kang H, Hwang CH, Oh SH, Kim CS, Chung JK, Lee MC (2006) The neural correlates of cross-modal interaction in speech perception during a semantic decision task on sentences: a PET study. *Neuroimage* 32 (1):423-431. doi:S1053-8119(06)00169-8 [pii]

10.1016/j.neuroimage.2006.03.016

Kilian-Hutten N, Vroomen J, Formisano E (2011) Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. *Neuroimage* 57 (4):1601-1607. doi:S1053-8119(11)00551-9 [pii]

10.1016/j.neuroimage.2011.05.043

Kimura M (in press) Visual mismatch negativity and unintentional temporal-context-based prediction in vision. *Int J Psychophysiol*

Kislyuk DS, Mottonen R, Sams M (2008) Visual processing affects the neural basis of auditory discrimination. *J Cogn Neurosci* 20 (12):2175-2184

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53 (2):279-292

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164 (1):177-190



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264 (5588):746-748

Mottonen R, Krause CM, Tiippana K, Sams M (2002) Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res* 13 (3):417-425

Munhall KG, Jones JA, Callan DE, Kuratate T, Vatikiotis-Bateson E (2004) Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological science* 15 (2):133-137

Naatanen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* 118 (12):2544-2590

Ponton CW, Bernstein LE, Auer ET, Jr. (2009) Mismatch negativity with visual-only and audiovisual speech. *Brain Topogr* 21 (3-4):207-215

Proverbio AM, Riva F (2009) RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience letters* 459 (3):142-146

Quian Quiroga R, Garcia H (2003) Single-trial event-related potentials with wavelet denoising. *Clin Neurophysiol* 114 (2):376-390

Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF (2007) Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience* 145 (1):162-184

Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2007) Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex* 17 (5):1147-1153

Rousselet GA (2012) Does Filtering Preclude Us from Studying ERP Time-Courses? *Frontiers in psychology* 3:131

1 Saint-Amour D, De Sanctis P, Molholm S, Ritter W, Foxe JJ (2007) Seeing  
2 voices: High-density electrical mapping and source-analysis of the multisensory  
3 mismatch negativity evoked during the McGurk illusion. *Neuropsychologia* 45  
4 (3):587-597  
5  
6  
7  
8  
9 Sams M, Alho K, Naatanen R (1984) Short-term habituation and dishabituation of  
10 the mismatch negativity of the ERP. *Psychophysiology* 21 (4):434-441  
11  
12  
13 Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J  
14 (1991) Seeing speech: visual information from lip movements modifies activity in  
15 the human auditory cortex. *Neurosci Lett* 127 (1):141-145  
16  
17  
18  
19  
20 Scarborough R, Keating P, Mattys SL, Cho T, Alwan A (2009) Optical phonetics  
21 and visual perception of lexical and phrasal stress in English. *Lang Speech* 52 (Pt  
22 2-3):135-175  
23  
24  
25  
26  
27  
28 Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, 'unisensory'  
29 processing. *Curr Opin Neurobiol* 15 (4):454-458  
30  
31  
32  
33 Schwartz JL, Berthommier F, Savariaux C (2004) Seeing to hear better: evidence  
34 for early audio-visual interactions in speech identification. *Cognition* 93 (2):B69-  
35 78  
36  
37  
38  
39  
40 Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration  
41 of ecologically valid audiovisual events. *J Cogn Neurosci* 19 (12):1964-1973  
42  
43  
44 Stekelenburg JJ, Vroomen J (2009) Neural correlates of audiovisual motion  
45 capture. *Exp Brain Res* 198 (2-3):383-390  
46  
47  
48  
49  
50 Strelnikov K (2007) Can mismatch negativity be linked to synaptic processes? A  
51 glutamatergic approach to deviance detection. *Brain Cogn* 65 (3):244-251  
52  
53  
54  
55 Strelnikov K (2010) Neuroimaging and neuroenergetics: brain activations as  
56 information-driven reorganization of energy flows. *Brain Cogn* 72 (3):449-456  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise.  
2 J Acoust Soc Am 26 (2):212-215  
3  
4 Tales A, Newton P, Troscianko T, Butler S (1999) Mismatch negativity in the  
5 visual modality. Neuroreport 10 (16):3363-3367  
6  
7 Ullsperger P, Erdmann U, Freude G, Dehoff W (2006) When sound and picture  
8 do not fit: Mismatch negativity and sensory interaction. Int J Psychophysiol 59  
9 (1):3-7  
10  
11 van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the  
12 neural processing of auditory speech. Proc Natl Acad Sci U S A 102 (4):1181-  
13 1186  
14  
15 Vanrullen R (2011) Four common conceptual fallacies in mapping the time course  
16 of recognition. Frontiers in psychology 2:365  
17  
18 Vatikiotis-Bateson E, Yehia HC (2002) Speaking mode variability in multimodal  
19 speech production. IEEE transactions on neural networks / a publication of the  
20 IEEE Neural Networks Council 13 (4):894-899  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Figure Legends

**Figure 1. Experimental protocol and analysis.** The upper part of the figure shows the successive presentation of the video clip (duration 2520 ms) in which an actress says the words “one-two-three” in French. A 400ms silent interval separates each presentation of the video. Our EEG analysis is restricted to a period between 700 and 1500 ms (lower enlarged video clip) that isolates the word “two” containing the visual emphasis. For this portion of the video clip, we quantified the differences (in pixels) between the facial features for the even and the emphasized “two” (lower panel). For the mouth opening, a distance was calculated for the mouth contour at the middle vertical line. For the eyebrows, a distance was calculated from the middle of the eyebrows to the upper frame border. The graph shows that the main feature of emphasis relates to the position of the eyebrows which are higher for the emphasized “two”. The zero time point for the word “two” corresponds to 700 ms from the beginning of the video.

**Figure 2. MMN-like component for the visually-evoked auditory predictive coding.**

**A.** Event-related potentials for the word "two" presented either emphasized (red) or non-emphasized (green). The critical t-value corresponding to  $p=0.05$  is 2.3 ( $df=9$ ). The spatial distribution of the t-values at peak is presented. The zero time point for the word “two” corresponds to 700 ms from the beginning of the video. The black bar indicates the t-values, which survived the cluster correction in the temporal domain. The right panel displays a statistical cluster plot of the negativities observed in the ERPs. Color values indicate the time intervals where the difference “deviant vs. standard” is significant at  $p<0.05$  in the point-wise

1 paired t-test. Electrode positions are arranged from frontal to posterior regions.  
2 Only statistically significant differences are depicted. In panel **B**, the significance  
3 of the results from the “AV-V” contrast is presented per subject in the fronto-  
4 central electrodes compared between standards and deviants. As in the previous  
5 analysis (see panel A), we found a robust negativity that precedes the apparition  
6 of the auditory stimulus.  
7  
8  
9  
10  
11  
12

13 **C. Spatial-temporal pattern of the MMN-like component.** The negativity  
14 during audio-visual stimulation originates from the occipito-parietal electrodes,  
15 propagates to the fronto-temporal electrodes, and then returns back to the occipital  
16 region. This loop occurred after the visual deviance but before the auditory  
17 stimulation. In the CSD analysis, the propagation involves the left frontal and the  
18 left posterior temporal sites.  
19  
20  
21  
22  
23  
24  
25  
26  
27

28 The colour scale of t-values is the same for all parts of the figure.  
29  
30  
31  
32  
33  
34

35 **Figure 3. MMN-like component during the visual-only conditions.** At  
36 the Fz electrode, we did not detect a MMN-like component, as observed during  
37 the audio-visual condition (see fig 2). However, the statistical cluster plot reveals  
38 a late negativity, restricted to the occipital electrodes corresponding to the visual  
39 cortex. Conventions including the colour code as in Figure 2.  
40  
41  
42  
43  
44  
45

46 **A.** Event-related potentials for the visual word "two" presented either  
47 emphasized (red) or non-emphasized (green).  
48  
49  
50

51 **B.** Statistical cluster plot of the negativities observed in the ERPs. Colour  
52 values indicate the time intervals where the difference “deviant vs. standard” is  
53 significant at  $p < 0.05$  in the point-wise paired t-test.  
54  
55  
56  
57  
58

59 **C.** CSD analysis indicates sources in the occipito-parietal areas.  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Figure 1  
[Click here to download high resolution image](#)

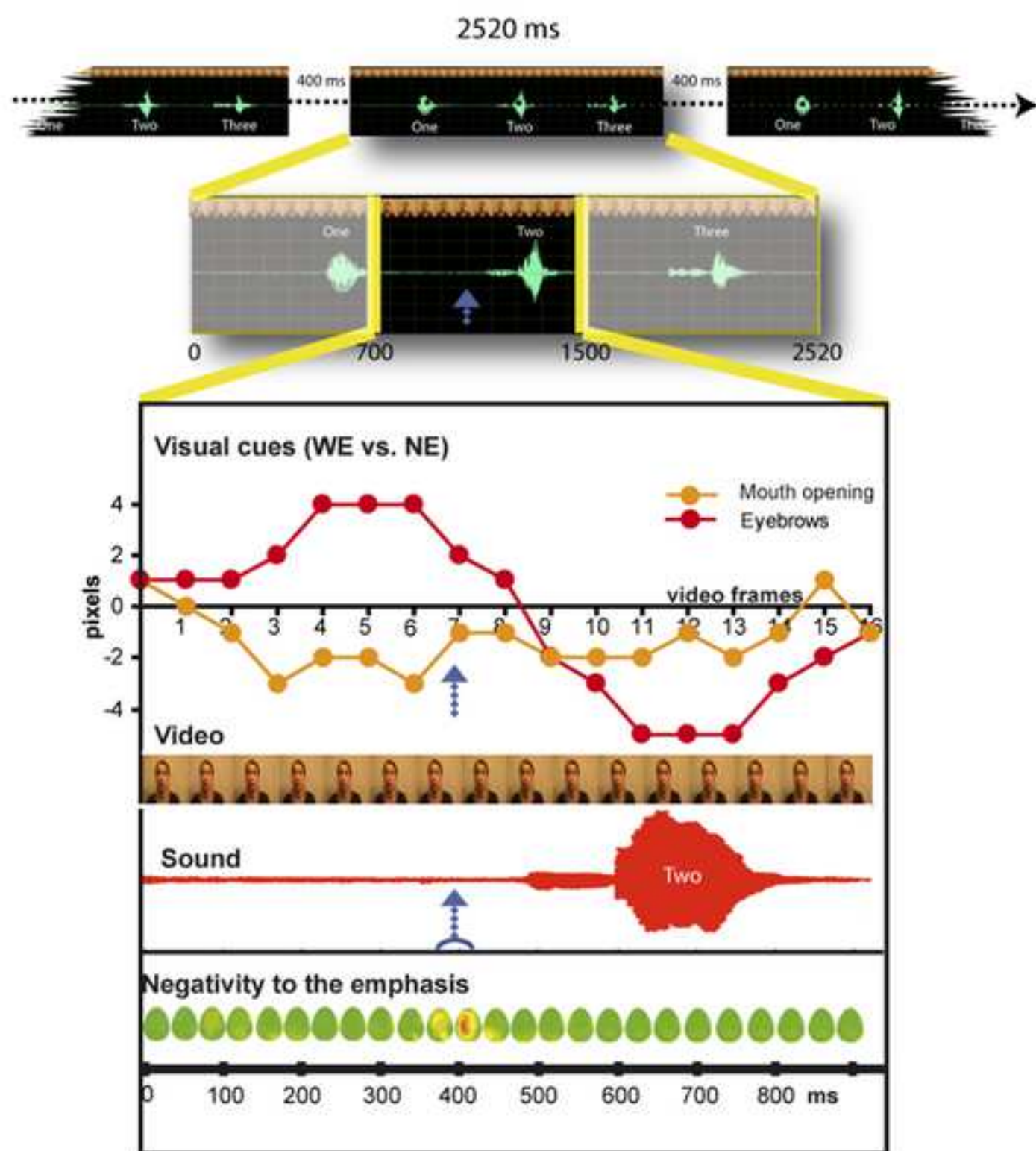
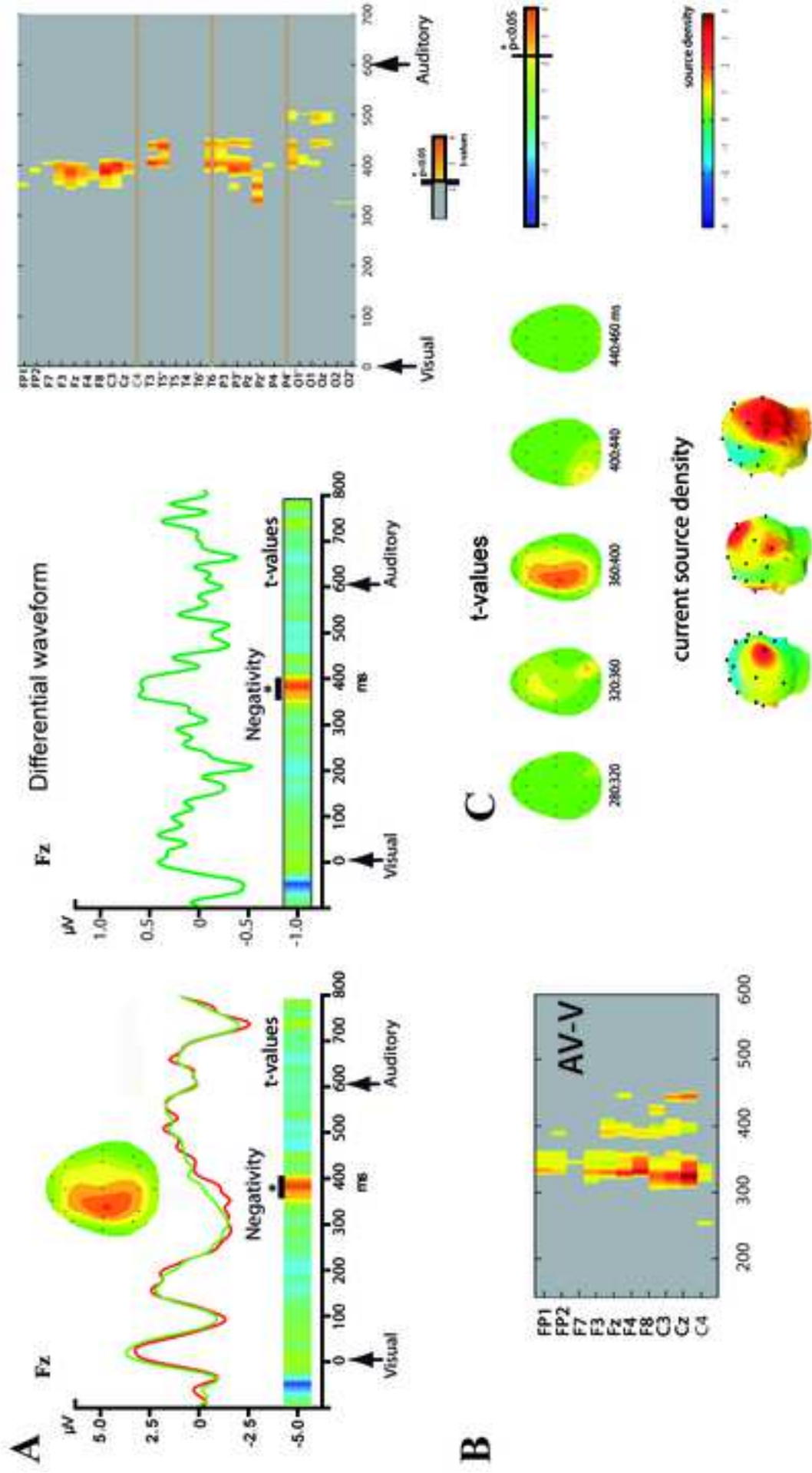


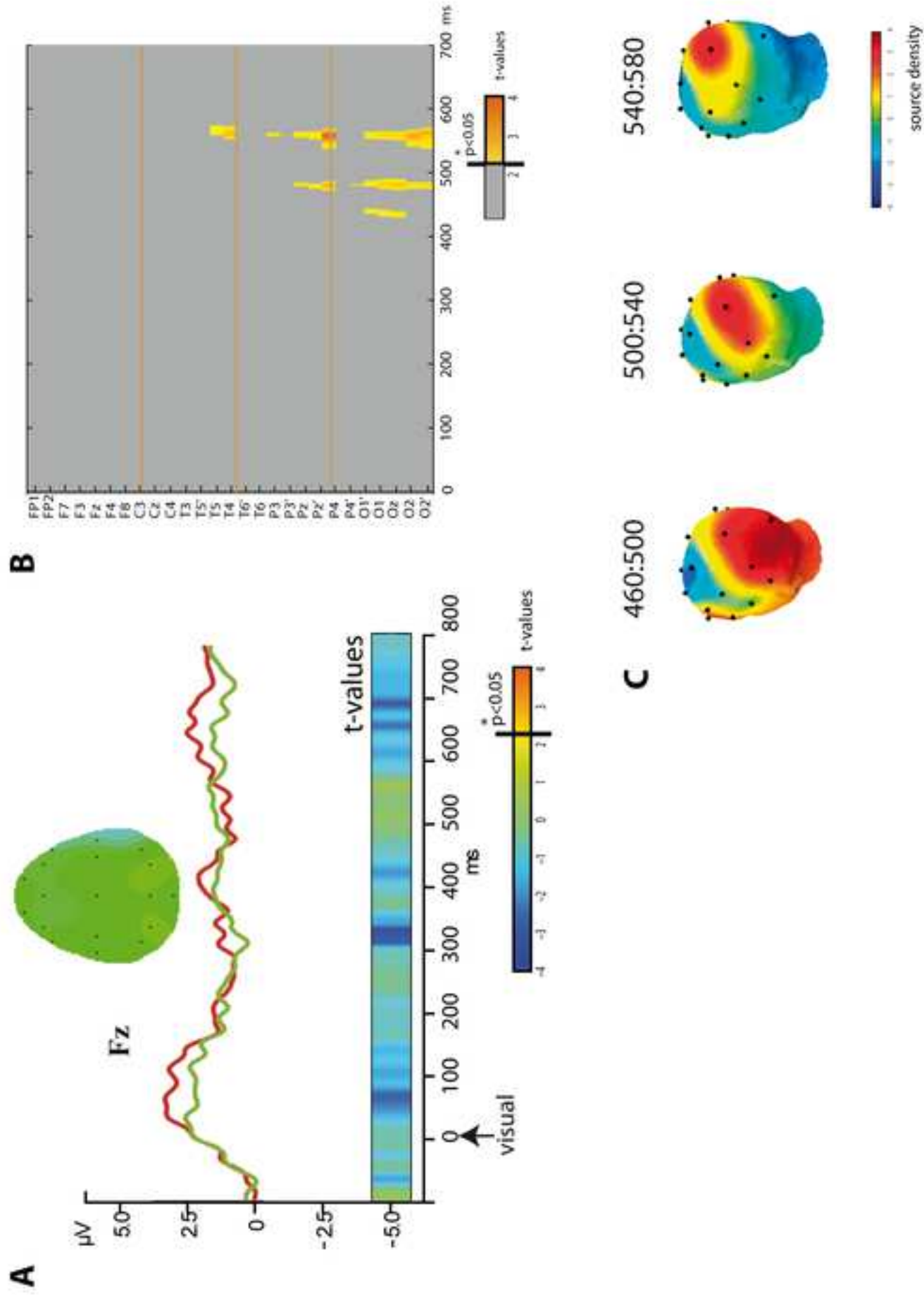
Figure 2  
[Click here to download high resolution image](#)

### Audio-visual speech





# Visual speech



## Résumé

Les bénéfices liés à l'implantation cochléaire sont connus pour la discrimination de la parole dans le silence. En revanche, leurs capacités perceptives pour les informations para linguistiques et extra linguistiques sont moins décrites. Nos travaux expérimentaux ont consisté à caractériser les éventuels déficits observés et expliquer leurs mécanismes pour la catégorisation auditive, pour la perception de la prosodie et de la musique. Nous avons dans ce but réalisé plusieurs études psychophysiques dans lesquelles leurs performances étaient comparées à celles de sujets contrôle normo entendant. Nos résultats font état d'un important déficit de perception auditive para et extra linguistique chez les sujets implantés cochléaires, vraisemblablement lié à deux principales raisons. La première réside dans la dégradation spectrale du signal acoustique par le processeur vocal de l'implant et l'insuffisance de restitution de l'information relative à la fréquence fondamentale. C'est pourquoi le déficit apparaît réduit voire absent chez les sujets bénéficiant d'une audition résiduelle associée à l'implant. La deuxième raison tient à la réorganisation corticale suivant la période de surdité, qui facilite le traitement de la parole mais pourrait se révéler délétère pour la perception des autres informations auditives.

## Perception of para and extra linguistic information by post lingually deafened cochlear implanted subjects

Cochlear implants have been shown to restore excellent speech recognition in quiet. However, post lingually deafened adults experience persistent handicap following cochlear implantation, which might be related to their perception abilities in other auditory fields than speech. In this report, we conducted several psychophysical experiments, which aimed at assessing the characteristics of their auditory categorization and their abilities for prosody and music perception. Normal hearing subjects were also tested in a control group. We found a strong and durable deficit, underpinned by two plausible mechanisms. First, the acoustic signal processing through the implant leads to an important spectral impoverishment and limits access to fundamental frequency information. Hence some cochlear implant recipients with substantial low-frequency residual hearing may achieve near normal performance. Brain plasticity following auditory deprivation facilitates the compensatory strategies for speech recognition but might also influence negatively the outcomes in para linguistic and extra linguistic information.