

« Étude de l'encodage des sons de parole par le tronc cérébral dans le bruit »

Par Céline RICHARD

Thèse de doctorat en Sciences cognitives

Sous la direction de Hung THAÏ-VAN

Présentée et soutenue publiquement le 17 décembre 2010

Membres du jury : Hung THAÏ-VAN, Professeur des Univ - Praticien Hosp., Université Lyon 1 Paul AVAN, Professeur des Univ - Praticien Hosp., Université Clermont-Ferrand 1 Annie MOULIN, Chargée de recherche, Université Lyon 1 Christian MARTIN, Professeur des Univ - Praticien Hosp., Université de Saint-Etienne Barbara TILLMANN, Directrice de recherche, Université Lyon 1 Olivier STERKERS, Professeur des Univ - Praticien Hosp., Université Paris 7

Table des matières

Contrat de diffusion	5
[Épigraphie]	6
Remerciements	7
Introduction générale	8
I. Du son de parole à son encodage neurophysiologique	10
1. Rappels anatomiques et physiologiques des voies auditives	10
1.1 La transduction périphérique	10
1.2. De la cochlée jusqu'au cortex :	13
1.3. Du cortex à la cochlée: les voies auditives efférentes	16
2. Caractérisation d'un son de parole	18
2.1. Le niveau acoustique :	19
2.2. Le niveau phonétique :	21
2.3. Le niveau phonologique :	25
2.4. Les autres niveaux :	25
3. La catégorisation de la parole	26
4. Propriétés du système auditif :	27
4.1. L'encodage temporel et spectral des sons par le système auditif :	27
4.2. La latéralisation corticale :	28
II. Les potentiels évoqués auditifs:	31
2.1 Les différents types de potentiels évoqués auditifs (PEA)	31
2.1.1. Potentiels évoqués auditifs précoce (PEAP) :	31
2.1.2. Potentiels évoqués de latence moyenne (PEALM): de 20 à 70ms	31
2.1.3. Potentiels évoqués de latence tardive : jusqu'à 500 ms	32
2.2. Technique de recueil des PEAP :	32
III. Analyse des sons par le tronc cérébral :	34
3.1. Du clic au son de parole...	34
3.2. Les potentiels évoqués auditifs en réponse à un son de parole:	35
3.2.1. Méthode d'acquisition :	35
3.2.2. Qu'est-ce qu'une réponse-type « PEA en réponse à un son de parole »?	36
3.2.3. Potentiels évoqués auditifs en réponse à un son de parole : artéfact ou réponse neurophysiologique ?	38
3.3. Où est générée la réponse ?	39
3.4. Intérêts neurophysiologiques des potentiels évoqués auditifs en réponse à un son de parole:	40
3.4.1. Encodage spatial et temporel de l'information auditive	40
3.4.2. Applications cliniques potentielles des PEAS	42
3.5. Que représentent le « what » et le « where » au niveau du système auditif ?	42
IV. Partie expérimentale :	45
Expérience 1 : Objectivation du traitement sous cortical d'un son de parole par rapport à celui d'un son complexe	46
1.1. Introduction	47

1.2. Matériel et méthode	47
1.3. Résultats :	55
1.4. Discussion :	59
Expérience 2 : Etude de l'encodage des sons de parole dans différentes conditions de bruit :	61
2.1. Introduction :	61
2.2. Matériel et méthodes	61
2.3. Résultats	65
2.4. Discussion :	69
Expérience 3 : Mise au point technique :	71
3.1. Optimisation de la chaîne de mesure :	72
3.2. Techniques d'analyse des traces PEASP :	75
Expérience 4 : Etude de l'encodage, d'un son de parole détérioré (bruité) au niveau du tronc cérébral	87
4.1. Introduction	87
4.2. Matériel et méthodes	88
4.3. Résultats	90
4.3. Discussion	94
Expérience 5 : Etude de l'encodage du voisement d'un son de parole par le tronc cérébral	96
5.1. Matériel et Méthodes :	96
5.2. Résultats :	99
5.3. Discussion :	106
Expérience 6 : Influence de l'intensité de stimulation sur le recueil des réponses de type PEAPSP	108
6.1. Matériel et méthode :	110
6.2. Résultats :	110
6.3. Discussion	114
Discussion, conclusion et perspectives	115
V. Références	118
VI. Communications à des congrès nationaux avec actes	131
VII. Communications à des congrès nationaux	147
VIII. Publications acceptées et n préparation	148

Contrat de diffusion

Ce document est diffusé sous le contrat *Creative Commons* « **Paternité – pas d'utilisation commerciale - pas de modification** » : vous êtes libre de le reproduire, de le distribuer et de le communiquer au public à condition d'en mentionner le nom de l'auteur et de ne pas le modifier, le transformer, l'adapter ni l'utiliser à des fins commerciales.

[Épigraphie]

« *Entendre ne veut pas dire écouter car l'ouïe est un sens mais l'écoute est un art* ».

Remerciements

Au Professeur Hung THAI-VAN, je vous remercie d'être présent ce jour pour la présentation de ce travail.

Au Docteur Annie MOULIN, je vous remercie pour l'ensemble de votre aide, vos conseils prodigués avec soin, l'ensemble du temps que vous m'avez accordé et le gout pour la recherche que vous m'avez donné. Veuillez accepter ma plus profonde reconnaissance.

Au Professeur Paul AVAN, je vous remercie d'accepter de juger ce travail, c'est un honneur de vous compter dans ce jury.

Au Professeur Olivier STERKERS, je vous remercie pour la qualité de la formation dont j'ai bénéficié à vos côtés, ainsi que pour vos précieux conseils. Je vous remercie d'avoir accepté de juger ce travail.

Au Professeur Christian MARTIN, je vous remercie de l'enseignement prodigué pendant ces nombreuses années, je vous en suis profondément reconnaissante.

A Madame Barbara TILLMANN, je vous remercie d'avoir accepté d'apporter votre jugement scientifique, de qualité, sur ce projet.

Au Professeur Lionel COLLET, Je vous remercie de m'avoir accueillie au sein de votre service pour ce travail de thèse.

A Arnaud JEANVOINE, sans qui le recueil des Potentiels évoqués auditifs par sons de parole n'existerait pas au sein du laboratoire. Avec mes remerciements pour toute l'aide apportée tout au long de cette thèse, et l'immense patience dont il a fait preuve devant mes débuts en Matlab^o.

A Fabien SELDRAN, mon plus fidèle sujet...

A ma famille, présente malgré les épreuves, m'apportant un réconfort de chaque instant.

A mon Gaston d'amour....

Merci tout particulièrement à mes cousins et cousines qui furent les principaux sujets de mes expériences...

A mes amis de toujours, Catherine, Magali, Sophie, Nathalie & Thierry, Clément, Lucille & Amandine, Valérie, Julie, Anne & Vincent, Catherine & Pierre, Carole, et tous ceux que je n'ai pas cités, me soutenant et me donnant envie de continuer. Merci à tous.

Merci à mes amis du Labo C.

Merci à l'ensemble du personnel du pavillon 50B, à Charlotte & Brigitte.

Introduction générale

De prime abord, la compréhension de la parole nous semble un phénomène simple et fluide. En effet, nous sommes capables dans la vie de tous les jours, de décrypter des messages parlés, ceci sans aucune difficulté, bien que la parole soit rarement utilisée en conditions « idéales », sous-entendu, en l'absence totale de bruit. Au contraire, nous sommes quotidiennement amenés à comprendre la parole malgré la présence de bruits extérieurs et de parasites sonores qui altèrent la qualité du signal émis. Même si le signal acoustique de la parole est dégradé, le signal reste intelligible pour l'auditeur. Nous avons donc la capacité de normaliser les multiples formes que peuvent prendre un même message, cette normalisation dépendant toutefois du degré de détérioration du signal et de la nature des indices acoustiques altérés. L'ensemble des détails du processus de compréhension de la parole ne sont toutefois pas encore parfaitement décryptés et cela malgré un grand nombre d'études menées depuis les cinquante dernières années.

Les travaux présentés dans cette thèse s'intéressent à la compréhension et notamment au traitement sous cortical de la parole dans des conditions d'écoute dégradée, cela au moyen d'un outil objectif et non invasif : les potentiels évoqués auditifs en réponse à un son de parole (PEASP).

Le principe des PEASP se base sur l'enregistrement de signaux physiologiques au moyen d'une technique s'apparentant à un électroencéphalogramme, sur un dispositif électrophysiologique de type Centor de Racine Alvar®. Cet appareil est couramment utilisé dans les services cliniques d'exploration fonctionnelle otorhinolaryngologique afin d'effectuer quotidiennement les enregistrements des potentiels évoqués auditifs de latence précoce (PEAP) en réponse à des clics. Les potentiels évoqués auditifs en réponse à un son de parole (PEASP), comme leur nom l'indique, se basent sur le même principe que le recueil des PEAP « standards », toutefois le stimulus va être remplacé par un son de parole, le plus souvent une syllabe. Cette technique est utilisée par l'équipe américaine de Nina Kraus, avec l'utilisation de la syllabe /da/ (Russo et al, 2004 ; Johnson et al, 2005). L'étude au moyen des PEASP, de la latence des pics enregistrés en réponse à cette syllabe donne des indications sur l'état du système auditif du sujet, notamment chez des enfants présentant des troubles de l'apprentissage, chez lesquels le déficit peut être mis en évidence, ainsi que l'efficacité des techniques de rééducation (Wible et al, 2004 ; Banai et al, 2007). Ainsi, nous souhaitons recréer ce système de recueil de PEASP et l'adapter aux syllabes françaises, avant de pouvoir procéder à l'étude de l'encodage sous cortical des sons de parole en condition adverse, dans un premier temps chez le malentendant.

Comme nous le verrons, ce travail thèse s'articule autour de plusieurs questions :

1. Quels sont les indices acoustiques encodés au niveau sous cortical?
2. Quel est l'encodage d'un son de parole dans différentes conditions de bruit de types : cocktail party, bruit blanc et speech noise?
3. Est-il possible d'obtenir des résultats exploitables de manière individuelle par l'élimination de tout parasitage et la mise au point d'une technique d'analyse automatisable?
4. Quel est l'effet du déphasage sur l'encodage d'un son de parole au niveau du tronc cérébral ?

5. Comment le voisement est-il encodé au niveau sous cortical ?
6. Quel est l'effet de l'intensité de stimulation sur les caractéristiques des PEASP ?

Pour introduire et situer le cadre de ces travaux de recherche, nous allons tout d'abord, dans un chapitre théorique, exposer le contenu des différents champs d'étude nécessaire à notre problématique. Comme tout processus cognitif, la compréhension du langage fait intervenir des processus ascendants (informations sensorielles remontant de la périphérie vers le système cognitif) et des processus descendants (informations intégrées influençant les niveaux de traitement plus précoce).

Concernant les processus ascendants, nous rappellerons l'anatomie du système auditif humain, nous verrons les principales voies auditives ascendantes, mais également le Système Efférent Olivo-Cochléaire Médian, rétrocontrôle du cortex auditif primaire jusqu'à la cochlée. Puis nous nous attacherons à la description d'un son de parole, selon différents niveaux (non exclusifs) : acoustique, phonétique, phonologique, morphologique, syntaxique, sémantique et pragmatique. Nous reviendrons ensuite sur certaines propriétés du système auditif, que son la latéralité corticale ainsi que sa capacité à encoder les caractéristiques temporelles et spectrales des sons.

Cette partie sera suivie par la description des techniques d'analyses dont nous disposons pour l'étude objective du traitement sous cortical. Ces techniques sont ici représentées par les potentiels évoqués auditifs (PEA) de latence précoce, moyenne et tardive. Le chapitre suivant s'attachera à la description des différentes caractéristiques et intérêts des PEASP, outil principal de cette recherche.

Nous passerons ensuite à la partie expérimentale, où seront reprises point par point les différentes questions posées ci-dessus.

Enfin, nous discuterons les principaux résultats obtenus au cours de ce travail, avant de conclure sur les possibles applications des PEASP.

I. Du son de parole à son encodage neurophysiologique

Un son est une onde qui se propage dans un milieu gazeux, liquide ou solide, sous forme de variations de pression. Cette onde sonore est perçue grâce à notre système auditif avec lequel elle va interagir. De l'organe périphérique, qui a le rôle de traducteur de l'information acoustique en information électrique, aux centres auditifs, qui ont le rôle du traitement et de l'interprétation de l'information électrique, l'information auditive passe par de multiples relais qui traitent l'information dans plusieurs dimensions (dimension fréquentielle, dimension temporelle) et qui interagissent, non seulement dans le sens afférent (fournissant l'information auditive aux relais supérieurs), mais également dans le sens efférent. En effet, les centres supérieurs pouvant exercer un rétrocontrôle sur tous les relais plus périphériques, à l'aide d'un système efférent auditif composé de multiples boucles, dont une permet de moduler directement l'information au sein de l'organe périphérique de l'audition (Suga, 2008).

1. Rappels anatomiques et physiologiques des voies auditives

1.1 La transduction périphérique

Le signal sonore, va tout d'abord être amplifié (de 5 à 20 dB pour les fréquences de 1.5 à 7 kHz) et modulé par les variations de direction que lui imprime la morphologie du pavillon de l'oreille externe, avant d'être transmis à travers l'oreille moyenne (membrane tympanique et osselets) jusqu'à l'oreille interne. Le rôle de l'oreille moyenne est l'adaptation d'impédance mécanique entre l'oreille externe, située en milieu aérien et l'oreille interne, afin de transmettre avec un bon rendement l'énergie vibratoire de l'air aux liquides de l'oreille interne (Fig 1).

Au niveau de l'oreille interne (organe de Corti), le son va être converti, d'un stimulus vibratoire en une forme interprétable par le système nerveux, ce phénomène est appelé la « transduction périphérique du son ». L'organe de Corti (Fig 2) constitue la partie neurosensorielle de la cochlée, il repose sur la membrane basilaire et contient les cellules ciliées internes et externes. Les cellules ciliées externes, environ trois fois plus nombreuses que les cellules ciliées internes, sont impliquées dans un mécanisme d'amplification actif ayant pour fonction d'amplifier les vibrations détectées par les cellules ciliées internes. Les cellules ciliées internes, reliées principalement à des fibres nerveuses afférentes, ont pour rôle la transduction du signal vibratoire en un signal électrique, transmis aux relais et voies centrales. Les cellules ciliées internes constituent des récepteurs sensibles aux mouvements liquidiens de l'oreille interne qu'elles transmettent sous forme de décharges nerveuses aux neurones bipolaires dont les axones forment le nerf auditif.

Une propriété de base de l'organe de Corti est qu'il effectue une analyse en fréquence des stimulations sonores. En réponse à une fréquence donnée, les différentes portions de l'organe de Corti, le long de la cochlée, vibrent avec des amplitudes différentes. Les régions basales répondent aux fréquences élevées, tandis que les régions proches de l'apex répondent aux basses fréquences : on parle de l'organisation tonotopique de la cochlée.

De manière historique, Helmholtz en 1885, inspiré des travaux de Fourier, formule l'hypothèse selon laquelle le codage des sons suivrait une théorie dite de « la place ». Ainsi, lorsqu'un stimulus parvient à la cochlée, les parties de la membrane basilaire dont les fréquences correspondent aux composantes de ce stimulus vont commencer à vibrer et induire un message nerveux. Cependant, Bekesy, 1947 a démontré que cette partition n'existe pas et a proposé une évolution de la théorie de « la place » en théorie de « l'onde propagée ». La distance parcourue par l'onde propagée le long de la membrane basilaire dépendrait de la fréquence du son. En effet, la structure de la membrane basilaire se modifie le long de la partition cochléaire. Ainsi, les sons de haute fréquence entraînent un déplacement d'amplitude maximal à la partie basale de la membrane basilaire, en dépensant beaucoup d'énergie et l'onde ne se propage pas très loin. Pour ceux de basse fréquence, l'onde peut se déplacer de la base à l'apex de la membrane basilaire avant que toute l'énergie soit épuisée. Cette relation liant la fréquence du son et la localisation entre la base et l'apex du codage est nommée carte cochléaire fréquentielle par Liberman (1982). De par l'existence de ce gradient, la membrane basilaire et par extension la cochlée, sont dites organisées de manière tonotopique. Propriété essentielle du système auditif, l'organisation tonotopique se maintient à travers les différents noyaux relais jusqu'au cortex. Bekesy, en 1949 a également démontré que les cellules de la cochlée codent préférentiellement une certaine fréquence mais également les fréquences voisines de façon moins spécifique, selon une courbe d'accord. Ces courbes d'accord deviennent plus étroites lorsque la fréquence augmente. Cet « accord en fréquence » serait étroitement lié à l'organisation tonotopique de la cochlée et de toutes les voies auditives. Ce mécanisme de codage de la fréquence serait particulièrement exploité lorsque les sons présentés sont courts ou/et de haute fréquence.

Par contre, en présence de sons purs de fréquence relativement basse (<4 kHz), les neurones déchargent préférentiellement lors des pics de pression de l'onde excitatrice (selon le principe dit du « calage de phase »). L'existence d'une période réfractaire des neurones -durée d'une milliseconde environ succédant à une décharge et pendant laquelle toute décharge est impossible- rend la synchronisation incomplète mais on peut toutefois déduire la fréquence d'un son en observant la cadence de décharge des fibres. En effet, celles-ci déchargent à des instants qui correspondent tous à un nombre entier de périodes. Ce type de codage de la fréquence pourrait bien être dominant en présence de sons relativement longs et surtout à des fréquences inférieures à 4-5 kHz (Rose *et al*, 1968; Moore, 1973). Au-dessus de 5 kHz, les neurones ne parviennent plus à suivre la cadence de son excitateur.

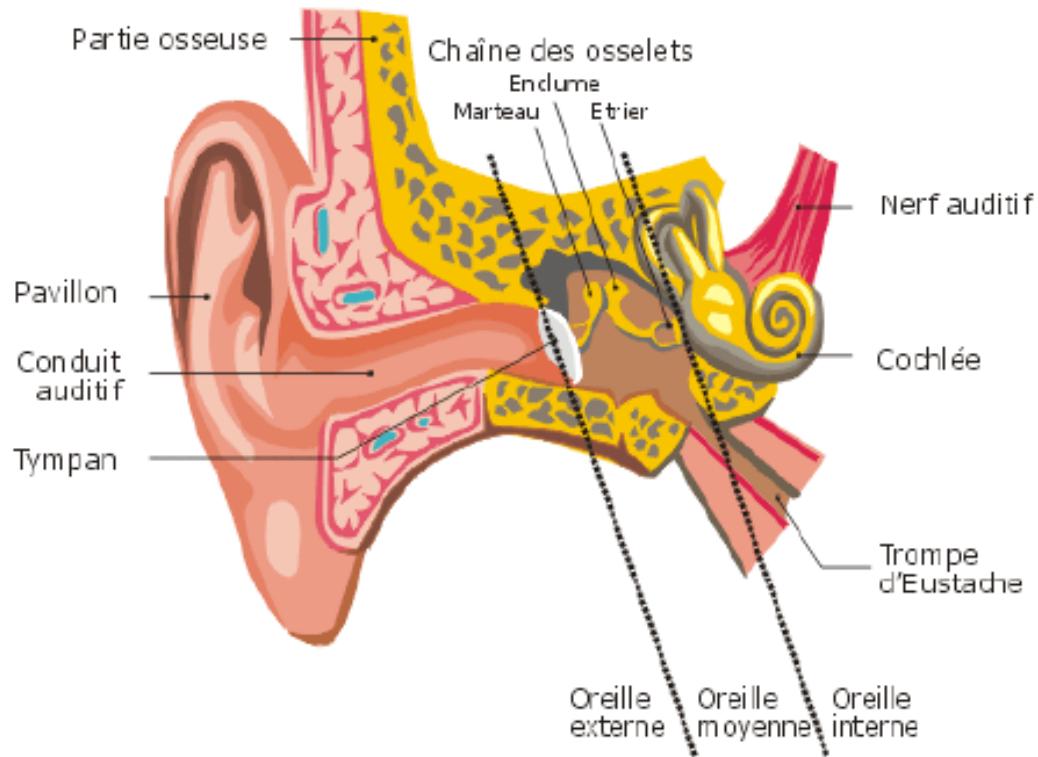


Figure 1 : Schéma de l'appareil auditif périphérique humain

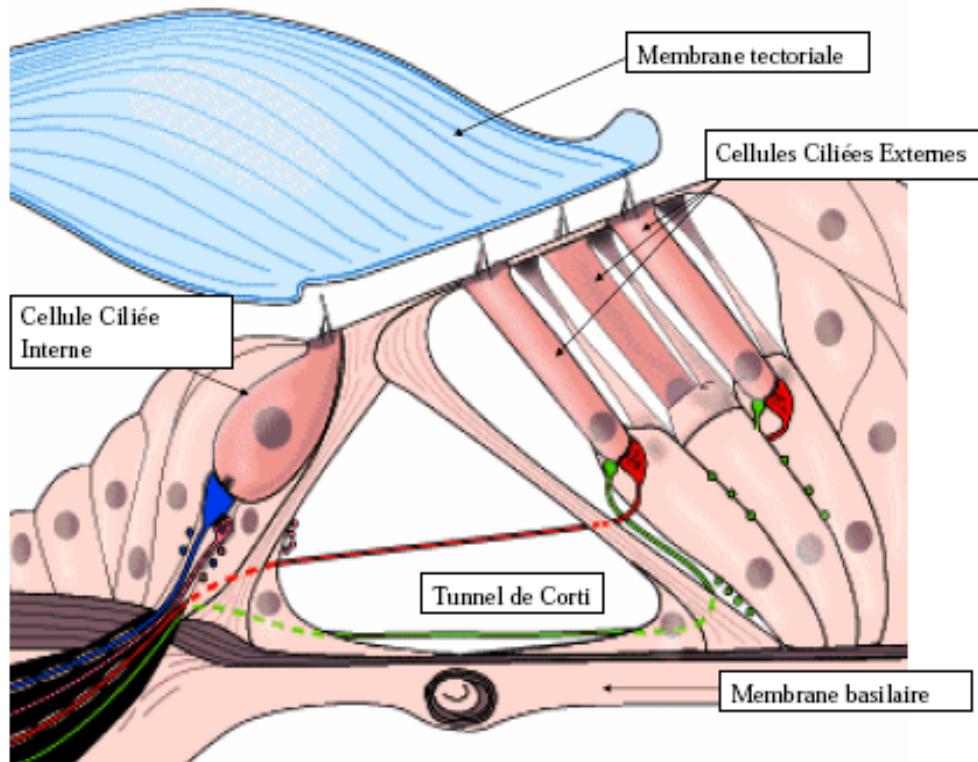


Figure 2 : L'organe de Corti.

Image créée par Stephan Blatrix, extraite de "promenade autour de la cochlée" par R. Pujol et coll., INSERM et Université Montpellier 1.

1.2. De la cochlée jusqu'au cortex :

Une des particularités du système auditif est que les fibres de l'oreille droite et de l'oreille gauche se projettent sur les deux hémisphères (ipsilatéral et controlatéral). Pour chaque oreille, deux faisceaux du nerf auditif se croisent au niveau du bulbe rachidien. Par conséquent, les informations sonores recueillies, par exemple par l'oreille droite, vont être traitées à la fois dans le cortex auditif gauche et dans le cortex droit, de même pour l'oreille gauche. Cependant, la projection est plus importante sur l'hémisphère controlatéral à l'oreille stimulée (Rosenzweig, 1951). De plus, le passage de l'information de la cochlée jusqu'aux centres auditifs se fait par l'intermédiaire de multiples structures (noyaux cochléaires, complexe olivaire supérieur, colliculus inférieur, corps genouillé médian...) qui ne se contentent pas d'un simple relais de l'information reçue, mais qui traitent l'information auditive et la modulent non seulement en fonction d'informations afférentes, mais également en fonction d'informations descendantes, provenant des structures plus centrales (Fig 3). Ces informations descendantes sont transmises grâce au système efférent auditif, qui est organisé sous forme de multiples boucles de rétroaction, et permet des rétroactions multiples entre les différents relais intermédiaires, et cela, du cortex jusqu'à l'organe le plus périphérique : la cochlée (Suga, 2008).

1.2.1. Le nerf auditif :

Le nerf auditif ou cochléaire est constitué d'environ 30000 neurones chez l'homme, ceux-ci ont leur corps cellulaire dans le ganglion spiral situé dans l'axe de la cochlée. Les neurones auditifs afférents ont leurs dendrites au contact avec les cellules ciliées et leurs axones se dirigent vers le noyau cochléaire. Il est important de noter que le nerf auditif contient aussi des faisceaux efférents, issus du complexe olivaire supérieur. Chaque fibre du nerf auditif produit une série d'impulsion dont les caractéristiques tiennent compte de la fréquence, de l'intensité et de la durée du stimulus. A chaque fibre auditive correspond sa courbe d'accord caractéristique : pour chaque fibre correspond une fréquence optimale, aussi appelée fréquence caractéristique déterminée par la position de la cellule ciliée qu'elle innervé. L'organisation tonotopique observée au niveau de la cochlée se poursuit donc le long du nerf auditif. Les fibres apicales, codant les basses fréquences, sont localisées dans la région centrale du nerf, alors que les fibres basales, codant les hautes fréquences sont plus localisées en périphérie (Kimura et al, 1964). En ce qui concerne le codage de l'intensité, celui-ci est codé, au sein d'une fibre, par le taux moyen par seconde de ses potentiels d'action. Différents types de fibres, différant entre elles par leur taux de décharges spontanées, codent pour différentes gammes d'intensité.

Par ailleurs, comme nous le verrons plus loin, les cellules du système auditif sont capables de se synchroniser sur la périodicité du stimulus acoustique, c'est le principe de « calage de phase » (ou phase-locking). Ce calage de phase se répercute tout le long des voies auditives afférentes en se limitant à des fréquences de plus en plus basses au fur et à mesure que le message nerveux est intégré : il intervient jusqu'à des fréquences de stimulation de 4 à 5 kHz au niveau du nerf auditif, et pour des fréquences inférieures à 80 Hz dans le cortex auditif (Palmer & Shamma).

1.2.2. Centres bulbo-protubérantiels :

1.2.2.1. Le noyau cochléaire :

C'est le premier relais auditif sur la voie ascendante et toutes les fibres afférentes du nerf auditif y font synapse. Le noyau cochléaire représente le premier relais sous-cortical, localisé sur la face dorso-latérale du tronc cérébral au niveau de la jonction du pont et de la medula. Il est le premier relai auditif sur la voie ascendante et toute l'information provenant de la cochlée doit obligatoirement traverser ce noyau. Deux régions principales composent le noyau cochléaire : le noyau cochléaire dorsal et le noyau cochléaire ventral, ce dernier étant lui-même subdivisé en noyau cochléaire antéro-ventral et noyau cochléaire postéro-ventral. Les fibres basales du nerf auditif vont pénétrer dans la partie dorsale du noyau cochléaire. La majorité des fibres apicales va aboutir dans le noyau cochléaire ventral, et quelques unes dans le noyau cochléaire dorsal. Ainsi, les fibres codant pour les basses fréquences se projettent dans les régions ventrales tandis que celles codant pour les hautes fréquences se projettent dans les régions dorsales, respectant ainsi la tonotopie cochléaire.

Le rôle principal du noyau cochléaire est de traiter les informations en provenance de la cochlée, de les modifier et enfin de les distribuer vers les centres nerveux supérieurs. Il existe trois voies de sortie, reliant le noyau cochléaire aux centres supérieurs : l'une vers les trois noyaux de l'olive supérieure (strie acoustique ventrale), la seconde vers les noyaux péri-olivaire (strie acoustique intermédiaire), et la dernière vers le lemnisque latéral et le colliculus inférieur controlatéral (strie acoustique dorsale dite de Monakow).

1.2.2.2. Le complexe de l'olive supérieure :

Constitué d'un ensemble de noyaux situés dans la région ventrale du tegmentum protubérantiel, il s'agit du deuxième grand relais des voies auditives centrales. Il reçoit les afférences directes des noyaux cochléaires ventraux et est composé de l'olive supérieure médiane (olive accessoire) et de l'olive supérieure latérale (olive principale). Le complexe olivaire supérieur, en recueillant les informations des deux cochlées constitue le premier niveau où se produit une interférence dichotique nécessaire à la localisation spatiale des sources sonores. Une partie importante du noyau latéral est consacrée aux moyennes et hautes fréquences (supérieures à 4 kHz), alors que dans le noyau médian, ce sont les basses fréquences (inférieures à 4 kHz) qui occupent plus de la moitié du volume.

1.2.2.3. Le lemnisque latéral :

Bordé dorsalement par le colliculus inférieur et ventralement par le complexe olivaire supérieur, le lemnisque latéral reçoit des afférences bilatérales de ces deux noyaux. Sur le plan de la cytoarchitecture, il est divisé en deux zones distinctes : le noyau dorsal et le complexe ventral.

1.2.2.4. Les centres mésencéphalo-thalamiques :

Le colliculus inférieur : appelé aussi tubercule quadrijumeau postérieur, se situe au niveau du toit du mésencéphale, dans la partie médiane du tronc cérébral. Les afférences provenant des noyaux précédents (noyau cochléaire, olive supérieure latérale, noyaux intermédiaire et dorsal du lemnisque latéral majoritairement) convergent vers le colliculus inférieur qui est un relai synaptique quasi obligé pour tous les neurones ascendants. Il se subdivise en trois noyaux : central, péricentral et externe. Il reçoit trois types d'afférences : des afférences ipsi et controlatérales de la voie montante par l'intermédiaire du lemnisque latéral, des afférences corticales et des afférences provenant du colliculus inférieur controlatéral. La plupart des axones quittant le colliculus inférieur se projettent sur le corps genouillé médian ipsilatéral mais quelques uns rejoignent le corps genouillé médian controlatéral par la commissure intercolliculaire.

Le thalamus : dernier relais avant l'entrée dans le cortex, il reçoit principalement ses afférences du colliculus inférieur. Il comprend trois régions distinctes : le corps genouillé médian, la partie latérale du groupe postérieur du thalamus et le secteur auditif du noyau réticulé du thalamus. Le noyau principal est le corps genouillé médian, lui-même subdivisé en trois zones : ventrale, dorsale et médiane. L'organisation tonotopique du thalamus est principalement présente au niveau de la division ventrale du corps genouillé médian. Les basses fréquences sont situées à un niveau latéral et dorsal alors que les hautes fréquences sont localisées à un niveau plus médian et ventral. La partie latérale du groupe postérieur du thalamus présente elle aussi une organisation tonotopique, les hautes fréquences étant codées dans sa partie caudale, et les basses fréquences dans sa partie rostrale.

1.2.3. Le cortex auditif :

Anatomie du cortex auditif chez l'humain :

Au sein du cortex auditif humain, on peut distinguer une aire primaire et une ceinture périphérique. Le cortex auditif primaire (A1) est situé dans le gyrus temporal supérieur du lobe temporal ; recevant des projections point par point en provenance de la division ventrale du corps genouillé médian et contient donc une carte tonotopique précise. Les aires de ceinture du cortex auditif reçoivent des afférences plus diffuses en provenance des aires de ceinture du corps genouillé médian ; elles ont de ce fait une organisation tonotopique moins

précise. Le cortex auditif primaire comporte une carte tonotopique, par rapport à laquelle, se trouve à angle droit, une disposition en bandes des propriétés binaurales. Les cellules appartenant à une bande donnée sont excitées par les deux oreilles tandis que les cellules de la bande adjacente sont excitées par une oreille et inhibées par l'autre. Les bandes de ces différentes cellules alternent. A un niveau plus latéral sont présentes les aires « autour de la ceinture » (parabelt area). L'architecture de cette zone reste peu connue, il est très difficile de la subdiviser même si deux régions semblent se distinguer : la zone rostrale et la zone caudale.

1.3. Du cortex à la cochlée: les voies auditives efférentes

Le système auditif efférent est constitué de boucles multiples de rétrocontrôle, organisées entre les structures corticales et chacun des centres de traitement sous-corticaux. Ce système de boucles permet une modulation constante des messages afférents auditifs à tous les niveaux : au niveau périphérique, où le système efférent olivo-cochléaire peut directement moduler l'amplificateur cochléaire que sont les cellules ciliées externes, et au niveau le plus central, où des stimulations des neurones du cortex auditif permettent d'altérer les réponses des neurones sous-corticaux du corps genouillé médian, grâce à une boucle cortico-thalamique (Suga, 2008).

Au niveau périphérique, les fibres efférentes cheminent avec le nerf vestibulaire et se décomposent en système efférent latéral et système efférent médian (SEOCM). Le système efférent latéral, provenant du noyau olivaire supérieur latéral et se connectant aux CCI de la cochlée par des fibres non-myélinisées, à conduction lente, ne joue qu'un rôle mineur et mal élucidé. En revanche, les fibres efférentes nombreuses et myélinisées qui constituent la boucle olivo-cochléaire médiane (initialement décrite chez le chat par Rasmussen, 1946), se projettent directement sur les CCEs cochléaires par l'intermédiaire de synapses cholinergiques, et sont capables de moduler l'activité de ces cellules. Cette boucle de rétroaction périphérique est facilement explorable chez l'humain, grâce au recueil des otoémissions acoustiques provoquées, qui sont des sons émis par la cochlée, en réponse à une stimulation sonore, et qui reflètent l'activité des cellules ciliées externes. L'application, dans l'oreille controlatérale, d'un bruit blanc d'intensité modérée, permet d'activer les neurones efférents au niveau du tronc cérébral (complexe olivaire supérieur), neurones qui vont exercer une action inhibitrice au niveau des cellules ciliées externes, et par là même, diminuer l'amplitude des otoémissions acoustiques. L'efficacité du système efférent médian peut alors être explorée de manière non-invasive, chez l'humain, par la diminution d'amplitude des otoémissions acoustiques provoquées lors de l'application d'une stimulation acoustique controlatérale.

Par ailleurs, il a été montré, chez l'humain, la possibilité d'avoir une modulation directe des centres auditifs, sur la périphérie auditive par les fibres efférentes, chez des patients souffrant d'accidents vasculaires cérébraux intéressant les centres auditifs (Khalfa, et al., 2001) et par stimulation électrique corticale chez des patients épileptiques (Perrot et al., 2006).

D'autres études s'intéressent au déficit de traitement auditif temporel par l'exploration du fonctionnement des voies auditives descendantes du système efférent olivocochléaire médian, mirent en évidence une influence de celui-ci sur la perception de la parole. Ainsi, un dysfonctionnement de cette voie descendante chez les dyslexiques pourrait expliquer leur trouble de perception de la parole (Veuillet et al, 1999).

Deux rôles majeurs sont classiquement attribués au système efférent olivo-cochléaire médian : un rôle de démasquage, et un rôle de protection contre les agressions sonores (Rajan, 2000). Plus récemment, il a été proposé que le rôle de protection contre les agressions sonores ne serait en fait qu'une conséquence du rôle plus général de démasquage (Christopher & Smith, 2003).

En effet, la fonction du SEOCM serait d'augmenter la discriminabilité d'un son bref dans du bruit. Des études, chez le chat, ont permis de mettre en évidence qu'une section du faisceau olivocochléaire réduisait les performances dans une tâche de discrimination de formant ou d'intensité pour des sons de haute fréquences dans du bruit (Hienz et al, 1998 ; May & Mc Quone, 1995). Un effet rapide du SEOCM (de l'ordre de 100 ms) a été observé dans un environnement bruyant. L'activation du SEOCM peut améliorer les réponses du nerf auditif à un son bref (Dolan & Nuttall, 1988). Si nous considérons que le taux de décharge du nerf auditif est élevé à basse intensité pour une réponse au bruit et que le taux de décharge du nerf auditif est diminué à forte intensité à cause de l'adaptation au bruit, par conséquent, un bruit continu va masquer partiellement la réponse du nerf auditif. La stimulation du SEOCM va inhiber la réponse au bruit, réduire l'adaptation ce qui va restaurer les réponses des fibres : le SEOCM a donc un rôle de démasquage (Kawase et al, 1993 ; Kawase & Liberman, 1993).

De ce rôle de démasquage découle un effet dans la perception de la parole dans le bruit. En effet, plusieurs études ont mis en évidence une corrélation entre l'efficacité du système efférent olivo-cochléaire médian, et les capacités de perception de sons dans un bruit environnant. Ainsi, chez l'Homme, le fonctionnement faible du SEOCM est corrélé à une détection défaillante des sons dans le bruit (Micheyl & Collet, 1996 ; Micheyl et al, 1997 ; Micheyl, Perrot, & Collet, 1997) et une perte d'intelligibilité de la parole dans le bruit chez l'adulte (Giraud et al, 1997) et chez l'enfant (Kumar & Vanaja, 2004). Le démasquage controlatéral, c'est-à-dire l'augmentation de l'intelligibilité de la parole dans le bruit par l'application d'un bruit blanc controlatéral d'intensité modérée, est supprimé chez des patients dont le nerf vestibulaire est sectionné, et qui n'ont plus d'effet du système efférent olivo-cochléaire (Giraud et al., 1997). Plus récemment, De Boer et Thornton (2008) ont montré d'une part, que les performances de perception de syllabes dans le bruit étaient corrélées à l'efficacité du rétrocontrôle efférent médian, et d'autre part, que l'amélioration des performances d'intelligibilité dans le bruit, par un entraînement, était corrélée à une amélioration de l'efficacité du rétrocontrôle efférent.

Plusieurs études ont montré un effet de l'attention sur l'activation du SEOCM : la détection d'un son pur d'une fréquence donnée dans du bruit est facilitée par l'audition préalable d'un son de fréquence similaire, permettant d'orienter l'attention du sujet sur cette fréquence. Cet effet facilitateur n'existe plus chez des patients dont le faisceau olivo-cochléaire médian a été sectionné lors d'une neurotomie vestibulaire (Sharf et al, 1997).

Le rétrocontrôle efférent a donc un rôle majeur dans la perception de la parole dans le bruit, à la fois par son rôle bien établi de démasquage, et probablement par un rôle de régulation de l'attention auditive.

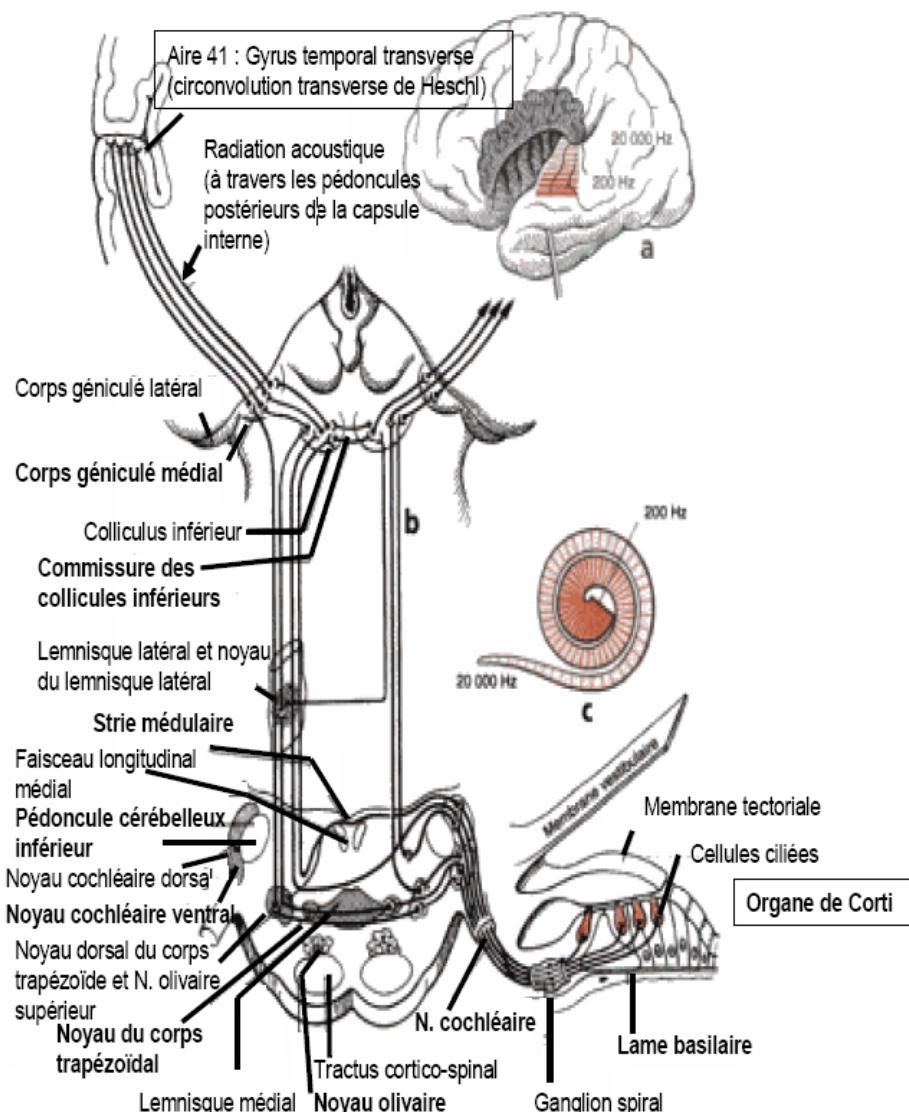


Figure 3 : Schéma des connexions centrales du nerf cochléaire.

D'après: Diagnostic neurologique: les bases anatomiques. Peter DUUS, 6ième édition. De Boeck Université. Thieme. p162.

2. Caractérisation d'un son de parole

Les signaux naturels comme la parole sont des sons complexes contenant de nombreuses composantes fréquentielles présentes simultanément. L'enveloppe de leur onde de pression peut également être très variable.

L'information portée par le signal de parole peut être analysée de multiples façons. On en distingue généralement plusieurs niveaux de description non exclusifs : acoustique, phonétique, phonologique, morphologique, syntaxique, sémantique et pragmatique.

2.1. Le niveau acoustique :

La parole apparaît physiquement comme une variation de la pression de l'air causée et émise par le système articulatoire. La phonétique acoustique étudie ce signal en le transformant dans un premier temps en signal électrique grâce au transducteur approprié : le microphone. De nos jours, le signal électrique résultant est le plus souvent numérisé. Il peut alors être soumis à un ensemble de traitements visant à mettre en évidence les traits acoustiques : sa fréquence fondamentale (F0), son énergie et son spectre. Chaque trait acoustique est lui-même intimement lié à une grandeur perceptuelle : sensation de hauteur (pitch), intensité et timbre.

La structure fréquentielle de la parole :

La fréquence fondamentale est un composant de basse fréquence de la parole, résultant de la vibration des cordes vocales, permettant la perception de la hauteur tonale de la voix d'un individu. Elle s'étend approximativement de 70 à 250 Hz chez les hommes, de 150 à 400 Hz chez les femmes, et de 200 à 600 Hz chez les enfants.

La structure harmonique est un phénomène acoustique dont les composants (les harmoniques) sont des multiples de la fréquence fondamentale. Les harmoniques, sont caractérisés par une fréquence ou hauteur (exprimée en hertz, Hz) et une amplitude ou intensité (exprimée en décibels, dB). L'interaction entre la fréquence de vibration des cordes vocales et les résonnances du tractus vocal d'un individu, privilégie certains harmoniques, proches de la fréquence de résonnance du tractus vocal : *les formants*. Les deux premiers formants (F1 et F2), jouant un rôle important dans la perception auditive des voyelles (Fig 4).

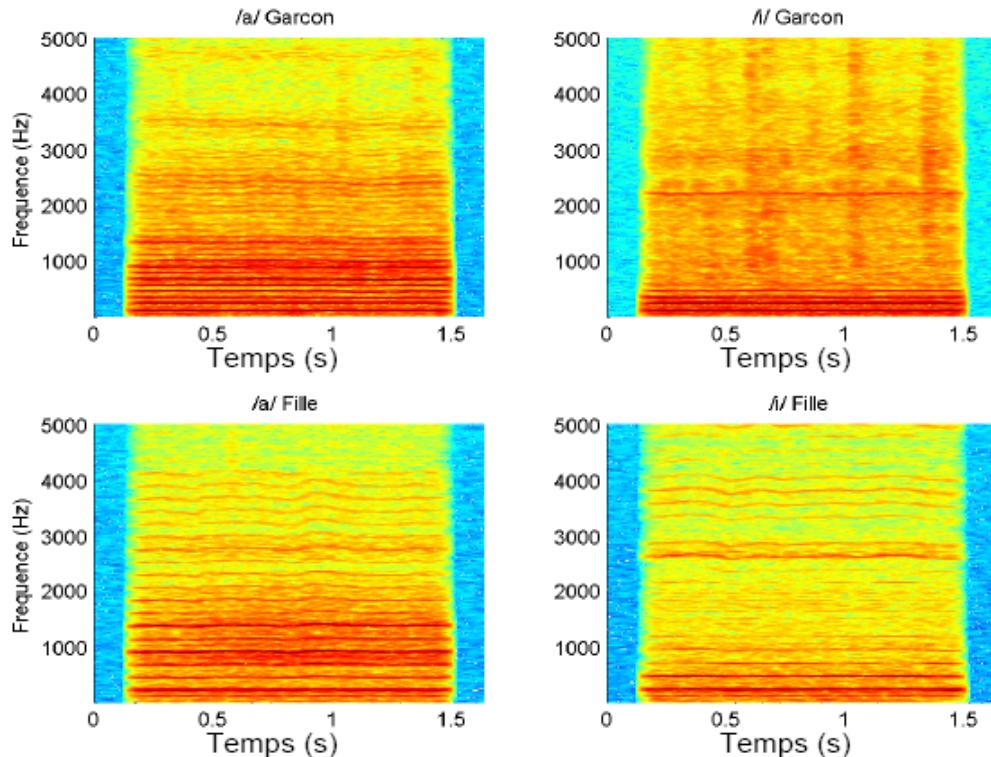


Figure 4 : Représentation spectrale de deux voyelles (/i/ et /o/).

Les bandes rouges fine correspondent aux harmoniques, tandis que les bandes rouges épaisses correspondent aux formants. A noter que les harmoniques sont beaucoup plus rapprochés chez l'homme que chez l'

- La structure temporelle de la parole : Les caractéristiques acoustiques temporelles jouent un rôle important dans la perception de la parole (Johnson et al, 2005 ; Skoe et al, in press). Elles interviennent en complément des informations fréquentielles (fréquence fondamentale et formants). On peut distinguer trois types d'information temporelle en fonction de la gamme de fréquence considérée :
- L'enveloppe temporelle (« contour du son ») qui véhicule les informations relatives au rythme global de la phrase, la délimitation des syllabes et concerne les variations temporelles de moins de 50 Hz (> 20 ms).
- La périodicité, qui reflète la prosodie (intonation de la voix), ainsi que le voisement (genre du locuteur, hauteur de la voix), et se situe entre 50 et 500 Hz.
- Enfin, la structure fine temporelle, qui se situe au-delà de 500 Hz (durées < 2 ms), et permet la perception, par exemple, des transitions formantiques (Rosen, 1992).

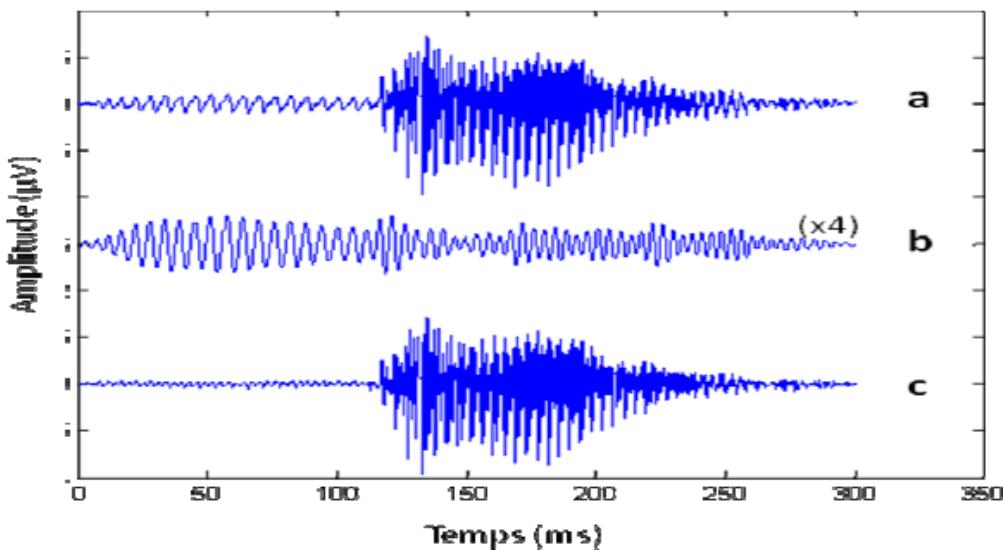


Fig. 5 : Illustration du type d'information temporelle contenue dans un signal de parole, ici la syllabe « BA ».

L'amplitude du signal recueilli au niveau du microphone est exprimée en fonction du temps en millisecondes (ms). La trace « a » représente la syllabe originale, la trace « b » représente la périodicité, entre 50 et 500 Hz (l'amplitude a été quadruplée pour une meilleure visibilité), et la trace « c » représente la structure fine temporelle, au-delà de 500 Hz.

2.2. Le niveau phonétique :

Ce niveau s'intéresse à la façon dont un son est produit par le système articulatoire et perçu par le système auditif.

La phonation :

La parole peut être décrite comme le résultat de l'action volontaire et coordonnée d'un certain nombre de muscles. Cette action se déroule sous le contrôle du système nerveux central qui reçoit en permanence des informations par rétroaction auditive et par les sensations kinesthésiques. La parole est donc obtenue par la modulation d'air de différentes manières, faisant intervenir plusieurs éléments corporels : de la cage thoracique aux éléments buccaux et faciaux. Tel les différents instruments composant un orchestre musical, les différents organes que sont les cordes vocales, la langue, les lèvres... peuvent être vus comme contribuant de manière conjointe et harmonieuse à l'exécution d'une partition. Toutefois, si chacun des instruments peut jouer sa partition seul, il n'en est pas de même pour les différents organes articulatoires, dont certains ne peuvent travailler qu'avec d'autres.

L'appareil phonatoire se décompose en *quatre éléments*, collaborant étroitement à la production de la parole : la soufflerie, le vibrateur, les résonateurs et le système articulateur.

La soufflerie consiste en un réservoir d'air : les poumons, actionnés par les muscles du thorax et de l'abdomen. *Le vibrateur* est constitué du larynx : ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée. Les cordes vocales sont placées en travers du larynx, elles peuvent le fermer complètement.

L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non voisés (ou sourds). Les sons voisés (ou sonores) résultent au contraire d'une vibration périodique des cordes vocales. *Les résonateurs* sont constitués du pharynx, de la cavité buccale et de la cavité nasale dont la mise en jeu dépend de la position du voile du palais. *Le système articulatoire* comprend la langue, les dents, les lèvres, les mâchoires et le voile du palais. La langue joue un rôle prépondérant dans le processus phonatoire : sa hauteur détermine la hauteur du pharynx, en effet plus la langue est basse plus le pharynx est court. Elle détermine aussi le lieu d'articulation, région de rétrécissement maximal du canal buccal, ainsi que l'aperture, écartement des organes au point d'articulation.

La phonétique articulatoire :

Il est intéressant de grouper les sons de parole en classes phonétiques, en fonction de leur mode articulatoire. On distingue généralement trois classes principales : *les voyelles, les semi-voyelles et les liquides, et les consonnes*.

Si le conduit vocal est suffisamment ouvert pour que l'air poussé par les poumons le traverse sans obstacle, il y a production d'une voyelle. Le rôle de la bouche se réduit alors à une modification du timbre vocalique. Si au contraire, le passage se rétrécit par endroit, ou même s'il se ferme temporairement, le passage forcé de l'air donne naissance à un bruit : une consonne est produite. La bouche est dans ce cas un organe de production à part entière. Les semi-voyelles [j, w], quant à elles, combinent certaines caractéristiques des voyelles et des consonnes. Enfin, les liquides [l, r] sont assez difficiles à classer.

Description des voyelles :

Comme nous venons de le voir, les voyelles se caractérisent par un libre passage de l'air à partir des cavités supra glottiques. Le « timbre » d'une voyelle dépend en partie de la résonance ou non de la cavité nasale, de la forme ainsi que du volume du résonateur buccal et de l'arrondissement ou non des lèvres. La configuration adoptée par ces différents facteurs détermine une fonction de transfert caractéristique. Ainsi, l'allure de l'onde excitatrice (onde glottale) sera modulée par cette fonction de transfert et fera émerger des zones fréquentielles de concentration d'énergie appelées formants.

On distingue ainsi selon la localisation les voyelles antérieure, moyennes et postérieures et selon l'écartement entre l'organe et le lieu d'articulation les voyelles fermées et ouvertes. La forme de cette fonction de transfert reste constante au cours de la production de la voyelle, la rendant assimilable à un signal quasi-stationnaire.

On peut approximativement représenter les voyelles dans le plan des deux premiers formants. Les voyelles sont donc identifiables par la position de leurs formants sur le spectre. Les premiers formants (notés F1 et F2) sont les plus importants pour la caractérisation de la voyelle. Delattre et al. (1951) proposent de caractériser les voyelles selon deux axes relatifs aux valeurs F1 et F2 (Fig 6).

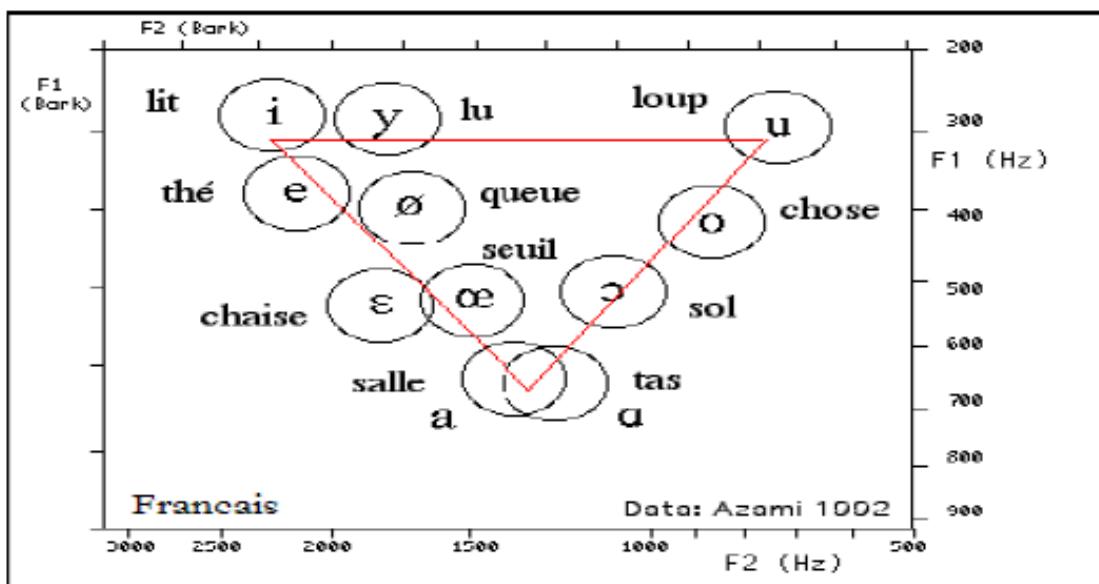


Figure6 : Représentation des voyelles françaises dans le plan (F1, F2).

En phonétique, on peut se limiter à l'étude des deux premiers formants puisqu'ils suffisent à caractériser et à identifier les voyelles prononcées. On représente alors, dans un plan, les voyelles en fonction des valeurs de F1 et de F2. L'ensemble des voyelles d'une langue se situe en général à l'intérieur d'un triangle formé par les voyelles [a], [i], [u] qu'on appelle le triangle vocalique.

Description des consonnes :

Contrairement aux voyelles, le flux d'air provenant des poumons est partiellement ou totalement obstrué au cours de la production d'une consonne, de ce fait l'énergie est généralement moins intense que pour les voyelles. On classe principalement les consonnes en fonction de leur mode d'articulation, de leur lieu d'articulation, et de leur nasalisation. Comme pour les voyelles, d'autres critères de différenciation peuvent être nécessaires dans un contexte plus général : l'organe articulatoire, la source sonore, l'intensité, l'aspiration, la palatisations et la direction du mouvement de l'air. En français, la distinction de mode d'articulation conduit à deux classes : les fricatives (ou constrictives) et les occlusives (ou plosives). Lorsque le passage de l'air est complètement fermé, le son résulte de son ouverture subite, on parle dans ce cas de consonnes occlusives ou plosives (/p/, /t/, /k/, /b/, /d/, /g/). Inversement, lorsque le passage du flux d'air bien que rétréci reste continu, on parle de consonnes constrictives (/s/ /z/ /f/ /v/ /ch/ /j/).

Le voisement :

Il s'agit d'un paramètre articulatoire retrouvé dans la plupart des langues, représentant l'intervalle entre la libération de la consonne (appelée onset) et le début des vibrations vocales (Lisker and Abramson 1964). Le voisement est long de quelques dizaines de millisecondes et se différencie selon les différentes langues.

L'étude du VOT (« Voice Onset Time ») fait encore l'objet de nombreux travaux. En français, le VOT est dénommé : Délai d'Établissement du Voiement (DEV), qui est défini comme l'intervalle de temps entre le début des pulsations périodiques régulières glottiques et la détente de l'occlusion supra-glottique. Ainsi, si le voisement débute avant la fin de l'occlusion le VOT peut-être négatif (cas d'un /ba/), il peut être nul si la synchronisation est parfaite et positif si le voisement commence un certain temps après la fin de l'occlusion (cas

d'un /pa/). Les consonnes aspirées ont un VOT plus long. En français, les occlusives voisées présentent une barre de voisement pendant la durée de l'occlusion, la valeur du voisement est donc négative. Ce voisement précède l'explosion de la consonne. Le voisement correspond aux oscillations périodiques de basse intensité qui précèdent l'explosion de la consonne (Snoeren et al, 2006). Les occlusives non voisées (/p/, /t/ /k/) montrent un VOT qui apparaît avec un certain retard après l'explosion : entre 10 et 30 ms pour le français. Serniclaes (1987) rapporte des valeurs moyennes de la durée du VOT en français et en anglais (Fig 7 et 8). Par ailleurs, il a été démontré, en anglais (Niyogi & Ramesh, 2003) que le VOT était un bon indice temporel pour discriminer sur le plan acoustique les consonnes voisées et non voisées.

Les différences de voisement (vibration des cordes vocales) et d'aspiration des consonnes de différents langages peuvent être caractérisées par des modifications du VOT, qui reflète les différences temporelles entre l'activité glottique et les événements supra laryngés. Par ailleurs, la perception du VOT serait sous le contrôle de l'hémisphère gauche (Liegeois-Chauvel et al, 1999). Toutefois, différentes études ont démontré des déficits limités dans la discrimination du VOT chez des patients présentant des lésions hémisphériques gauches, ceci suggérant soit une perception du VOT à la fois au niveau sous cortical et cortical, soit une participation de l'hémisphère droit dans la perception du VOT (Liegeois-Chauvel et al, 1999).

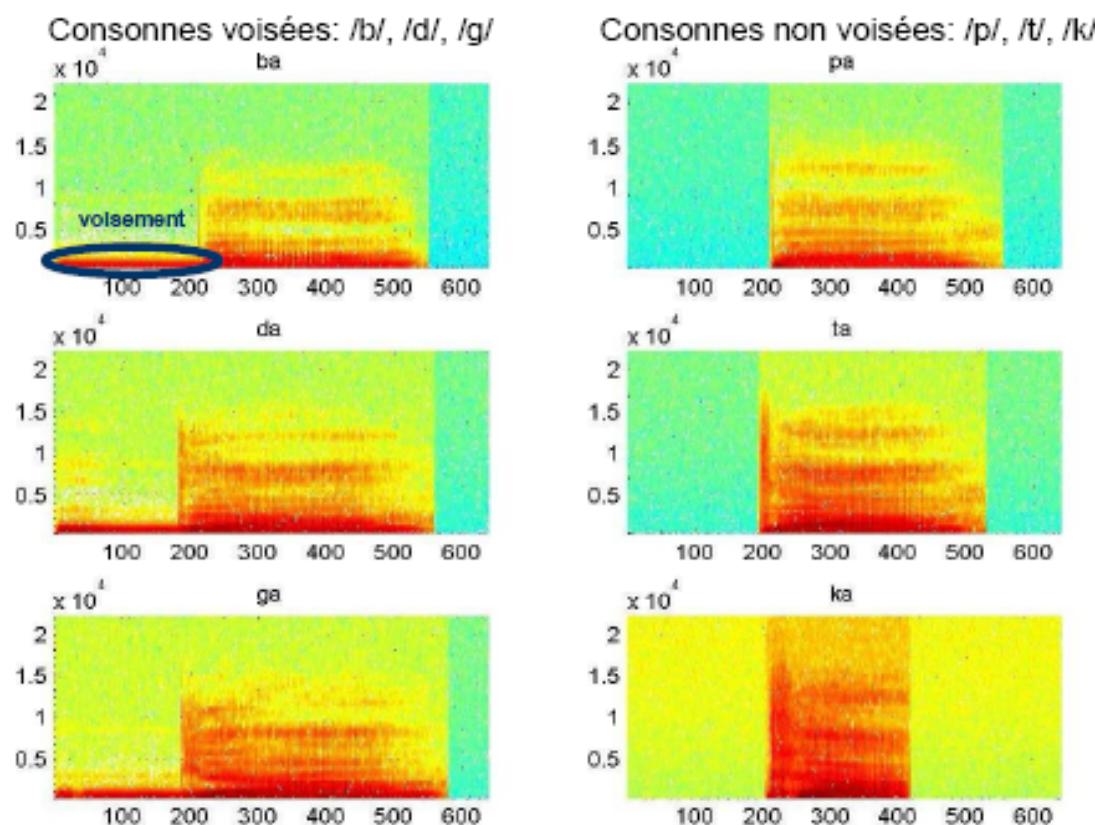


Figure 7 : Représentations spectrographiques (temps fréquence énergie) des sons [ba, da, ga] et [pa, ta, ka] en français.

On remarque la barre de voisement pour les sons de la première colonne alors qu'il n'y en a pas pour ceux de la seconde : En langue française, le VOT, ou DET (relation temporelle entre le relâchement de la consonne et le début de la vibration des cordes vocales) des consonnes voisées (/b/, /d/, /g/) est négatif, alors que dans le cas des consonnes non voisées (/p/, /t/, /k/), celui-ci est positif.

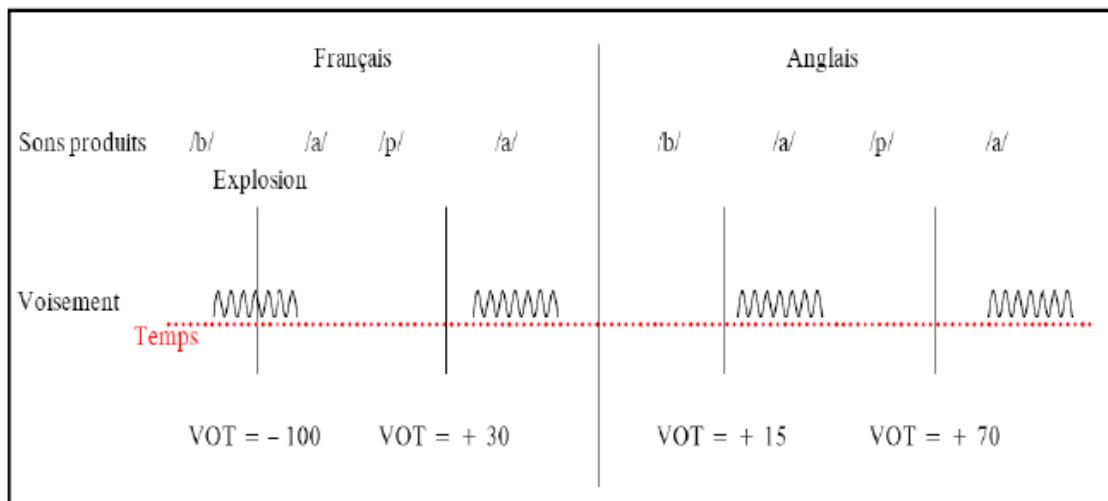


Figure 8 : VOT en français et en anglais (valeurs de VOT en ms, provenant de Serniclaes (1987). Figure d'après la thèse de Christophe Dos Santos (2007).

Il existe une variation temporelle dans la perception et la production du VOT parmi les différents langages. Par exemple, en anglais, pour les consonnes voisées, le voisement coïncide à peu près avec le relâchement de la plosive. En effet une consonne anglaise voisée va avoir le début du voisement peu avant le début du phonème, alors que dans le cas d'une consonne non voisée, il existe une aspiration initiale, suivie par un court silence jusqu'au début du voisement de la voyelle suivante, cette période de silence représentant le VOT anglais (Fig 8).

2.3. Le niveau phonologique :

Il représente l'interface nécessaire entre la phonétique et les descriptions linguistiques de niveau plus élevé. La phonologie introduit la notion d'unité abstraite du discours : le phonème. Le phonème est la plus petite unité phonique fonctionnelle, c'est-à-dire distinctive. Les phonèmes n'ont pas d'existence indépendante : ils constituent un ensemble structuré dans lequel chaque élément est intentionnellement différent de tous les autres, la différence étant à chaque fois porteuse de sens.

Une description phonologique ne peut être complète si elle ne permet pas de rendre compte de la durée, de l'intensité et de la fréquence fondamentale des phonèmes, ces trois composantes étant collectivement désignées sous le terme de prosodie. La durée des silences et des phones détermine le rythme de la phrase, tandis que l'évolution de la fréquence fondamentale constitue sa mélodie.

2.4. Les autres niveaux :

Pour le citer simplement, le **niveau morphologique** correspond à la branche de la linguistique qui étudie comment les formes lexicales sont obtenues à partir d'un ensemble réduit d'unités porteuses de sens, appelées morphèmes. L'importance de la morphologie en traitement de la parole tient à ce que la catégorie grammaticale et la prononciation des mots peuvent être expliquées dans une large mesure par leur composition morphémique.

Le **niveau syntaxique** fait appel à un ensemble de règles de structures précises. La liste des phrases admises, bien qu'infinie dans les langues naturelles, est restreinte par leur syntaxe. La syntaxe est toutefois différente de la grammaire, qui permet pourtant de décrire l'organisation hiérarchique des phrases et donc leur structure syntaxique.

La différence entre niveau syntaxique et **niveau sémantique** reste parfois floue. L'étude des significations des mots, de la façon dont elles sont liées les unes aux autres et des bases du choix lexical fait l'objet de la sémantique lexicale.

Le **niveau pragmatique** ou niveau du discours, au contraire du sens sémantique que l'on qualifie souvent d'indépendant du contexte, est défini comme dépendant du contexte. Tout ce qui se réfère au contexte, souvent implicite, dans lequel une phrase s'inscrit et à la relation entre le locuteur et son auditoire, a quelque chose à voir avec la pragmatique.

3. La catégorisation de la parole

La reconnaissance de la parole consiste à partir du signal continu qui arrive à nos oreilles, de segmenter ce signal continu en unités discrètes, les phonèmes, puis d'apparier cette représentation infra-lexicale avec les entrées du lexique. Les phonèmes sont perçus sur un mode catégoriel. La question posée ici est comment nous passons du signal acoustique continu et variable à des représentations catégorielles. Plusieurs études ont été réalisées sur ce phénomène de perception catégorielle (Eimas & Corbit, 1973 ; Liberman, Harris, Hoffman, & Griffith, 1957 ; Repp, 1984). La perception catégorielle se définit comme la discrimination des seules différences entre phonèmes et non des variantes acoustiques d'un même phonème (Liberman et al, 1957). Dans une expérience d'identification d'un stimulus le long d'un continuum qui va d'une catégorie perceptive à une autre, les réponses des sujets basculent brusquement d'une catégorie à l'autre. La tâche d'identification détermine donc la frontière entre les catégories phonétiques. Alors que dans une tâche de discrimination de deux stimuli le long d'un continuum, les sujets sont meilleurs lorsque les deux stimuli appartiennent à deux catégories différentes que lorsque les deux stimuli sont des variantes d'une même catégorie. Cela signifie que l'auditeur est peu sensible aux différences entre les sons appartenant à une même catégorie phonémique. La tâche de discrimination permet donc de dire si la discrimination est bonne à travers les frontières des phonèmes (inter-catégorielle) ou faible à l'intérieur des catégories (intra-catégorielle). Cela correspond à la perception catégorielle. Par exemple, la modulation de l'indice acoustique, le *Voice Onset Time* (VOT), peut altérer la perception catégorielle des phonèmes (Abramson & Lisker, 1970 ; Lisker & Abramson, 1964).

La frontière phonétique dans le signal de parole change avec les variations de VOT. Un continuum artificiel fabriqué pour donner toute une gamme d'intermédiaires acoustiques entre deux sons, par exemple, entre un [ba] et un [pa]. La frontière phonétique divise le stimulus auditif en deux catégories : [b] et [p]. Quand on demande à des sujets naïfs de retranscrire (tâche d'identification) ce qu'ils entendent, on obtient généralement une catégorisation très franche : la plupart des stimuli sont perçus soit comme [pa], soit comme

[ba] et la transition entre les deux catégories est très abrupte (Fig 9). Si on présente maintenant deux stimuli assez peu différents sur le continuum entre [ba] et [pa] et que l'on demande aux sujets d'effectuer une tâche de discrimination (les deux stimuli sont-ils identiques ou non ?), on obtient ici encore des réponses assez tranchées. La plupart des paires de stimuli sont très difficiles à discriminer, sauf à l'endroit de la frontière catégorielle (pic de discrimination), ou la performance devient très bonne.

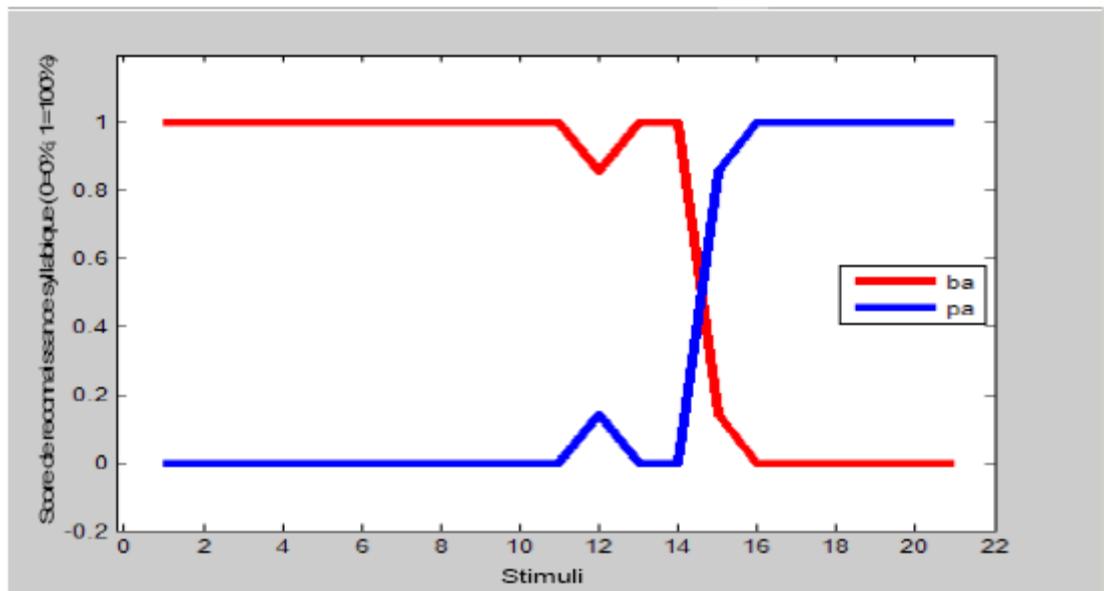


Figure 9 : Exemple du résultat d'un test de perception catégorielle utilisant un continuum de 21 signaux variant du /ba/ au /pa/.

La frontière perceptive est représentée par le croisement des deux courbes (ici au niveau du stimulus 15).

4. Propriétés du système auditif :

4.1. L'encodage temporel et spectral des sons par le système auditif :

Dans la plupart des situations naturelles d'écoute, un ensemble de sons, provenant de différentes personnes constituant une scène auditive, parviennent simultanément à nos oreilles. En général, nous parvenons assez aisément à isoler, identifier les différentes sources sonores en présence, ainsi que le genre des locuteurs, en dépit de leur enchevêtrement au niveau acoustique. Les mécanismes qui président à cette "analyse" des scènes sonores complexes par le système auditif humain font l'objet d'un intérêt croissant depuis plusieurs années (voir Bregman, 1990, pour une revue). Les mécanismes mis en œuvre font appel, notamment, à l'exploitation des informations contenues aussi bien dans le domaine temporel (les fluctuations d'amplitude ou d'enveloppe des sons) que dans le domaine spectral, dans le but d'analyser les scènes sonores complexes.

Le spectre d'un son est un indice essentiel à l'identification de la source sonore (voix/instrument) et notamment à la reconnaissance des voyelles. L'analyse spectrale (ou

fréquentielle) d'un son dépend en grande partie du filtrage de la cochlée. Ainsi, quel que soit le son complexe, que ce soit le son d'un instrument de musique ou une voyelle, celui-ci va subir cette analyse dès les premières étapes du traitement auditif. La cochlée fonctionne comme un ensemble de filtres passe-bandes ayant des fréquences centrales allant des basses fréquences (20 Hz) jusqu'aux hautes fréquences (20000 Hz), la largeur de ces filtres, conditionnant le degré de précision de l'analyse fréquentielle du signal par la cochlée. En effet, la largeur de ces filtres augmente en fonction de leur fréquence centrale, ainsi, l'analyse spectrale se dégrade dans les hautes fréquences (Grimault et al, 2000).

Le code temporel intervient en complément du code spectral : dans chaque noyau du système auditif, certaines cellules ont la capacité de synchroniser leur cadence de décharge de potentiels d'action sur la périodicité du stimulus : c'est le principe du « *calage de phase* » (appelé « phase-locking » dans la littérature). Aux fréquences inférieures à 4-5 kHz, les décharges des neurones sont synchronisées sur une phase particulière de l'onde de pression incidente. Cela permet au système auditif d'encoder assez précisément la période, et donc la fréquence du stimulus. En effet, même si la période réfractaire des neurones auditifs ne leur permet pas d'émettre un potentiel d'action à chaque cycle, les intervalles inter-décharges sont des multiples entiers de la période. Ce phénomène dit de phase-locking débute dès la sortie de la cochlée et se répercute tout le long des voies auditives afférentes en se limitant à des fréquences de plus en plus basses au fur et à mesure que le message nerveux est intégré. On cite les valeurs limites de 4 kHz dans le nerf auditif, 1 kHz dans le noyau cochléaire, 300 Hz dans le colliculus inférieur, 200 Hz dans le corps genouillé médian du thalamus et en dessous de 80 Hz dans le cortex auditif. Il est intéressant de noter que le mode de synchronisation change au niveau du colliculus inférieur, qui est par ailleurs le noyau de convergence des voies auditives périphériques bilatérales. Le rôle du codage temporel reste à préciser, même s'il semble intuitif dans le sens où il renforcerait l'encodage cochléaire spatial (dérivé de la tonotopie cochléaire) pour certains sons basse-fréquence (les plus utiles pour la communication humaine et les vocalisations animales). Pour des revues neurophysiologiques complètes sur le code temporel, voir Eggermont (2001), Joris et al, (2004).

Mettre en évidence ces deux types de codage reste le sujet de nombreux travaux, notamment par l'utilisation de l'imagerie par résonnance magnétique fonctionnelle (IRMf). Cette technique a permis de montrer l'existence de schémas visant à recoder temporellement le signal dès le noyau cochléaire et ce, jusqu'au cortex auditif (Griffiths, 2003 ; Kumar et al, 2007). Ces données sont corroborées par des enregistrements de potentiels évoqués sous corticaux en réponse à des stimuli langagiers, l'augmentation de la vitesse de stimulation entraînant une augmentation de la latence de l'onset, une augmentation d'amplitude des hautes fréquences (1400Hz) ainsi qu'une diminution d'amplitude des fréquences correspondant à la fréquence fondamentale (Krizman et al, 2010).

Par ailleurs, l'importance relative des analyses spectrale et temporelle apparaît être très variable suivant les individus. Il faut cependant savoir que les filtres auditifs des sujets malentendants peuvent être quatre fois plus larges que ceux des normo-entendants (Moore, 1995). Par conséquent, l'analyse spectrale des signaux est dégradée par un facteur quatre pour ces sujets et l'analyse temporelle, souvent non dégradée (Bacon & Gleitman, 1992) revêt alors sans aucun doute une importance particulière.

4.2. La latéralisation corticale :

Chacun des hémisphères cérébraux est le siège de la perception sensorielle et de la régulation de la motricité volontaire du côté opposé du corps. La structure du cortex des deux hémisphères cérébraux est presque symétrique, mais ils ne sont pas tout à fait égaux sur le plan fonctionnel.

La démonstration par Broca (1865), de la localisation frontale gauche des lésions responsables d'aphasie a constitué la première preuve que les hémisphères ne sont pas équivalents quant à leur rôle fonctionnel dans la cognition humaine. Ces constations corroborées par l'étude de nombreux patients cortico-lésés présentant une aphasie de Broca ou de Wernicke par lésion hémisphérique gauche, mettent en avant le rôle dominant de cet hémisphère dans le codage de la parole (Efron, 1963 ; Geschwind, 1970 ; Tallal et al 1993). Cette hypothèse est soutenue par l'imagerie fonctionnelle, mettant en avant le rôle de l'hémisphère gauche dans le traitement des variations formantiques rapides du langage (Fiez et al, 1996 ; Belin et al ; 1998). Des arguments supplémentaires ont démontré ce rôle, par la mise en évidence de l'existence d'une résolution temporelle meilleure dans l'hémisphère gauche par rapport à l'hémisphère droit (Liégeois-Chauvel et al, 1999, 2001 ; Zatorre et al, 2001). La parole étant un signal acoustique présentant des changements rapides, celle-ci sera préférentiellement encodée au niveau de l'hémisphère gauche. Chez la grande majorité de la population (environ 90%) l'hémisphère gauche est celui qui exerce le plus d'influence sur le langage, les habiletés mathématiques et la logique. C'est donc cet hémisphère qui intervient lorsque vient le temps d'écrire.

L'hémisphère droit, quant à lui, serait responsable, des habiletés spatio-visuelles, de l'intuition, de l'émotion, de l'appréciation de l'art et de la musique. En effet, la musique, phénomène sonore présentant des variations plus lentes, et plus précises en fréquence (variations de l'ordre de $> 1/12$ à $1/6$ d'octave) est encodée de manière prépondérante dans l'hémisphère droit, où la résolution fréquentielle y est meilleure (Hornickel et al, 2009 ; Zatorre, 2002 ; Tremblay & Kraus 1997 ; Johnsrude et al, 2000).

N'oublions pas toutefois, que les hémisphères sont reliés entre eux par différentes connections parmi lesquelles le corps calleux, et qu'ils ne travaillent pas comme « deux cerveaux distincts », mais bien en inter-connection permanente (Liégeois-Chauvel et al, 2004 ; Friederici & Alter, 2004).

L'ensemble de ces données sont représentées dans le tableau 1.

Hémisphère gauche	Hémisphère droit
Filtrage Passe-Haut	Filtrage Passe-Bas
Codage des fluctuations rapides [20 - 50] Hz	Codage des fluctuations lentes [3 - 5] Hz (= syllabes)
Fenêtre d'analyse [20 - 40] ms	Fenêtre d'analyse [150 - 300] ms
Perception 'locale' = détail = HF	Perception 'globale' = grossière = BF
	Phase-locking speech envelope
Traitements temporels	Traitements fréquentiels
Traitements de la parole	Traitements de la musique

Tableau 1 : Tableau récapitulatif des différentes caractéristiques fonctionnelles hémisphériques.

II. Les potentiels évoqués auditifs:

Un certain nombre de relais synaptiques existent entre la cochlée, lieu de la conversion de l'onde mécanique du son en courant électriques, et le cortex. Afin d'étudier la transmission de l'information auditive au niveau sous cortical, nous disposons de différentes méthodes, parmi lesquelles les potentiels évoqués auditifs. Les potentiels évoqués auditifs sont une méthode permettant d'enregistrer l'activation des différents relais des voies auditives, ceci de manière objective et non invasive.

2.1 Les différents types de potentiels évoqués auditifs (PEA)

Les PEA peuvent être enregistrés à proximité de la cochlée (électrocochléographie) ou à distance au niveau du scalp (potentiels évoqués auditifs précoce, de moyenne latence, de longue latence). Les PEA enregistrés au niveau du scalp sont distingués en fonction de leur latence. Les potentiels précoce sont enregistrés pendant les dix premières millisecondes après le stimulus. Les potentiels de moyenne latence concernent les réponses enregistrées entre 10 et 50 millisecondes, et les PEA tardifs ou corticaux concernent les réponses apparaissant après 50 millisecondes (Fig 10).

2.1.1. Potentiels évoqués auditifs précoce (PEAP) :

Depuis les travaux de Jewett (1970), Sohmer, on sait qu'il est possible de détecter au niveau du scalp l'activité électrique de populations de neurones situées à distance du point d'enregistrement (potentiels de champ lointains) (Sohmer et al, 1977). Les PEAP sont enregistrés dans les dix millisecondes qui suivent le début de la stimulation. A forte intensité sonore, les PEAP sont constitués de 5 à 7 pics correspondant à l'activation successive du nerf auditif et des différents relais du tronc cérébral, les pics les plus importants en clinique étant les pics I, III et V.

Le pic I provient de la décharge synchronisée des neurones intra cochléaires, l'onde II aux noyaux cochléaires, l'onde III à l'olive supérieure, l'onde IV au lemnisque latéral et enfin l'onde V au colliculus inférieur. Cette partie de la réponse auditive a la propriété d'être reproductible à la fois en morphologie et en latence chez un même individu et dans une certaine mesure entre les individus. Cette remarquable stabilité lui donne beaucoup d'importance en clinique.

2.1.2. Potentiels évoqués de latence moyenne (PEALM): de 20 à 70ms

D'origine thalamique et sous corticale, ils sont enregistrés sur 50ms avec un maximum d'amplitude sur les régions fronto-centrales du scalp, les PEALM sont décrits comme une succession de six ondes No, Po, Na, Pa, Nb et Pb (Picton et al 1974). Le complexe Na/

Pa (18/30ms environ) est facilement observable chez tous les individus, contrairement aux autres ondes de latence moyenne (No, Po, Nb et Pb), beaucoup plus instables et variables en latence comme en morphologie (Picton et al, 1974).

2.1.3. Potentiels évoqués de latence tardive : jusqu'à 500 ms

Les potentiels évoqués tardifs constituent les ondes les plus amples de la réponse évoquée auditive. Une de leurs caractéristiques est leur extrême variabilité intra et inter individuelle, qui peut s'expliquer par de nombreux facteurs relatifs à l'état du sujet. La première onde des potentiels tardifs est appelée N100 ou N1. Elle correspond à une déflexion négative qui culmine au vertex entre 70 et 130 ms après le début de la stimulation. La deuxième onde P200 est décrite comme une large positivité centrale dont la latence est comprise entre 150 et 220 ms. Lorsque la stimulation se prolonge au-delà de 200 ms, on voit apparaître immédiatement après l'onde P200, un potentiel négatif fronto-central qui se maintient tant que dure le stimulus. Cette déflexion est baptisée potentiel de longue durée, elle est suivie par une réponse « off » marquant la fin de la stimulation. Les autres pics de la réponse tardive sont appelés onde N2 (autour de 280 ms) et onde P300. Ces deux dernières composantes appartiennent à la catégorie des ondes lentes et sont largement endogènes.

L'étude des ondes P1-N1-P2-N2 chez des groupes d'enfants, d'adolescents et d'adultes met en avant la persistance d'une maturation auditive corticale au cours de l'adolescence, les réponses étant de type adulte à l'âge de 22 ans (Sussman, 2008).

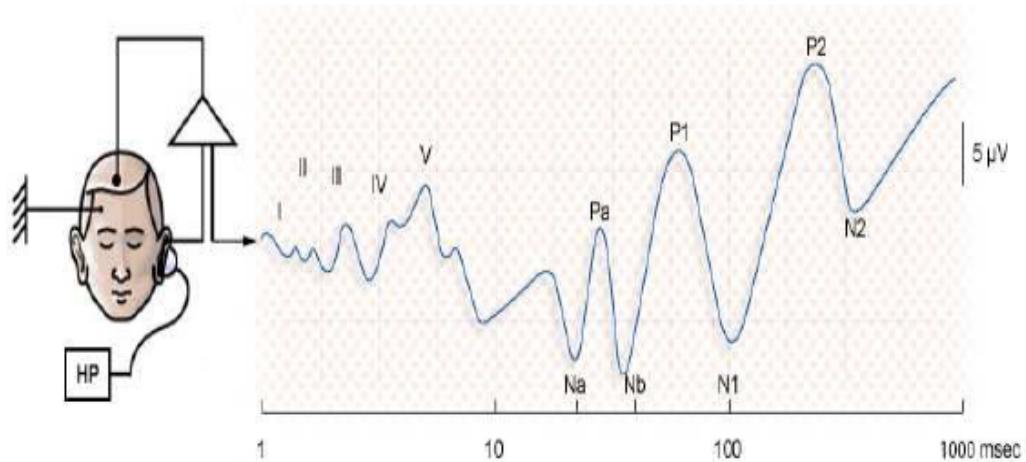


Figure 10 : Schéma de référence des voies auditives permettant de repérer le site anatomique des différentes ondes du PEA.

Dessin : S. Blatrix, P. Minary. Tiré de « Audition Promenade autour de la Cochlée ».

2.2. Technique de recueil des PEAP :

Les signaux enregistrés à l'aide des potentiels évoqués auditifs ont une amplitude de l'ordre de quelques microvolts. Cette faible amplitude pose un problème de parasitage du signal par des rayonnements électromagnétiques externes, que l'on peut diminuer par le moyennage

de plusieurs enregistrements. Comme nous le verrons au cours des chapitres suivants, les problèmes de parasites électromagnétiques sont particulièrement importants dans le cas des potentiels évoqués auditifs en réponse à des sons de parole.

Les PEAP sont largement utilisés en pratique clinique quotidienne, aussi bien chez le nouveau-né que chez la personne âgée. Le montage utilisé en clinique présente certaines similitudes avec celui de recueil des potentiels évoqués auditifs en réponse à un son de parole avec toutefois certains problèmes supplémentaires dans le cas des PEASP, nous ayant amené à repenser complètement la chaîne de recueil des speech ABR, comme détaillé dans la partie expérience 3.

Le sujet est placé au repos dans une pièce insonorisée. Les stimulations utilisées sont des clics envoyés de manière monaurale en conduction aérienne ou osseuse au sujet. Les variations de potentiels sont enregistrées à l'aide 4 électrodes cutanées placées respectivement sur chacune des mastoïdes, le vertex (électrode positive) et le front (masse) (Fig 11). Les traces sont moyennées, et les latences des différents pics sont analysées.

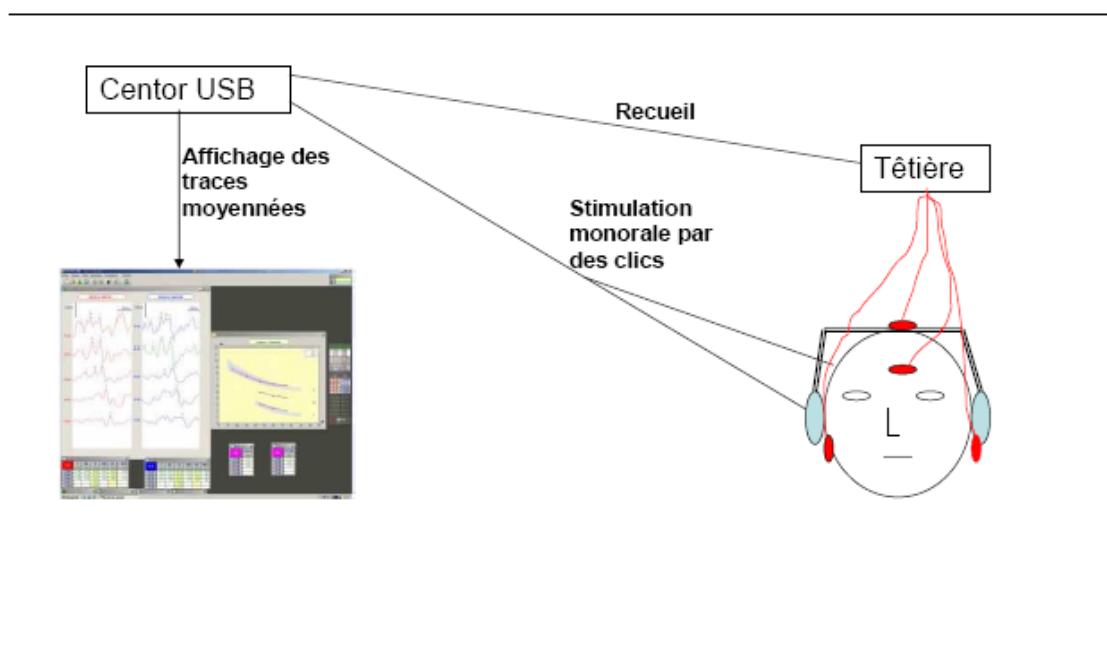


Figure 11 : Représentation schématique du système de recueil des PEAP, ici avec le système Centor USB Racia Alvar®.

III. Analyse des sons par le tronc cérébral :

3.1. Du clic au son de parole...

Historiquement, les premiers enregistrements électrophysiologiques du système auditif datent des années 1930 (Wever & Bray, 1930). Toutefois, la mise en évidence du potentiel évoqué auditif en réponse à des clics comme une mesure objective de la fonctionnalité des voies auditives ne débutera que 40 ans plus tard (Jewett et al, 1970) (PEA). Elle sera suivie par les premiers enregistrements d'une réponse soutenue à la fréquence (RSF définie ci-dessous) ou « Frequency Following Response », *in situ* chez l'animal (Worden & Marsh, 1968), et en champ lointain chez l'homme (Marsh et al, 1975). L'utilisation des potentiels évoqués auditifs en pratique clinique débute dès les années 1980, avec une répartition en différents sous-groupes, définis en fonction de leurs latences d'apparition après le début du stimulus: les PEA de latences précoces ou PEAP (recueillis dans les 10 premières millisecondes), moyennes (entre 10 et 50 millisecondes) et longues (entre 50 à 80 millisecondes).

Toutefois les PEA en réponse à des clics présentent des limites. Ils ne permettent pas d'explorer l'ensemble de la gamme fréquentielle représentative du spectre auditif humain, ni l'exploration spécifique de l'encodage des informations temporelles et spectrales contenues dans des stimuli écologiquement valides, comme la parole (Song et al, 2006).

La réponse soutenue à la fréquence (RSF) est un phénomène naturel reflétant chez l'humain la synchronisation de l'activité des neurones auditifs sur la fréquence d'un stimulus extérieur. D'un point de vue physiologique, la RSF est un potentiel évoqué illustrant l'activité globale de verrouillage de phase dans les cellules du tronc cérébral, entre le noyau cochléaire et le corps genouillé médian du thalamus (Moushegian et al, 1973 ; Daly et al, 1976 ; Faingold & Caspary, 1979 ; Hall, 1979). Au niveau du colliculus inférieur, l'activité de verrouillage de phase est retrouvée jusqu'à 1000 Hz environ (Greenberg et al, 1987 ; Galbraith et al, 1995 ; Krishnan et al, 2005 ; Musacchia et al, 2007) et est composée de pics calqués sur la périodicité d'un stimulus complexe contenant des harmoniques à la fréquence fondamentale (F0) (Worden & Marsh, 1968; Krishnan et al, 2004 ; Marsh & Worden, 1970; Smith, 1975). L'activité de verrouillage de phase contenue dans la RSF permet donc l'encodage des fluctuations fréquentielles rapides du son de parole.

La RSF neuronale (de latence supérieure à 6 ms) obtenue en enregistrant la différence de potentiel entre une électrode placée sur le lobe auriculaire et une autre sur le scalp, est à différencier du potentiel microphonique cochléaire (de faible latence), que l'on peut faire disparaître en faisant la moyenne des réponses obtenues à des paires de stimuli en opposition de phase (Sohmer et al, 1977). L'information contenue dans la RSF neuronale serait transmise le long de fibres provenant du tour apical de la cochlée et, après 3 à 5 synapses, atteint le colliculus inférieur (tronc cérébral) où la RSF neuronale est générée (Marsh et al, 1975 ; Moushegian et al, 1973; Sohmer et al, 1977; Delgutte, 1980).

La RSF reproduit une oscillation à une fréquence qui s'apparente, d'un point de vue perceptif, à la sensation de hauteur tonale perçue (Greenberg, 1987). Il est intéressant de noter qu'un stimulus composé d'un son pur modulé en amplitude ou en fréquence (modulation inférieure à 100 Hz), peut générer une réponse électrophysiologique, s'apparentant à une RSF : les PEA stationnaires multiples (PEASM) ou « auditory steady state response ».

Les PEASM sont générés à travers le système nerveux auditif par un stimulus continu dans le temps et sont représentés par un potentiel évoqué neuronal qui suit l'enveloppe d'un stimulus complexe périodique avec une activité de « verrouillage de phase ». Différents stimuli peuvent être utilisés pour l'étude des PEASM, variant par leur fréquence de modulation (90Hz (Rance et al, 1998), (80 Hz (Menard et al, 2008), ou bien composés de plusieurs fréquences porteuses (500 Hz, 1000 Hz, 2000 Hz, 4000 Hz), chacune modulée à une fréquence différente (John et al, 2002). Le stimulus vient exciter de façon continue une zone précise de la cochlée à une certaine fréquence, que l'on retrouve dans la réponse PEASM. L'analyse des fréquences recueillies est effectuée de manière individuelle afin de déterminer la présence d'une réponse à chaque fréquence audiométrique. Il a été par ailleurs mis en évidence, une corrélation entre l'amplitude des PEASM et la perception subjective d'intensité sonore ou sonie (Menard et al, 2008).

Les PEASM viennent en complément des PEA, notamment dans la prédiction des seuils auditifs sur l'ensemble de la gamme fréquentielle, chez le nouveau-né et l'enfant sédaté/endormi (Picton et al, 2003). De plus, les PEASM possèdent des champs d'investigation vastes tels que l'étude de l'encodage du fondamental des voyelles, (Aiken & Picton, 2006, 2008), l'objectivation de la fonction de sonie (Ménard et al, 2008), ou encore de la dys-synchronisation auditive dans le cadre des neuropathies auditives (Rance et al 1998).

Cependant, bien que les PEASM permettent une étude plus approfondie de l'audition d'un sujet, le traitement sous cortical de l'encodage temporel de la parole n'est pas analysable et les stimuli utilisés restent peu écologiques. Or, avec la mise en évidence de troubles du traitement de la structure fine de l'information temporelle, à l'origine de troubles auditifs en particulier en environnement bruyant, il apparaît nécessaire de disposer d'un outil diagnostique permettant une étude objective de cet encodage temporel : les PEASP répondent à cette nécessité (Johnson et al, 2005 ; Purcell et al, 2004; Wible et al, 2004).

3.2. Les potentiels évoqués auditifs en réponse à un son de parole:

3.2.1. Méthode d'acquisition :

Les stimuli les plus couramment utilisés par les différentes équipes sont des syllabes (par exemple /da/ (Johnson et al, 2005); /ba/ (Akhoun et al, 2008) contenant en moyenne 5 formants, dont la durée varie de 40 millisecondes à plusieurs centaines de millisecondes. Ces sons complexes permettent l'analyse des plus petites unités individualisable, à une fréquence élevée avec ainsi un temps d'acquisition diminué par rapport à l'utilisation d'un mot complet. Ces stimuli sont le plus souvent présentés de manière monaurale à droite, l'hémisphère gauche encodant préférentiellement les sons de parole chez les individus

gauchers. Les intensités utilisées varient de 50 dB SL (Daly et al, 1976) à 90,3 dB SPL (Hornickel et al, 2008) suivant les études. Le son est délivré au sujet via des inserts afin d'éviter l'artefact électromagnétique, cet artefact représentant l'une des difficultés d'acquisition des PEASp.

Les réponses EEG induites par les stimuli sont enregistrées à l'aide d'électrodes cutanées placées : sur le vertex (électrode positive), le front (masse) et chaque mastoïde (Krishnan & Parkinson, 2000 ; Akhoun et al, 2008 ; Galbraith et al, 2000 ; Johnson et al, 2005). Les réponses doivent être moyennées afin de diminuer l'effet des artefacts, le moyennage (nombre de répétitions du stimulus) utilisé varie de 999 (Moushegian et al, 1973 ; Gerken et al, 1975) à 6000 (Wible et al, 2004) répétitions collectées pour chaque polarité du stimulus, avec une fenêtre d'enregistrement de réponse variant de 52ms (Wible et al, 2004) à plus de 340 ms selon les stimuli (Song et al, 2008).

Comme nous le verrons dans la partie expérimentale, le principe des montages de recueil utilisés sont sensiblement identiques et s'apparentent à celui représenté sur la figure 29 de l'expérience 3.

3.2.2. Qu'est-ce qu'une réponse-type « PEA en réponse à un son de parole »?

Les PEA en réponse à une syllabe peuvent être divisées en différentes parties (fig 12):

- la première partie est représentée par la réponse impulsionale ou «onset response », précédée par une phase d'enregistrement du bruit ambiant appelée « pré-moyennage ». La réponse impulsionale (ondes V-A) est une réponse transitoire, encodant les variations temporelles rapides inhérentes à la consonne, similaire à la réponse observée à un click et apparaissant approximativement 7 ms après le début du stimulus dans le cas de voyelles anglaises (Banai et al, 2007), et de 7 à 9 ms dans le cas de voyelles françaises.
- la seconde partie représentée par la réponse soutenue à la fréquence (RSF) ou « frequency following response (FFR) », apparaît en réponse aux harmoniques et à la structure périodique de la voyelle à la fréquence de la source (voisement) ; elle présente une activité soutenue, la périodicité des pics de cette phase correspond à l'information fréquentielle contenue dans le stimulus (Russo et al, 2004; Marsh, 1975). L'onde C est le reflet du début du voisement, tandis que la période entre les pics D, E, F de la FFR (activité de verrouillage de phase) correspond à la fréquence fondamentale du stimulus (F0), et que les pics compris entre les ondes D, E, F représentent l'activité de verrouillage de phase à la fréquence du premier formant F1 (Banai et al, 2007). Une dernière onde apparaît ensuite: l'onde O qui correspond à l'arrêt du son, et temporellement à la fin du stimulus (Johnson et al, 2005).

Selon les équipes et les protocoles, les méthodes d'analyse varient et sont répertoriées dans les tableaux (1) et (1bis) de l'article de revue Richard et al, 2010, en annexe.

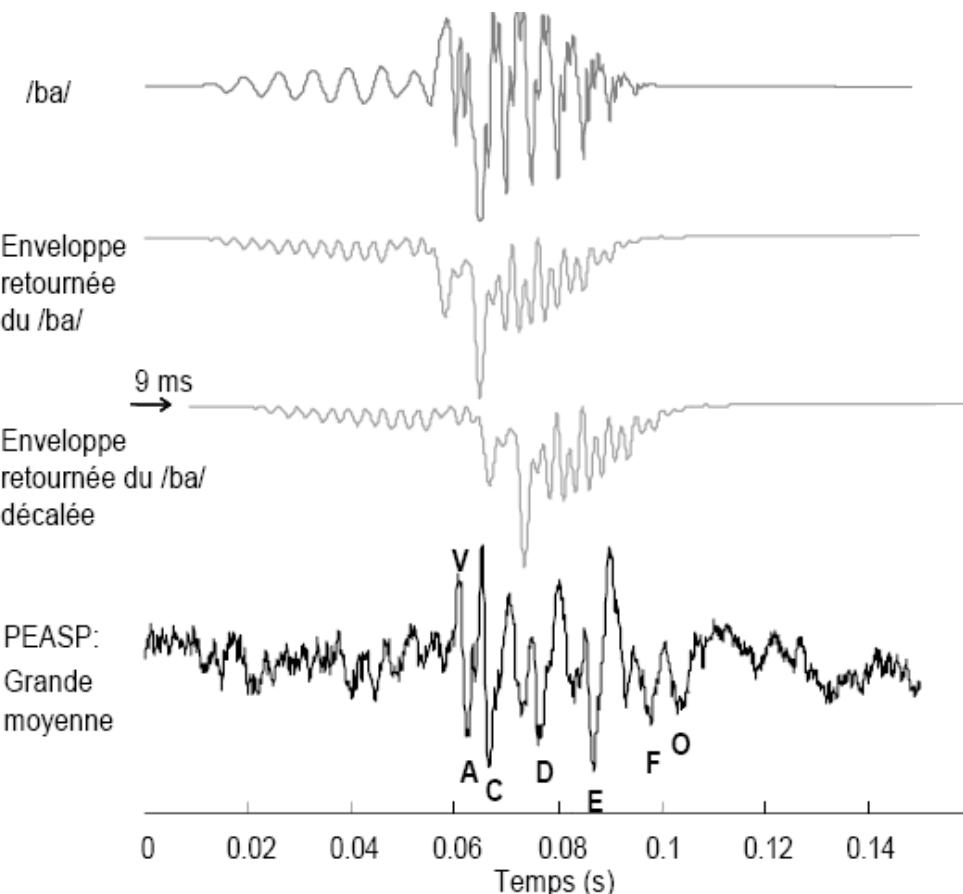


Figure 12 : Caractéristiques temporelles d'un PEASP.

De haut en bas : Représentation temporelle du stimulus /ba/ ; Représentations temporelles de l'enveloppe retournée du /ba/, et de celle-ci décalée de 9 ms, ce qui correspond au délai moyen d'apparition de la RSF. Représentation temporelle de la trace PEASP obtenue en réponse au /ba/ (intensité de stimulation : 55 dBDSL). On retrouve un calage de phase (ou « phase locking ») entre cette trace et l'enveloppe retournée et décalée du stimulus. Les pics V-A représentent l'onset, tandis que les pics C à O représentent la RSF.

Différents travaux menés par l'équipe américaine de Nina Kraus, sur les PEASP en réponse à un /da/ ont permis d'isoler certaines caractéristiques des différentes parties constitutives des PEASP : l'onset (onde V-A) et la RSF.

Ces différentes données sont représentées dans le tableau 2.

Tableau 2 : Valeurs normatives de PEASP obtenues à partir de 88 enfants de 8 à 12 ans, en réponse à un /da/, à une intensité de stimulation de 80 dBHL (d'après Banai et al, 2007).

Mesures PEASP		
Partie transitoire : l'onset	Latence des pics (ms)	Amplitude des pics (µV)
onde V	6.68 +/- 0.25	0.31 +/- 0.15
onde A	7.59 +/- 0.31	- 0.67 +/- 0.17
onde C	17.86 +/- 0.48	- 0.32 +/- 0.13
onde D	22.29 +/- 0.43	- 0.33 +/- 0.17
onde E	30.99 +/- 0.44	- 0.39 +/- 0.13
onde F	39.54 +/- 0.44	- 0.44 +/- 0.19
onde O	47.95 +/- 0.52	- 0.19 +/- 0.11
Partie soutenue: la RSF	Amplitude	
RMS	0.20 +/- 0.03	
FFT F0 (µV)	0.081 +/- 0.032	
FFT F1 (µV)	0.034 +/- 0.009	

3.2.3. Potentiels évoqués auditifs en réponse à un son de parole : artéfact ou réponse neurophysiologique ?

Une des difficultés rencontrées lors du recueil de PEASPeST la simultanéité entre la stimulation et l'enregistrement d'une part des PEASP, et la forte similitude entre la forme du stimulus et le pattern du signal recueilli d'autre part, induisant un risque d'artéfact électromagnétique particulièrement important. Cet artéfact, résultant de la superposition du signal neuronal (RSF) et du couplage électromagnétique des transducteurs, est lié aux problèmes de compatibilité électromagnétique entre la chaîne de stimulation (en majeure partie les transducteurs acoustiques) et les électrodes de recueil. En effet, la précision, souvent spectaculaire, de la synchronisation entre la RSF et la voyelle, pose la question de l'authenticité de cette réponse neuronale, qui pourrait n'être que le reflet du couplage électromagnétique entre la chaîne de stimulation et les électrodes de recueil. Cet artéfact n'existe pas avec les PEA en réponses à des clics, car le PEA est enregistré après que le stimulus impulsionnel ait été délivré, et peut être facilement éliminé dans les PEASMS. Les PEA en réponse à un son de parole sont, quant à eux, difficiles à distinguer de l'artéfact, les premiers pics fréquentiels du stimulus correspondant à ceux de la réponse soutenue à la fréquence. De plus, dans le cas des PEA en réponse à un son de parole, la polarité alternée (stimulus envoyé alternativement de manière inversée et non inversée (Johnson et al, 2005) à elle seule ne suffit pas à annuler les réponses artéfactuelles. Afin de diminuer les risques d'artéfacts différents paramètres doivent être pris en compte.

L'un de ces paramètres dépend du type de transducteur utilisé. Les transducteurs les moins susceptibles de donner lieu à du couplage électromagnétique sont les écouteurs intra auriculaires dont la totalité de l'habitacle (hormis les tubes d'air), ainsi que les fils, sont encastrés dans une cage de Faraday, reliée à la masse électrique (unique) du dispositif expérimental. En effet, tous les autres transducteurs risquent d'engendrer des réponses contaminées, pour les intensités de stimulation habituellement utilisées. Les casques habituellement utilisés en audiologie (TDH 39 ou Sennheiser 250) sont donc à proscrire dans l'enregistrement de Speech ABR. Enfin, la comparaison des résultats de mesures effectuées chez des sujets normo-entendants et chez des patients présentant une cophose (surdité complète) unilatérale confirme que les réponses enregistrées sont bien d'origine neuronale.

3.3. Où est générée la réponse ?

Une fois le caractère non-artéfactuel de la réponse établi, encore faut-il définir les générateurs des PEASP. Les neurones du système auditif ont des caractéristiques fonctionnelles spécifiques, telles que des taux de décharges spontanées, des fréquences de réponse, ainsi que des types de décharge (transitoires, périodiques synchronisées, périodiques mais à une périodicité intrinsèque). De plus, chaque filtre auditif répond avec un décalage temporel (en raison du délai de propagation de la vibration de la membrane basilaire), et chaque noyau cellulaire répond également au son au fur et à mesure (avec des latences progressives d'une milliseconde en sortie de la cochlée, jusqu'à une dizaine de millisecondes dans le colliculus inférieur). A priori, aucune réponse identifiable ne devrait parvenir aux électrodes sur le scalp, ou alors noyée dans le bruit de fond résultant de l'ensemble de l'activité neuronale. Toutefois, les phénomènes électriques expliquant la propagation du message nerveux jusqu'au scalp ont récemment été décrits dans le cas des PEA en réponse à des clics pouvant être étendus aux PEA en réponse à un son de parole. En effet, d'un point de vue physique, le champ électromagnétique composite accompagnant le signal électrophysiologique peut être transmis jusqu'aux électrodes de scalp, lorsque le signal change de milieu biologique (tissu neuronal, liquide céphalo-rachidien, os, peau...) et lorsque suffisamment de neurones déchargent de façon synchrone (Joris et al, 2004). Les neurones thalamiques, à la différence de ceux du colliculus inférieur, ne semblent pas impliqués dans la genèse des PEASP.

Les travaux portant sur le modèle animal ont démontré une sensibilité des neurones du nerf auditif et du noyau cochléaire aux propriétés langagières du signal, comme la structure formantique (Delgutte 1980), la transition formantique (Delgutte et Kiang 1984), et le VOT (Clarey et al 2004). Les données cliniques ont toutefois mis en évidence le rôle du tronc cérébral, et notamment du colliculus inférieur dans le traitement de la parole chez l'humain, notamment par la description du cas d'un patient présentant une atteinte bilatérale des deux colliculi inférieurs associé à une agnosie auditive alors qu'il était indemne de toute lésion du lobe temporal (Johkura et al 1998). D'autres études viennent corroborer ces données, notamment à l'aide de modèles animaliers, chez lesquels ces régions semblent sensibles à la structure temporelle et spectrale des stimuli complexes (Eggermont et Ponton 2002) et seraient donc à même de jouer un rôle dans le traitement du signal de parole chez l'humain (Moller 1989).

Un argument supplémentaire en faveur de l'origine sous corticale de la RSF, est la différence d'amplitude entre les réponses d'origine corticale et sous corticale, en outre, les réponses corticales sont obtenues après un moyennage de 75–100 traces, moyennage bien inférieur à celui nécessaire pour obtenir une RSF exploitable. De plus, les réponses corticales se réduisent en amplitude avec l'augmentation du nombre de répétitions du stimulus, phénomène dénommé adaptation neurale (Grill-Spector et al, 2006) tandis que la RSF reste stable malgré le grand nombre de répétition (Johnson et al, 2008). Ces différences sont en accord avec la notion qu'à l'échelle du neurone, l'adaptation dépendante du stimulus est plus présente au niveau des neurones corticaux que sous corticaux (Ulanovsky et al, 2003). Par ailleurs, des variations de latence 10–25 ms sont considérées comme normales au niveau cortical, alors qu'une variation de moins de 1 ms est considérée comme anormale au niveau du tronc cérébral et associée à des difficultés de perception dans le bruit ou des troubles de l'apprentissage (Banai et al., 2005; Cunningham et al., 2001; King et al, 1999; Wible et al., 2004; Cunningham et al., 2002; Banai et al., 2009; Hornickel et al., 2009).

Un aspect universel concernant la maturation neurale est que les structures centrales mettent pour de temps à arriver à maturité en comparaison des structures périphériques. Or, des enregistrements de RSF semblable à celles d'un adulte peuvent être obtenus chez le nouveau-né, tandis que les potentiels évoqués corticaux recueillis chez les enfants ne sont pas semblables à ceux des adultes (Gardi et al, 1979). En effet, les potentiels corticaux ne sont pas complètement matures avant la fin de l'adolescence (Sussman et al, 2008; Suzuki & Hirabayashi, 1987). L'ensemble de ces données sont donc en faveur de l'origine sous corticale de la RSF et donc des PEASP.

Les PEASP apportent donc un aperçu précis du fonctionnement du colliculus inférieur. Ceci serait en particulier très utile pour faire le diagnostic différentiel des surdités centrales ou des neuropathies auditives (Delgutte, 1984).

3.4. Intérêts neurophysiologiques des potentiels évoqués auditifs en réponse à un son de parole:

3.4.1. Encodage spatial et temporel de l'information auditive

Quelles que soient leurs caractéristiques (voyelles constantes, présence d'un accent de langue, voyelles apériodiques - chuchotées -, voyelles fricatives sans voisement (Delgutte, 1980, 1984), les stimuli langagiers sont encodés de manière fidèle au niveau du tronc cérébral, l'activité de verrouillage de phase reflétée dans la FFR ayant été démontrée aussi bien pour des sons purs (Moushegian, 1973), des stimuli complexes (Greenberg et al, 1987) que pour des stimuli variant dans le temps (Krishnan & Parkinson, 2000 ; Lee et al, 2009).

L'étude des PEA en réponse à un son de parole a permis de caractériser de manière plus précise ce codage temporel et spatial de l'information: à faible intensité, la réponse temporelle à chaque stimulus est maximale dans les fibres nerveuses auditives ayant une fréquence caractéristique proche de celle du stimulus, suggérant une spécificité spatiale de la réponse. Ainsi, en augmentant l'intensité de stimulation, il existerait une diffusion spatiale jusqu'aux unités de hautes fréquences pour le codage des formants (Krishnan, 2002). Une étude récente a permis de mettre en évidence un effet de l'intensité du son de parole sur le temps de latence du signal PEASP. En effet la diminution de l'intensité de stimulation (entre 60 et 0 dB SL) entraîne un allongement de temps de latence des PEASP : de 7 à 10 ms pour la réponse impulsionnelle et de 15.6 à 23 ms pour la RSF. Toutefois, l'allongement du temps de latence de la réponse impulsionnelle (-0.6 ms pour 10 dB) est moins important que celui observé pour la RSF (-1.4 ms pour 10 dB), suggérant l'existence de deux voies séparées de traitement de l'information au niveau du tronc cérébral, avec des générateurs différents pour la réponse impulsionnelle et la RSF (Akhoun et al, 2008).

Le nerf auditif et les fibres du noyau cochléaire augmentent leur verrouillage de phase (aux harmoniques des formants) mais il apparaît que des populations neuronales différentes encoderaient les premier et deuxième formants (Johnson et al, 2008) qui seraient suffisants à l'identification de la plupart des voyelles (Krishnan, 2002). Un « seuil » d'encodage spatial et de verrouillage de phase de l'information auditive dans les situations de la vie courante pourrait donc être obtenu pour chaque sujet sur l'ensemble de la gamme fréquentielle (Johnson et al, 2005). Toutefois, l'activité de verrouillage de phase serait limitée à la

fréquence de coupure au niveau du colliculus inférieur, qui est d'environ 1000 Hz (Batra et al, 1986 ; Galbraith et al, 1998).

Un même phonème peut revêtir, selon le contexte, une multitude de patterns acoustiques et le phénomène de co-articulation est prégnant surtout pour les consonnes, les voyelles montrant un pattern acoustique plus stable (Liberman et al, 1967). D'un point de vue électrophysiologique, l'enregistrement des réponses neuronales dans le bruit à différents sites sous-corticaux et corticaux chez l'animal (cobaye) montre un encodage sous-cortical plus robuste de la voyelle dans le bruit, les voyelles ayant une périodicité plus longue et plus forte que les consonnes (Cunningham et al, 2002 ; Russo et al, 2004). En condition de bruit, la transition formantique et les voyelles (RSF) restent encodées au niveau du tronc cérébral, alors que la réponse impulsionnelle est très réduite ou éliminée. Les problèmes perceptifs d'identification de la consonne dans le bruit peuvent donc être attribués à une diminution de la synchronisation neuronale de la réponse impulsionnelle, alors que la préservation de la RSF permettrait la perception de la voyelle (Russo et al, 2004). Cette résistance de la RSF dans le bruit par rapport à la réponse impulsionnelle laisse supposer une certaine différence dans leur processus d'encodage au niveau du tronc cérébral. De plus, l'analyse des PEASP montre que l'encodage de la fréquence fondamentale (qui est importante pour identifier le locuteur, la hauteur tonale -pitch - et percevoir la prosodie) persiste dans le bruit, alors que l'Onset-Response et le premier formant (essentiel à la perception phonémique) se dégradent (Russo et al, 2004).

La perception de la prosodie (intonation, stress), liée à F0, peut être déficiente dans certaines pathologies comme l'autisme. En effet, Russo et al. (2004) ont mis en évidence, sur des PEASP recueillis dans un groupe de 21 enfants présentant un syndrome autistique, une diminution de l'encodage de F0, ce qui pourrait être lié à une moins bonne perception de significations linguistiques (par exemple: affirmation ou question) et de l'émotion véhiculée par la parole.

Pour certains, la détection de la hauteur tonale par le tronc cérébral serait expérience-dépendante, malléable et liée au traitement d'autre facteurs cognitifs comme le langage et la musique : les musiciens encoderaient F0 de manière plus forte et avec un verrouillage de phase plus résistant en comparaison à des sujets non musiciens (Russo et al, 2008 ; Song et al, 2008 ; Wong et al, 2007). En effet, la pratique et l'écoute régulière de la musique feraient notamment intervenir des réseaux cognitifs supérieurs exerçant un effet modulateur corticofuge (modulation top-down) ayant pour conséquence une amélioration des capacités d'encodage au niveau sous-cortical. Cette intervention corticofuge expliquerait en partie, sur le plan neurophysiologique, les facilités d'apprentissage du langage présentes chez les musiciens (Lee et al, 2009). Enfin, l'asymétrie hémisphérique fonctionnelle liée aux traitements du langage a été mise en évidence également sur les réponses de type PEASP (Hornickel et al, 2008). En effet, l'existence d'une asymétrie fonctionnelle du traitement du langage est bien établie au niveau cortical (Habib et Robichon, 1998) : l'hémisphère gauche serait spécialisé dans le traitement de la parole, de par les caractéristiques acoustiques de celle-ci (événements acoustiques rapides associés aux transitions temporelles courtes, alors que la musique, dont les événements temporaux et spectraux sont plus soutenus et associés à des transitions plus lentes, serait traitée au niveau de l'hémisphère droit (Hornickel et al, 2008). Au niveau sous-cortical, des études électrophysiologiques ont montré des réponses d'apparition plus rapide et comportant des harmoniques correspondant au premier formant du stimulus plus robustes, lors d'une stimulation langagièrre monaurale droite par rapport à une stimulation controlatérale (Strait et al, 2009). Cette asymétrie périphérique a été également mise en évidence au niveau

cochléaire et pourrait être liée au rétro-contrôle cortico-sous-cortical mis en évidence chez l'humain sur les oto-émissions acoustiques (Perrot et al, 2006). Ainsi, l'hypothèse d'une asymétrie périphérique présente dans les PEASP apparaît tout à fait plausible.

3.4.2. Applications cliniques potentielles des PEASP

L'intérêt des PEASP réside dans les nombreuses informations théoriques et cliniques que ceux-ci nous donnent sur le traitement des stimuli auditifs complexes au niveau du tronc cérébral (Johnson et al, 2005). Les PEASP peuvent ainsi être utilisés comme marqueurs biologiques d'une déficience d'encodage des sons au niveau du tronc cérébral (King et al, 2002 ; Wible et al 2005). Ainsi, le recueil des PEASP dans le bruit permet d'objectiver la dégradation de l'encodage des sons de parole, avec une moins bonne résistance de la consonne par rapport à la voyelle, chez des individus normaux (Liberman et al, 1967) et par extension chez des personnes présentant des troubles de l'intelligibilité.

Les difficultés de traitement du signal au niveau central observées chez des enfants présentant des troubles des apprentissages, se traduisent sur les PEASP par une diminution de la réponse RSF, une amplitude plus faible et une durée allongée de la réponse impulsionnelle (complexe V-A) (Wible et al, 2004). Ceci suggère qu'une représentation appauvrie des composants-clés du signal de parole au niveau du tronc cérébral, contribuerait aux difficultés du traitement du même signal à un niveau plus central. Différents groupes peuvent ainsi être individualisés au sein de ces mêmes enfants, selon qu'une corrélation est retrouvée ou non entre les mesures effectuées au niveau de leur tronc cérébral (PEASP) et celles effectuées au niveau cortical. Lorsque cette corrélation existe, les PEASP sont un outil précieux pour le clinicien, puisqu'ils permettent de juger de l'efficacité des programmes de réhabilitation (Russo et al, 2005). L'absence d'anomalie des PEASP chez certains enfants souffrant de troubles des apprentissages a été interprétée comme la conséquente modulation corticofuge du traitement de la parole par le tronc cérébral (Song et al, 2006, 2008). Un système de mesure dédié à l'enregistrement des réponses du tronc cérébral à des sons complexes (BioMAP: Biological Marker of Auditory Processing) a été proposé par Kraus et collègues afin d'identifier, chez les enfants, les troubles des apprentissages accessibles à une rééducation.

3.5. Que représentent le « what » et le « where » au niveau du système auditif ?

La notion de « what » et « where » fut en premier lieu appliquée au système visuel. En effet, depuis maintenant plus de vingt ans, deux schémas de traitement ont été décrits au niveau du système visuel : la voie dorsale et la voie ventrale (Ettlinger, 1990). Ces deux voies sont responsables de l'identification d'un objet : le « what » et de sa localisation : le « where », cette séparation en deux voies ne se limitant pas à la rétine, mais se prolongeant jusqu'au cortex visuel. Une dichotomie fonctionnelle similaire fut proposée pour le système auditif (Rauschecker, 1998), avec des axes de recherche similaires, à savoir la localisation ou la composition d'un « objet » auditif (Romanski et al, 1999 ; Kaas & Hackett, 1999).

Bien que les voies du « what » et du « where » soient toutes deux impliquées dans la perception de la parole, il existe une certaine tendance à reléguer la perception de la

localisation d'un objet au niveau de la voie du « where », tandis que l'identité de l'objet auditif est attribuée au « what ». Belin et Zatorre (2000) ont proposé un rôle de la voie du « where » dans la localisation ou les mouvements dans l'espace ainsi que sur une variation spectrale : le « where » en fréquence. Ils dessinèrent un parallèle entre les mouvements de la lumière à travers le champ visuel et le mouvement des stimulations acoustiques à travers la carte fréquentielle de la membrane basilaire dans la cochlée. Cela pose les fondations de la distinction voyelle, consonne et phonème composant le message.

La littérature sur la perception de la parole, nous propose elle aussi une dichotomie quant aux caractéristiques acoustiques de la parole. Le modèle source versus filtre décrit la vibration des cordes vocales comme la source tandis que le tractus vocal, la cavité orale, la langue, les lèvres et les joues seraient à l'origine du filtre (Fant, 1960). De manière schématique, les voyelles et les consonnes, c'est-à-dire le contenu linguistique, dépendraient de la formes du filtre reposeraient sur celui-ci, tandis que les informations non linguistiques (telles que le ton de la voix, sexe du locuteur, état émotionnel...), reposeraient sur les caractéristiques de la source.

Le filtre serait assimilé à la voie du « where », tandis que la source dépendrait de la voie du « what ».

L'étude des PEASP a permis de mettre en évidence l'encodage de la source (F0) et du filtre (onset, onde O, et transition formantique), caractéristiques contenues dans le signal de parole, proposant ainsi l'idée que les voies de traitement du "what" et du "where" (Belin & Zatorre, 2000; Kaas & Hackett, 1999; Romanski et al., 1999) pourraient avoir une origine sous corticale, au niveau du tronc cérébral (Kraus & Nicol, 2005) (Fig 13).

D'un point de vue clinique, il a été montré que la représentation du filtre au niveau du tronc cérébral puisse être altérée chez certaines personnes présentant des troubles d'apprentissage, tels que la dyslexie (Hornickel et al, 2009; Banai et al, 2009; Banai et al, 2007; Cunningham et al, 2001; King et al, 2002; Wible et al, 2004). A contrario, une atteinte préférentielle de la voie du traitement du « what » est retrouvée chez les enfants présentant des troubles d'ordre autistique (Russo et al, 2008).

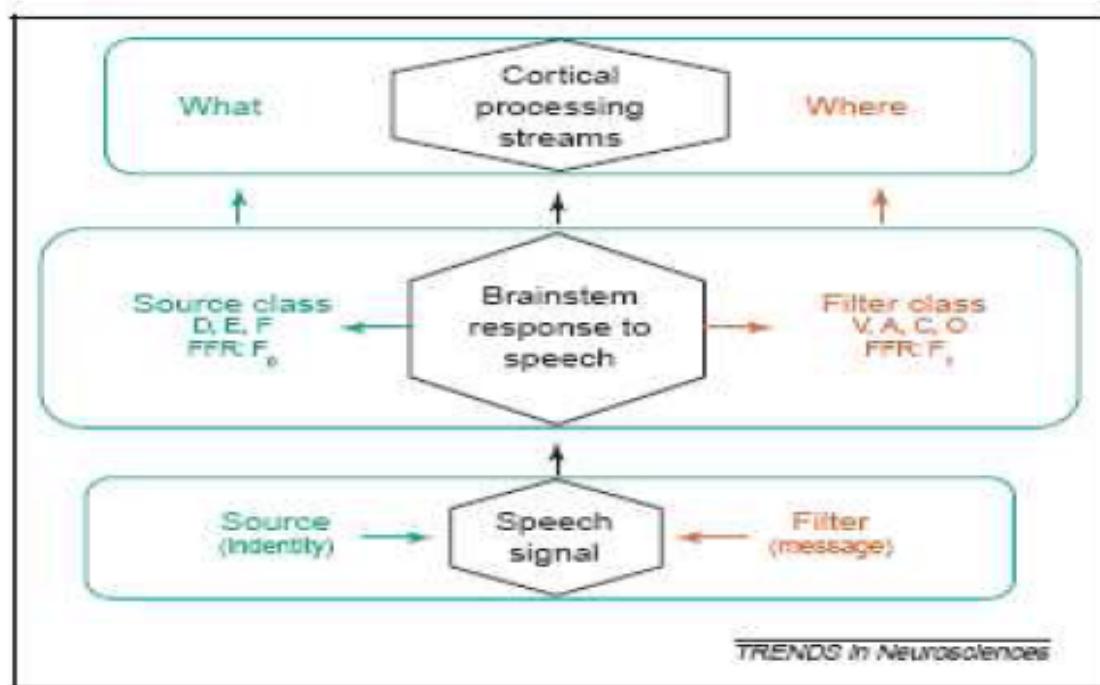


Figure 13: Tirée de Kraus et al, 2005. Représentation schématique du tronc cérébral sous la forme d'un médiateur entre propriétés acoustiques du signal et traitement cortical.

En bas : un locuteur encode ses pensées en sons audibles par l'utilisation de son tractus vocal qui va filtrer le son émis par la vibration des cordes vocales (la source). Au milieu : le son est transformé en signal électrique au niveau de la cochlée du sujet écoutant le message, le nerf auditif transmet le signal au tronc cérébral où les composants de la source et du filtre du signal sont préservés. Cette préservation étant quantifiable dans les différentes réponses des PEASP. La source correspond aux mesures de F0, tandis que le filtre est représenté par l'onset, l'offset et F1. La source est résistante au bruit et au taux de stimulation et n'est pas perturbée chez les sujets présentant des troubles de l'apprentissage, tandis que le filtre y est vulnérable. En haut: séparation des composants de la parole dans le tronc cérébral, comme précurseurs des voies what et where corticale.

IV. Partie expérimentale :

La perception de la parole est un phénomène complexe qui fait appel à différentes étapes d'intégration et d'analyse du signal acoustique tout le long du tractus auditif, étapes dont la réalisation peut être rapidement altérée dans la vie de tous les jours, notamment par le bruit environnant.

Les différentes expériences réalisées et rapportées ici ont pour but la mise en évidence de cette dégradation de la perception de la parole en condition adverses et ce au moyen des PEASP. Le but final était d'obtenir des données individuelles mettant en évidence l'implication du traitement sous cortical des sons de parole dans la perception dégradée de ceux-ci en situation de bruit chez des personnes normo-entendantes, afin de pouvoir, à terme, étendre cette recherche à des personnes présentant des troubles de l'intelligibilité de la parole.

La première expérience avait pour but l'étude de l'encodage d'un son langagier signifiant par rapport à un homologue, ainsi que l'encodage du voisement au niveau sous cortical. Le but étant la compréhension du rôle de l'enveloppe temporelle d'un son langagier dans la genèse des PEASP. Les résultats obtenus lors de cette expérience mettent en évidence que la RSF ne contiendrait pas que l'information liée à l'enveloppe temporelle.

Mon second travail était l'objectivation de la modification de l'encodage des signaux langagiers dans différentes conditions de bruit. Les résultats montrent un meilleur encodage des signaux de parole dans le bruit blanc par rapport aux autres conditions, toutefois ces résultats durent être nuancés, les différentes conditions de bruit ne pouvant être énergétiquement comparables. Par ailleurs, lors de cette seconde expérience, il persistait un problème que nous n'arrivions pas à résoudre jusqu'alors : l'analyse des résultats à l'échelon individuel. Il était donc nécessaire d'effectuer une optimisation de la chaîne d'acquisition des PEASP, de même que la mise au point d'une méthode d'analyse automatisable avant de poursuivre toute expérimentation.

Ainsi, le troisième volet de mon travail a eu pour objectif l'obtention de résultats individuels exploitables, ainsi que la mise au point d'un système d'analyse automatique, venant en complément de l'analyse visuelle. Le fruit de ce travail fut visible lors des expériences suivantes, même si quelquefois certaines traces n'ont encore pu être totalement exploitées.

Cette mise au point effectuée, il m'a été possible de reprendre le sujet premier de ma thèse, à savoir l'encodage de la parole dans le bruit. Pour cette 4^{ème} expérience, il a été décidé d'effectuer l'étude de d'un son de parole dans le silence et de comparer son encodage sous cortical lorsque celui-ci était « parasité » par un autre son non signifiant, ainsi que d'étudier l'influence de l'intensité de stimulation sur les caractéristiques des PEAPSP. Les résultats montrèrent la nette dégradation des PEASP dans le bruit avec certaines particularités concernant l'effet de l'intensité. En effet, chez certains sujets, on identifia facilement une réponse de type PEASP avec un stimulus /ba/, alors que cela n'était plus possible avec un stimulus /ba+ba/. Ces observations nous amenèrent à proposer un protocole d'étude systématique des fonctions entrée/sorties, protocole décrit dans l'expérimentation 6.

L'expérience 5 quant à elle reprenait le projet d'étude de l'encodage du voisement au niveau sous cortical, projet qui avait été évoqué lors de la première expérience. Les résultats obtenus mirent en évidence un encodage du voisement dès l'étape sous corticale avec un rétrocontrôle cortical.

La dernière expérience fut donc la suite logique de l'expérience 4, le but étant la connaissance du seuil de détection d'une réponse de type speech ABR chez le normo-entendant avant de pouvoir disposer d'une application clinique chez les patients malentendants.

Expérience 1 : Protocole expérimental
Objectivation du traitement sous cortical d'un son de parole
par rapport à celui d'un son complexe

Problématique :

Evaluation de l'importance de l'enveloppe temporelle dans les PEASP : rôle de la structure fine et de l'enveloppe dans leur genèse.

Mise en évidence d'un encodage éventuel de la consonne dans la RSF.

Matériels et méthodes :

•Population étudiée :

Quinze sujets droitiers normo-entendants, âgés de 20 à 30 ans.

•Tests comportementaux :

* Audiométrie vocale utilisant les listes de Lafon

* Un test de perception catégorielle du voisement :

-utilisant le continuum de 18 *signaux de parole /ba/ /pa/*.

-utilisant le continuum de 18 *signaux homologues langagiers* du continuum /ba/ /pa/.

•Tests neurophysiologiques : les PEASP (durée d'acquisition par sujet : 70 minutes)

Les deux signaux extrêmes des continuums de stimuli utilisés lors des tests de perception catégorielle seront utilisés après raccourcissement de la voyelle, ainsi que des homologues non langagiers correspondants (durée de 230 ms chacun).

Stimulation binaurale à un niveau de 45 dB SL, par blocs de 1200, à une cadence de 6,6 stimuli par seconde.

Recueil des traces à l'aide du centor USB Racia Alvar®, sur 2 voies d'enregistrements, via 4 électrodes d'argent (Nicolet®, 10mm), fixées sur le scalp du sujet au niveau du vertex (électrode active), du front (masse), et des deux mastoides (droite pour la voie 1 et gauche pour la voie 2).

Résultats :

Les RSF en réponse aux homologues sont moins bien synchronisées que celles observées en réponse au /ba/ et au /pa/.

La RSF ne contient pas que l'information liée à l'enveloppe temporelle.

Expérience 1 : Objectivation du traitement sous cortical d'un son de parole par rapport à celui d'un son complexe

1.1. Introduction

La compréhension du langage apparaît comme un processus rapide et automatique, faisant intervenir des processus corticaux et sous corticaux. Comme nous venons de le voir, il est possible d'objectiver de manière non invasive le traitement sous cortical des sons de parole au moyen des PEASP. Toutefois l'encodage des différentes caractéristiques intrinsèques des stimuli reste encore méconnu. Ainsi, mon premier travail porta sur l'étude des différentes parties des PEASP en réponses à des syllabes /ba/ et /pa/ ainsi qu'à leurs homologues. Conjointement nous souhaitions pouvoir corrélérer ces résultats neurophysiologiques avec ceux obtenus lors de tests de perception, afin de pouvoir obtenir des valeurs de références et des corrélats neurophysiologiques.

Deux approches sont possibles pour évaluer l'intelligibilité de la parole : soit en mesurant *subjectivement* la perception auditive du sujet à travers des tests psycho-acoustiques, soit en enregistrant *objectivement* l'activité neurophysiologique évoquée par un stimulus sonore dans le système auditif.

Dans cette étude, nous avons souhaité évaluer l'importance de l'enveloppe temporelle dans le PEASP. Est-ce que le PEASP n'est que le reflet pur-et-simple de l'enveloppe, ou bien est-ce que la structure fine joue également un rôle dans la génération de la RSF ? Est-ce que seule la voyelle est codée par la RSF, ou est-ce que la consonne intervient également ? Pour élucider ces questions, nous avons comparé les PEASP engendrés par des sons de parole réels /ba/ et /pa/ (qui ont rigoureusement la même voyelle), ainsi que par leurs homologues. Les homologues furent synthétisés comme une somme de sinusoïdes à la fréquence des formants, modulés par l'enveloppe temporelle des sons de parole. Les sons de parole /ba/ et /pa/ furent choisis plus naturels que ceux, synthétiques, utilisés dans les études précédentes, en effet les signaux utilisés sont ceux enregistrés à partir de voix humaines par Cazals, avec une durée supérieure à ceux précédemment utilisés (durée de 230 millisecondes). Cette étude évalue l'importance de la nature du stimulus pour les composants du PEASP. La comparaison des PEASP en réponse à des sons de parole et leurs homologues permet d'apporter des précisions sur la possibilité éventuelle d'une spécificité du traitement de la parole dès le tronc cérébral. En effet, il existe deux hypothèses divergentes : selon la première, le langage est traité dans les aires intégrées corticales (spécifiques de l'homme), ainsi le traitement sous-cortical ne permettrait pas une analyse des caractéristiques des sons de parole (Zatorre & Gandour, 2008). La seconde s'appuie sur des travaux portant sur l'animal ayant mis en évidence une sensibilité de certaines cellules sous-corticales à des sons mimant les voyelles (vowel-like sounds, Palmer & Shamma), et par extrapolation à l'être humain, pourrait laisser supposer un traitement en partie sous cortical de la parole. L'utilisation des homologues ou « analogues » avait pour but de disposer de sons complexes ayant la même enveloppe temporelle que /ba/ et /pa/ tout en ayant une structure fine appauvrie, stationnaire et ne reproduisant pas la dynamique formantique nécessaire à la bonne intelligibilité du son de parole.

1.2. Matériel et méthode

1.2.1. Population étudiée :

Quinze sujets, âgés de 20 à 30 ans ont participé à l'étude. Les sujets inclus ne présentaient aucun antécédent médical particulier notamment otologique, et tous avaient un niveau d'étude supérieur. Chaque sujet présentait une audiométrie tonale normale (seuils normaux entre 125 et 8000 Hz, testés à l'aide d'un Audiomètre AC 40 d'Interacoustics®), ainsi que

des potentiels évoqués auditifs en réponse à des clicks (centor USB Racia alvar®) normaux (élimination d'une éventuelle pathologie retro cochléaire). L'audiométrie vocale, réalisée à l'aide des listes de mots de Lafon, était normale. Ces tests ont été réalisés en accord avec la loi Huriet (accord du comité d'éthique de Lyon, CCPPRB # 05-024) et un consentement éclairé a été obtenu auprès de chaque sujet.

1.2.2. Tests comportementaux :

Un test de perception catégorielle du voisement :

- utilisant le continuum de *signaux de parole /ba/ /pa/*.
- utilisant le continuum de *signaux homologues langagiers* du continuum /ba/ /pa/.

La frontière perceptive de chaque sujet a été recherchée à l'aide du test de perception catégorielle utilisant le continuum de 18 signaux allant de /ba/ à /pa/, la durée de chaque signal étant décroissante à partir du premier /ba/ (300ms). Le continuum est présenté, en stimulation binaurale via un casque Sennheiser linear 250®, cinq fois pour chaque sujet, l'ordre de présentation étant pour chaque présentation, randomisé. Le sujet doit en réponse au stimulus entendu cliquer sur un écran placé en face de lui où sont visibles deux mots : « ballon » et « papillon ». Le sujet doit, soit cliquer sur le mot « ballon » s'il lui semble avoir entendu /ba/, soit cliquer sur « papillon » s'il lui semble avoir entendu /pa/. L'ensemble du continuum /ba/ /pa/ est représenté en temporel, sur la fig. 14.

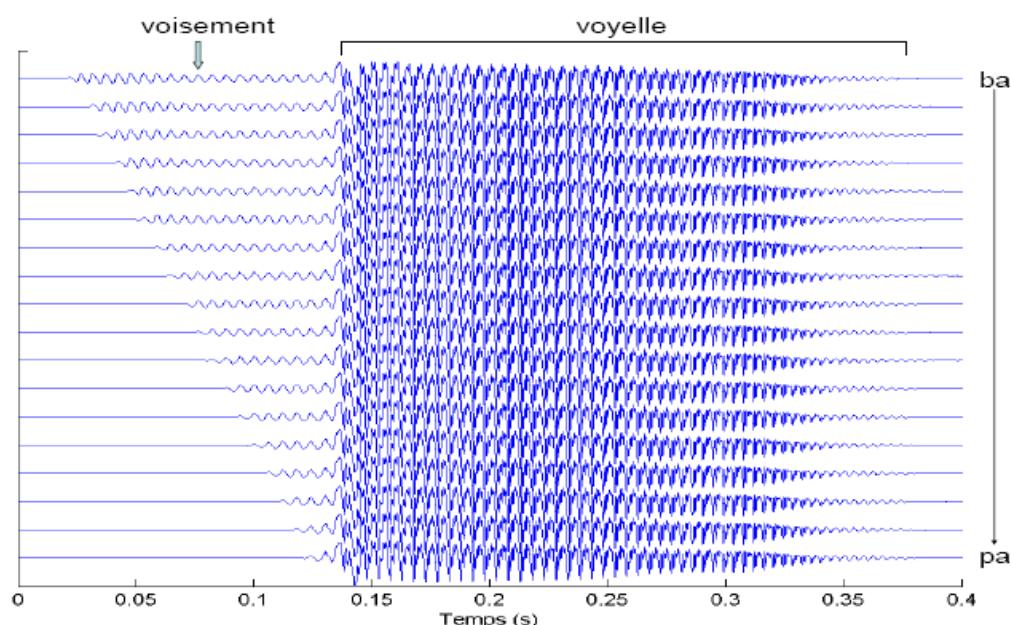


Figure 14 : Représentation temporelle des 18 signaux constitutifs du continuum langagier ba-pa.

Le premier /ba/, situé tout en haut sur le continuum, possède la durée maximale pour ce continuum (350ms). La voyelle est strictement identique le long du continuum, seul, le délai de voisement est modifié de manière progressive.

1.2.3. Tests neurophysiologiques :

1.2.3.1. Stimuli et système de stimulation:

Les deux signaux extrêmes des continuums de stimuli utilisés lors des tests de perception catégorielle ont été utilisés : le contraste /ba/-/pa/, en diminuant toutefois la durée de la voyelle afin de ne pas avoir un temps de passation trop long.

Les syllabes /ba/ et /pa/ contiennent la même voyelle et ne diffèrent que par la durée de leur consonne respective (délais de voisement respectifs de -55ms et 10ms). Chaque stimulus dure 230ms, avec comme dans le continuum utilisé dans le test de perception catégorielle le début de la voyelle à 140ms. La fréquence fondamentale de /a/ est stationnaire autour de 200 Hz et ses quatre formants (F1 à F4) varient quant à eux autour de 750, 1660, 2800 et 4400 Hz.

Les homologues non langagiers correspondants ont été obtenus par extraction de l'enveloppe du /ba/ de Cazals (/ba/ de référence et /ba/ correspondant à la frontière perceptive du sujet), multipliée par la somme des cinq sinusoïdes, correspondant à F0 et aux différents formants de la voyelle /a/ de /ba/ et /pa/, chacune pondérée selon la hauteur du formant par rapport à F0, et modulées par l'enveloppe temporelle des syllabes, (traitement inspiré par les travaux de Serniclaes et al 2001)

soit selon des sinusoïdes pures :

$$\text{signal} = A1 * \sin(2\pi f1 t) + A2 * \sin(2\pi f2 t) + A3 * \sin(2\pi f3 t)$$

soit avec l'addition d'une modulation de faible amplitude :

$$\text{signal(one period)} = [A1 * \sin(2\pi f1 t) + A2 * \sin(2\pi f2 t) + A3 * \sin(2\pi f3 t)] * e^{(-t/0.05)}$$

Ainsi la fréquence de chaque sinusoïde correspond aux fréquences respectives de F0, F1, F2, F3 et F4 de la voyelle /a/. Le stimulus est ensuite filtré passe-bas à 500 Hz par un filtre à correction de phase de Butterworth d'ordre 4 (car ce filtre possède un gain plat dans la bande passante avec une pente d'ordre 4 pour couper le plus rapidement possible). Le filtrage passe bas a été choisi par rapport à la fréquence de coupure du PEASP qui dans notre équipe a précédemment été observée à 500Hz (Akhoun et al, 2008).

Les homologues obtenus ne ressemblent perceptivement pas à la syllabe de départ (cela étant du en grande partie à l'appauprissement spectral au-delà de 500 Hz). Les stimuli naturels et leurs homologues sont représentés sous forme temporelle et spectrale (fig 15 et 16).

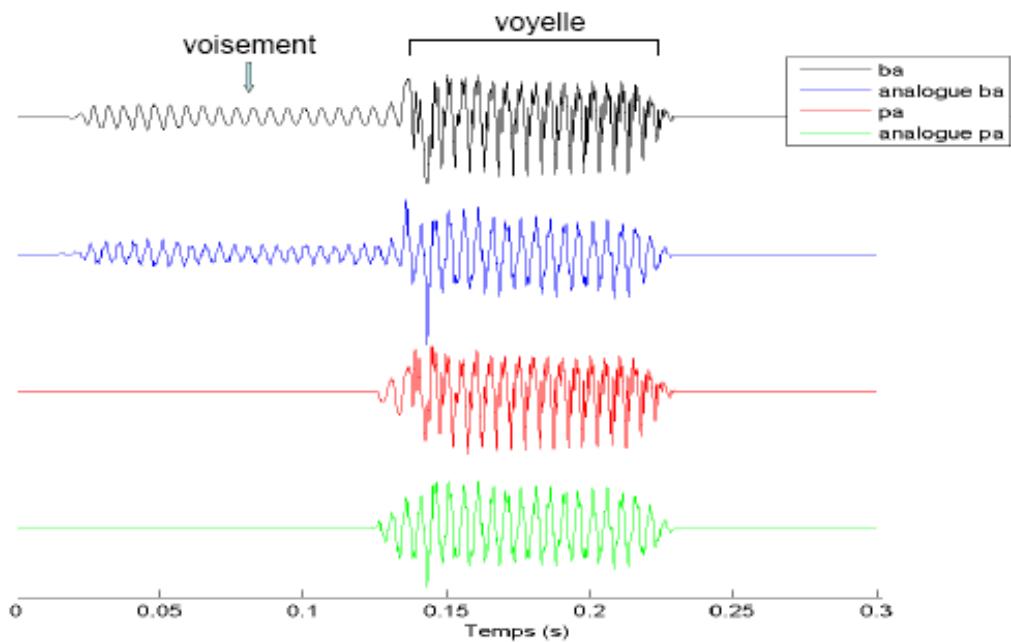


Figure 15: Représentation temporelle des stimuli langagiers et homologues (ou « analogues »).

En partant de la partie supérieure de la figure, les deux premiers stimuli correspondent au /ba/ et à son analogue, de durées similaires (230ms). Les deux stimuli suivants correspondent au /pa/ qui n'est autre que le dernier stimulus du continuum utilisé lors du test de perception catégorielle. Ces stimuli sont tous tirés du continuum langagier et non langagier du test de perception catégorielle avec une durée raccourcie par rapport à ceux-ci, au niveau de la voyelle.

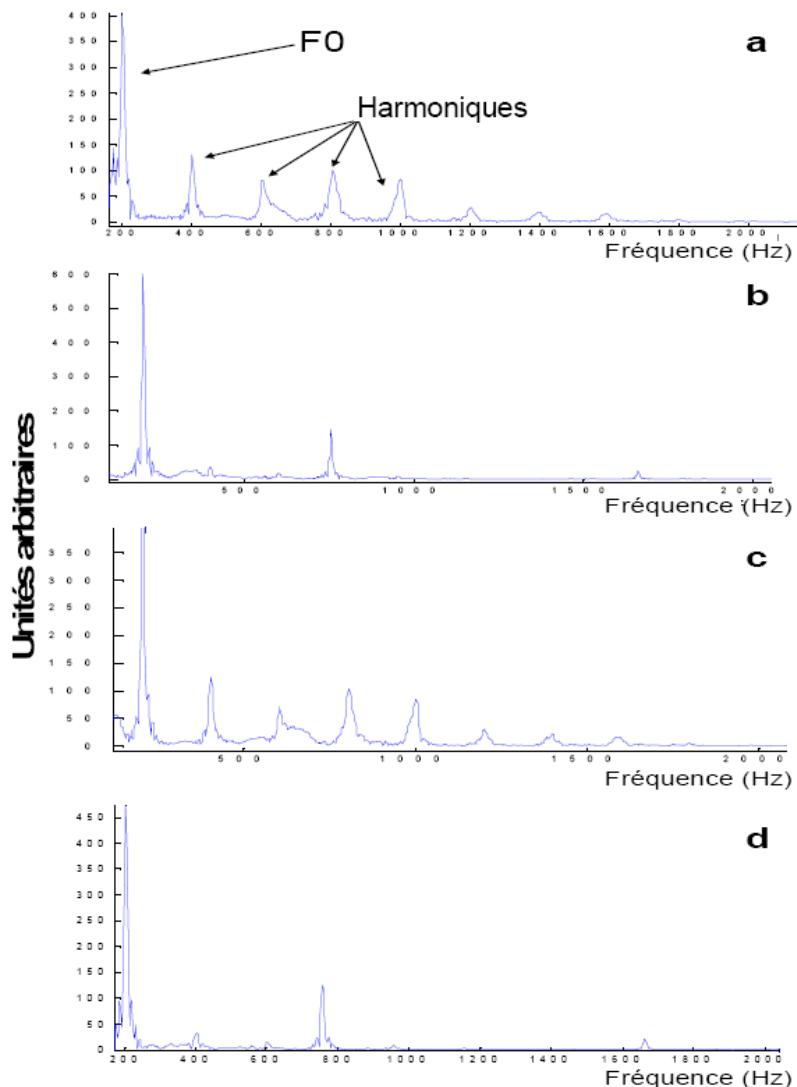


Figure 16 a,b,c,d: Représentation fréquentielle des stimuli langagiers et non langagiers.

a : stimulus /ba/ ; b : « homologue » du /ba/ ; c : stimulus /pa/ ; d : « homologue » du /pa/. Les fréquences sont représentées en abscisse, tandis qu'en ordonnée est représentée l'amplitude des pics obtenue au moyen de la Transformée de Fourier rapide (ou « Fast Fourier Transform, FFT »)

Comme nous pouvons le voir sur la figure 15, dix millisecondes de pré moyennage précédait le début du phonème /ba/ contre cent quarante millisecondes dans le cas du phonème /pa/, ceci nous permettant de quantifier le bruit ambiant sur les traces recueillies, lors de ces 10 ms. Les stimuli utilisés étaient présentés à l'aide du logiciel Audacity®, (logiciel de traitement des sons disponible gratuitement sur internet), pour jouer les sons stockés en format *.wav. Les sons étaient joués par blocs de 1200, à une cadence de 6,6 stimuli par seconde. La stimulation était binaurale à un niveau de 45 dB SL (dB *Sensation Level*, au-dessus du seuil du sujet), niveau sonore confortable, voire fort, mais non gênant. L'intensité sonore était ajustée par un audiomètre AC-40 (Interacoustics®), qui relie la carte-son Soundblaster Audigy® aux transducteurs acoustiques : des inserts-

earphones (Biologic®). Les inserts sont équivalents au casque acoustique dans la qualité de transmission du son au sujet, mais permettent une diminution de l'artefact électromagnétique par rapport au casque. Les inserts sont blindés par une cage de Faraday reliée à la masse électrique du dispositif expérimental afin d'éliminer le risque d'artefact électromagnétique ; le son étant diffusé au bout d'une paire de tubes de caoutchouc. Le transfert de la stimulation acoustique dans les tubes d'air long de 30 cm induit 1ms de délai, qui est ensuite retranché dans les enregistrements.

1.2.3.2. Système de recueil :

Les différentes acquisitions ont été effectuées à l'aide du centor USB Racia Alvar®, sur 2 voies d'enregistrements, via 4 électrodes d'argent (Nicolet®, 10mm), fixées sur le scalp du sujet au niveau du vertex (électrode active), du front (masse), et des deux mastoïdes (droite pour la voie 1 et gauche pour la voie 2). Pour chaque expérimentation, la valeur des impédances de chaque électrode était maintenue en dessous de 5 Ω . Le dispositif expérimental utilisé est représenté sur la figure 17. Le signal neural recueilli sur les électrodes est ensuite pré-amplifié et traité par le Centor USB, dont les paramètres d'enregistrement (filtrage lors de l'acquisition, fréquence d'échantillonnage (13.3kHz), taille de la fenêtre d'acquisition, inclusion ou exclusion des rejets, nombre de moyennages souhaité), peuvent être réglés sur le PC de contrôle. La fenêtre d'acquisition est de 300 millisecondes. Les PEASP ainsi enregistrés sont filtrés en temps réel entre 80 et 3200 Hz, afin de se concentrer sur la gamme d'activité des potentiels du tronc cérébral auditif. Chaque trace trop ample (au-delà de 75 μ V) est assimilée à un artefact musculaire et de ce fait rejetée directement sur le Centor USB®. L'acquisition des Speech ABR est synchronisée sur l'envoi des stimuli par un module de synchronisation « triggering » reliant la carte son à l'entrée synchronisation du Centor USB, à partir d'une impulsion placée sur la voie de gauche du fichier wav, tandis que le stimulus, envoyé vers l'audiomètre, était sur la voie de droite du même fichier stéréo de stimulation (.wav). La chaîne de stimulation est contrôlée depuis le même PC que pour l'acquisition.

Dans le but d'éliminer tout artefact lié au potentiel microphonique cochléaire, nous avons enregistré les PEASP en polarité alternée, c'est-à-dire le stimulus et son opposé (800 blocs en polarité positive, et 800 blocs en polarité négative).

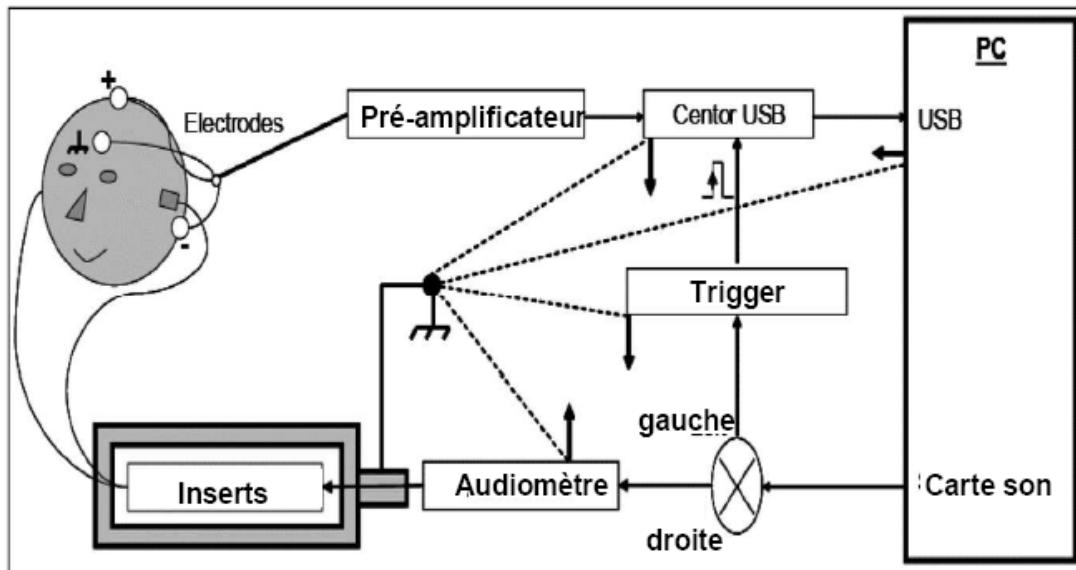


Figure 17 : schéma général de stimulation/recueil des PEASP, tiré de Akhoun et coll., 2008.

Le même PC contrôle la stimulation et l'acquisition. A la sortie du PC, la voie de droite transmet le stimulus à l'audiomètre, puis au sujet via les inserts, alors que la voie de gauche rejoint un module de synchronisation « triggering » reliant la carte son à l'entrée synchronisation du Centor USB permettant ainsi l'acquisition des Speech ABR de manière synchronisée sur l'envoi des stimuli. Le recueil des traces est effectué à l'aide de 4 électrodes cutanées placée sur le sujet au niveau de chaque mastoïde, du front (masse) et du vertex (électrode positive), dont les traces sont envoyées au PC via le pré-amplificateur et le centor USB.

1.2.3.3. Traitement des données

Dans cette première expérience, le traitement des données utilisé s'inspire des travaux de l'équipe de Kraus et al. (Cunningham et al, 2002 ; Russo et al, 2004 ; Wible et al, 2005 ; Song et al, 2006 ; Banai et al, 2007 ; Hornickel et al, 2008), qui utilise à la fois des méthodes subjectives avec identification des différentes caractéristiques des traces par des observateurs individuels, et des méthodes mathématiques.

L'ensemble de la visualisation des données et des analyses mathématiques sont effectuées à l'aide du logiciel Matlab®.

1. Analyse dans le domaine temporel : Du fait de la difficulté d'obtenir des traces individuelles avec une réponse claire dans le domaine temporel, les moyennes des traces obtenues pour chaque stimulus, pour tous les sujets testés, sont calculées (grandes moyennes), et permettent une analyse visuelle. Les grandes moyennes représentent le résultat de la somme algébrique des données normalisées (soustraction de la composante continue et correction des erreurs standards par loi normale centrée) de chaque sujet dans une polarité donnée (positive, négative, alternée).

2. Analyse dans le domaine fréquentiel : Les résultats individuels, ainsi que les grandes moyennes sont représentés sous forme de spectrogrammes d'énergie. Les spectrogrammes sont obtenus avec Matlab en calculant la transformée de Fourier sur des

fenêtres glissantes (1024 points pour les PEASP et 2048 pour les stimuli) à partie des traces temporelles normalisées (Fig 18). De la même manière que pour l'analyse temporelle, la grande moyenne des spectrogrammes fut obtenue par la somme des spectrogrammes individuels. Les spectrogrammes permettent d'identifier la bande de fréquence privilégiée de la RSF, au cours du temps.

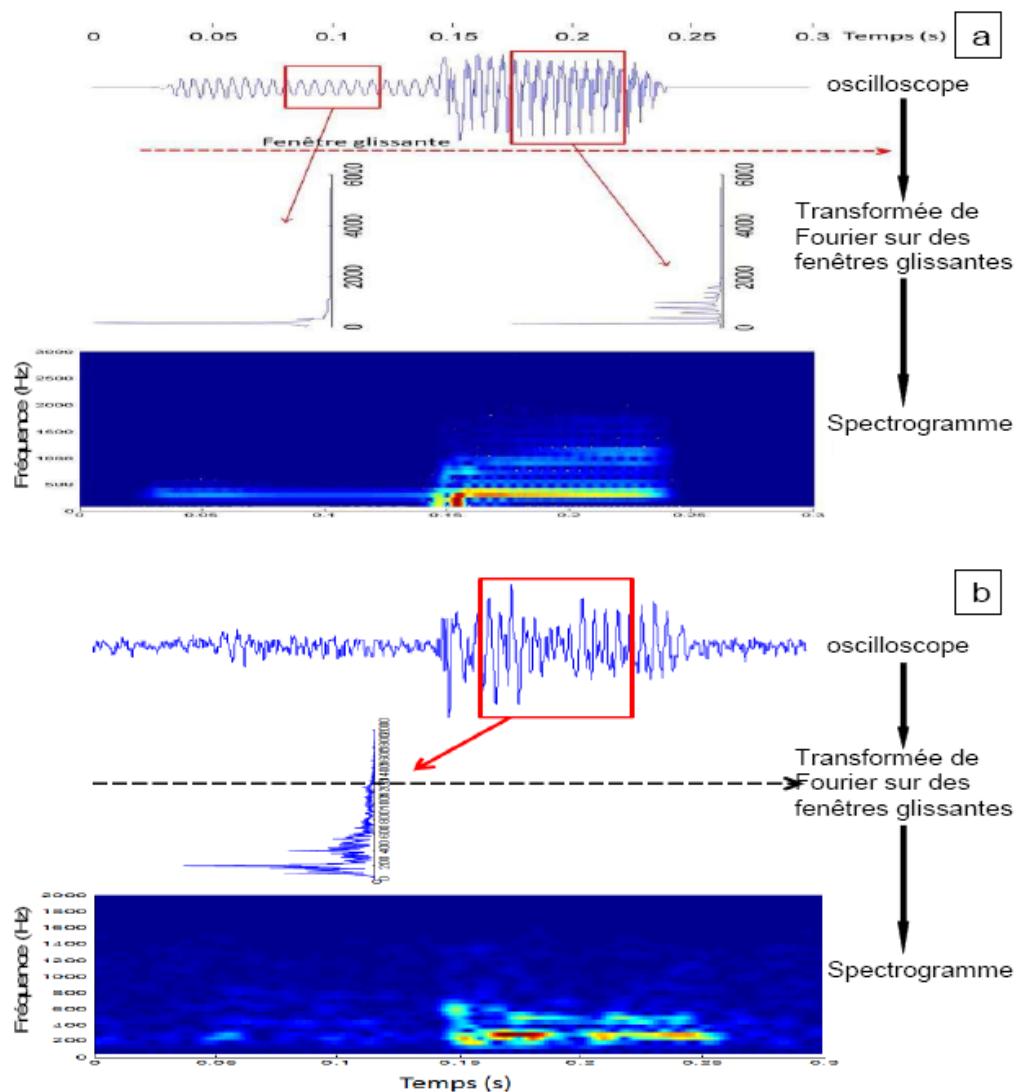


Figure 18 a et b: Représentation du traitement effectué pour l'obtention d'un spectrogramme.

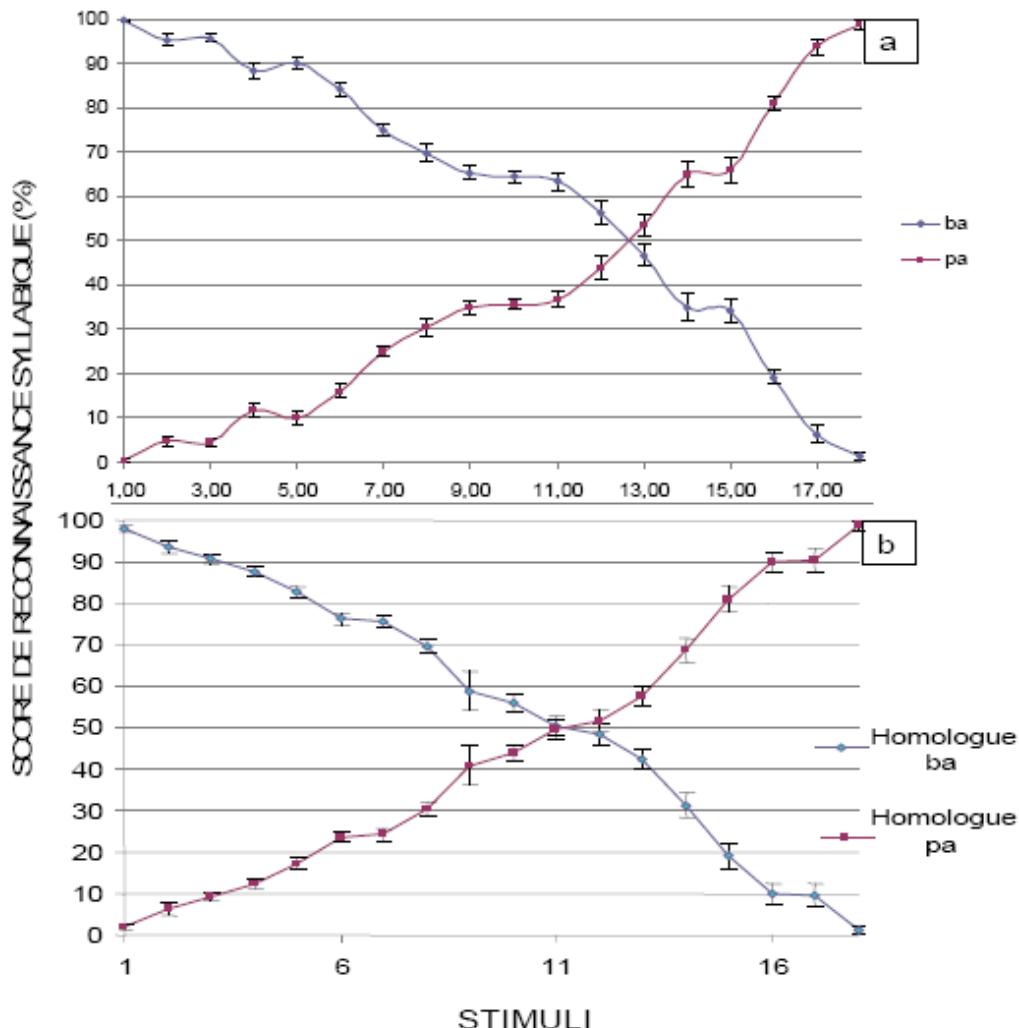
A partir d'une trace temporelle (a : celle du stimulus /ba/ ; b : celle d'un PEASP), on effectue des transformées de Fourier (FFT) successives sur une fenêtre de 20 ms de large, glissante le long de l'axe temporel. Le spectrogramme permet la visualisation de la succession de ces FFT dans l'axe temporel, donnant une représentation temps/fréquence. L'ensemble du traitement est ici effectué à l'aide du logiciel Matlab®. Le spectrogramme est visible en bas des figures a et b, il s'agit d'une représentation temps/fréquence avec l'amplitude des pics traduit sous forme d'intensité de couleur : bleu=négatif, rouge=positif.

1.3. Résultats :

1.3.1. Tests de perception catégorielle :

Pour l'ensemble des sujets, les stimuli extrêmes du continuum langagier /ba/ /pa/ furent reconnus à 100%, permettant à la fois d'éliminer tout trouble de l'apprentissage chez les sujets testés, ainsi que la validation des stimuli utilisés pour la partie neurophysiologie, ceux-ci étant identifiés de manière distincte.

La frontière perceptive moyenne des sujets pour le continuum langagier /ba/-/pa/ se situe entre le 12^{ième} et le 13^{ième} stimulus du continuum, tandis que les résultats obtenus avec le continuum non langagier montre une variation de cette frontière perceptive, celle-ci se situant entre le 11^{ième} et le 12^{ième} stimulus du continuum (Fig 19).



a : avec le continuum langagier /ba/-/pa/ ; b : avec le continuum non langagier /homologue ba/-/homologue pa/. La frontière perceptive correspond à un VOT de 0 ms, pour chaque

son, nous avons un décalage du VOT de 5,5 ms donc le VOT du ba 1 (/ba/) est de -72 ms et celui du ba 18 (/pa/) est de +27 ms sur le continuum langagier. Les 18 stimuli différents sont mentionnés à l'aide de leur numéro d'ordre sur l'axe des abscisses et en ordonnée est représenté le score moyen de réponses correctes pour l'ensemble des sujets

1.3.2. Tests neurophysiologiques : les PEASP

Le traitement temporel et fréquentiel n'ayant pas montré de différence entre les différentes polarités, les résultats ont été analysés en polarité alternée. Les traces les mieux représentatives temporellement et fréquentiellement ont été retrouvées au niveau de la voie 1 (dipôle mastoïde droite-vertex), chez l'ensemble de nos sujets, tous droitiers.

Concernant la réponse impulsionale, l'analyse visuelle de la représentation temporelle figure 20 montre un décalage dans l'apparition de celle-ci sur les grandes moyennes, de 0,098 ms entre /ba/ et son homologue, et de 0,048 entre /pa/ et son homologue. Les résultats sont ici présentés sur les grandes moyennes, les traces individuelles n'étant pas analysables.

Une réponse soutenue en fréquence (RSF ou frequency following response) bien synchronisée est visible sur la représentation temporelle des grandes moyennes du /ba/ et du /pa/ alors que celle-ci n'est pas individualisable sur les traces PEASP de leurs homologues respectifs. Les représentations fréquentielles mettent en avant la caractéristique passe-bas des PEASP, toujours en dessous de 1000 Hz (fig 21). Sur les spectrogrammes, on retrouve des courbes d'énergie similaires entre les enveloppes de /ba/ et de son homologue, ainsi que pour /pa/ et son homologue à 200 Hz, avec une différence à 400 et 600 Hz. La comparaison des énergies au niveau de F0 sur les spectrogrammes de /ba/, /pa/ et de leurs homologues montre une énergie supérieure pour la bande spectrale de F0 de /ba/ par rapport à celle de son homologue (Anova : $F(1,118) = 4,811, p = 0,03$), ainsi que pour /pa/ (Anova : $F(1,118) = 16,6, p < 0,001$), mais aussi à 400 et 600 Hz. Sur l'analyse visuelle des traces temporelles, les RSF en réponse à /ba/ et /pa/ sont semblables. Ceci a été corroboré par le calcul d'énergie à partir des spectrogrammes des grandes moyennes. L'énergie choisie afin de comparer les réponses /ba/ versus /pa/ a été choisie au niveau de F0, là où elle était le plus visible. Celle-ci variait de -30 à -9 dB pour /ba/ sur l'ensemble des 15 sujets ($M = -18 \text{ dB}, \text{std} = 6 \text{ dB}$), et de -31 à -9 dB pour /pa/. Statistiquement, l'utilisation d'un test-t sur les calculs d'énergies de la RSF, entre /ba/ et /pa/, contenue dans la bande spectrale de F0 est non significatif ($p=0,35$). Les FFR en réponse à /ba/ et à /pa/ sont semblables, ce qui n'est pas surprenant étant donné qu'ils possèdent la même enveloppe, la même voyelle et donc une périodicité identique.

Ces courbes effectuées sur la représentation spectrographique des PEASP obtenus (Fig 22), mettent donc en évidence la synchronisation de la FFR sur le fondamental des PEASP en réponse au /ba/ et /pa/, synchronisation conditionnée par la structure fine (comme Greenberg 1987 et Dau 2003), alors que les réponses PEASP homologues ne montrent pas une FFR aussi bien synchronisée que celles observée en réponse au /ba/ et au /pa/.

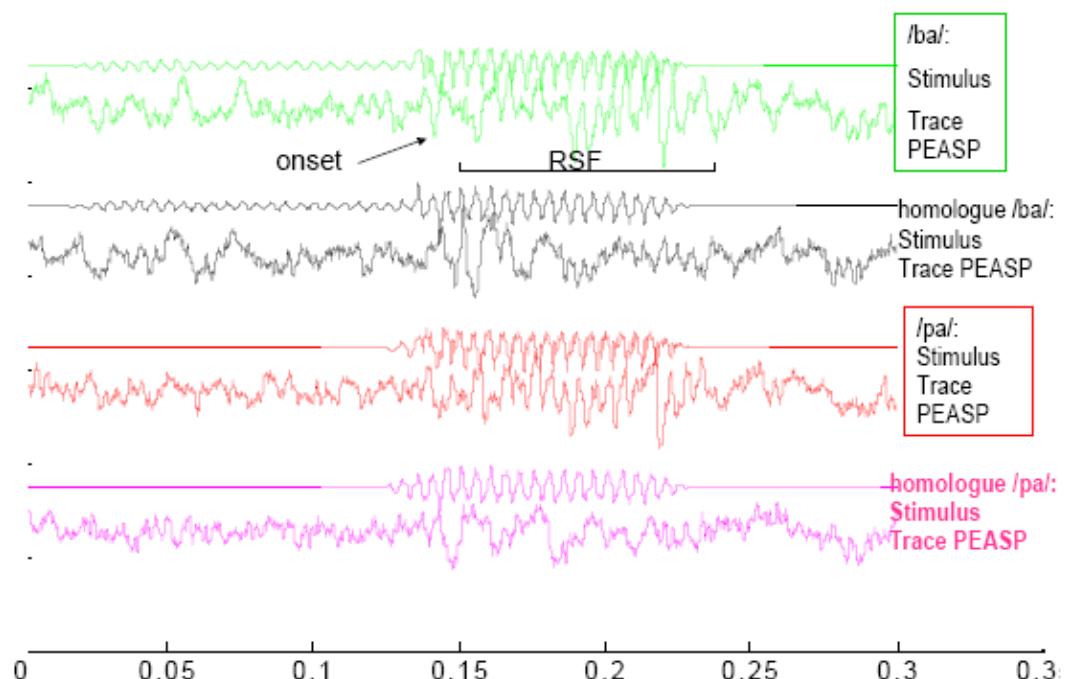


Figure 20 : Représentation temporelle des stimuli /ba/, /pa/, de leurs homologues et des grandes moyennes PEASP obtenues en réponse à ces différents stimuli.

De haut en bas : Sur la trace PEASP en réponse au /ba/ sont visibles l'onset et la RSF. Sur la trace PEASP en réponse à l'homologue du /ba/, l'onset est visible, mais la RSF n'apparaît pas identifiable. Il en est de même pour les traces PEASP en réponse au /pa/ et à son homologue.

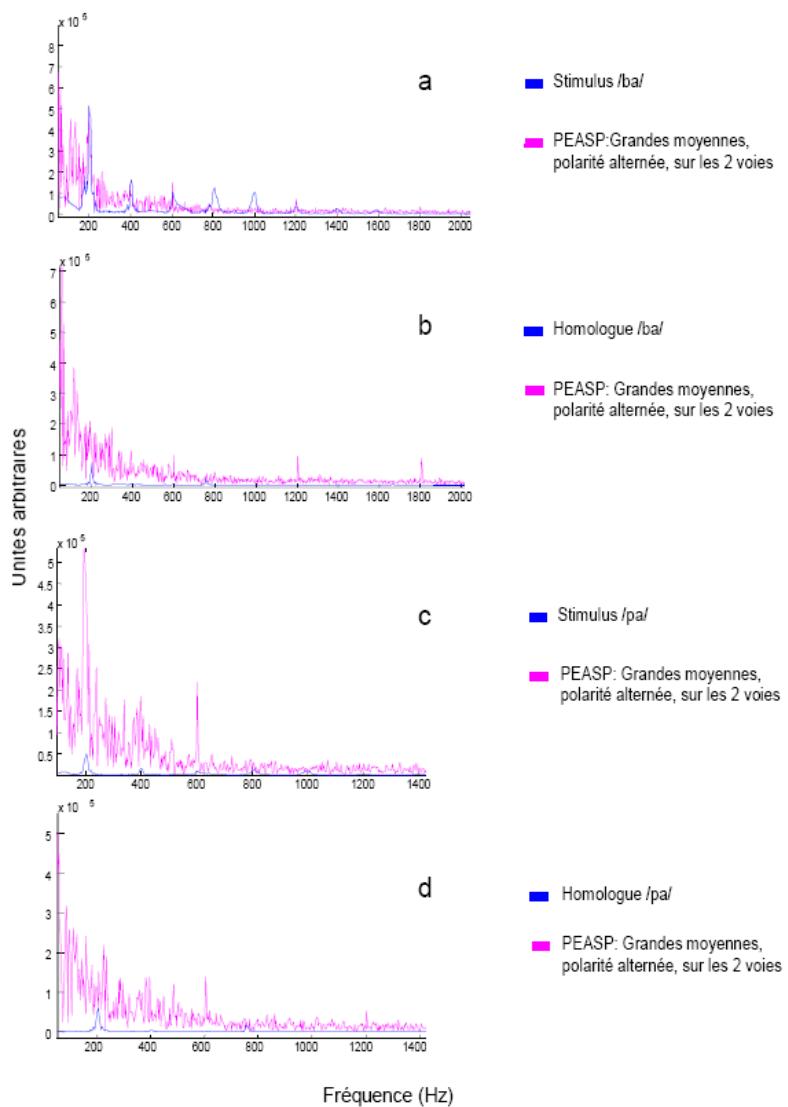


Figure 21 a, b, c, d: Représentation fréquentielle des différents stimuli et des traces PEASP correspondantes.

a : stimulus /ba/ ; b : homologue du stimulus /ba/ ; c : stimulus /pa/ ; d : homologue du stimulus /pa/.

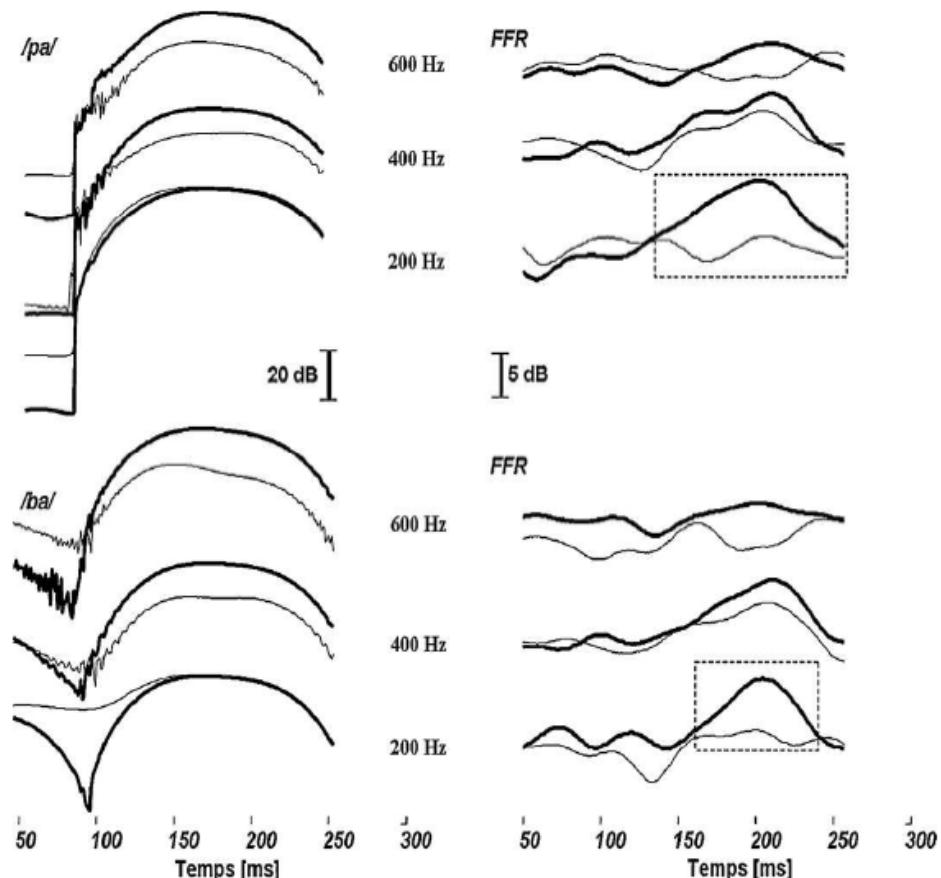


Figure 22 : Tirée de Akhoun et al, 2008 : Comparaison des coupes des spectrogrammes à différentes bandes de fréquences : F_0 (200 Hz) et aux deux premiers harmoniques (400 et 600 Hz).

Les signaux concernant la syllabe sont en gras et en trait fin pour l'homologue. Les stimuli se situent à gauche et la RSF à droite. Enfin, ce schéma concernant le /ba/, en haut, est repris à l'identique pour le /pa/ en bas. La superposition des courbes d'énergies des stimuli à 200 Hz (/ba/ ou /pa/) montre la similarité des enveloppes des homologues et des syllabes. Cette distribution est différente pour les harmoniques supérieurs. L'énergie contenue dans la RSF à 200 Hz est beaucoup plus importante en réponse aux syllabes (rectangles en pointillées), tandis qu'elle est semblable dans les deux conditions 400 et 600 Hz (harmoniques).

1.4. Discussion :

Certaines équipes suggèrent un rôle de la dynamique spectrale dans la genèse de la RSF surtout au niveau de F_0 , dynamique supprimée dans le cas des homologues (obtenus

avec une modulation par une enveloppe filtrée à basse fréquence, 500Hz), et pouvant ainsi expliquer la disparition de la RSF sur les PEASP en réponse à ceux-ci. Du point de vue psychoacoustique, la partie 500 - 1000 Hz correspond à la partie basse de la structure fine, qui est la zone de dominance de la perception du pitch.

Les travaux précédents de Greenberg et al (1987) montrent la relation entre le pitch et la RSF, celle-ci pouvant même représenter le corrélat neurophysiologique de la hauteur tonale.

Ainsi, F0 contient-elle l'ensemble des informations d'un son ? L'enveloppe temporelle contiendrait elle les informations nécessaire à la perception de la hauteur tonale ? Une précédente étude avait mis en évidence une reproduction fidèle de l'enveloppe du stimulus (redressée filtrée à 500 Hz) au sein de la RSF jusqu'à une fréquence de coupure de 500 Hz. Toutefois le problème de l'origine de la RSF persiste. En effet, les homologues et les syllabes naturelles possédaient la même enveloppe temporelle, or aucune RSF synchronisée n'a été retrouvée en réponse aux homologues. Cela sous-entendrait que la RSF ne contient pas que l'information liée à l'enveloppe temporelle.

Expérience 2 : Protocole expérimental Etude de l'encodage des sons de parole dans différentes conditions de bruit

Problématique :

Pouvoir disposer de données neurophysiologiques chez des sujets normo-entendants, afin d'objectiver la modification d'encodage des signaux langagiers par les différentes conditions de bruit

Matériels et méthodes :

• Population étudiée :

Six sujets droitiers normo-entendants, âgés de 20 à 30 ans.

• Tests comportementaux :

Audiométrie vocale utilisant les listes de Lafon, en binaural à 40 dB SL
- dans un bruit blanc avec un rapport signal/bruit de + et -6dB
- dans un bruit de type speech noise avec un rapport signal/bruit de +6 et -6dB
- dans un bruit de type cocktail party avec un rapport signal/bruit de +- et -6dB

• Tests neurophysiologiques : les PEASP (durée d'acquisition par sujet : 1h 45min)

Stimulus : /ba/ de Cazals de 230 millisecondes, présenté dans différentes conditions :
-sans bruit
-noyé dans une condition de bruit de type speech noise à SNR +6 et -6 dB.
-noyé dans une condition de bruit de type cocktail party à SNR +6 et -6 dB.
-noyé dans une condition de bruit de type bruit blanc à SNR +3 et +6 dB
Stimulation binaurale à un niveau de 45 dB SL, par blocs de 1200, à une cadence de 6,6 stimuli par seconde.
Recueil des traces à l'aide du centor USB Racia Alvar®, sur 2 voies d'enregistrements, via 4 électrodes d'argent (Nicolet®, 10mm), fixées sur le scalp du sujet au niveau du vertex (électrode active), du front (masse), et des deux mastoides (droite pour la voie 1 et gauche pour la voie 2).

Résultats :

Il existe un meilleur encodage dans le bruit blanc par rapport aux autres conditions. Toutefois les conditions ne sont pas comparables entre elles.

Expérience 2 : Etude de l'encodage des sons de parole dans différentes conditions de bruit :

2.1. Introduction :

Dans notre environnement quotidien, tenir une conversation à plusieurs peut parfois s'avérer un véritable défi. L'intégration des sons de parole en éléments porteurs de sens, quelles que soient les conditions de bruit, est une nécessité pour la communication de tous les jours. Toutefois, dans une scène auditive contenant plusieurs sources sonores, nous sommes capables de sélectionner un signal sonore qui nous intéresse. Il existe un effet de l'attention sur la perception de la parole. La perception de la parole dans le bruit consiste à séparer le signal acoustique pertinent du bruit (Bronkhorst & Adelbert, 2000). Notre système auditif applique des stratégies de ségrégation qui utilisent des règles précises pour distinguer les différents flux de la scène auditive (Cherry, 1953 ; Bronkhorst & Adelbert, 2000).

Le démasquage de la parole (l'amélioration substantielle de l'intelligibilité de la parole dans un bruit fluctuant par rapport à un bruit stationnaire, observée chez les sujets entendants), est systématiquement réduit voire aboli chez les personnes présentant une perte cochléaire et les personnes porteuses d'un implant cochléaire. Ceci suggère qu'aux niveaux de bruit de fond pour lesquels les personnes entendantes atteignent une intelligibilité parfaite, la plupart des personnes malentendantes ne comprennent quasiment rien (Shinn-Cunningham, 2009). De nombreuses études indiquent que le phénomène de démasquage dépend de la capacité à écouter le signal de parole dans les vallées spectro-temporelles du bruit de fond, où le rapport signal-sur-bruit est localement amélioré. Parmi les différents déficits sensoriels associés à une perte cochléaire, la dégradation de la sélectivité fréquentielle et un trouble de la perception des informations de structure temporelle fine semblent participer à la réduction substantielle du démasquage observée chez les sujets malentendants. Cette hypothèse est motivée par le fait qu'une lésion cochléaire ne semble pas affecter la capacité à suivre les fluctuations d'amplitude lentes, alors qu'elle dégrade quasi systématiquement la capacité à accéder aux informations spectrales et de structure temporelle fine, et ce parfois pour des pertes légères (< 30 dB HL) (Lorenzi et al, 2006).

Des études précédentes mettent en évidence une vulnérabilité accrue des consonnes par rapport aux voyelles dans le bruit, similitudes se retrouvant dans le PEASP. En effet, le PEASP ou « speech ABR » peut être divisée en 2 parties : l'onset response et la RSF ou « FFR », la première étant engendrée par la consonne contenue dans le stimulus, apparaît être dégradée de manière plus nette dans le bruit par rapport à la FFR (Boston & Moller 1985 ; Cunningham et al, 2001 ; Russo et al, 2004). Par ailleurs, Russo et al 2004, ont mis en évidence le rôle des parties impulsionales et soutenues, quant à l'encodage des différents aspects du stimulus le long du tronc cérébral, suggérant une représentation indépendante pour ces deux types de réponses.

L'objectif de cette expérience est de pouvoir disposer de données neurophysiologiques chez le normo entendant, afin d'objectiver la modification d'encodage des signaux langagiers par les différentes conditions de bruit, ceci dans le but futur de réitérer l'expérience chez des personnes souffrant de troubles de l'intelligibilité dans le bruit.

2.2. Matériel et méthodes

2.2.1. Population étudiée :

Ce travail fut mené sur 6 sujets droitiers, âgés de 20 à 30 ans, tous ne présentaient aucun antécédent médical particulier notamment otologiques. Tous les sujets étaient issus de formations de niveau d'étude supérieur. Avant toute inclusion, chaque sujet a bénéficié d'un examen otoscopique ainsi que d'une audiometrie tonale, dont les seuils devaient être strictement normaux entre 125 et 8000 Hz (Audiomètre AC 40, Interacoustics®). Chaque sujet a eu un examen par potentiels évoqués auditifs en réponse à des clicks (centor USB Racia alvar®) dont les latences normales ont permis d'éliminer toute pathologie retro cochléaire. Ces tests ont été réalisés en accord avec la loi Huriet (comité d'éthique de Lyon (CCPPRB # 05-024). Un consentement éclairé a été obtenu auprès de chaque sujet.

2.2.2. Les différents types de bruit :

- Le bruit blanc :

Il s'agit de la réalisation d'un processus aléatoire dans lequel la densité spectrale de puissance est la même pour toutes les fréquences.

- Le speech noise :

Il s'agit du bruit blanc filtré passe bas à 800 Hz avec une pente d'ordre 1 (Hoen et al, 2005).

- Le bruit de type cocktail party :

Le bruit de type cocktail party a été utilisé car celui-ci est représentatif des conditions qu'un individu peut rencontrer dans la vie de tous les jours (Bronkhorst & Adelbert, 2000). Il est constitué d'un mixage de 8 voix différentes (cocktail party mis au point par Nicolas Grimault de l'équipe CNRS NSCC Lyon 1). L'utilisation de ce bruit a pour but second la mise en évidence de l'effet cocktail party, qui correspond à la possibilité de détecter un flux de parole parmi plusieurs flux simultanés (Cherry, 1953)

2.2.3. Tests comportementaux :

Audiométrie vocale utilisant les listes de Lafon : (listes monosyllabiques permettant d'enlever la suppléance mentale, composées de syllabes équilibrées sur le plan phonétique)

- noyées dans un bruit blanc avec un rapport signal/bruit de +- et -6dB
- noyées dans un bruit de type speech noise avec un rapport signal/bruit de +6 et -6dB
- noyées dans un bruit de type cocktail party avec un rapport signal/bruit de +- et -6dB

Les différents stimuli sont présentés en binaural à travers un casque Sennheiser linear 250®, de manière aléatoire, à une intensité confortable de 40dBSL.

2.2.4. Tests neurophysiologiques : les PEASP

2.2.4.1. Stimuli :

Le signal langagier utilisé est un /ba/ de Cazals de 230 millisecondes, présenté dans différentes conditions :

- sans bruit afin d'apprécier l'encodage du /ba/ en conditions optimales et s'en servir de référence (Fig 23).
- noyé dans une condition de bruit de type speech noise à deux rapports signal/bruit : +6 et -6 dB.

- noyé dans une condition de bruit de type cocktail party à deux rapports signal/bruit: +6 et -6 dB. Les représentations spectrales des différents stimuli sont visibles sur la figure 24.
- noyé dans une condition de bruit de type bruit blanc à deux rapports signal/bruit: +3 et +6 dB, la condition avec un rapport signal/bruit à -6 dB n'étant pas supportable. L'énergie des différents bruits a été équilibrée, et l'intensité de stimulation est de 45 dB SL, avec une fréquence de stimulation égale à 3Hz.

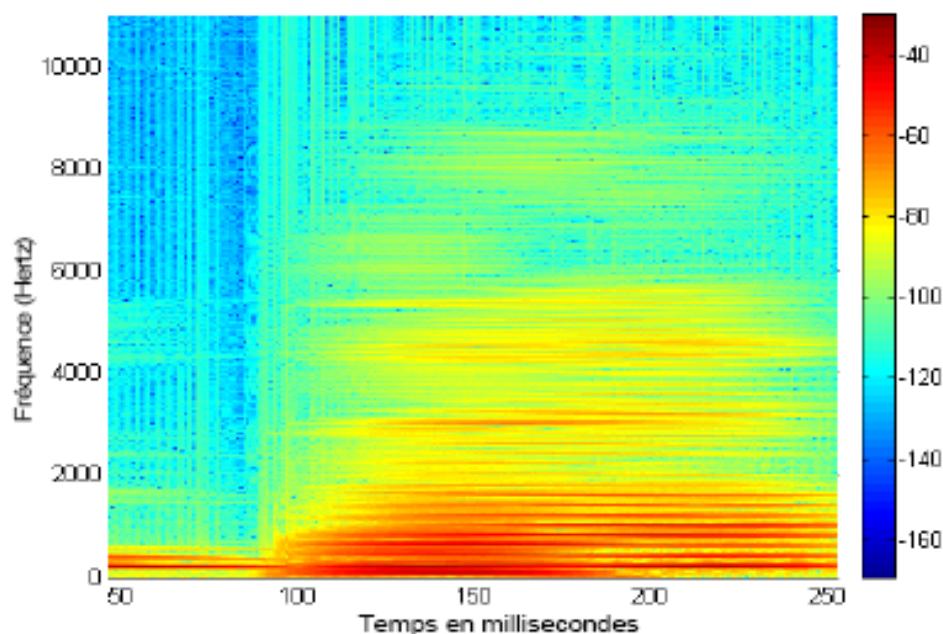


Figure 23 : Représentation spectrale du stimulus /ba/ obtenue à l'aide du logiciel Matlab®.

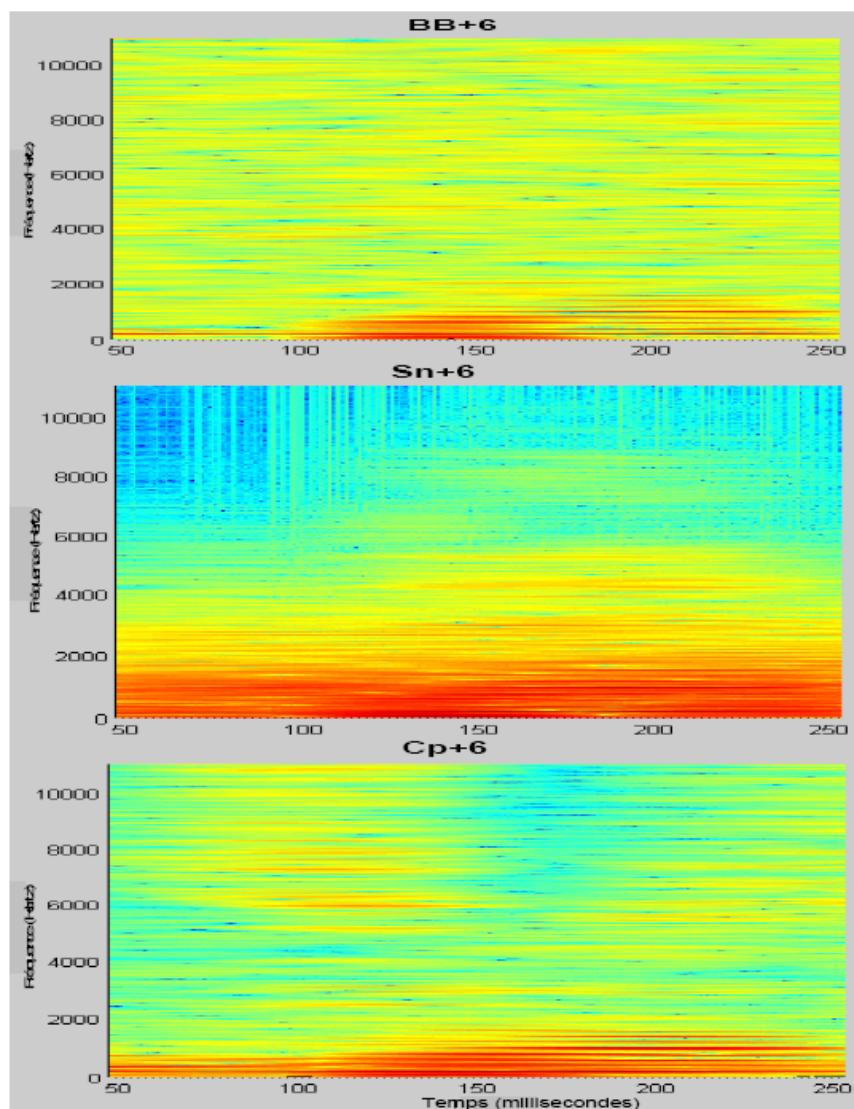


Figure 24: Représentation spectrale du stimulus /ba/ noyé dans les différents bruits.

De haut en bas : Le stimulus /ba/ noyé dans un bruit de type bruit blanc (SNR +6 dB). Les amplitudes spectrales maximales se situent en dessous de 1000 Hz et principalement entre 100 et 180 millisecondes. Le stimulus /ba/ noyé dans un bruit de type speech noise (SNR +6 dB). Les amplitudes spectrales sont maximales en dessous de 2000 Hz et sur toute la durée du stimulus. Le stimulus /ba/ noyé dans un bruit de type cocktail party (SNR +6dB). Les amplitudes spectrales sont maximales en dessous de 1000 Hz et entre 100 et 200 ms.

Chaque stimulus est envoyé de manière binaurale au sujet, l'ordre des stimuli étant randomisé.

2.2.4.2. Système de recueil:

Le système de recueil est identique à celui présenté dans la partie « système de recueil » de l'expérience 1 (Fig 17). Chaque stimulus est moyenné 2400 fois (1200 polarités positives, 1200 polarités négatives), avec une fenêtre d'enregistrement de 300 ms et une fréquence d'échantillonnage de 13,3 kHz.

2.2.4.3. Traitement des données PEASP :

De la même manière que nous l'avons vu dans l'expérience 1, le traitement des données fait appel à une analyse visuelle, couplée à une analyse par bandes de fréquences sur les spectrogrammes.

2.3. Résultats

2.3.1. Tests comportementaux:

Les différentes conditions avec un rapport signal/bruit défavorable ne montrent pas de différences statistiquement significatives. Toutefois, un effet d'amélioration des résultats est observé dans les conditions de bruit avec un rapport signal/bruit favorable par rapport aux conditions avec des rapports signal/bruit défavorables (-6 dB) (fig 25).

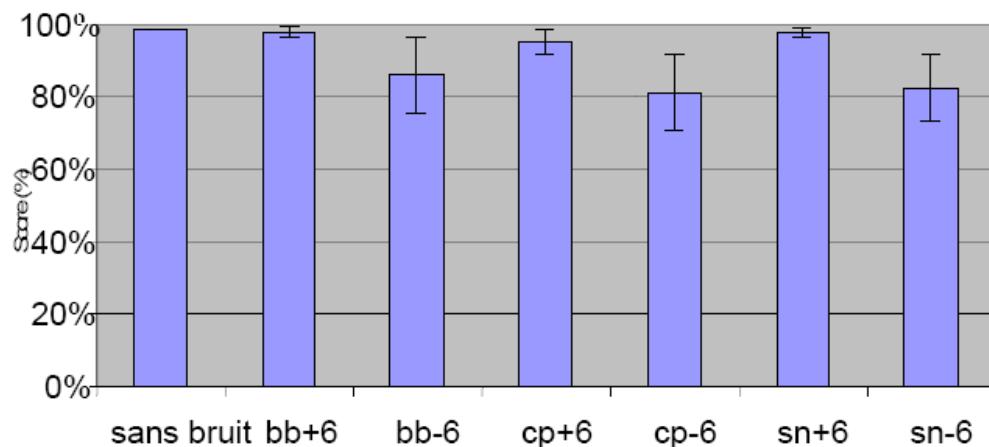


Figure 25 : Moyennes (+/-DS) des scores des 6 sujets obtenus en audiométrie vocale, dans différentes conditions de bruit (bb pour bruit blanc, CP pour bruit de cocktail party, sn pour « speech noise »), avec des rapports signal/bruit de 6 dB et -6 dB (signal à 40 dB SL, et bruit à 46 dB SL).

2.3.2. Tests neurophysiologiques :

- Différence voie 1 versus voie 2 :

L'analyse visuelle sur les traces temporelles représentant les grandes moyennes de l'ensemble des 6 sujets, suggère une meilleure synchronisation sur la voie 1 (dipôle mastoïde droite-vertex) que sur la voie 2 (dipôle mastoïde gauche-vertex).

- L'onset :

Sur les traces temporelles, l'onset est difficilement individualisable dans les conditions de bruit, excepté pour les conditions de bruit blanc et speech noise +6 dB où celle-ci apparaît mais de façon retardée (1,7 et 2 millisecondes respectivement) par rapport à la condition sans bruit de référence (Fig 26).

En revanche, sur les spectrogrammes, l'onset n'est pas individualisable, apparaissant totalement noyée dans le bruit (Fig 27).

- La réponse soutenue en fréquence :

Celle-ci apparaît être bien synchronisée dans les deux conditions de bruit blanc, alors qu'elle n'est pas identifiable dans les autres conditions de bruit. La RSF apparaît même être « améliorée », d'amplitude plus importante dans les deux conditions de bruit blanc par rapport à la condition sans bruit de référence.

Sur les spectrogrammes, la fréquence coupure des PEASP étant comme nous l'avons vu précédemment de 500 Hz (Akoun et al. 2009, ceux-ci apparaissent difficilement individualisables du bruit de type speech noise et cocktail party. Toutefois, une réponse PEASP est visible dans la condition BB+6 avec un pic d'énergie au niveau de la RSF.

Les résultats obtenus sur les spectrogrammes étant difficilement analysables de manière visuelle, des bandes de fréquences ont été choisies à 200 Hz (F0), 400 Hz (premier harmonique) et 600 Hz (deuxième harmonique).

L'analyse de ces bandes de fréquences sur les spectrogrammes à la fréquence de 200 Hz, a permis de retrouver une énergie supérieure dans le cas de la condition bruit blanc +6dB par rapport aux autres conditions de bruit possédant le même rapport signal/bruit. Concernant les conditions sans bruit et bruit blanc +6 dB, les énergies ne sont pas statistiquement différentes ($p=0,12$), alors que l'on constate une dégradation de cette énergie pour les autres conditions de bruit ($p=0,031$), (Fig 28). Les coupes à 400 Hz, retrouvent une diminution de l'énergie, ceci de manière semblable entre les différentes conditions de bruit ($p=0,023$).

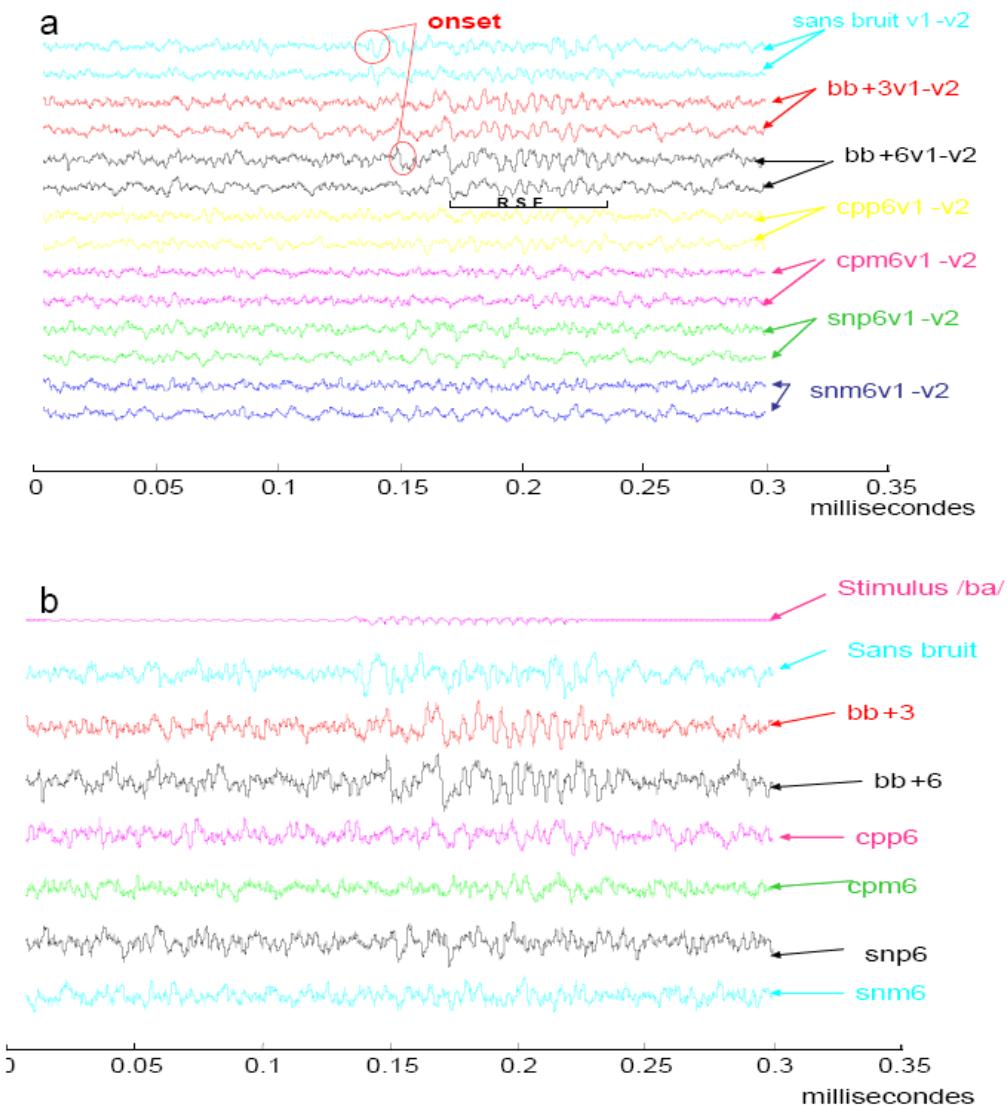


Figure 26 a et b : Représentation temporelle des grandes moyennes des traces PEASP en réponse aux différents stimuli.

a : traces recueillies sur la voie 1 et la voie 2 ; b : traces recueillies sur la voie 1.
 bb=bruit blanc (bb+3=SNR +3dB, bb+6= SNR +6dB) ; cp=cocktail party (cpp6=SNR +6dB, cpm6=SNR -6dB);sn=speech noise (snp6=SNR +6dB, snm6=SNR -6dB).

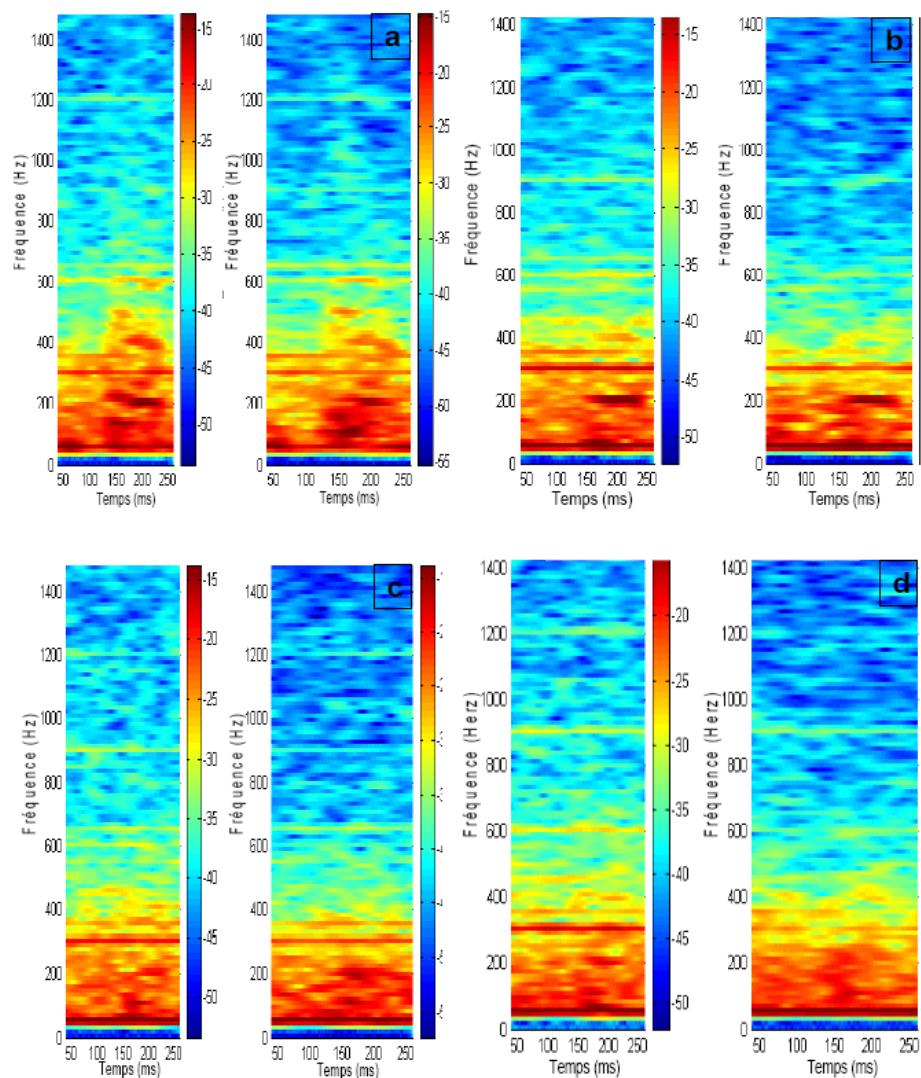


Figure 27 a, b, c, d: Représentation spectrographique de la grande moyenne des traces PEASP.

a : traces recueillies dans la condition sans bruit (voie 1 à gauche sur l'image, et voie 2 à droite). b: traces recueillies dans la condition bruit blanc SNR +6dB (voie 1 à gauche sur l'image, et voie 2 à droite). c : traces recueillies dans la condition cocktail SNR +6dB (voie 1 à gauche sur l'image, et voie 2 à droite). d : traces recueillies dans la condition speech noise SNR +6dB (voie 1 à gauche sur l'image, et voie 2 à droite)

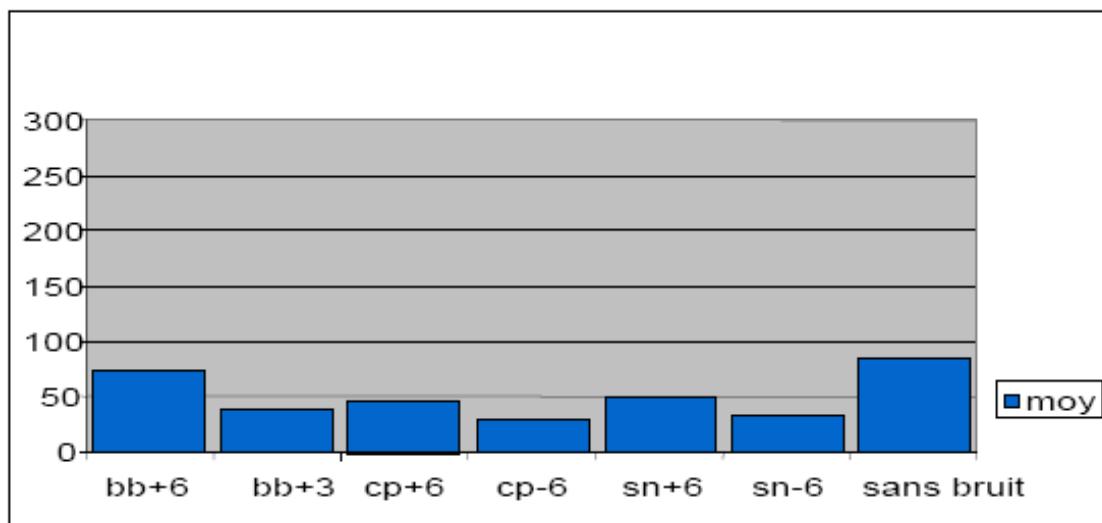


Figure 28 : Représentation de l'énergie obtenue en réponse aux différents stimuli à 200 Hz (voie 1+2) pour les 6 sujets).

L'énergie présente sur chaque spectrogramme à 200 Hz est représentée sous forme d'amplitude. abscisses : stimuli utilisés ordonnées : amplitude de l'énergie sur la bande de fréquence 200Hz (unités arbitraires).

2.4. Discussion :

Le recueil des traces PEASP apparaît être meilleur sur la voie 1 (dipôle vertex-oreille droite), ce qui concorde avec les données précédentes de Hornickel et al (2009), le son envoyé au niveau de l'oreille droite étant encodé préférentiellement au niveau du cortex gauche. En effet, les éléments temporaux et harmoniques (représentant le filtre) de la parole sont préférentiellement encodés selon le trajet reliant l'oreille droite à l'hémisphère gauche (Hornickel et al, 2009).

Par ailleurs, on observe un encodage des sons de parole préservé dans les conditions de bruit blanc, se dégradant faiblement avec la diminution du rapport signal/bruit, aussi bien sur les résultats temporels que les spectrogrammes (SNR +6 et +3 dB). Les conditions de bruit utilisées étaient, même avec un rapport signal/bruit de -6, plutôt favorables, (dans la mesure où, lors des résultats comportementaux, on arrive à des scores de plus de 80% pour les scores les plus faibles), ce qui explique la relative faible dégradation des PEAPSP.

Nous n'avons pas observé de diminution plus importante de l'onset (codant la consonne) par rapport à la RSF (codant la voyelle), à la différence de Cunningham et al (2002), certainement par le masquage moins efficace de nos différents bruits (résultats visibles sur les tests subjectifs).

Toutefois, les différentes conditions de bruit ne peuvent être comparées entre elles. En effet, les conditions bruit blanc, speech noise et cocktail party ne sont pas énergétiquement comparables : les énergies globales sont identiques mais leur répartition sur le spectre fréquentiel est différente : ainsi, l'énergie du speech noise sur les fréquences habituellement dévolues à la parole, sera beaucoup plus importante que l'énergie du bruit blanc pour les mêmes fréquences. Ainsi, la bonne synchronisation de la RSF observée dans les conditions de bruit blanc s'expliquent par une répartition homogène du bruit sur l'ensemble de la

gamme fréquentielle et donc un « masquage » inférieur pour une fréquence donnée à celui produit par le cocktail party ou le speech noise.

Il aurait été intéressant de poursuivre cette étude en incluant des personnes gênées dans le bruit afin de comparer leurs résultats avec ceux des sujets normo-entendants de cette étude.

Par ailleurs, le fait de ne pas pouvoir analyser les traces de PEASP en termes de résultats individuels, ne nous permet pas de faire une corrélation entre les résultats obtenus sur le plan comportemental, et ceux des PEASP. En effet, il serait important de voir si un sujet se dégradant particulièrement en situation de bruit adverse au niveau comportemental, verrait ses traces de PEASP fortement dégradées. Une première approche de ce type a été réalisée par l'équipe de N. Kraus, qui a montré une perturbation de l'encodage neural par le bruit, celui-ci étant corrélé avec l'incapacité à percevoir la parole dans des conditions adverses (Anderson et al, 2010). Afin de s'approcher de résultats interprétables à l'échelon individuel, en utilisant des stimuli écologiques Français, nous avons décidé d'essayer d'améliorer la technique de recueil utilisée. Toutefois, en reprenant les résultats de Anderson et al, 2010, on peut remarquer que le diagnostic de troubles d'audition dans le bruit, basé sur les PEAPSP, n'est pas encore réalisable à l'échelon individuel.

Expérience 3 : Mise au point technique
Problématique :

Nécessité de disposer d'un outil objectif, fiable et reproductible à l'échelon individuel.
Mise au point d'une analyse automatisée des traces PEASP, applicable à tout individu.

A. Optimisation de la chaîne de mesure :

Matériels et méthodes :

Stimuli : la durée des stimuli est diminuée à 150 millisecondes.

Les inserts : diminution de leur impédance à 50 Ω (vs 300 Ω).

La chaîne de stimulation :

- passage de l'ordinateur en mode batterie.
- utilisation d'une carte son externe.
- nombre d'acquisition minimum de 2500 acquisitions en polarité alternée, ce qui fait une durée de recueil de l'ordre de 15 minutes pour une trace.
- re calibration de l'ensemble de la chaîne de stimulation au moyen de l'oreille artificielle.

Résultats :

Diminution du bruit de fond et amélioration du rapport signal/bruit.

Traces interprétables à l'échelon individuel.

B. Techniques d'analyse des traces PEASP :

Confrontation de l'analyse visuelle effectuée par 3 lecteurs différents aux modèles créés.

Création d'une onset type : deux triangles inverses, avec une durée et une pente moyennes.

Création d'une RSF type à partir de RSF identifiées précédemment, avec une périodicité et une amplitude moyenne dérivées de celles-ci.

Comparaison de ces modèles aux traces obtenues.

Expérience 3 : Mise au point technique :

L'utilisation d'une technique telle que les PEASP dans un but ultime clinique a des exigences de reproductibilité, doit pouvoir s'appliquer à la grande majorité des sujets, et doit pouvoir constituer un test avec une sensibilité et une spécificité correctes. Ces exigences demandent de pouvoir enregistrer des traces reproductibles, à l'échelon individuel, et ce, chez la majorité des sujets. La durée plus longue des stimuli en Français (plus de 100 ms pour des stimuli écologiques), par rapport aux stimuli américains utilisés par N. Kraus (40ms), a pour conséquence des traces plus bruitées, et qui demandent des temps de recueil beaucoup plus longs. Nous avons alors entrepris, avec l'aide de A. Jeanvoine (doctorant, ingénieur en Génie Biologique et Médical), une mise au point technique, avec des altérations substantielles de la chaîne de mesure, dans le but d'obtenir des traces

reproductibles à l'échelon individuel, et ceci, chez la majorité des sujets, étape qui nous semblait indispensable avant de poursuivre les expérimentations.

3.1. Optimisation de la chaîne de mesure :

L'ensemble des résultats obtenus jusqu'ici étaient encore trop artéfactés pour obtenir des traces reproductibles à l'échelon individuel. En effet, l'artefact électromagnétique a été fortement diminué par l'utilisation d'insert et le relai de la cage de faraday à une masse, mais l'utilisation de l'audiomètre AC 40 dans la chaîne de stimulation induisait un bruit surajouté au stimulus audible à l'oreille nue. De même, il persistait des problèmes de pollution par du 50 Hz au niveau de l'envoi de la stimulation. Il s'avérait donc nécessaire d'isoler l'ordinateur en le remplaçant par un ordinateur fonctionnant sur batterie et donc isolé du réseau. Comme nous allons le voir plus en détails, de nombreuses petites modifications ont du être effectuées afin d'optimiser les traces PEASP recueillies.

Stimuli :

La durée des stimuli a été diminuée car en diminuant celle-ci on diminue non seulement le temps d'acquisition, mais aussi sa sensibilité au 50 Hz. Sachant que la durée du stimulus ne doit pas être identique à l'une des cadences de répétition du 50 Hz. Ainsi la durée des stimuli est diminuée à 150 millisecondes.

Les inserts :

Il est nécessaire de vérifier les inserts avant toute expérience, la moindre altération de ceux-ci entraîne une variation de 10 dB dans l'intensité du stimulus transmise au sujet. Par ailleurs les inserts précédemment utilisés présentaient une Impédance très forte (300Ω vs 50Ω pour les nouveaux) entraînant ainsi une diminution rapide de l'autonomie des piles, problème résolu par le changement de ceux-ci.

La chaîne de stimulation :

L'ordinateur est utilisé en mode batterie afin de l'isoler du réseau électrique, diminuant ainsi la contamination artéfactuelle électromagnétique. Une carte son externe remplace celle de l'ordinateur afin d'éviter le problème de diaphonie, la diaphonie (cross over) étant le recueil de signal du côté gauche alors que celui avait été envoyé à droite et vice versa. Ainsi l'impulsion du trigger, pouvait se retrouver ajoutée au stimulus et être perçue comme un clic surajouté à celui-ci. La carte son externe n'étant pas assez puissante pour permettre un gain suffisant en sortie des inserts, un amplificateur avec gain variable sur batterie a été rajouté.

Afin d'optimiser la qualité du signal recueilli, le nombre d'acquisition est augmenté à 2500 acquisitions en polarité alternée, ce qui fait une durée de recueil de l'ordre de quinze minutes pour une trace.

L'ensemble de la chaîne de stimulation a été re-calibrée de manière précise au moyen de l'oreille artificielle, ceci à 5 reprises, avec un résultat d'intensité en sortie des inserts stables lors de chaque test/re-test.

Chaque stimulus étant précédé de quelques millisecondes de silence, le bruit ambiant peut ainsi être apprécié sur les traces PEASP.

La nouvelle chaîne de stimulation est représentée sur la figure 29.

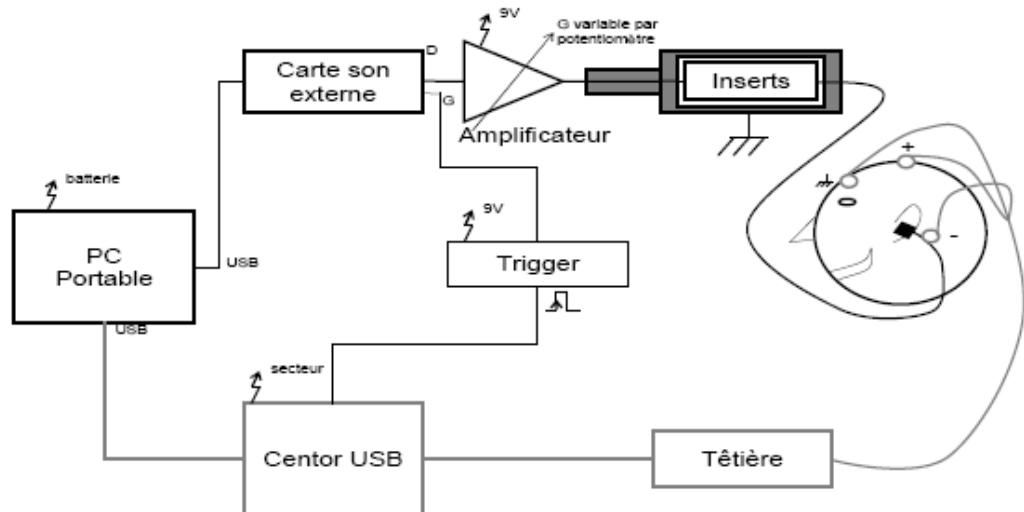
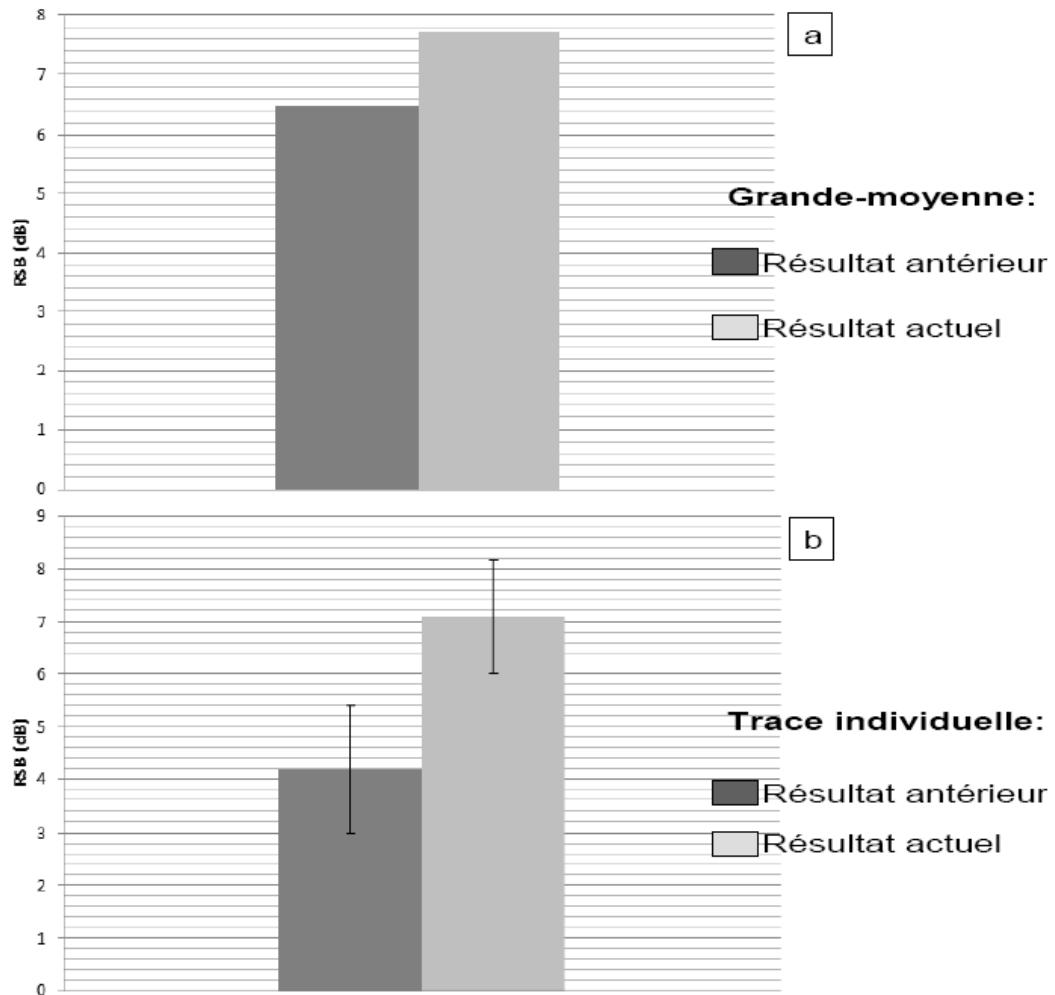


Figure 29 : Schéma de la nouvelle chaîne d'acquisition.

La nouvelle chaîne d'acquisition des signaux PEASP est entièrement portable, afin d'éviter au maximum les problèmes du secteur et de rayonnement. Ainsi, l'ordinateur fixe est remplacé par un ordinateur portable, l'audiomètre a été remplacé par un amplificateur sur batterie, une carte son externe est utilisée afin d'éviter la diaphonie entre la voie de gauche et la voie de droite. En effet, ces 2 voies n'ont pas les mêmes fonctions, il ne faut pas que l'impulsion au niveau du trigger se retrouve dans le son que l'on envoie à la personne car il y a risque d'obtenir une réponse à un clic sur les enregistrements. La carte son externe permet d'éviter ce problème en séparant les deux voies et en ayant des amplificateurs indépendant.

La chaîne d'acquisition des données est restée la même, à la différence que le stockage des données se fait dans l'ordinateur portable.

Par ces modifications de la chaîne d'acquisition des PEASP, nous avons pu observer une amélioration du rapport signal sur bruit (Figure 30). Des exemples de traces PEASP obtenues avec la nouvelle chaîne d'acquisition sont visibles figures 31 et 32.



a : le résultat du RSB antérieur est tiré de ceux obtenus avec le montage précédent (Akhoun et al, 2008), pour un même stimulus. Ce résultat est représenté à partir de la grande moyenne des sujets. b : le résultat du RSB antérieur est tirés de ceux obtenus par Akhoun et al, toutefois ici sont représentés les résultats à l'échelon individuel (6 sujets issus du montage antérieur et 6 du nouveau montage).

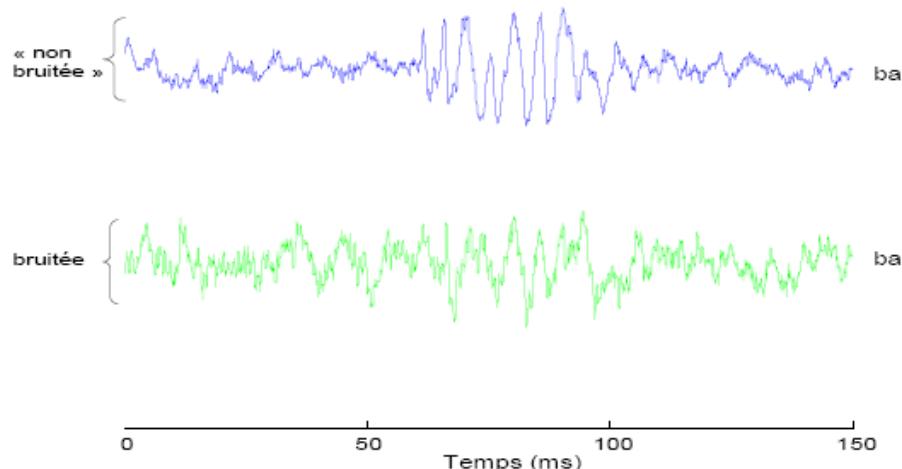


Figure 31 : Exemple de traces individuelles en réponse à un même stimulus /ba/.

Les courbes vertes représentent les traces PEASP obtenues avec le montage utilisé antérieurement, tandis que les courbes bleues, obtenues avec le nouveau montage, permettent de mieux visualiser la synchronisation de la RSF par la diminution du bruit de fond, visible sur les premières millisecondes de recueil.

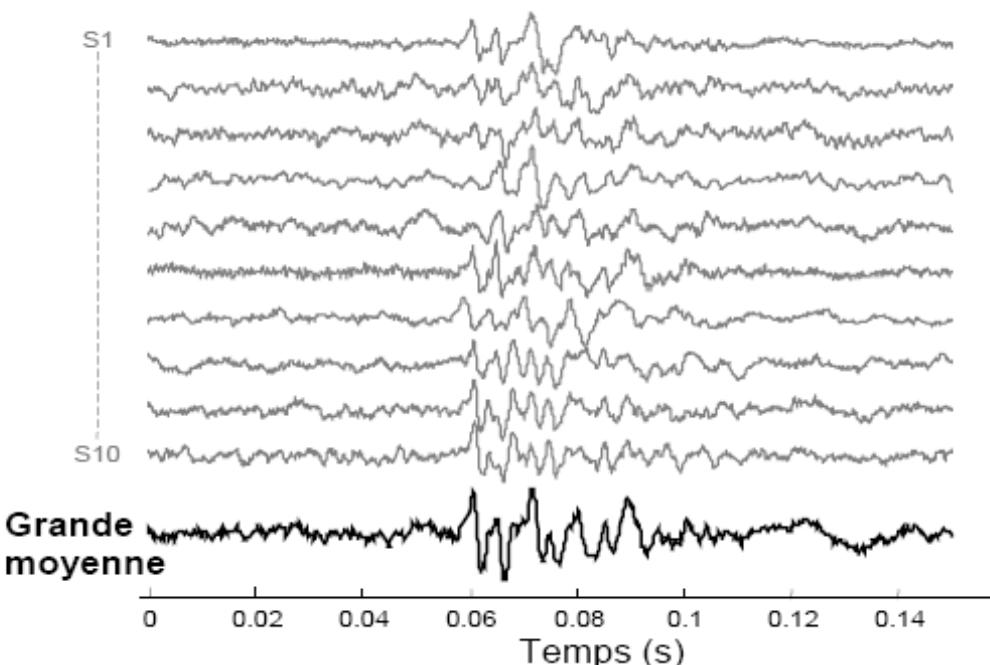


Figure 32 : Représentation temporelle de résultats individuels et la grande moyenne en résultant.

On obtient visuellement une disparition du bruit de fond lors du calcul de la grande moyenne.

3.2. Techniques d'analyse des traces PEASP :

« Étude de l'encodage des sons de parole par le tronc cérébral dans le bruit »

Parallèlement à cette mise au point, une revue de la littérature a été nécessaire afin de faire le point sur les techniques d'analyse des PEASP les plus usitées (article de revue Richard et al, 2010, complété par le tableau 3 a, b, c, d, e, f).

REFERENCE	Schomer et al 1977	Hoermann et al 1994	Batra et al 1986	Galbraith et al 1995	Galbraith et al 1996	Krishnan et al 1998	a
OBJECTIF	origine de la RSF	effet de l'attention sur la RSF	étude de la RSF	étude du dipôle vertical/horizontal	effet de l'attention, du côté stimulé et de F0	étude du composant d'interaction (coup d'œil de la RSF fonction de la différence interaurale d'intensité)	
STIMULUS sous cortical	cirrusoides et bouffées toniques	bouffées toniques	sens continu	mot(s) de durée max=640ms	voyelles anglaises /a/ et /e/	bouffées toniques d'intensité variable	
LATENCE							
AMPLITUDE							
PENTE							
aire							
DUREE							
LATENCE	1er pic de la RSF	1er pic de la RSF	1er pic de la RSF	1er pic de la RSF	1er pic de la RSF	RSF pour perte de conscience	
Meure de la FFT				RSF	RSF et surtout étude de F0 (oreille droite=haute fréquence et parole, gauche=basse fréquence, musique)	RSF pour perte de conscience + du composant d'interaction bilatéral	
Spectrogramme							
Cross-correlation							
Corrélation simple							
Autocorrelation et autocoréogrammes							
RMS							
Meure amplitude, représentation temporelle, en jv		RSF	RSF		RSF et représentation spectrale de F0	RSF	
RESULTATS	latence RSF neurale= 7ms	la latence de la RSF diminue de 45ms en fonction de l'attention du sujet, pas d'effet sur l'amplitude	latence RSF = 8.2±0.1ms	la latence de la RSF varie selon que le dipôle est horizontal (œuf aillé) ou vertical (TC) (5.3ms et 4.19)	amplitude RSF augmente lorsque la voyelle est attendue	diminution du composant d'interaction de la RSF avec l'augmentation de l'amplitude d'intensité interaurale: les neurones bilatéraux du TC utilisent les indices des deux oreilles	

Tableau 3 a, b, c, d, e et f : Différentes caractéristiques des PEASP, d'après une revue non exhaustive de la littérature.

REFERENCE	Krishnan A et al 2000	Galbraith et al 2000	Cunningham et al 2001	King et al 2002	Krishnan 2002	Cunningham et al 2002	b
OBJECTIF	étude du phase locking de la RSF	étude du dipôle vertical/horizontal	les enfants avec un trouble de l'apprentissage ont une mauvaise représentation neurophysiologique du langage de le bruit effet de l'entraînement	comparaison ABR/PEASp chez enf N et TA	démontrer que l'encodage tonique est préservé dans PEASp et peut être dégradé chez les malentendants	étude encodage parole ds bruit chez cochon d'inde	
STIMULUS sous cortical	échelles tonales de fréquence variant pour un même stimulus	sens purs sinusoïdeux	syllabe de 5 formants /da/	click/transit formantique (da/sans la voyelle)	voyelle /u/ et /a/	deux /ada/ de 5 formants chacun	
LATENCE			onde V	onde V, onde A			
AMPLITUDE						mesure pic à pic et RMS pour l'onset	
PENTE							
AIRE							
DUREE							
LATENCE							
Meure de la FFT	amplitude spectrale de la RSF sur la FFT	amplitude spectrale de la RSF sur la FFT	amplitude spectrale de la RSF sur la FFT et courtout des premiers formants	acc C et F de la RSF (dépendant du FO)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1...)	
Spectrogramme	RSF spécifique de puissance en fonction du temps, amplitude et fréquence des pics spectraux en fonction de l'intensité et du temps						
Cross-correlation							
Corrélation simple			stimulus-RSF		stimulus-RSF		
Autocorrelation et autocorélogrammes							
RMS							
Meure amplitude, représentation temporelle, en μ V	RSF					de la réponse tonique, qui se situe dans les 20 ms suivant onset	
RESULTATS	codage tonique dans la RSF, l'amplitude est inférieure pour les sens descendants/sens ascendants, système auditif plus sensible aux sens ascendants	diminution du phase locking, plus on monte dans le système auditif	les sujets ont des ABR normaux (onde V entre 6 et 8 ms), latence onde V=7.7 C=17.7 F=30 Schiz N et certains TA ont début aug de V dans l'onde C et les sens TA ont des latences V.C et F normales l'entraînement ne fait pas varier l'amplitude de la discrimination et la résistance à la dégradation de le bruit est améliorée pour les enfants chez les TA qui ont des latences augmentées	réduction du spectre des hautes fréquences, baisse du phase locking et allongement latence onde V chez les enfants TA, effet de l'entraînement	diminution de l'amplitude avec C=17.7 F=30 Schiz N et certains TA ont début aug de V dans l'onde C et les sens TA ont des latences V.C et F normales l'entraînement ne fait pas varier l'amplitude de la discrimination et la résistance à la dégradation de le bruit est améliorée pour les enfants chez les TA qui ont des latences augmentées	diminution de l'amplitude avec C=17.7 F=30 Schiz N et certains TA ont début aug de V dans l'onde C et les sens TA ont des latences V.C et F normales l'entraînement ne fait pas varier l'amplitude de la discrimination et la résistance à la dégradation de le bruit est améliorée pour les enfants chez les TA qui ont des latences augmentées	diminution de l'amplitude (μ V) l'onset et de la transition tonique dans la RSF (fig 10b, 7). La voyelle apparaît être respectée dans le bruit (steady state)

« Étude de l'encodage des sons de parole par le tronc cérébral dans le bruit »

REFERENCE	Hayes et al 2003	Kraus et al 2003	Nicol et al 2004	Russo et al 2004	Krishna et al 2004	Wible et al 2004	C
OBJECTIF	étude de la plasticité du système auditif chez les enfants TA	neuronal animal: spécialisation corticale gauche pour parole, étude des TA	mise au point	établir des valeurs normatives en PEASP	RSF contient les informations du pitch	PEASP chez enfants N/TA	
STIMULUS sous cortical	ABR/PEASP /da/ /ga/ se différenciant par la fréq de F3	/da/ (5 formants)	/da/ (5 formants)	/da/ (5 formants)	4 syllabes chinoises spectralement différentes mais avec F0 équivalents	/da/ (5 formants)	
LATENCE	onde V, onde A	onde A	onde A	onde V, onde A		onde V, onde A	
AMPLITUDE			RMS	onde V, A (par méthode ondlettes)		onde V, A	
PENTE				complexe VA		complexe VA	
AIRE				complexe VA			
DUREE				complexe VA			
LATENCE		1er pic RSF		pic G, F			
Meure de la FFT	amplitude spectrale de la RSF sur le FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur le FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur le FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur le FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur le FFT (division en segments: F0, F1...)		
Spectrogramme				RSF pour déterminer la composition spectrale et les amplitudes			
Cross corrélation				latences C, F			
Corrélation simple	stimulus/RSF	stimulus/RSF	stimulus/RSF				
Autocorrelation et autocorrélogrammes					extraction du pitch de la RSF (calcul durée et amplitude) et étude intervalles inter pics de la RSF		
RMS				amplitude			
Meure amplitude, représentation temporelle, en uv							
RESULTATS	latence onde A=7.5ms, 1er pic RSF à 17.7ms, les TA ont diminution du phase locking de le bruit et une diminution de l'énergie de la RSF, diminution de l'onde /RSF et des composants basse fréq de bruit (effet plus marqué au niveau cortical /TC)	latence onde A=entre 8 et 10ms, C=16 à 20ms, F=40ms	latence C=19ms, modification des latences avec l'entraînement		RSF phase lockée sur F0 et les harmoniques	RSF entre 11.4 et 45 msms, F0=114Hz, F1=229-686, F2 à F5 dépassent les 1000Hz et non visibles sur RSF, latence onde V=2.5ms, cédé A=7.2ms, C= entre VA=0.88/Vrms, pente VA est diminuée mais durée augmentée et amplitude diminuée chez TA, moins bon encodage des hautes fréq chez TA,	

REFERENCE	Russe et al 2005	Abrams et al 2006	Johnson et al 2005	Wible et al 2005	Kraus et al 2005	Song et al 2006	d
OBJECTIF	effet de l'entraînement sur le PEASP des enfants TA	déficit de l'encodage temporel niveau TC a un impact négatif sur le traitement cérébral	encodage des caractéristiques acoustiques de la source et du filtre la partie vocale de la parole évoque la partie périodique (phase locking) de la RSF	effet du bruit sur les réponses corticales chez les N et TA,	identité du locuteur ainsi que la tonalité du message seraient contenues dans différents composants du PEASP : source-inhat et filter-inhat	relation entre PEAP et PEASP chez enfants N et TA	
STIMULUS sous cortical	/da/ (5 formants)	/da/ (5 formants)	/da/ (5 formants)	/da/ (5 formants)	/da/ (5 formants)	/da/ (5 formants)	
LATENCE	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	
AMPLITUDE	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	onde V, onde A	
PENTE AIRE			complexe VA	complexe VA	complexe VA	complexe VA	
DUREE			complexe VA	complexe VA	complexe VA	complexe VA	
LATENCE	latence O (offset) et pics D E F (F0)	latence O (offset) et pics D E F (F0)			1er pic RSF		
Meure de la FFT	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1.)	
Spectrogramme							
Cross-correlation							
Corrélation simple			stimulus/RSF				
Autocorrelation et autocorralogrammes			RSF				
RMS			amplitude				
Meure amplitude, représentation temporelle, en μ V							
RESULTATS	complexe VA apparaît à 6.5ms après le début du stimulus. Phase locking sur F0: ondes DEF (entre 20 et 45ms), latence V-7.8, A-E latence onde O (offset)-49ms, les pics VAO reflètent des mécanismes neural différents de pics DEF, les pics DEF sont un peu plus tardifs qu'en hemisphere g dominant (timing de onset et offset corrélate et non RSF) discernement mieux /da/ de /ga/ TA ont une augmentation des latences O (ou O moins pas des latences DEF, TA ont diminution amplitude de F1 mais pas de F0	complexe VA apparaît entre 2.5 et 7.2 ms après le début du stimulus, la durée du complexe VA est allongée et atteint 7.2 ms chez les TA, pas plus profonde chez les TA, pas de modification dans la latence, l'onde C débute du tronc cérébral, onde C débute de la partie, onde C débute du stimulus;	source=F0 (= pics D E F), filter-formants (petites fluctuations de l'onde face entre D et E), amplitude VA début du stimulus niveau tronc cérébral, onde C débute de la partie, onde C débute du stimulus;	source=F0 (= pics D E F), filter-formants (petites fluctuations de l'onde face entre D et E), amplitude VA début du stimulus niveau tronc cérébral, onde C débute de la partie, onde C débute du stimulus;	processus neurologiques différents entre une réponse à une tâche et à des sons de parole, et que seul les PEASP sont différents chez les TA, TA: 5.8ms, V=6.97±0.3ms N=5 à 85 chez TA, amplitude de VA (pic négatif)= 0.37 chez N et 0.31 chez TA, TA: 5.8ms, V=6.97±0.3ms, augmentation de la latence d'apparition de V aussi bien chez N que TA, N=10, TA=10, diminution du pic négatif de V (0.19±0.07)		

« Étude de l'encodage des sons de parole par le tronc cérébral dans le bruit »

REFERENCE	Musacchia et al 2007	Wong et al 2007	Warner et al 2004	Song et al 2008	Johnson et al 2007	Song et al 2008	e
OBJECTIF	la pratique musicale influe sur l'encodage sous cortical	encodage sous cortical du pitch	étude trouble encodage des TA surtout ds le bruit	certaines TA présentent des défauts d'encodage temporel PEASP, ces défauts sont-ils engendrés par le TA ou des niveaux inférieurs et si ils associes à des modif onde III	les TA ont surtt des pb de résolution temporel dont on veut étudier l'impact sur l'encodage de la parole	étude de l'entraînement sur la capacité à détecter la hauteur tonale	
STIMULUS sous cortical	/dəl/ (5 formants)	stimuli mandant différents en F0	/dəl/	/dəl/	/dəl/	/m/ avec 3 pitch différents	
LATENCE	onde V, onde A	onde V, onde A		onde V, onde A	onde V, onde A		
AMPLITUDE	onde V, onde A			onde V, onde A	onde V, onde A		
PENTE				complexe VA	complexe VA		
AIRE							
DUREE				complexe VA	complexe VA		
LATENCE		1er pic RSF	pic D E F O (à 48ms) par cross correlation de 19 à 40ms		1er pic RSF		
Meure de la FFT	amplitude spectrale de la RSF sur la FFT (division en segments F0, F1..)	amplitude spectrale de la RSF sur la FFT (division en segments F0, F1..)	amplitude spectrale de la RSF sur la FFT (division en segments F0, F1..)		amplitude spectrale de la RSF sur la FFT (division en segments F0, F1..)	amplitude spectrale de la RSF sur la FFT (division en segments F0, F1..)	
Spectrogramme			RSF		RSF	RSF	
Cross-correlation			onset RSF		onset RSF	onset RSF	
Corrélation simple			stimulus-RSF				
Autocorrelation et autocorélogrammes			extraction du pitch de la RSF (calcul durée et amplitude) et durée intervalles entre pics de la RSF	RSF	RSF	pour l'étude du pitch, pitch tracking en onsetspectral dominance de la pitch/noise ratio	
RMS			amplitude RSF	amplitude RSF	amplitude RSF		
Meure amplitude, représentation temporelle, en μ V		RSF	RSF		RSF		
RESULTATS	les musiciens présentent une latence de l'onset plus courte, plus ample et un meilleur phase locking niveau RSF (surtout niveau F0)		effet positif de la musique à long terme sur la perception du pitch	certaines TA ont anomalies des PEASP qui peuvent diminuer avec l'entraînement	les anomalies visibles en PEASP ont pour origine une modulation continue. Il y a de différences au niveau de l'onde I et II. certains TA ont des latences V retardées ou un caractère VA imprécis (division de la pente, durée prolongée de VA)	latence TA (V=7.78, A=0.64, C=18.8, D=0.19, E=0.95, F=0.6, O=48.86 et pente VA=1.16) les latences augmentent chez certains TA : ondes ACO-filter, diminution du phase locking sur les htes fréq chez certains TA; certains TA encodent moins bien les indices du filtre et normalement la source	il existe une plasticité, même à l'âge adulte

REFERENCE	Johnson et al 2008	Russo et al 2008	Hannikainen et al 2009	Johnson et al 2008	Russo et al 2009	Gkoe et al., 2010	f
OBJECTIF	étude de la maturation des clics et PEASP chez enfants de 3 à 12 ans	encodage du pitch chez les autistes	étude de l'existence d'une latéralité au niveau sous cortical	encodage des indices temporaux et spectraux au niveau sous cortical	étude du déficit de l'encodage temporel (timing) et fréquentiel (phase locking) chez les autistes	revue de l'approche méthodologique du recueil et de l'analyse des PEASP	
STIMULUS sous cortical	/da/	/yal/ avec pitch modifié	/dal/	/ga'/ /da/ /ba/ différent selon les fréquences	/da/ dans silence et bruit	voyelles, syllabes, synthétiques/naturelles/hybrides	
LATENCE	apparition onde V-A en moy à 6.5ms		onde V, onde A		onde V, onde A	onde V, onde A	
AMPLITUDE	onde V, onde A		onde V, onde A		onde V, onde A	onde V, onde A	
PENTE AIRE	complexe VA		complexe VA		complexe VA	complexe VA	
DUREE	élargissement du complexe				complexe VA	complexe VA	
LATENCE	pic D E F O (à 48ms)		des 7 pics de la RSF		des 7 pics de la RSF	des 7 pics de la RSF	
Mesure de la FFT	de la RSF de 21.9 à 40.6 ms		amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1...)		amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1...)	amplitude spectrale de la RSF sur la FFT (division en segments: F0, F1...)	
Spectrogramme						RSF	
Cross-correlation						onset RSF	
Corrélation simple						stimulus/RSF	
Autocorrelation et autocorréogrammes	pour l'étude du pitch, pitch tracking, error/spectral dominance de F0/pitch noise ratio					RSF	
RMS					amplitude RSF	amplitude RSF	
Meure amplitude, représentation temporelle, en JV	de la RSF de 21.9 à 40.6 ms par RMS						
RESULTATS	la latence de l'onde V est mature dès 2 ans, la représentation de l'onset varie selon l'âge, l'amplitude de F0 et de ses harmoniques est aussi harmonique est diminuée chez le petit enfant	moins bon encodage des harmoniques et de F0 chez certains autistes	le pic C encoderaient la transition entre la partie aperiodique et la transition formantique, les réponses en provenance de l'oreille droite apparaissent toujours plus tard que celles de gauche au niveau des premiers pics de la RSF, avec une meilleure robustesse au niveau de F1, A=7.6, D=22.52, E=30.96, F=39.33, G=48.14, amplitude spectrale F0=6.03, F1=0.8	les latences PEASP sont plus courtes pour /ga'/ puis /da/ puis /ba/	retard latence VA chez autiste et élargissement VA de même pour D et F	partie transitoire apparaît 6 à 10ms après le début du stimulus, des algorithmes (équilibrage par les ondeslettes/filtre passe haut) recherchant un équilibrage entre différents pics (et confirmé visuellement); l'analyse de la partie soutenue fait appel à des techniques d'analyses stationnaires et glissantes selon une technique de RMS cross correlation Fourier et représentation sous forme de correlogrammes	

Ainsi, l'analyse est tout d'abord effectuée de manière visuelle par deux lecteurs entraînés, aveugles des conditions d'acquisition des traces. Chacun des lecteurs identifie chaque pic de l'onset et de la RSF afin que puissent être effectuées des mesures de latence, amplitude et durée. Ces résultats sont ensuite comparés entre eux au moyen de cross correlation. Toutefois, notre but était de pouvoir disposer d'un algorithme permettant une analyse automatisée au niveau individuel, afin de s'affranchir de la subjectivité des observateurs extérieurs. Ainsi, l'ensemble des caractéristiques de l'onset tout d'abord, puis de la RSF ont été modélisés afin de créer une « onset » et une « RSF » type.

Comme décrit dans la présentation : « Détection Automatique d'Indices Pertinents dans les Réponses Electrophysiologiques de Type Potentiels Evoqués Auditifs recueillis en réponse à un son de parole (Speech ABR) » présentée en annexe, deux méthodes ont été utilisées.

La première méthode se base sur le calcul du coefficient directeur minimum et de l'énergie maximale. Ainsi, partant de l'hypothèse que l'onset response est un signal haute fréquence, pour chaque trace PEASP on applique un filtre entre 300 et 550Hz. Puis on

prend la dérivée du résultat obtenu afin d'obtenir les coefficients directeurs à chaque instant. On prend le plus grand coefficient se dirigeant vers le bas, ce qui caractérise l'onset, qui va nous donner la localisation de l'impulsion. De même, on calcul l'énergie du signal filtré, le pic d'énergie maximale permettant la localisation temporelle de l'onset. En couplant l'analyse du coefficient directeur minimum à celle de l'énergie maximale, on peut localiser temporellement l'onset response. Toutefois cette méthode n'est pas applicable en cas de trace fortement parasitée, car dans un tel cas, il existe trop d'impulsions parasites pouvant être identifiées à tort comme une onset (Fig 33 et 34).

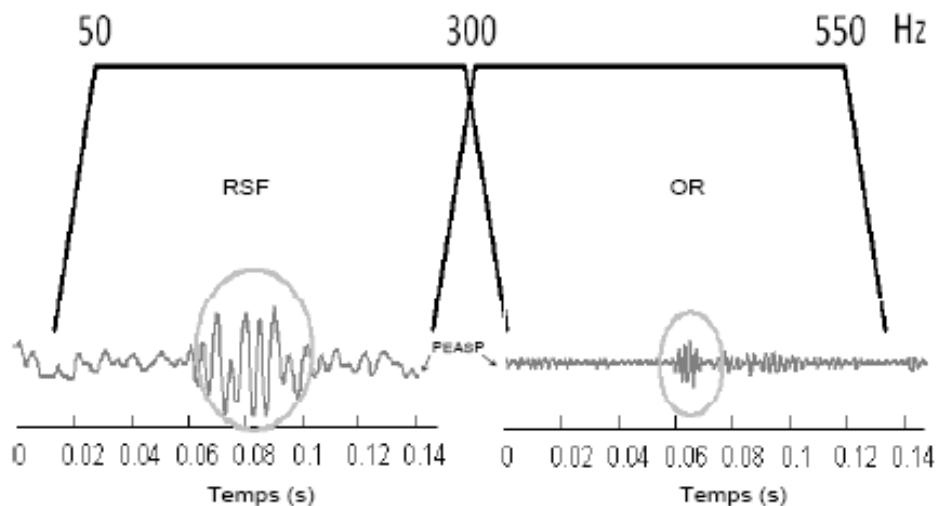


Figure 33 : Représentation temporelle de deux traces PEASP avec les caractéristiques temporelles de l'onset response (OR) et de la réponse soutenue à la fréquence (RSF).

La RSF se situe au niveau des basses fréquences, entre 50 et 300 Hz, alors que l'onset response appartient plutôt aux hautes fréquences entre 300 et 550 Hz.

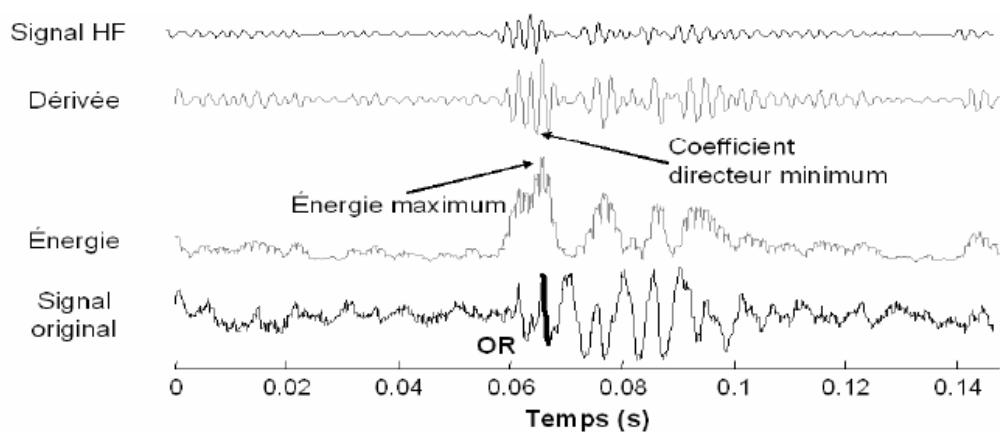


Figure 34 : Illustration de la première technique utilisée pour localiser temporellement l'onset response (OR).

Représentation temporelle, de haut en bas: PEASP filtré avec filtre haute fréquence. Dérivée de ce signal filtré permettant de trouver le coefficient directeur minimum.

Représentation temporelle de l'énergie du signal filtré permettant de localiser là où l'énergie est maximale. A l'aide du coefficient directeur minimum et de l'énergie maximale, en recoupant ces données on localise temporellement l'onset response.

La même méthode de traitement au moyen du calcul de la dérivée et de l'énergie est appliquée pour identifier la RSF. Toutefois, la RSF est un signal de plus basse fréquence que l'OR, et l'énergie sera maximale non plus sur une durée brève mais sur une durée plus constante (Fig 35).

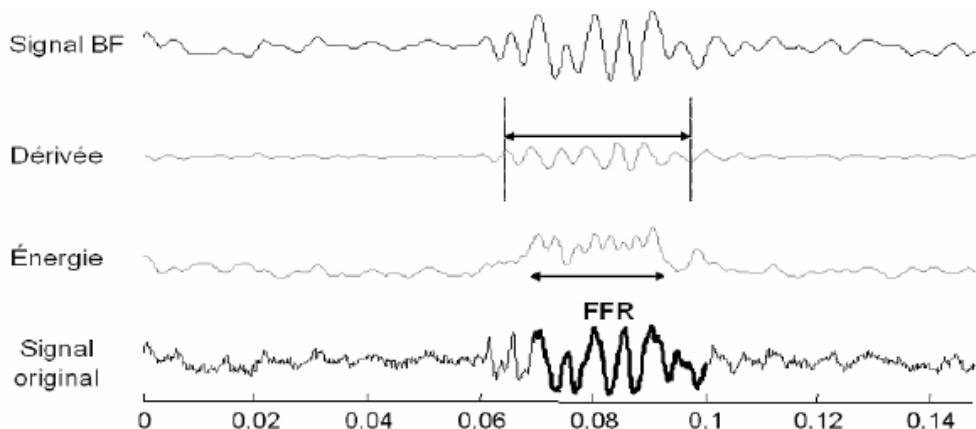


Figure 35 : Illustration de la première technique utilisée pour localiser temporellement la RSF (réponse soutenue en fréquence, ou FFR).

De haut en bas: PEASP filtré avec un filtre basse fréquence. Dérivée de ce signal filtré permettant de trouver des amplitudes régulières. Représentation temporelle de l'énergie du signal filtré permettant de localiser là où l'énergie maximale et constante sur plusieurs millisecondes. A l'aide de la dérivée et du calcul de l'énergie, en recoupant ces données on localise temporellement la RSF.

La seconde méthode, fait appel à la création de modèles d'onset et de RSF. Au préalable, il était nécessaire de définir les caractéristiques principales de l'onset et de la RSF (tableau 4). Ainsi, l'onset type a été créée à partir de deux triangles inverses, avec une durée et une pente moyennes tirées de la moyenne de durée et de pente de plusieurs onsets obtenues sur des traces précédentes et clairement identifiées comme « onset » par trois lecteurs (CR, AJ et AM), de manière indépendante (Fig 36).

De même, la RSF type fut créée à partir de RSF identifiées précédemment, avec une périodicité et une amplitude moyenne dérivées de celles-ci (Fig 37).

théorie	amplitude	impulsion	régime		fréquences		Énergie	Périodicité	latence		durée
			transitoire	stationnaire	basse	haute			breve	longue	
OR	+	++	+	-	-	+	++	-	+	-	-
FFR	+	-	-	+	+	-	+	++	-	+	+

Tableau 4 : Tableau rassemblant les caractéristiques des PEASP.

OR : onset response (ondes V-A) FFR : RSF réponse soutenue à la fréquence

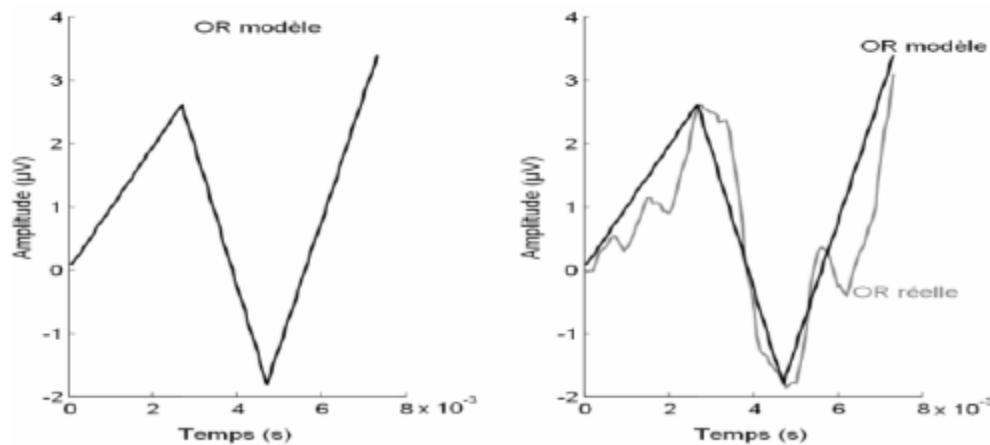


Figure 36 : Modélisation de l'onset response.

La caractérisation de l'OR a été faite en isolant l'OR sur un grand nombre de personnes, trois droites ont ainsi été créées afin de ressembler au maximum à l'OR réelle.

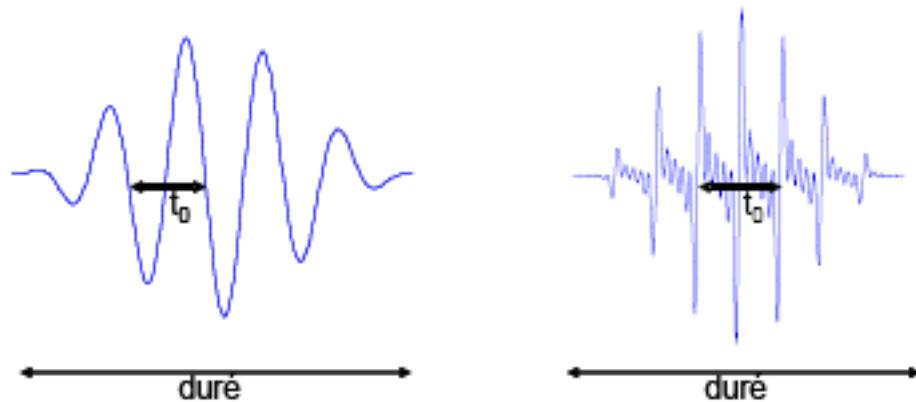


Figure 37 : Modélisation d'une RSF type. Pour caractériser la RSF, nous avons créé un signal complexe harmonique se rapprochant au plus de la FFR réelle.

abscisse : durée du signal ; ordonnée : amplitude du signal

Une fois l'onset type et la RSF type créées, il faut trouver la localisation de celles-ci sur des vraies traces. Une trace PEASP correspond à un certain nombre de points successifs temporellement et ayant chacun une fréquence caractéristique. A partir de là on va prendre des bandes de fréquences, c'est-à-dire on va découper une bande de fréquence de 12 Hz et on va l'appliquer à la trace PEASP : on effectue ce filtrage entre 80 et 3200 Hz. Une fois le signal filtré par toutes les bandes, nous appliquons une fenêtre glissante contenant l'onde correspondante (OR ou RSF) et nous obtenons un coefficient de corrélation (cross correlation). On obtient donc en sortie une trace avec le coefficient de corrélation pour toutes les bandes correspondantes. Ensuite peut-être créer un spectrogramme qui donne en amplitude la valeur de la corrélation pour chaque bande. L'onset response correspond alors au pic d'énergie maximale sur toutes les fréquences (Fig 38), alors que l'isolation de la FFR

se fait différemment. En effet, pour la RSF, la somme de toutes les bandes de fréquences sur le spectrogramme permet de connaître l'énergie à chaque instant. Les deux harmoniques de la fréquence fondamentale du signal sont visibles lorsque la RSF est présente, l'isolation de celle-ci sur la bande d'énergie nous permet d'en déduire la caractérisation de la FFR (Fig 39).

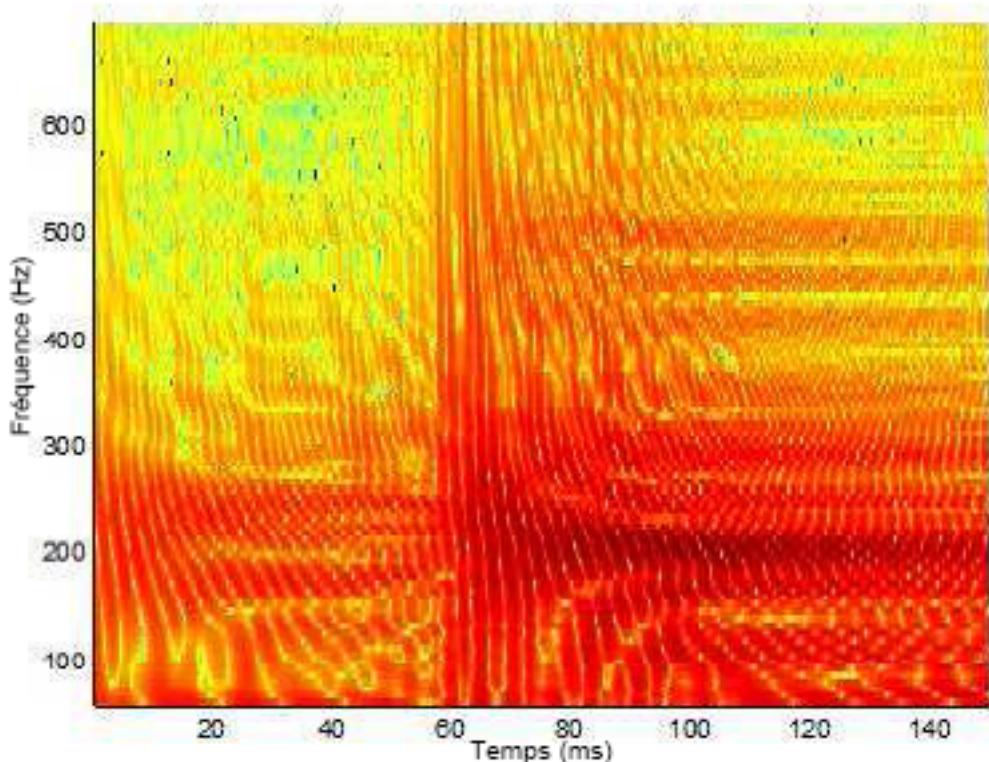


Figure 38 : Représentation spectrographique de l'onset response après application de la seconde méthode d'analyse.

Le pic d'énergie maximale correspond à l'onset response (entre 6 et 7ms).

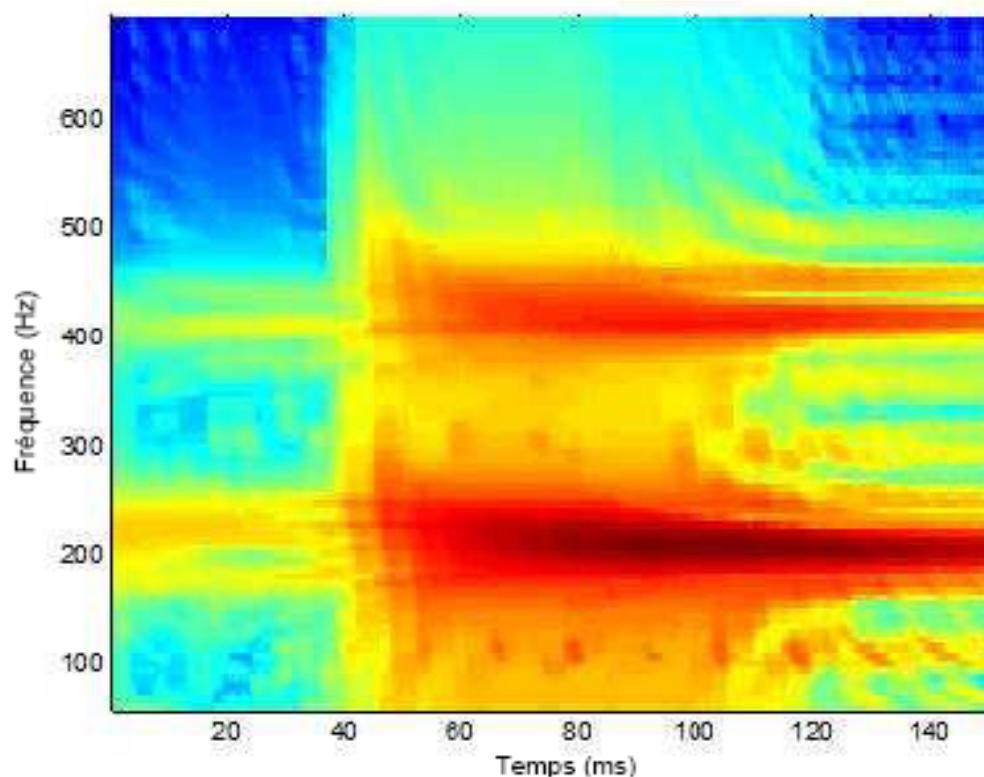


Figure 39 : Représentation spectrographique permettant de localiser la RSF.

Lorsque les deux harmoniques sont présentes, on en déduit la durée de la FFR.

Cette mise au point était nécessaire à l'avancée des travaux. En effet, il n'était pas envisageable de réaliser un nouveau protocole si la chaîne de mesure ne permettait pas d'obtenir des traces individuelles non parasitées. Par ailleurs, la mise au point d'un système d'analyse automatique, venant en complément de l'analyse visuelle nous paraissait être un atout majeur.

Toutefois, toutes ces mises au point ont de nouveau nécessité la réalisation de PEASP sur des sujets volontaires, ceci afin de s'assurer de l'intérêt et de la validité des modifications effectuées.

Le fruit de ce travail sera visible lors des expériences suivantes, même si quelquefois certaines traces n'ont encore pu être totalement exploitées.

Expérience 4 : Protocole expérimental
Etude de l'encodage
d'un son de parole détérioré (bruité)
au niveau du tronc cérébral.

Problématique :

Encodage d'un son de parole par le tronc cérébral en condition adverse.
 Effet de l'intensité sur l'encodage sous cortical.

Matériaux et méthodes :

- Population étudiée : 12 sujets
- Tests neurophysiologiques : les PEASP (durée d'acquisition : 70 minutes par sujet)
- Stimuli : ils sont au nombre de cinq
 - (1) un /ba/ de 100 ms ; (2) l'enveloppe du signal /ba/ redressée et filtrée passe bas à 500 Hz, modulée avec le speech noise ; (3) le /ba/ + l'enveloppe ; (4) /ba+ba/, ce qui équivaut à une augmentation de l'intensité perçue de 6 dB HL ; (5) l'enveloppe doublée.

Stimuli	Condition(s) iso-énergétique(s)	Intérêts
1	BA	2
2	ENV	1
3	BA+ENV	4 et 5
4	BA+BA	3 et 5
5	ENV+ENV	3 et 4

Stimulation binaurale à un niveau de 45 dB SL, par blocs de 1200, à une cadence de 6,6 stimuli par seconde. Recueil des traces à l'aide du centor USB Racia Alvar®, sur 2 voies d'enregistrements, via 4 électrodes d'argent (Nicolet®, 10mm), fixées sur le scalp du sujet au niveau du vertex (électrode active), du front (masse), et des deux mastoides (droite pour la voie 1 et gauche pour la voie 2). Chaque trace est moyennée 2500 fois en polarité alternée.

Résultats :

Les stimuli 1 et 2 ont permis une première approche de l'influence de l'intensité de stimulation sur les caractéristiques des PEAPSP. Se pose la question de la réelle influence de l'intensité de stimulation sur les caractéristiques des réponses de type PEASP, latence, synchronisation et amplitudes.

Expérience 4 : Etude de l'encodage, d'un son de parole détérioré (bruité) au niveau du tronc cérébral.

4.1. Introduction

D'après Van Tasell et al 1987 ; Rosen 1992, d'importantes informations fréquentielles sont encodées dans l'enveloppe temporelle d'un son.

Toutefois, la manière dont les variations dynamiques de la parole sont codées ou représentées dans le système auditif reste encore source à spéculation. Il semble bien que la phase locking et la vitesse d'excitation jouent un rôle dans la discrimination de l'intensité, d'autres facteurs interviennent.

Le but premier de cette expérience est l'étude de l'encodage d'un son voisé signifiant dans le silence et en condition adverse, la condition adverse étant ici représentée par un signal composé de l'enveloppe temporelle du /ba/ avec les mêmes caractéristiques spectrales que le /ba/ avec une phase aléatoire, ainsi le son obtenu est sur le plan perceptif non reconnu comme un /ba/.

Ainsi l'objectif de cette expérience est la mise en évidence de l'ensemble des caractéristiques dont notamment les caractéristiques temporelles et fréquentielles de l'encodage de signaux de parole chez des sujets normaux. Dans ce but, nous avons utilisé un signal /ba/ non dégradé dont nous avons comparé les réponses avec ce même signal désynchronisé par du bruit, afin de mieux comprendre l'impact du bruit sur les caractéristiques nécessaires à l'encodage d'un son de parole.

4.2. Matériel et méthodes

4.2.1. Population étudiée :

Douze sujets furent inclus, tous droitiers, et d'âge compris entre 20 et 30 ans. Avant inclusion, l'absence d'antécédent médical était vérifiée ainsi que leur état otologique clinique par un examen otoscopique et une audiometrie tonale, dont les seuils devaient être strictement normaux entre 125 et 8000 Hz (Audiomètre AC 40, Interacoustics®). Ceci complété par la réalisation de PEAP afin d'éliminer tout problème rétro cochléaire. Comme les expériences précédentes, ces tests ont été réalisés en accord avec la loi Huriet et elles furent approuvées par le comité d'éthique de Lyon (CCPPRB # 05-024). Un consentement éclairé a été obtenu auprès de chaque sujet.

4.2.2. Tests neurophysiologiques :

- Stimuli utilisés : Les stimuli utilisés sont de durée inférieure à ceux utilisés lors des expériences 1 et 2. Les stimuli se composaient :

Tout d'abord d'un /ba/ de 100 millisecondes, précédé et suivi de 10 millisecondes de silence. Ce /ba/ était envoyé en condition binaurale à une intensité de 45 dBSL, à une fréquence de 3Hz. Ce stimulus, voisé et signifiant est analysé dans le silence.

Le second stimulus est l'enveloppe du signal /ba/ décrit ci-dessus, redressée et filtrée passe bas à 500 Hz (ce qui permet de garder le voisement) et modulée (multipliée) avec le speech noise (bruit blanc filtré passe bas 800 Hz). Au contraire du premier stimulus, l'enveloppe représente un son non voisé et non signifiant. Toutefois, l'enveloppe est un /ba/ avec des informations spectrales qui sont déphasées, ce n'est donc pas une « vraie enveloppe ». Toutefois, le stimulus étant synthétisé à partir de l'enveloppe du /ba/ celui-ci reste synchronisé sur le /ba/.

Le troisième stimulus se composait du premier stimulus : le /ba/ auquel est ajouté le second stimulus : l'enveloppe. Le but de ce stimulus ainsi créé est de permettre l'étude de la synchronisation (voisement) dans le bruit, et de détériorer, à l'entrée, le stimulus signifiant /ba/.

Les quatrième et cinquième stimuli ont été créés afin de disposer de stimuli comparables énergétiquement au troisième stimulus ba+enveloppe et pouvoir ainsi comparer ces conditions en enlevant l'effet de l'énergie.

Le quatrième stimulus est composé de ba+ba, afin de pouvoir étudier l'évolution des latences de l'onset et de la RSF en fonction de l'intensité. Le stimulus était donc le premier stimulus /ba/ mais que l'on avait doublé, ce qui équivalait à une augmentation de l'intensité perçue de 6 dB HL. Le stimulus /ba+ba/ possédait la même durée que /ba/ et était présenté aux mêmes intensités et fréquences.

Le cinquième stimulus est composé de l'enveloppe doublée.

Les différents stimuli sont représentés dans le Tableau 5, ainsi qu'en temporel (fig 40).

Tableau 5 : Représentation des différents stimuli utilisés.

Stimuli		Condition(s) iso- énergétique(s)	Intérêts
1	BA	2	étude encodage son voisé, signifiant
2	ENV	1	étude encodage son non voisé, non signifiant
3	BA+ENV	4 et 5	étude de la détérioration du signal de parole
4	BA+BA	3 et 5	étude de l'effet de l'intensité, condition contrôle de 3 et 5
5	ENV+ENV	3 et 4	condition contrôle de 3 et 4

Ainsi sur le plan énergétique, le premier et le second stimulus sont iso-énergétiques. De même, les troisième, quatrième et cinquième stimuli sont iso-énergétiques entre eux. La comparaison des conditions ba et ba+env permet de comparer les réponses de type speech ABR entre un signal écologique et un signal détérioré, avec une condition iso-énergétique pour le total de la stimulation : /ba+ba/, et une condition iso-énergétique pour le seul signal signifiant /ba/ et ba+env, et des conditions de contrôle permettant de vérifier l'absence de réponse pour le signal détérioré seul (env et env+env), (Tableau 5).

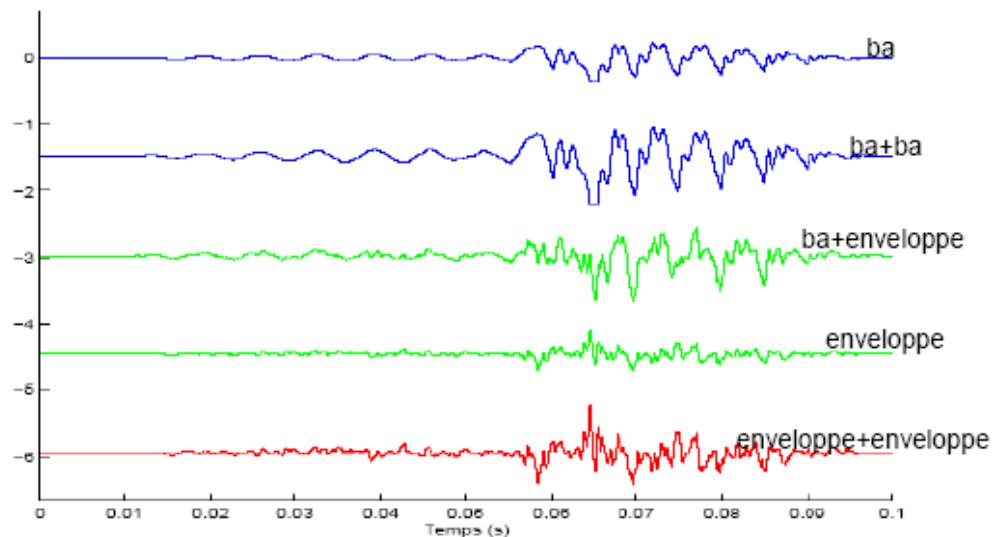


Figure 38 : Représentation temporelle des différents stimuli utilisés.

L'amplitude du stimulus /ba+ba/ apparaît être deux fois supérieure à celle du /ba/.

Recueil des réponses:

Le système de recueil utilisé pour cette expérience et la suivant est identique à celui présenté sur la Fig 29.

Le positionnement des électrodes était identique lors du recueil des PEAP et des PEASP. La fenêtre d'acquisition est de 150 ms et chaque trace est moyennée 2400 fois en polarité alternée, pour une durée d'acquisition de 14 minutes par trace, soit environ 70 minutes par sujet. Afin de s'assurer de la bonne reproductibilité des traces, deux sujets ont été testés deux fois (soit un total de passation de 2h20 pour chacun d'eux).

4.3. Résultats

4.3.1. Obtention de réponse de type PEASP :

On obtient une onset et une RSF visibles sur les PEASP en réponse au / ba/ et /ba+ba/, alors que celles-ci ne sont pas obtenues en réponse aux stimuli /env/ et /env+env/.

La trace temporelle en réponse au ba+env montre une dégradation de l'onset et de la RSF supérieure à ce que l'on aurait pu s'attendre (Fig 41).

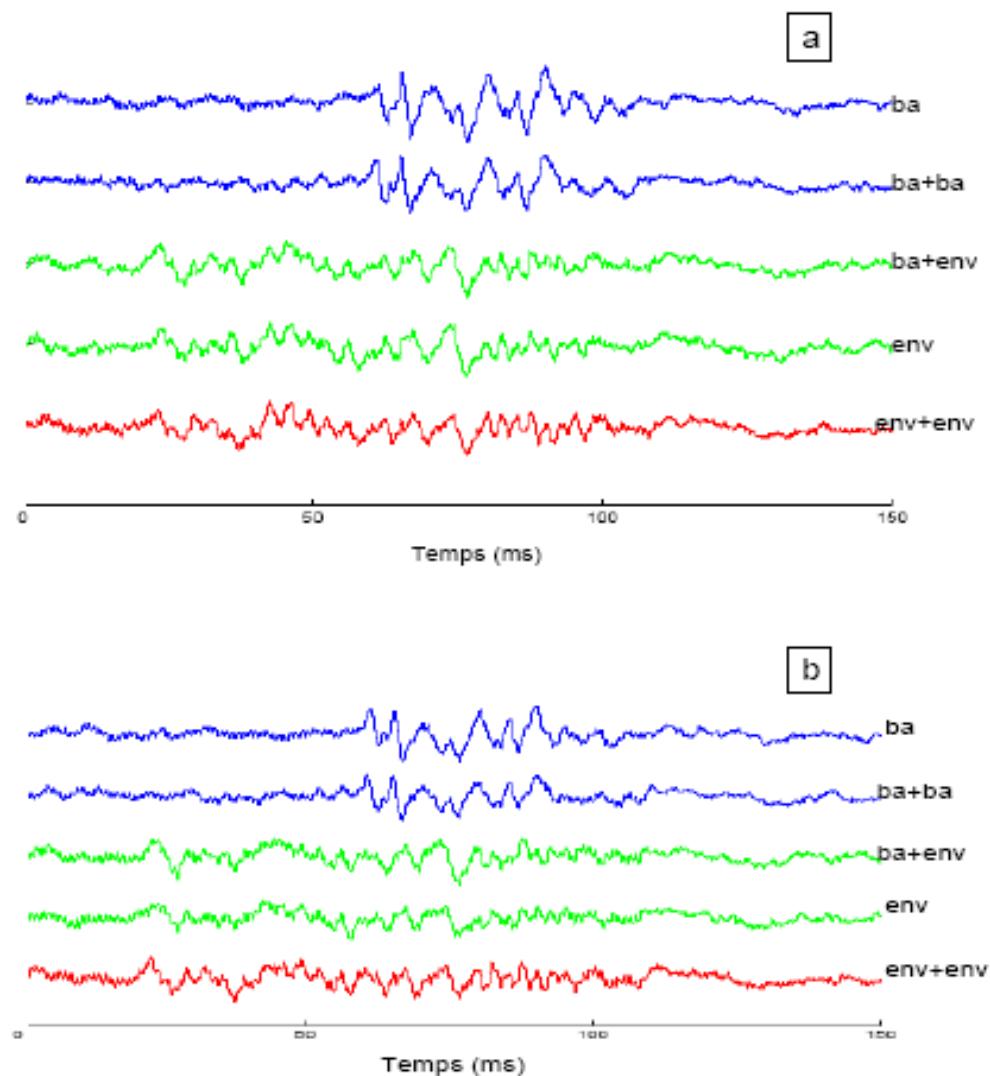


Figure 41 : Représentation temporelle des grandes moyennes en réponse aux différents stimuli.

a : voie 1 ; b : voie 2.

4.3.2. Caractéristiques des réponses en fonction du type de stimulus.

En utilisant la technique d'analyse automatique et en appliquant aux différentes traces, on observe des différences de latences entre les stimuli /ba/ et /ba+ba/, alors que nous n'avons aucune modification en réponse aux autres stimuli (Fig 42 et 43). En effet, les résultats analysés avec une ANOVA par mesures répétées retrouve une différence significative au niveau de la latence de l'onset entre ba et ba+ba : différence moyenne=0,375 millisecondes (erreur standard=0.09 ; p=0,0027) alors qu'il n'existe qu'une différence de 6dB résultats en accord avec ceux précédemment trouvés par akhoun et al.

A contrario, aucun effet n'a été observé entre le côté droit et le côté gauche (différence de 0,07 millisecondes ; p=0,21), Fig 44.

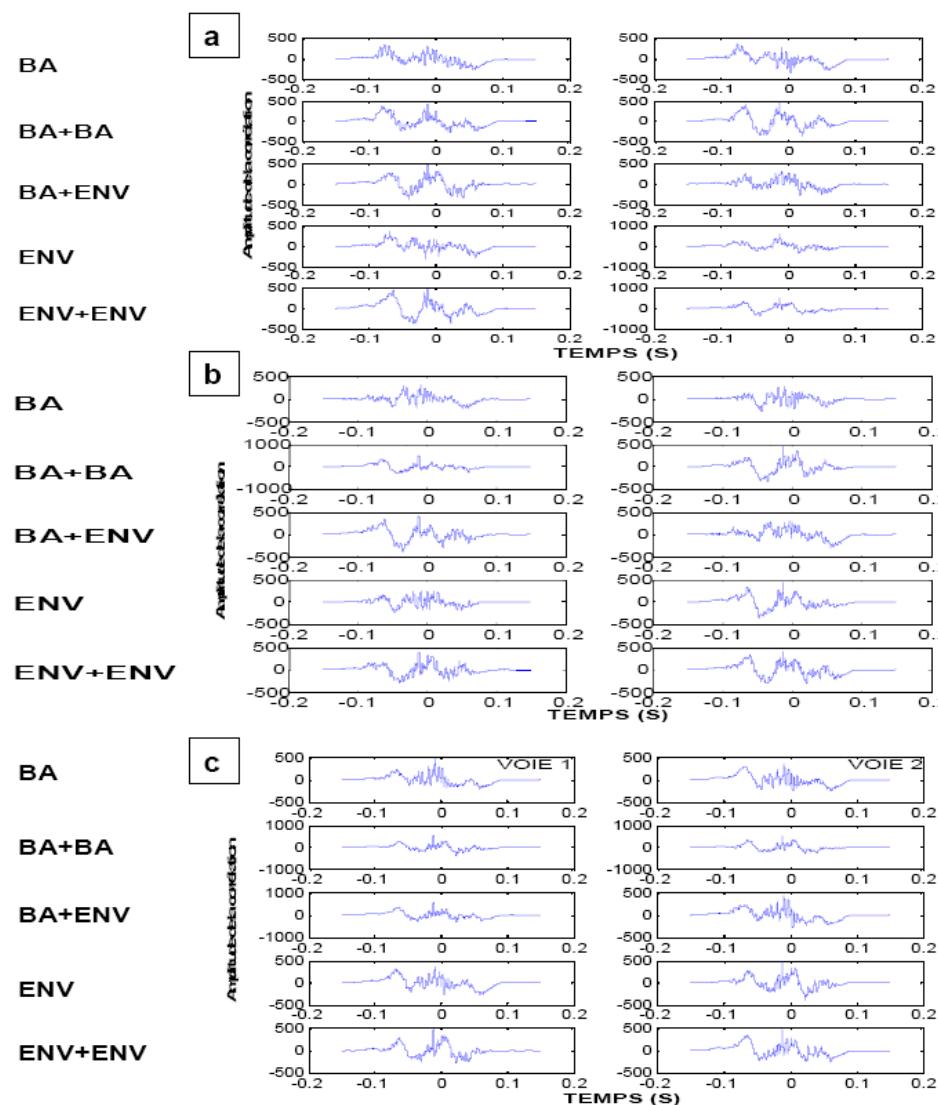


Figure 42 a, b, c : Représentation des latences de l'onset en utilisant la fonction cross correlation.

a : au niveau de l'oreille gauche ; b : au niveau de l'oreille droite ; c : au niveau des deux oreilles.

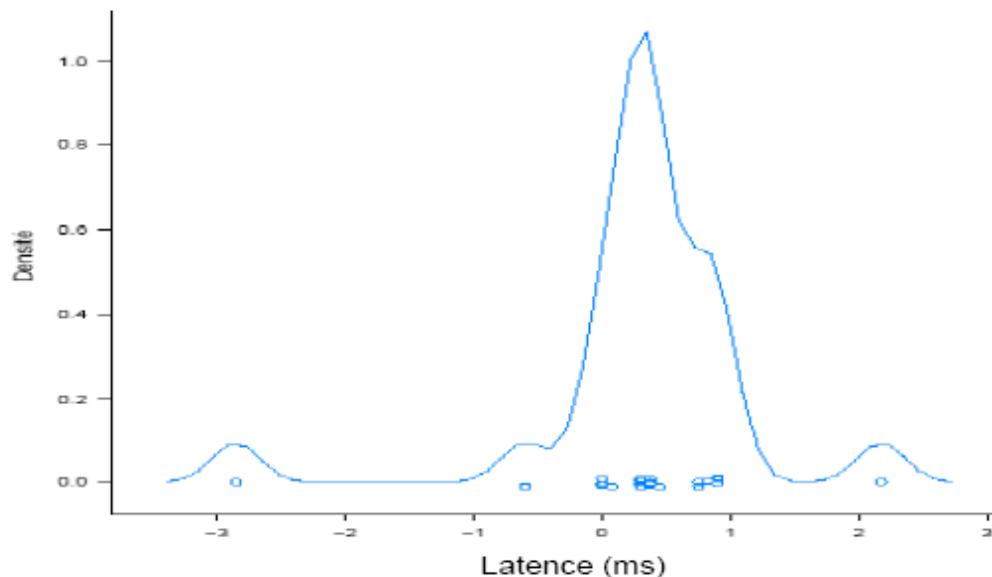


Figure 43 : Représentation des différences de latences de l'onset entre ba et ba+ba.

La différence des latences de l'onset des traces PEASP en réponse à ba et à ba+ba obtenue par cross correlation sont ici représentées pour chaque sujet (ronds bleus) et interprétées sous forme de densité de localisation (courbe bleue). Ainsi la différence maximale retrouvée chez la plupart des sujets est de 0,375 millisecondes en moyenne.

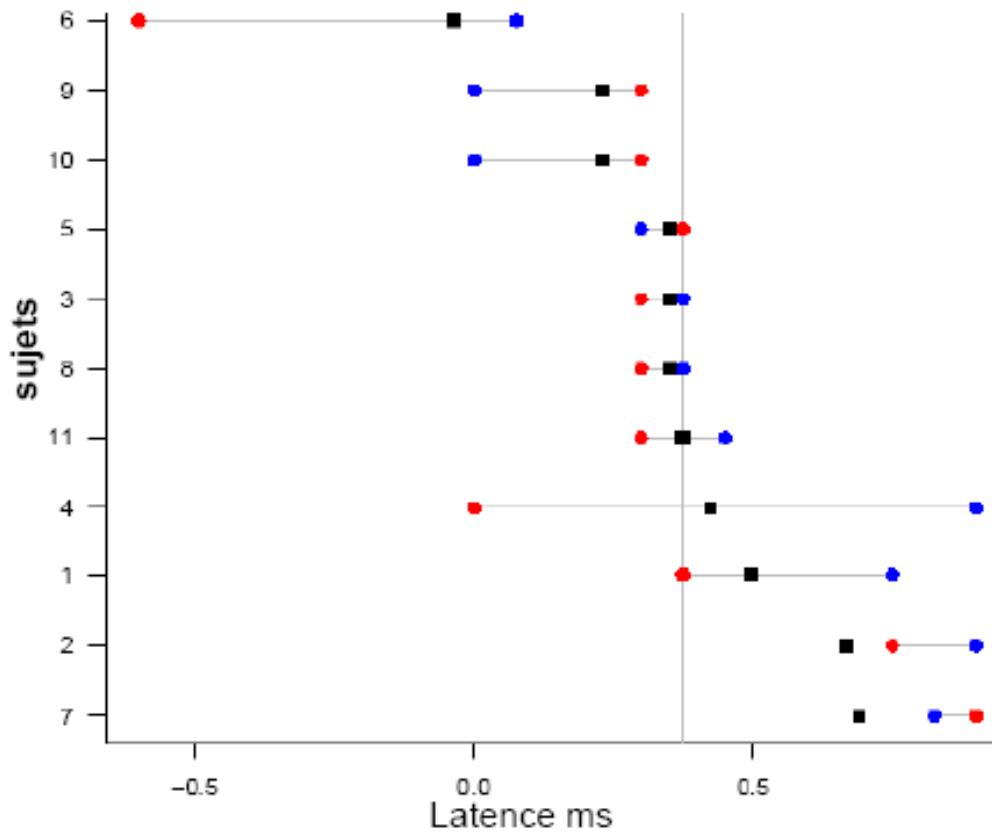


Figure 44 : Représentation de la différence de latence de l'onset entre l'oreille droite et gauche pour chaque sujet.

4.3. Discussion

Les réponses aux stimuli 1 et 4 diffèrent peu entre elles, notamment au niveau de l'amplitude de la FFR, malgré la différence d'intensité de 6 dB entre les stimuli. En revanche, le stimulus 3, /BA+env/ a mis en évidence des réponses de type PEASP très dégradées par rapport au stimulus 4 (ce qui n'est guère surprenant, le stimulus 4 ne comportant qu'un signal signifiant à une intensité double de celle du stimulus 3), mais également très dégradée par rapport au stimulus 1, alors que la partie « signifiante » du stimulus, la syllabe /BA/ était identique à celle du stimulus 1. L'importance de cette dégradation est quelque peu surprenante, au regard de l'absence de différence entre les réponses de type FFR aux stimuli 1 et 4. En aucun cas, cette dégradation n'est équivalente à une diminution d'intensité de stimulation, et met en avant la fragilité de la synchronisation temporelle sur le signal. Enfin, comme on pouvait s'y attendre, pour les conditions contrôles, (2 et 5), aucune réponse de type PEASPn'a été obtenue, les stimuli 2 et 5 (enveloppe, et enveloppeX2), ne comportant que l'enveloppe du BA, avec un bruit aléatoire (phase randomisée) à l'intérieur, ne permettant aucune synchronisation.

Les stimuli 1 et 4 ont permis une première approche de l'influence de l'intensité de stimulation sur les caractéristiques des PEAPSP. La plupart des études sur les

PEASPutisent des stimuli d'intensité forte, avec une intensité unique. La seule étude s'intéressant aux effets de l'intensité sur les PEASPut celle de Akhoun et al. (in press, voir annexe), effectuée sur des grandes moyennes, avec l'ancien système de recueil que nous utilisions, et qui avait permis d'établir un comportement différent des latences de l'onset response et de la FFR.

Certains résultats individuels, où chez un même sujet, on identifie facilement une réponse de type PEASP avec un stimulus /ba/, et où la réponse avec un stimulus /ba+ba/ n'est plus identifiable, nous amène à poser la question de la réelle influence de l'intensité de stimulation sur les caractéristiques des réponses de type PEASP, latency, synchronisation et amplitudes. Cette question nous amené à proposer un protocole d'étude systématique des fonctions entrée/sorties, protocole décrit dans l'expérimentation 6.

Expérience 5 : Protocole expérimental
Etude de l'encodage du voisement d'un son de parole par le tronc cérébral

Problématique :
Existe-t-il un encodage du voisement dès l'étape sous corticale du traitement du signal, au niveau du tronc cérébral?

Matériels et méthodes :
•Population étudiée : 21 sujets (1ère partie : 8 sujets, 2ième : 13 sujets), tous droitiers.
Test de latéralité manuelle d'Edinburgh et test de latéralité auditive (dichotique).
•Test comportementaux :
Perception catégorielle : continuum de 21 signaux.
•Tests neurophysiologiques : les PEASP (L'expérience fut divisée en deux parties de 80 et 95 minutes respectivement).
Stimuli (250 ms): stimulation binaurale à 45 dBSL.
1ère partie de l'étude : ba1, ba6, ba11, ba16 et ba 21 du continuum
2ième partie de l'étude : ba1, ba6, ba11, ba16, ba19 et ba 21 du continuum
Système de recueil identique à celui de l'expérience 4.

Résultats :
Il existe un encodage du voisement dès l'étape sous corticale, avec de plus, un effet corticofuge.

Expérience 5 : Etude de l'encodage du voisement d'un son de parole par le tronc cérébral

La perception de la parole requiert des mécanismes capables d'analyser et d'encoder à la fois les modifications spectrales (fréquence) et temporales du signal acoustique.

Des études comportementales chez le singe ont mis en évidence le rôle des propriétés spécifiques du système auditif dans les phénomènes catégoriels et phonétiques. Les réponses temporelles enregistrées au niveau du cortex auditif primaire (A1) reflètent à la fois un encodage différent des consonnes voisées et non voisées ainsi que la frontière perceptive dont elle serait un corrélat neurophysiologique (Steinschneider et coll., 1995 & 1999). L'étude temporelle des réponses à des sons de parole au niveau du cortex auditif montrent l'importance des mécanismes de traitement temporel dans la perception du voisement. Des études effectuées chez le singe révèlent des schémas d'activités différents : en effet, les consonnes non voisées induisent en langue anglaise une réponse de type « double pic » au niveau de A1 (les bouffées étant verrouillées à la fois sur la libération de la consonne et le début de la vibration des cordes vocales) alors qu'une consonne anglaise voisée n'induit qu'un monopic (la seule bouffée de réponse étant verrouillée sur la libération de la consonne). De plus, le verrouillage de phase sur l'onset augmenterait lorsque l'on passe sur le plan perceptif d'une consonne voisée à une consonne non voisée. Ainsi, la perception d'une consonne voisée serait facilitée quand une seule réponse évoquée est engendrée au niveau cortical, alors que la perception de consonne non voisée serait engendrée par deux réponses individualisables : l'une en réponse à la libération de la consonne et l'autre en réponse au début de la mise en vibration des cordes vocales. La limite entre ces deux schémas de réponse représente la frontière perceptive (Steinschneider et coll., 2005).

Chez l'humain, il existe différentes méthodes d'exploration du traitement du signal acoustique, principalement au niveau cortical, dont l'IRM et la mesure de l'activité d'un neurone font partie. Certaines études utilisant le stéréoélectroencéphalogramme ont mis en évidence un encodage de la structure temporelle du langage par l'hémisphère gauche humain. En effet, les consonnes voisées et non voisées sont traitées différemment au niveau de l'hémisphère gauche, suggérant un traitement temporel spécifique avec l'activation successive des différents neurones, celui-ci n'étant pas retrouvé dans les structures hémisphériques droites. Ainsi, le traitement temporel de l'information acoustique apparaît être plus développé dans l'hémisphère gauche, chez l'humain, alors que ce mécanisme est bilatéral chez le singe (Liégeois et coll., 1999).

Au niveau sous cortical, la seule technique permettant une analyse temporelle et fréquentielle précise, est les PEASP, par sa capacité à donner en temps réel et sur une large population neuronale un aperçu de l'encodage de la parole (Kraus et al, 1998).

L'intérêt de cette expérience repose sur la mise en évidence d'un encodage du voisement dès l'étape sous corticale du traitement du signal, au niveau du tronc cérébral. Un mauvais traitement de la parole, et donc de ses indices acoustiques, pouvant être à l'origine de troubles de la lecture, et de l'écriture (Giraud et al, 2008).

5.1. Matériel et Méthodes :

5.1.1. Population étudiée :

21 sujets, âgés de 20 à 30 ans ont participé à l'étude, les sujets inclus ne présentaient aucun antécédents médicaux particuliers notamment otologiques, et tous avaient un niveau d'étude supérieur. Avant toute inclusion, l'état otologique de chaque sujet était vérifié au moyen d'un examen otoscopique, et d'une audiometrie tonale, dont les seuils devaient être strictement normaux entre 125 et 8000 Hz (Audiomètre AC 40, Interacoustics®). Des potentiels évoqués auditifs en réponse à des clicks (centor USB Racia alvar®) ont été réalisés, dont les latences normales ont permis d'éliminer toute pathologie retro cochléaire. Ces tests ont été réalisés en accord avec la loi Huriet et elles furent approuvées par le comité d'éthique de Lyon (CCPPRB # 05-024). Un consentement éclairé a été obtenu auprès chaque sujet. Avant toute inclusion, la latéralité droite de chaque sujet était vérifiée par un test d'Edinburgh (Oldfield, 1971), ainsi que par un test de latéralité auditive (dichotique : listes de mots différentes dans chaque oreille délivrées à niveau de sonie équivalente, d'où audition symétrique nécessaire). Le test dichotique a été réalisé de deux façons: la première permettant l'étude de l'intégration binaurale (il faut répéter les deux listes), et la seconde l'étude de la séparation binaurale (il ne faut répéter qu'une seule liste c'est à dire soit les mots entendus dans l'OG soit les mots entendus dans l'OD). Lors de la verbalisation des items, les items entendus dans l'oreille droite sont souvent énoncés avant ceux entendus dans l'oreille gauche (car le traitement sémantique se fait à gauche).

5.1.2. Test comportemental :

5.1.2.1. Perception catégorielle :

La perception catégorielle permet de tester la capacité de catégoriser des syllabes formant un continuum. Par exemple, dans le cas du continuum /ba/ - /pa/, les syllabes s'échelonnent entre un /ba/ et un /pa/ (Fig 45), avec un VOT variant par pas de quelques millisecondes. Les sujets sont sensés percevoir la moitié des stimuli comme des /ba/, l'autre comme des /pa/, avec une zone d'incertitude entre les deux (Veuillet et al, 2007).

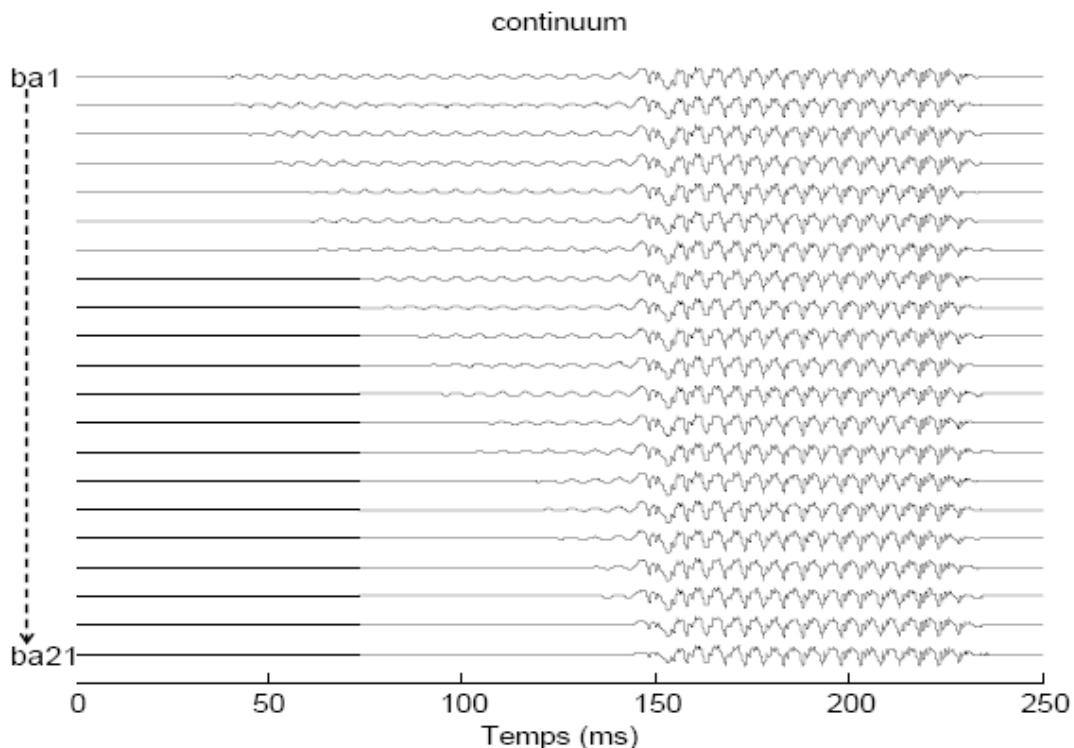


Figure 45 : Représentation temporelle des 21 signaux constitutifs du continuum /ba/- /pa/.

Ba1 (230ms) est reconnu comme un /ba/ et ba21 comme un /pa/. La voyelle reste stable tout le long du continuum, seul le voisement varie (ba1=-93 ms ; ba6=-70 ms ; ba11= -39 ms ; ba 16=-8 ms ; ba19=+9 ms ; ba 21=+17 ms ; soit un VOT long↔/ba/ ; un VOT court↔/pa/).

5.1.2.2. Tests neurophysiologiques :

- Stimuli et système de stimulation:

L'expérience fut divisée en deux parties.

Dans la première partie de l'étude, portant sur 8 sujets, les stimuli suivants furent utilisés : le ba1, ba6, ba11, ba16 et ba 21 du continuum utilisé lors du test de perception catégorielle. Le ba 1 étant entendu et reconnu comme un vrai /ba/, et le ba 21 comme un /pa/.

Chaque stimulus durait donc 250 millisecondes au total avec des périodes de silence incluses avant et après (afin d'apprécier le bruit ambiant), la fenêtre d'enregistrement était de 300ms et la fréquence de répétition de 3Hz.

La stimulation était envoyée en binaural à 45 dBSL.

Dans la seconde partie, 13 sujets furent inclus. Cette seconde partie s'était avérée nécessaire car la frontière perceptive moyenne des sujets se situait entre 16 et 20, nous voulions donc disposer d'un point supplémentaire s'approchant de celle-ci. C'est pourquoi, en plus des stimuli sus-cités, le stimulus ba19 fut rajouté.

Les stimuli possédaient les mêmes caractéristiques que ceux de la première partie, et la stimulation fut aussi effectuée de manière binaurale.

Le système de stimulation est identique à celui détaillé sur la figure 29.

- Système de recueil :

Celui-ci est identique à celui mis au point dans l'expérience 3.

- Analyse des résultats :

Les résultats PEASP furent analysés visuellement ainsi que par l'utilisation de la méthode des ondelettes (Bertrand et al, 1994).

Pour obtenir de l'information à la fois temporelle et fréquentielle sur un signal, il faut le décomposer sur des fonctions concentrées à la fois en temps et en fréquence. Pour définir ces fonctions, il faut partir d'une fonction appelée "ondelette mère" si elle est bien localisée et si elle est oscillante (elle ressemble à une onde, mais étant localisée, c'est une ondelette). La condition de localisation s'écrit sous la forme habituelle de décroissance en amplitude rapide quand la fréquence augmente. La condition d'oscillation suggère qu'elle vibre comme une onde : on demande donc que l'intégrale sur le temps soit nulle. La mère des ondelettes engendre les autres ondelettes de la famille par dilatation et par translation dans le temps, l'énergie doit être constante sur chaque ondelette. Les coefficients de l'ondelette mesurent, en un certains sens, les fluctuations du signal autour du point à l'échelle fournie. Quelle que soit l'échelle et quel que soit l'endroit, l'analyse du signal se fait avec la même fonction.

La transformée de Fourier est un cas particulier d'ondelette, l'ondelette est une sinusoïde (amplitude constante sur la durée du signal).

Les données ont ensuite été analysées à l'aide de tests statistiques de type : PCA (analyse en composantes principales) et LDA (analyse en composantes linéaires discriminantes).

5.2. Résultats :

5.2.1. Test comportemental :

Chaque stimulus extrême fut catégorisé correctement (soit un /ba/, soit un /pa/), avec une frontière perceptive moyenne située entre les stimuli 16 et 17. Cela permit à la fois la validation des sujets inclus, et celle des stimuli utilisés pour les tests neurophysiologiques (Fig 46).

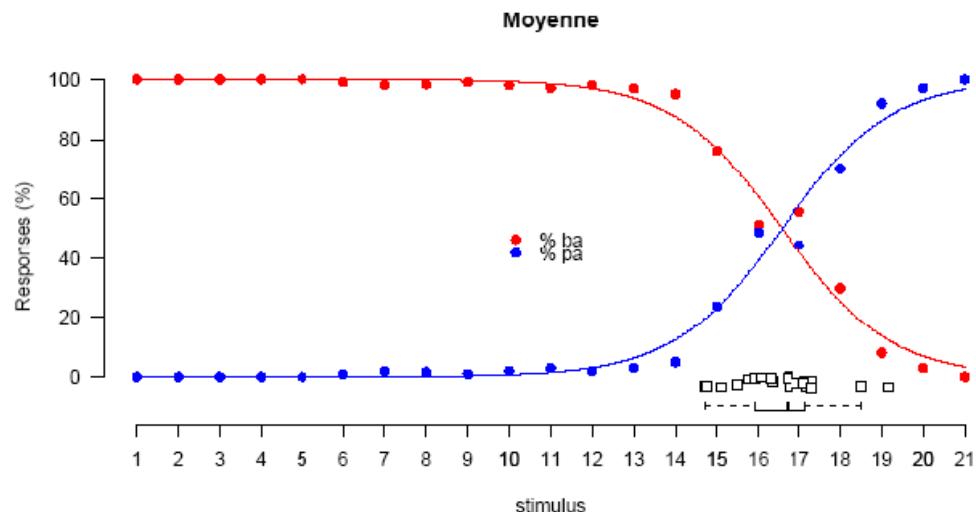


Figure 46 : Test de perception catégorielle.

Les points représentent les scores de reconnaissance moyen parmi les sujets. Les courbes représentent les ajustements des scores moyens par des fonctions logistiques. Les couleurs rouge et bleue correspondent aux réponses /ba/ et /pa/, respectivement. Les résultats individuels ont été également ajustés avec la fonction logistique. Les seuils de reconnaissance à 50% pour chaque sujet sont représentés par les carrés. Une petite variation dans la position verticale des carrés a été ajoutée pour améliorer la visualisation du graphique. Les quartiles de la distribution de ces seuils sont représentés dans le boxplot au dessous des carrés.

5.2.2. Test neurophysiologiques : les PEASP :

5.2.2.1. Représentation temporelle

La représentation temporelle des grandes moyennes PEASP en réponse aux différents stimuli (Fig 47), met en évidence les deux parties du PEASP : l'onset et la RSF, excepté pour le ba21, où l'onset est difficilement individualisable. L'amplitude de l'onset diminue avec la variation du VOT de ba1 à ba 21 (Fig 47).

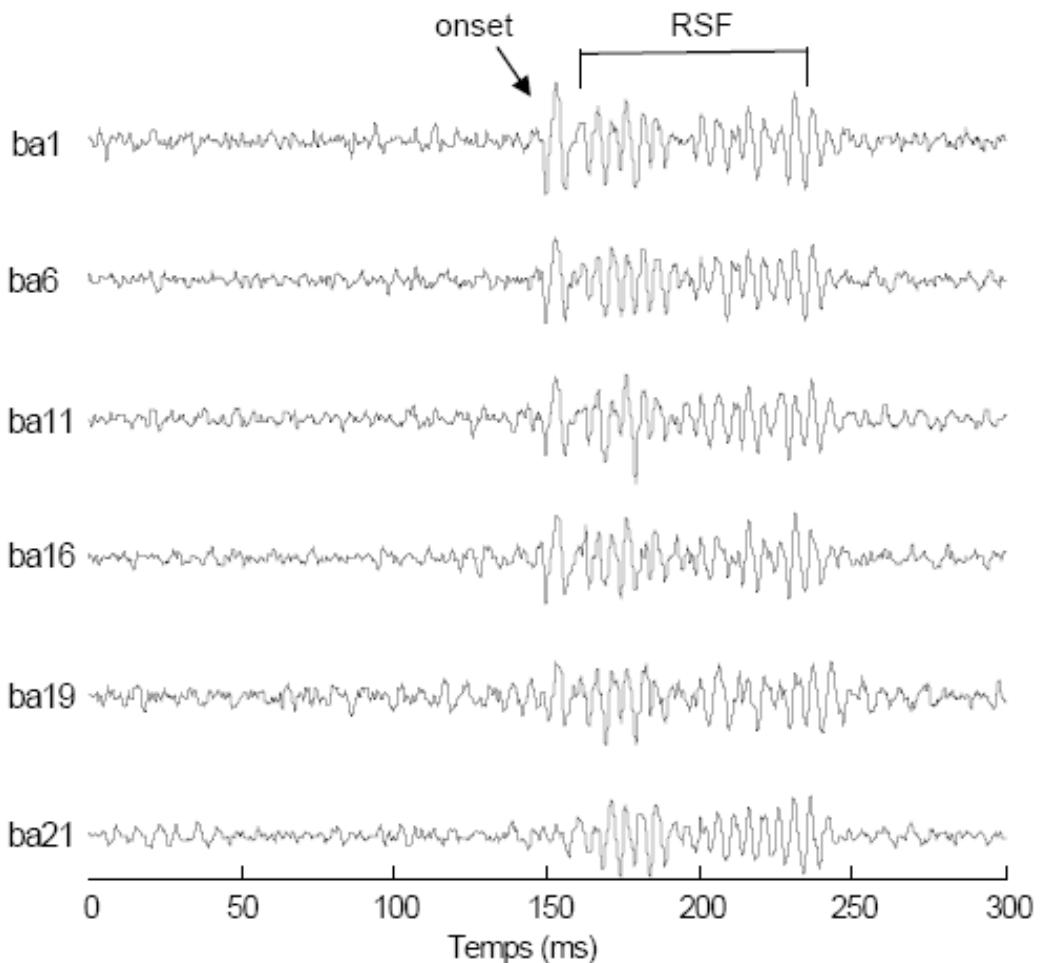


Figure 47 : Représentation temporelle des traces PEASP en réponse aux différents stimuli.

5.2.2.2. Représentation en ondelettes

La décomposition discrète en ondelettes (Bertrand et coll. 1994) permet la représentation du signal PEASP dans le domaine temps/fréquence avec un nombre de coefficients égal au nombre d'échantillons présents dans le signal à analyser. Dans les cas de cette expérience, nous avons utilisé 2048 échantillons sur la durée de signal. Chaque trace PEASP peut donc être représentée par un vecteur de dimension 2048 dans l'espace des coefficients temps/fréquence.

Le but de notre analyse sera de trouver dans cette espace des coefficients temps/fréquence les corrélats neurophysiologiques du traitement du VOT. À l'aide d'une analyse en composantes linéaires discriminantes, nous allons trouver la direction dans l'espace des coefficients temps/fréquence qui permet une discrimination maximale des vecteurs représentant les réponses pour ba1 et ba21. Cependant, l'espace de représentation a une dimension beaucoup trop élevée par rapport à la taille de nos données. Il est donc

nécessaire de procéder à une réduction de dimensionnalité de l'espace avant de passer à la LDA.

Cette réduction de dimensionnalité se réalise en deux temps. Tout d'abord une *réponse d'intérêt* est définie dans l'espace des coefficients correspondant aux coefficients temps/fréquence les plus énergétiques des traces PEASP. La Figure 48 illustre la région d'intérêt obtenue pour une trace spécifique, qui se situe dans la partie FFR de la réponse évoquée et dans les bandes de fréquence de 80 à 500 Hz, puis sur l'ensemble des traces (Fig 49). Cette région d'intérêt permet en outre la définition d'un filtre temps/fréquence qui permet une représentation lissé de la trace PEASP (Fig. 50). Les régions d'intérêt trouvées pour chaque stimulus, chaque côté et chaque sujet seront utilisées pour calculer la région d'intérêt globale qui sera utilisée dans la suite de analyses. Pour ce faire, l'histogramme de contage des coefficients présents dans les régions d'intérêt individuelles est calculé. Les coefficients qui apparaissent au moment dans 50% sont retenus.

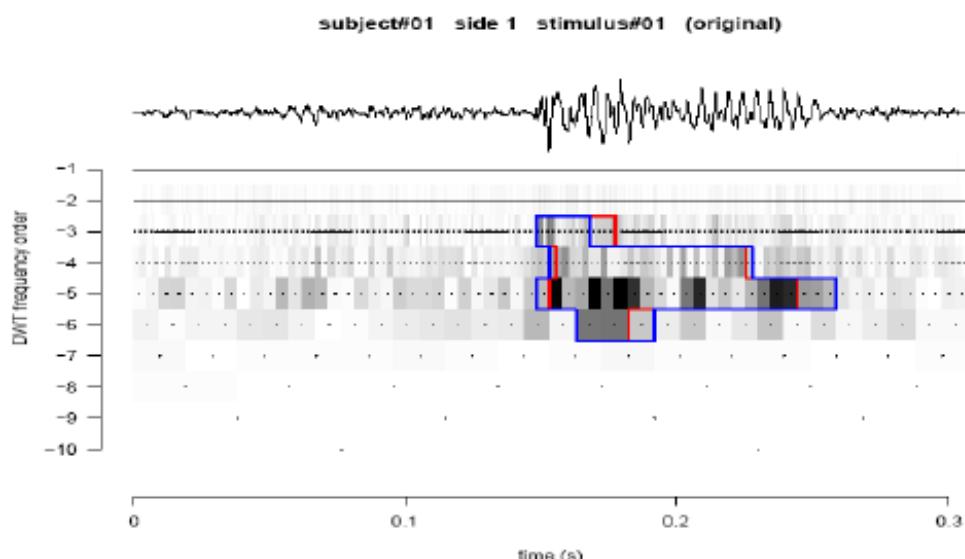


Figure 48 : Représentation de la décomposition du signal PEASP.

Calcul de l'énergie la plus proche du signal. abscisse : temps, ordonnées : bande de fréquence, en z : intensité de la réponse correspondant à une bande de fréquence donnée (des basses fréquences en bas vers les hautes fréquences en haut).

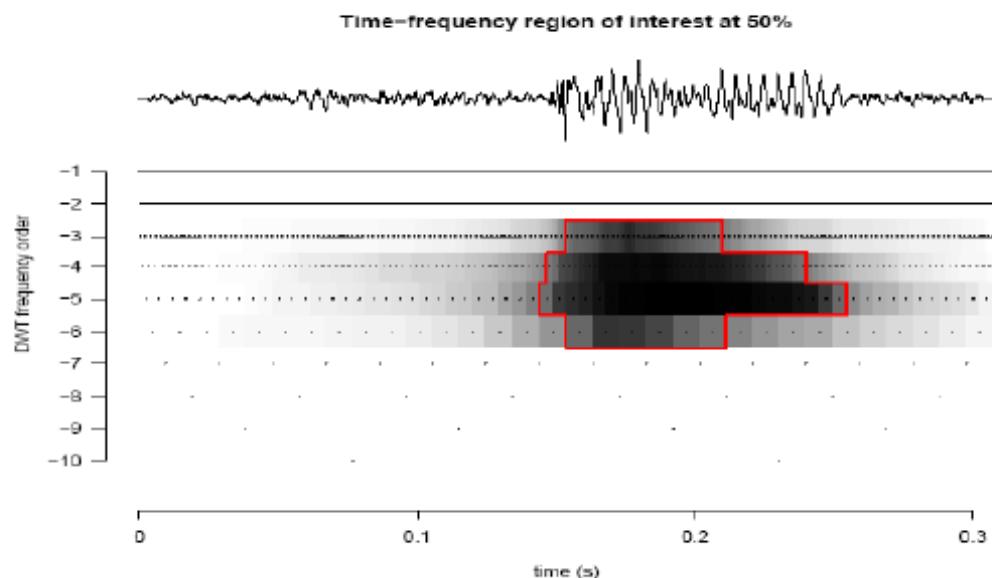


Figure 49 : Représentation de la décomposition du signal PEASP.

Région d'intérêt à 50% calculé sur la totalité des stimuli, délimitée par le tracé rouge. Les niveaux de gris représentant le nombre de fois que chaque coefficient temps/fréquence est apparue dans les régions d'intérêt individuelles, plus c'est foncé, plus le coefficient correspondant était présent dans les traces. Une trace PEASP représentative est montré en haut.

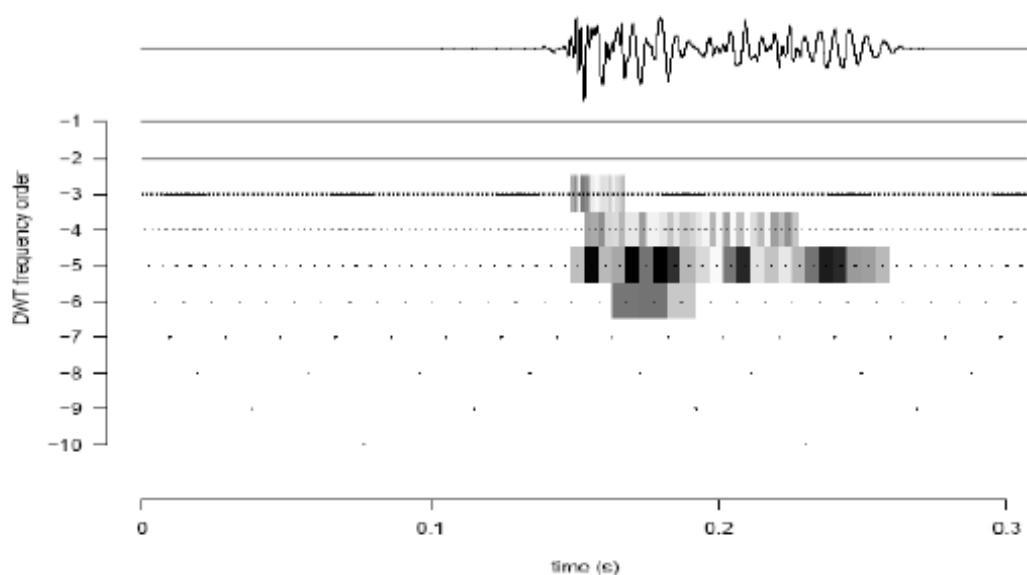


Figure 50 : Représentation de la trace PEASP filtré par la région d'intérêt.
La trace PEASP est visible en haut du schéma, tandis qu'en dessous sont représentées les énergies des différents coefficients temps/fréquence.

La trace PEASP est visible en haut du schéma, tandis qu'en dessous sont représentées les différentes énergies.

L'utilisation de l'analyse statistique en composante principale (PCA) nous a permis de mettre en évidence une différence statistiquement significative entre les résultats PEASP obtenus en réponse au ba 1 versus ceux obtenus en réponse au ba 21. Il faut enfin analyser le composant principal par rapport à la variance de tous les signaux, cela va nous donner une répartition de l'ensemble des vecteurs correspondant chacun à une trace PEASP, ce qui permet de choisir un composant principal (ici à 80%), vecteur de séparation, contenant la majorité des informations des PEASP (Fig 51). Ainsi, en utilisant le composant principal, et en appliquant ce vecteur aux traces PEASP en réponse à ba 1 et celles obtenues en réponse à ba 21 (pa) ont montre que celles-ci sont bien statistiquement différentes (Fig 52).

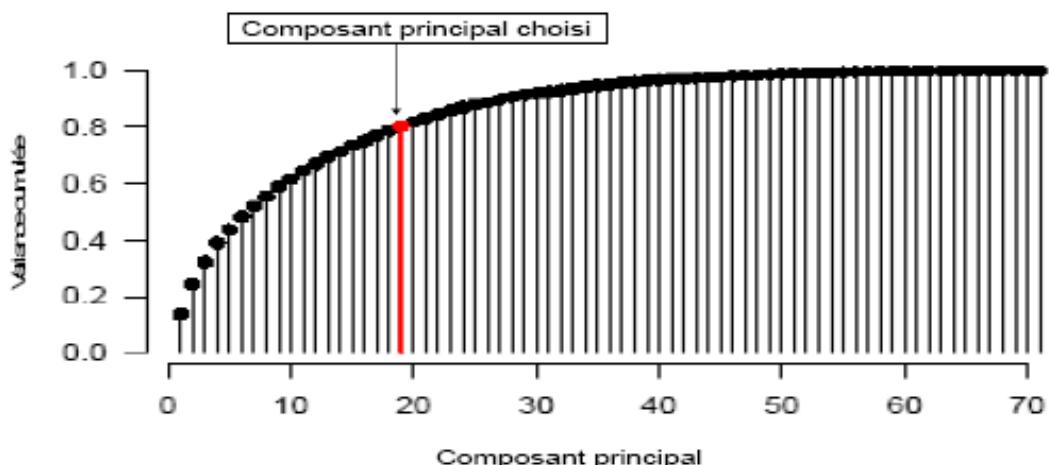


Figure 51 : Représentation du composant principal.

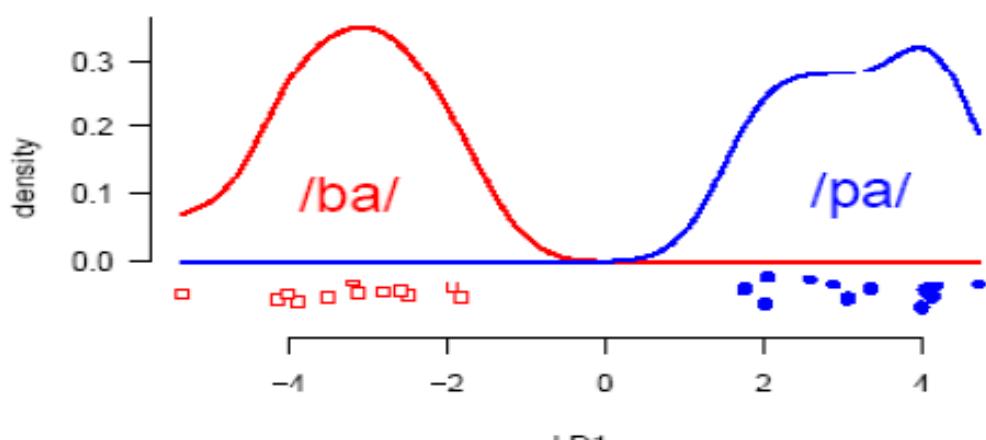


Figure 52 : Représentation statistiques des réponses PEASP au ba 1 et ba 21.

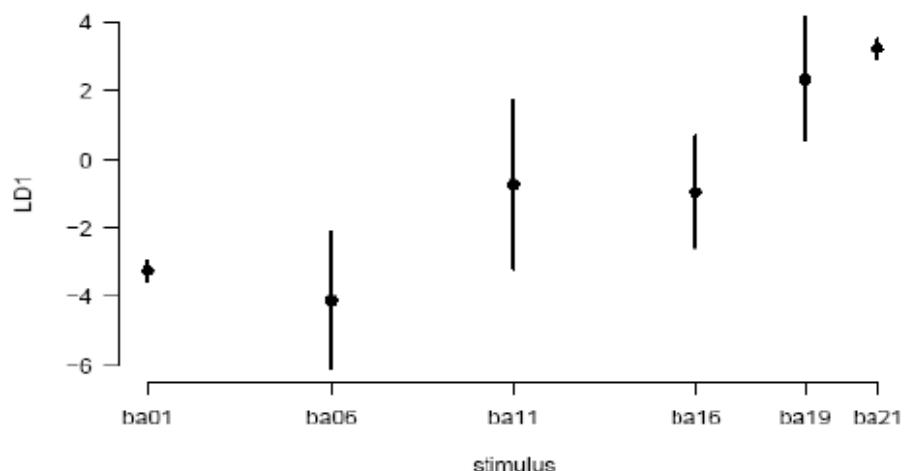


Figure 53 : Représentation statistiques des réponses PEASP aux différents stimuli (côté gauche).

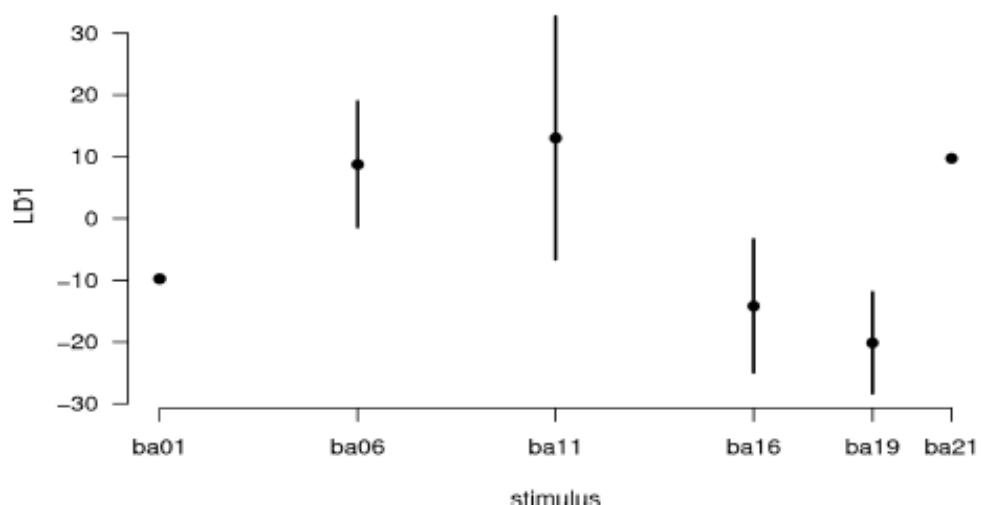


Figure 54 : Représentation statistiques des réponses PEASP aux différents stimuli (côté droit).

De même, l'analyse statistique des traces PEASP recueillies, mettent en évidence une différence entre ba1 et ba 21, cela de manière statistiquement significative (Fig 53 et 54).

Autre élément, mis en avant, lorsque l'on recherche le vecteur discriminant entre la trace PEASP obtenue à partir de ba 1 et celle obtenue à partir de ba 21 (Fig55), on trouve un vecteur dont l'enveloppe possède une fréquence de 40Hz. Laissant supposer l'existence d'un effet corticofuge.

L'ensemble de ces résultats est en faveur d'un encodage du voisement déjà, en partie au niveau sous cortical un encodage du voisement, même si son traitement sera complété

au niveau cortical. De plus, il existerait une boucle corticofuge intervenant lors de l'encodage du voisement.

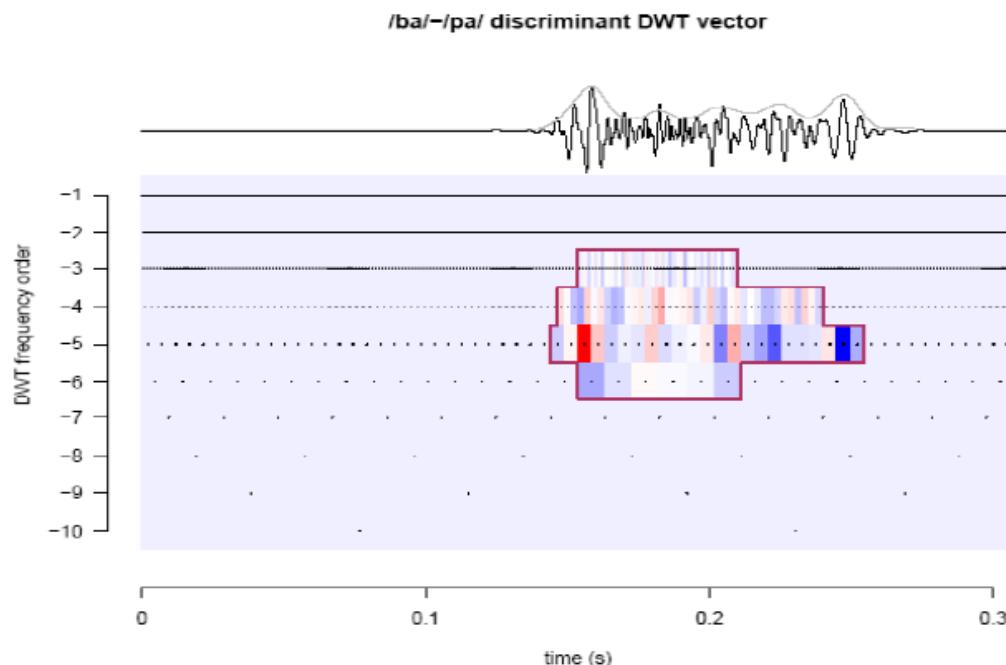


Figure 55 : Le vecteur discriminant entre ba1 et ba 21 est recréé.

couleur : rouge amplitude de l'ondelette positive, bleu amplitude de l'ondelette négative

La méthode décrite ci-dessus est ensuite appliquée à chaque stimulus.

Au niveau du vecteur ainsi créé, on obtient une périodicité de 40 Hz au niveau de la RSF, ce qui sous-entendrait une rétroaction cortico sous-corticale.

5.3. Discussion :

Des travaux portant sur le modèle animal ont démontré une sensibilité des neurones du nerf auditif et du noyau cochléaire aux propriétés langagières du signal, comme la structure formantique (Delgutte 1980), la transition formantique (Delgutte et Kiang 1984), et le VOT (Clarey et al 2004). Il a notamment été montré des schémas de réponse temporelle spécifiques au sein du cortex auditif primaire du singe, reflétant les limites perceptives du VOT et pouvant représenter un corrélat physiologique de la perception catégorielle des paramètres phonétiques (Steinschneider et al, 1995, 2005).

En effet, les schémas de réponses reflétant le VOT sont présents depuis la périphérie et sont transmis tout le long du système auditif jusqu'au cortex, par exemple, une accentuation du phase locking au début du voisement serait notamment visible au niveau du nerf auditif (Sinex & Mc Donald, 1989 ; Sinex et al, 1991). Par ailleurs, Sinex et al (Sinex & McDonald 1988 ; Sinex et al 1991) montrèrent une modification de comportement neuronal touchant principalement les neurones basses fréquences inférieure à 1kHz lorsque l'on étudie les réponses aux sons voisés et non voisés. Les neurones basses fréquences ne déchargent plus durant la période non voisée et recommencent à décharger en présence de voisement. Au niveau cortical, Steinschneider et al, 1994, mirent en évidence la transformation des schémas de réponse reflétant le VOT entre le tronc cérébral et les réponses évoquées intracorticale au niveau du cortex auditif primaire, mettant ainsi en avant

un rôle complémentaire du cortex auditif primaire dans le codage temporel du VOT, rôle observé aussi chez l'humain (Kurtzberg, 1989).

Au niveau du colliculus inférieur, des données obtenues sur un modèle animal (chinchilla, Chen et al (1996)) viennent corroborer les résultats obtenus dans notre expérience. En effet, Chen et al concluent que les réponses des neurones au VOT de syllabes synthétiques présentées en monaural droit étaient semblables à celles observées au niveau du nerf auditif, le type de réponse des neurones du colliculus inférieur (réponse soutenue ou transitoire) dépendant du type de VOT.

L'autre point mis en avant dans cette expérience est l'effet corticofuge, qui concorde avec certaines données précédentes mettant en avant son existence. En effet, il est admis que 30% des enfants présentant des troubles de l'apprentissage ont des résultats PEASP différents (augmentation de la latence de l'onset, pente V-A moins profonde et diminution de l'amplitude de la RSF) des autres enfants du même âge (Cunningham et al, 2001 ; King et al, 2002 ; Wible et al, 2004 ; Banai et al, 2005), résultats pouvant être modifié par l'entraînement : témoin d'un effet corticofuge, améliorant le traitement au niveau sous cortical. En effet, cet encodage neural déficient chez certains de ces enfants serait lié aux processus corticaux, mettant en exergue l'importance du tronc cérébral et surtout l'ensemble des connexions cortico sous corticales sous jacentes (Abrams & Kraus, 2009). De plus, afin de lier précisément le codage des indices acoustiques par le tronc cérébral aux fonctions linguistiques, il a été montré qu'un encodage anormal au niveau du tronc cérébral est lié à des réponses anormales mesurées à l'étage cortical. Par ailleurs, certains enfants présentant une représentation imprécise des sons à l'étage du tronc cérébral ont aussi démontré des représentations anormales des stimuli simples (Ahissar et al, 2000) et complexes (Kraus et al, 1996 ; Bradlow et al, 1999 ; Wible et al, 2002 ; Banai et al, 2005) à l'étage cortical. Ainsi, l'encodage d'indices acoustique au niveau du tronc cérébral représente un phénomène complexe relié aux capacités linguistiques et aux fonctions corticales.

Un autre argument en faveur d'un effet corticofugal est l'existence de connexions cortico sous corticale (Kral & Eggermont, 2007), en effet, ces connexions efférentes entre les couches du cortex auditif et le tronc cérébral sont le support d'un contrôle excitateur ou inhibiteur sur le colliculus inférieur (Keuroghlian & Knudsen, 2007). Pour certains, la mise en évidence de latence de réponse augmentées au niveau de la partie dorsale et périphérique du colliculus inférieur chez le cochon d'inde (Liu et al, 2006), suggèreraient que les neurones situés dans ces régions reflèteraient l'influence du cortex auditif sur le tronc cérébral. Notion corroborée par une diminution de la protéine Fos dans ces régions dorsale et latérale du colliculus inférieur lorsque le cortex est inhibé, sans qu'il y ait de variation du taux de cette protéine dans la région centrale du colliculus inférieur (Sun et al, 2007).

Expérience 6 : Protocole expérimental

Influence de l'intensité de stimulation sur le recueil des réponses de type PEASP

Problématique :

Les premiers résultats obtenus dans l'expérience 4 n'étant pas comparables entre ba+enveloppe et les autres conditions iso-énergétiques, l'expérience a été complétée par une étude paramétrique plus fine comportant des PEAP et des PEASP recueillis à différentes intensités.

Connaitre le seuil de détection d'une réponse de type speech ABR chez le normo-entendant avant de pouvoir disposer d'une application clinique chez les patients malentendants.

Matériels et méthodes :

•Population étudiée : 10 sujets droitiers (test de latéralité manuelle d'Edinburgh).

•Test comportementaux :

Identification du seuil auditif des sujets pour le stimulus : procédure de type « up & down », par pas de 2 dB.

•Tests neurophysiologiques :

♦PEAP : réalisés par pas de 10 dB de 0dBHL à 60 dB HL au niveau de l'oreille droite.

♦PEASP :

Le stimulus utilisé était le même /ba/ que dans l'expérience 4, toutefois les intensités de stimulation furent de 0, 10, 20, 30, 40, 46 et 50 dB SL, le stimulus étant envoyé de manière monaurale droite, fréquence de stimulation de 5 Hz.

Le système de recueil est identique à celui de l'expérience 4, fréquence d'échantillonnage de 13.3 kHz avec une fenêtre d'acquisition de 150 ms.

Résultats :

L'évolution des latences en fonction de l'intensité de stimulation est plus importante pour les réponses impulsionales que pour l'onde V des PEAP. La RSF montre un comportement non linéaire avec l'intensité de stimulation, avec un seuil de recueil de la réponse PEASP de 20 dB SL, et un seuil de plus de 40 dB SL pour avoir une réponse complète. Aux intensités de stimulation modérées (20-30 dB SL), des réponses bilatérales ont été obtenues, et s'unilatéralisent du côté stimulé pour les intensités plus importantes (40-50 dB SL).

Conclusion :

Les PEASP montrent d'importantes différences, sur la réponse impulsionale, sur l'encodage de F0 et sur la RSF, en fonction de l'intensité de stimulation, avec deux seuils de réponses : un seuil d'obtention de 20 dB SL, et un seuil de réponse « complète » de 40 dB SL.

Expérience 6 : Influence de l'intensité de stimulation sur le recueil des réponses de type PEAPSP

Lorsque l'on s'interroge sur la définition même de l'intensité d'un son, on se rend compte que celle-ci apparaît être variable selon les circonstances. On peut la décrire comme la sensation auditive correspondant à la mesure physique de l'intensité sonore, bien que cette définition ne soit pas valable en toutes circonstances. Une définition plus appauvrie en serait la description psychologique de l'amplitude de la sensation auditive (Fletcher & Munson 1933). La complexité de cette définition, s'accompagne d'une complexité dans le traitement de l'information tout le long du système auditif. En effet, même si cette capacité à identifier les différences d'intensité des sons, capacité dite de discrimination d'intensité, nous paraît

tout à fait simple et naturelle, celle-ci fait intervenir différentes structures du système auditif aussi bien au niveau cortical que sous cortical.

L'étude de l'encodage de l'intensité sous-entend de s'intéresser à la manière dont l'intensité physique d'un son est représentée en termes de patterns d'activité des fibres nerveuses au sein du système auditif. Le codage de l'intensité est donc, de manière inhérente, corrélé à la discrimination de cette même intensité. Lors de l'écoute de deux sons d'intensité différente, notre habileté à différencier ces deux sons vient de notre capacité à représenter cette différence à tous les étages du système auditif, de la cochlée jusqu'à l'établissement du processus de décision. La fidélité des mécanismes d'encodage va déterminer la plus petite différence détectable. Un des modèles permettant de mettre en évidence le codage de cette intensité est basée sur le fait qu'il faut prendre en compte la sélectivité fréquentielle de la cochlée, si bien que l'intensité est encodée de manière indépendante pour différents canaux de fréquence, même si cette information est regroupée à des étages plus haut situés. L'hypothèse simpliste selon laquelle l'intensité de n'importe quelle région fréquentielle est simplement codée par le taux de décharge des fibres correspondant à cette région fréquentielle, plus l'intensité est élevée, plus le taux de décharge est élevé doit être nuancée.

Carlyon and Moore (1984) ont suggéré que dans certaines circonstances, l'intensité peut être codée par une activité de phase locking des fibres du nerf auditif. Une augmentation de l'intensité d'un son pur en présence de bruit peut produire une augmentation de la synchronisation sur la structure fine du son pur, s'éloignant de la structure fine du bruit même si le taux global de décharge des fibres ne varie pas : en d'autres mots la fibre nerveuse a atteint son taux de saturation (Javel 1981).

Les premiers résultats n'étant pas comparables entre ba+enveloppe et les autres conditions iso-énergétiques, l'expérience a été complétée par une étude paramétrique plus fine comportant des PEAP et des PEASP recueillis à différentes intensités.

En effet, la grande majorité des études sur les « Speech ABR » utilisent des intensités de stimulation relativement fortes, de 60 à 90 dB SPL, soit 50 à 70 dB au dessus du seuil du sujet. Le problème des intensités relativement élevées, est le risque de contamination des réponses obtenues par l'artefact électromagnétique des transducteurs (cf. chapitre 3 « mise au point technique »). Ceci peut rapidement devenir un problème pour le recueil de réponses de type PEASP chez le patient malentendant: en effet, pour des stimuli d'audibilité identique, le patient malentendant aura besoin d'une intensité supérieure, avec un risque de contamination artéfactuelle rapidement accru. Utiliser des intensités de stimulation plus modérées (par exemple 30 dB au-dessus du seuil du sujet au lieu de 60 dB) amène à poser la question de la caractérisation précise des speech ABR, c'est-à-dire caractériser non seulement les latences des 2 types principaux de réponses (onset réponse et réponse de type FFR), mais également les amplitudes et durées obtenues pour toute une gamme d'intensités. Connaître le seuil de détection d'une réponse de type speech ABR semble indispensable à une ultérieure application clinique, en particulier chez les patients malentendants. En effet, il semble probable que les différents composants des Speech ABR soient affectés différemment par les intensités de stimulation utilisées (Skoe and Kraus, in press). Enfin, il semble important de caractériser la variabilité inter-individuelle des différentes paramètres en fonction de l'intensité de stimulation, élément qu'il nous était impossible de faire avant la mise au point de la technique d'enregistrement, qui nous permet aujourd'hui le recueil de réponses à l'échelon individuel.

La stimulation est monaurale droite, en effet, une étude précédente menée par Hornickel et al 2009, met en évidence la latéralisation du traitement du langage vers

l'hémisphère gauche. En effet, l'envoi d'un /da/ en monaural à l'OD puis à l'OG met en évidence un retard dans les réponses de l'oreille gauche par rapport aux réponses de l'oreille droite avec des réponses spectrales de la RSF plus amples dans l'oreille droite. Par contre, il n'y aurait pas de différence entre l'oreille droite et l'oreille gauche au niveau de l'onde V en réponse à un clic. Ainsi, la préférence de l'oreille droite pour le codage des sons de parole commence dès l'arrivée du message dans le tronc cérébral.

En utilisant le même montage d'électrodes, nous recueillons les réponses auditives de type PEAP, et PEASP à 7 intensités différentes, ce qui nous permettra de comparer l'évolution de l'onde V des PEAP en fonction de l'intensité de stimulation, à celle de la réponse de type « Onset Response » des Speech ABR. En effet, la réponse de type « onset » est classiquement comparée à l'onde V des PEAPs. Si une excellente corrélation est obtenue entre la fonction Entrée/Sortie des PEAP et celle des Speech ABR, le PEAP permettrait de déterminer avec précision à quelle intensité le PEASP ou « speech ABR » pourrait être effectué. En effet, un seuil de PEAP peut-être effectué en une dizaine de minutes, alors qu'une seule trace de PEASP demande entre 10 minutes et 20 minutes. Pour des applications cliniques notamment chez l'enfant, où la rapidité du test est importante, pouvoir choisir l'intensité de stimulation optimum sans avoir à répéter plusieurs traces de PEASP, semble un paramètre important.

6.1. Matériel et méthode :

Sujets inclus : 10 sujets droitiers (test de latéralité d'Edinburgh).

Le seuil auditif des sujets pour le stimulus est obtenu par une procédure de type « up & down », par pas de 2 dB, ceci permettant, par ailleurs, de montrer la variabilité inter-individuelle des seuils de perception du stimulus utilisé chez un groupe de sujets normo-entendant.

Chaîne de stimulation/recueil :

PEAP :

Des PEAP (clics) ont été réalisés aux intensités variant par pas de 10 dB de 0dBHL à 60 dB HL au niveau de l'oreille droite, avec le même montage que celui utilisé pour le recueil des PEASP (Fig29). Un calcul des latences des ondes III et V aux différentes intensités était effectué.

PEASP :

Le stimulus utilisé était le même /ba/ de 100 ms, que dans l'expérience 4, toutefois les intensités de stimulation furent de 0, 10, 20, 30, 40, 46 et 50dBSL, le stimulus étant envoyé de manière randomisée, en monaural droit. La fenêtre d'acquisition était de 150 ms, avec une cadence de répétition de 5Hz, avec un échantillonnage de 13kHz.

6.2. Résultats :

Les mesures des latences des ondes III et V, obtenues sur les PEA en réponse à des clics mettent en évidence une décroissance parallèle et linéaire de leurs latences. Il est à noter que les latences obtenues lors des mesures des PEA en réponse à des clics avec le montage utilisé pour les PEASP retrouvent des latences dont les valeurs sont dans la normalité. L'évolution des latences des ondes III et V sont similaires, mais diffèrent de celle des latences de l'onset (Fig 56). En effet, lorsque l'on regarde la représentation temporelle

du stimulus /ba/ et de la réponse PEASP obtenues à différentes intensités de stimulation (Fig 57), on remarque que pour une voyelle française, l'onset (onde V-A) correspondrait à au passage du /ba/ au /a/. La sinusoïde du signal n'est pas encodée sur la réponse PEASP, l'encodage commence lors de la transition du /b/ au /a/.

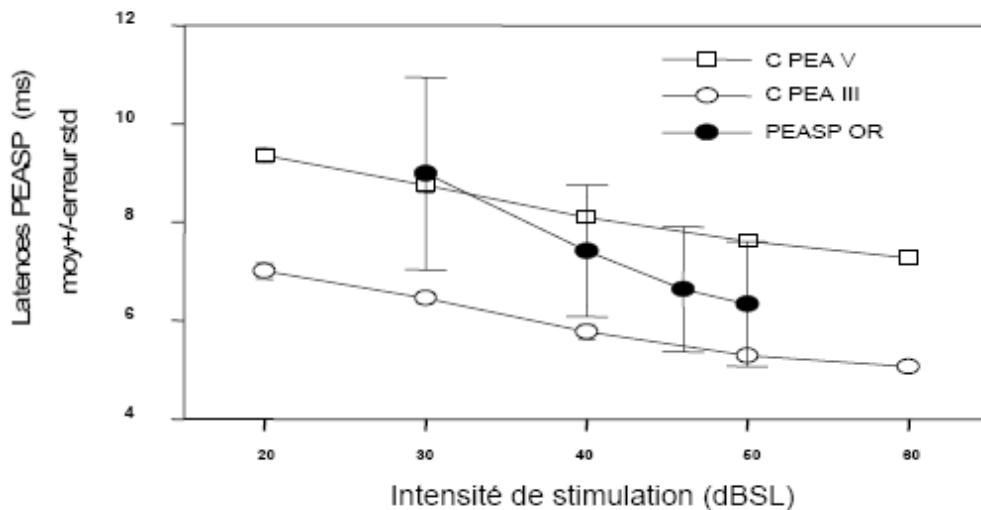


Figure 56 : Potentiels évoqués auditif (PEA, ms) en fonction de l'intensité de stimulation (dBSL).

C PEA V: latences de l'onde V sur les PEA en réponse à des clics C PEA III: latences de l'onde III sur les PEA en réponse à des clics PEASP OR: latence de l'onset sur les PEASP Les latences des ondes III et V diminuent de manière parallèle avec l'augmentation de l'intensité de stimulation. Tandis que la latence de l'onset diminue de manière plus importante que celles des ondes III et V.

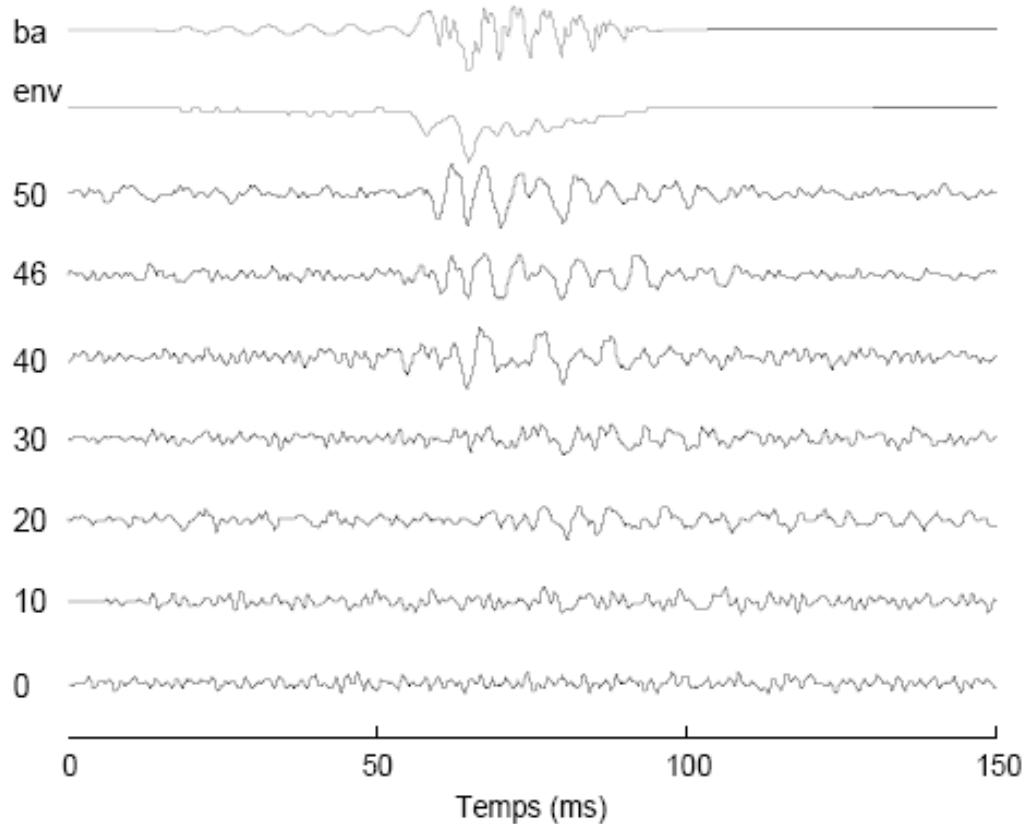


Figure 57 : Représentation temporelle des grandes moyennes des traces PEASP obtenues en réponse au /ba/, à différentes intensités de stimulation.

Pour Kraus et al (Song et al, 2008 ; Johnson et al, 2008) L'amplitude de la FFT (fast fourier transform, représentation fréquentielle du signal) sur la RSF serait le reflet de la synchronisation neurale. De manière visible (Fig 58), les amplitudes de la FFT sur l'ensemble des fréquences de la RSF, augmentent au fur et à mesure de l'augmentation de l'intensité de stimulation, évoluant d'un facteur 10 entre 0 et 50 dBSL.

La RMS, qui donne la valeur efficace du signal, c'est-à-dire l'amplitude moyenne du signal au carré (ou énergie), permet de mettre en évidence des valeurs supérieures obtenues sur la voie 1 (de droite) en comparaison de la voie 2 (Fig 59). Cela étant en accord avec des notions précédentes selon lesquelles l'oreille droite est reliée préférentiellement à l'hémisphère gauche, hémisphère principal dans le traitement du langage.

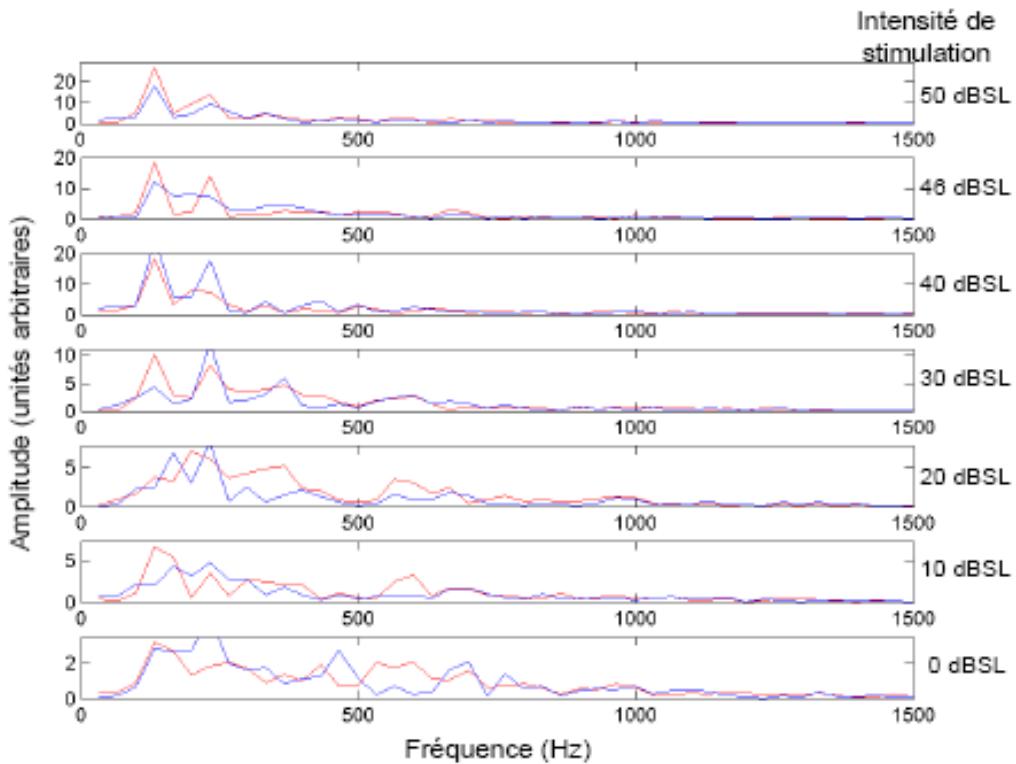


Figure 58 : Représentation fréquentielle des grandes moyennes PEASP en fonction de l'intensité de stimulation.

Voie 1 en rouge ; voie 2 en bleu

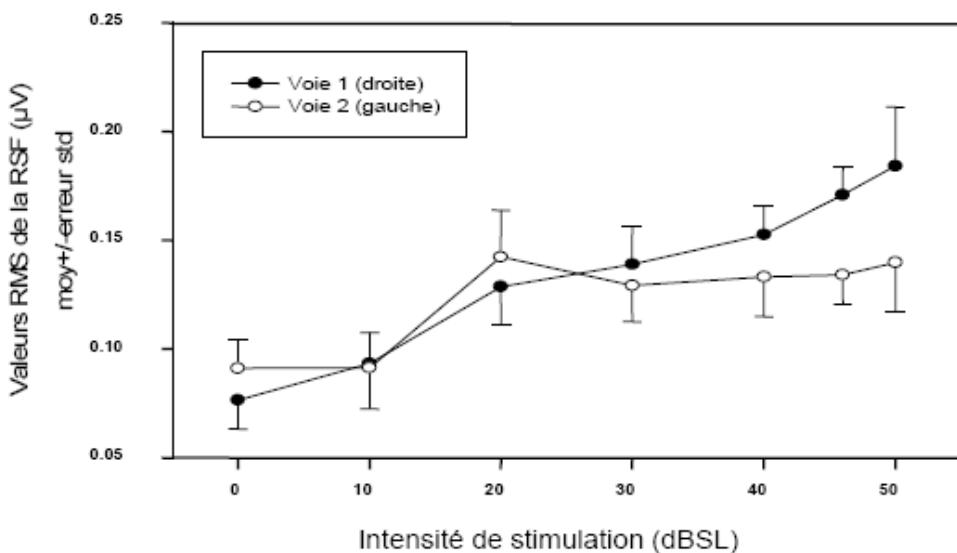


Figure 59: Valeurs RMS (Root Mean Square value, μ V) calculées sur la RSF en fonction de l'intensité de stimulation (dBSL).

L'amplitude des valeurs de la RMS augmente avec l'intensité de stimulation.

6.3. Discussion

Une étude précédente a montré une influence du niveau sonore sur les latences des composants du PEASP. Conformément aux résultats bien connus sur les clics-ABR, ces latences de l'OR et de la RSF diminuent lorsque le niveau sonore augmente. Parallèlement, la forme des PEASP varie très peu pour les intensités variant entre 10 et 50 dB SL, avec une forme assez peu définie alors que le PEASP est quasi-plat au niveau du seuil auditif, et il était très bien défini à 60 dB-SL, avec une OR bien marquée et une RSF très accentuée. De même, la fréquence de coupure du PEASP semble fixe aux alentours de 300 Hz. Cette précédente étude met en évidence une décroissance plus rapide de la latence de la RSF par rapport à celle de l'OR, avec -0.6 ms pour 10 dB supplémentaires, pour l'OR, contre -1.4 pour la RSF. Ces premiers résultats ont amené à supposer l'existence de deux générateurs cellulaires distincts dans le tronc cérébral pour l'OR et pour la RSF (Akhoun et al, 2008).

Dans cet article, la fréquence de coupure était de 500Hz, ce qui diffère des données obtenues par l'équipe de Kraus et al, dans lesquelles la fréquence de coupure des PEASP apparaît avoisiner les 1000Hz. Une des explications avancées est la non utilisation d'inserts, ceci induisant une possible contamination par l'artefact électromagnétique.

L'onset response est bien sensible à l'intensité de stimulation, (Junius & Dau, 2005 ; Akhoun et coll., 2008), les latences de l'onset response sont bien en accord avec un générateur au niveau du colliculus inférieur (Picton et al 1974, Wible et al 2004). Les latences de l'onset observées ainsi que celles de la RSF sont supérieures aux latences habituellement retrouvées dans la littérature (Russo et al 2004, Johnson et al 2005), expliqué en partie par les stimuli différents (langue anglaise versus langue française) et des populations test différentes (adulte versus enfants). Toutefois la maturation du tronc cérébral est achevée dès l'âge de 3 ans avec toutefois des modifications concernant les PEASP tout au long de l'enfance. Des études animales retrouvent des variations dans les durées de la RSF entre différents stimuli synthétiques /ada/ variant entre 5 et 9 millisecondes (Cunningham et al 2002).

L'évolution non linéaire de la décroissance des latences de l'onset et de la RSF sous entend des générateurs distincts, toutefois, la RSF ne peut provenir d'une origine corticale étant donné la fréquence de coupure corticale qui avoisine les 80Hz (Eggermont, 2001). L'onset response serait le résultat direct de la réponse des cellules du colliculus inférieur en analogie avec les PEAP (Picton et al 1974). : supportant l'idée d'un traitement de la RSF dans le même noyau du tronc cérébral que l'onset, mais selon des patterns différents (Wible et al 2004, Purcell et al 2004, Smith et al 1975).

D'après Kraus et al 2005, la source (le what ou l'identité du locuteur) et le filtre (le where) seraient encodés dans le PEASP de manière distincte. La source, liée aux paramètres temporels, à la hauteur tonale, à l'identification du locuteur sont encodés dans la RSF. Le filtre, convoie les patterns spectraux du message, l'information formantique est encodé dans l'onset et l'offset, alors que la RSF et l'onset n'ont pas la même robustesse dans le bruit (Kraus & Nicol 2005 ; Johnson et al 2005). Supposant des mécanismes différents entre l'onset et la RSF.

Les études portant sur des stimuli français retrouvent une fréquence de coupure de la RSF inférieure à 500Hz.

Les neurones du colliculus inférieur collectent des axones provenant de différentes sources au niveau du tronc cérébral, incluant le noyau cochléaire, et le complexe olivaire supérieur, de même que les axones descendants du cortex et du thalamus (Oliver et Huerta

1992). Bien que les propriétés fonctionnelles de ses neurones aient été étudiées de manière extensive (Irvine 1992, Ehret 1997) la connaissance des mécanismes de traitement des stimuli complexes au niveau des colliculi reste encore incomplète. D'après Delgutte et al 1998, le système auditif serait composé d'un filtre passe bande suivi d'un modèle de compression afin d'extraire l'enveloppe de la réponse, suivi d'un second filtre modélisant la modulation de sensibilité des neurones. Ce modèle fait intervenir la fonction de modulation de transfert MTF, résultat de l'étude du gain sur la modulation fréquentielle, le gain étant défini comme la sensibilité des neurones à toute modulation. Cette MTF apparaît être similaire en différents points du système auditif, excepté au niveau du colliculus inférieur, qui apparaît être moins sensible aux hautes fréquences de modulation (au dessus de 100Hz, Joris et al 2004).

Discussion, conclusion et perspectives

La parole est un ensemble d'éléments acoustiques produits à un taux moyen de trois à six syllabes par seconde (Laver, 1994). Notre capacité à décoder ces éléments de manière compréhensible impliquent des traitements à de multiples niveaux du système auditif. Les modèles s'intéressant aux bases neurales sous jacentes à la perception de la parole se sont tout d'abord intéressées au cortex auditif (Bennett & Hacker, 2006; Hickok & Poeppel, 2007; Poeppel & Hickok, 2004; Poeppel, Idsardi, & vanWassenhove, 2008; Scott & Johnsrude, 2003; Scott & Wise, 2004). Toutefois, avant que la parole puisse être perçue et intégrée, stockées de manière stable au travers de ses représentations linguistiques, les indices acoustiques doivent être représentés à travers un code neural tout le long du système auditif et délivré au cortex auditif avec une précision temporelle et spectrale parfaite par les structures sous corticales (Eggermont, 2001; Hickok & Poeppel, 2007; Poeppel & Hickok, 2004; Poeppel et al, 2008).

Un certain nombre d'équipes se sont intéressées à l'encodage des sons de parole au niveau du nerf acoustique (Delgutte, 1980 ; Sachs & Young, 1980 ; Miller & Sachs, 1983), au niveau du noyau cochléaire (Caspary et al, 1977 ; Keilson et al 1997), et au niveau du tronc cérébral (Galbraith et al, 1995, 1997 ; Krishnan, 1999, 2002). Toutefois, la réalisation de ce travail de thèse a été motivée par la question de l'encodage des sons de parole au niveau du tronc cérébral et notamment dans le bruit. En effet, la parole est un signal complexe composé d'un grand nombre de traits acoustiques, tous nécessaires à la compréhension normale de celle-ci. La perception « normale » de ces traits acoustiques reposant sur leur encodage neural, fil conducteur de cette thèse.

Ainsi, nous avons essayé de mieux comprendre le traitement sous cortical de la parole dégradée, notamment par la caractérisation de l'importance de certains traits acoustiques dans la perception de la parole normale. Nous nous sommes appuyés sur des mesures neurophysiologiques du système auditif, représentées par les potentiels évoqués auditifs en réponse à un son de parole (PEASP), système non invasif de mesure de l'activité électrophysiologique du tronc cérébral, ceci dans l'objectif de pouvoir décrire les mécanismes impliqués, permettant de reconstruire de la parole dégradée et de comprendre le message parlé.

D'une part nous nous sommes intéressés aux conditions de recueil nécessaires à l'obtention de traces PEASP optimales (chaîne d'acquisition, intensité de stimulation optimale) et à leur analyse automatique, d'autre part, nous avons étudié l'effet de certaines

caractéristiques d'un son, à savoir, l'enveloppe temporelle, et le voisement sur les PEASP. Cette nécessité de mise au point fut d'ailleurs l'une des limitations de cette thèse, en effet, la nécessité d'optimiser le système d'acquisition des PEASP, fut une étape incontournable ayant demandé de nombreuses heures et sujets supplémentaires.

L'ensemble de ce travail a permis de mettre en évidence l'importance du rôle de certains indices acoustiques contenus dans le message parlé, pour la perception correcte de celui-ci. Ainsi, une synchronisation de la RSF des PEASP en réponse aux stimuli langagiers fut mis en évidence sur la fréquence fondamentale, synchronisation conditionnée par la structure fine ; synchronisation qui ne fut pas retrouvée sur les PEASP en réponse aux analogues.

Initialement, nous avions adopté une attitude se focalisant sur l'encodage de la parole par les voies ascendantes, nous avions souvent tendance à considérer l'encodage neural comme un mécanisme afférent. Toutefois, on ne peut concevoir l'étude de l'encodage de la parole sans l'étude conjointe de cet encodage au niveau cortical et sous cortical. En effet, lors de l'étude de l'encodage du voisement, indice dont le traitement a précédemment été mis en évidence au niveau cortical, et qui se révèle être en partie encodé dès le tronc cérébral, nous avons observé l'existence d'une interaction cortico sous corticale. L'existence de boucles de modulation cortico-sous-corticales était déjà connue (Xia & Suga, 2002 ; Perrot et al, 2006) et avait notamment été mise en évidence, par l'équipe de Nina Kraus lors de travaux menés chez des enfants présentant des troubles de l'apprentissage. En effet, l'entrainement linguistique effectué par ces enfants, permit l'initiation d'un « feedback » cortical, mis en évidence sur le traitement de certains indices acoustiques tels que la hauteur tonale, cette plasticité étant présente jusqu'à l'âge adulte (Song et al, 2008 ; Russo et al, 2005). Par ailleurs, il serait possible que l'activité de la RSF reflète de manière concomitante l'activité sous-corticale, l'activité corticale modulatrice efférente et l'activité corticale évoquées par un son de parole (Musacchia et al, 2008). De récentes études ont ainsi étendu la théorie hiérarchique inverse (Ahissar & Hochstein, 2004) mettant en avant l'action bénéfique de l'effet corticofugal (« top down »), initiant ou modulant la plasticité de la représentation de certaines caractéristiques acoustiques des sons au niveau du tronc cérébral (Suga, 2008 ; Luo et al, 2008 ; Tzounopoulos & Kraus, 2009). Cette plasticité cérébrale, favorisée par la pratique de la musique de manière professionnelle, pourrait être l'un des éléments à l'origine de la limitation des effets négatifs du bruit sur l'encodage neural des sons de parole (Parbery-Clark et al, 2009 ; Kraus et al, 2009 ; Strait et al, in press).

Pour l'équipe américaine de Nina Kraus, l'encodage des informations linguistiques constituerait un vrai marqueur biologique de la fonctionnalité du système auditif, notamment chez les enfants présentant des troubles de l'apprentissage. Ainsi, cette équipe développa un système d'exploitation clinique (BioMap), permettant l'objectivation des anomalies de traitement de la parole chez des enfants d'âge scolaire (8 à 12 ans) (Johnson et al, 2008 ; Song et al, 2008). Toutefois, l'ensemble des résultats présentés par cette équipe porte sur des groupes de sujets et non au niveau individuel. L'ensemble des résultats obtenus dans cette thèse, grâce à l'adaptation du système d'acquisition aux syllabes françaises, va d'ailleurs dans ce sens. En effet, même si une optimisation du recueil des traces a pu être obtenue, les résultats ne sont pas encore analysables à l'échelon individuel pour chaque sujet.

En effet, de récentes études s'intéressant aux potentiels évoqués auditifs en réponse à un son de parole (PEASP) ont révélé une caractéristique précieuse du tronc cérébral : celui-ci est capable de représenter avec une très grande fidélité les éléments acoustiques de base de la parole (pour revue : Kraus & Nicol, 2005). Ces PEASP nous offrent donc une méthode non invasive de tester l'intégrité du fonctionnement des structures sous corticale

dans le traitement de stimuli complexes tels que la parole (Galbraith, Jhaveri, & Kuo, 1997; Glaser, Suter, Dasheiff, & Goldberg, 1976; Moushegian, Rupert, & Stillman, 1973; Russo, Nicol, Musacchia, & Kraus, 2004). Les PEASP offre un pointe de vue unique dans la compréhension de la représentation des éléments clés du signal acoustique, par le tronc cérébral.

Ainsi, les processus auditifs sous corticaux sont eux aussi dynamiques, et non figés comme nous l'avions longtemps pensé. Leur rôle dans la perception de la parole, sont encore loin d'être totalement élucidés, mais les données obtenues au cours de ce travail suggèrent de riches interactions sous corticale et corticale.

Cela permet d'envisager de nombreuses perspectives, telles que l'extension de l'étude de l'encodage dans le bruit à des sujets présentant des troubles de la perception, l'étude conjointe du système efférent auditif couplé à l'analyse des PEASP, ou bien encore l'enregistrement concomitant des PEASP et de l'activité corticale par électroencéphalographie.

V. Références

- Abrams D, Kraus N. (2009) [Auditory pathway representation of speech sounds in humans](#). In: *Handbook of Clinical Audiology*, Katz J, Hood L, Burkard R, Medwetsky L. (eds.) 611-626.
- Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Crosslanguage tests. Paper presented at the Sixth International Congress of Phonetic Sciences, Prague: Academia.
- Ahissar M, Hochstein S, (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8:457-464.
- Ahissar M, Protopapas A, Reid M, Merzenich MM, (2000). Auditory processing parallels reading abilities in adults. *Proc Natl Acad Sci U S A* 97:6832-6837.
- Aiken SJ, Picton TW, (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245:35-47.
- Aiken SJ, Picton TW, (2006). Envelope following responses to natural vowels. *Audiol Neurotology* 11:213-232.
- Akhoun I, Gallego S, Moulin Aet al, (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 119:922-933.
- Akhoun I, Moulin A, Jeanvoine Aet al, (2008). Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. *J Neurosci Methods* 175:196-205.
- Anderson S, Skoe E, Chandrasekaran B, Kraus N (in press). Neural timing is linked to speech perception in noise. *J Neurosci* 30:4922-4926.
- Bacon, S.P., and Gleitman, R.M. (1992). Modulation detection in subjects with relatively flat hearing losses. *J. Speech Hear. Res.* 35, 642-653.
- Banai K, Abrams D, Kraus N, (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *Int J Audiol* 46:524-532.
- Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N, (2009). Reading and subcortical auditory function. *Cereb Cortex* 19:2699-2707.
- Banai K, Nicol T, Zecker SG, Kraus N, (2005). Brainstem timing: implications for cortical processing and literacy. *J Neurosci* 25:9850-9857.
- Batra R, Kuwada S, Maher VL, (1986). The frequency-following response to continuous tones in humans. *Hear Res* 21:167-177.
- Bekesy GV, (1949). *Experientia* 5:326.
- Békésy, G. Von (1947). The variation of phase along the basilar membrane with sinusoidal vibrations, *J. Acoust. Soc. Am.* 19, 452-460.

- Belin P, Zilbovicius M, Crozier Set al, (1998). Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 10:536-540.
- Belin, P. and Zatorre, R.J. (2000) 'What', 'where' and 'how' in auditory cortex (commentary). *Nat. Neurosci.* 3, 965–966
- Bennett MR, Hacker PM, (2006). Language and cortical function: conceptual developments. *Prog Neurobiol* 80:20-52.
- Bertrand O, Bohorquez J, Pernier J, (1994). Time-frequency digital filtering based on an invertible wavelet transform: an application to evoked potentials. *IEEE Trans Biomed Eng* 41:77-88.
- Boston JR, Moller AR, (1985). Brainstem auditory-evoked potentials. *Crit Rev Biomed Eng* 13:97-123.
- Bradlow AR, Kraus N, Hayes E, (2003). Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res* 46:80-97.
- Bradlow AR, Kraus N, Nicol TG et al, (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *J Acoust Soc Am* 106:2086-2096.
- Bregman, A.S. (1990). Auditory Scene Analysis: The perceptual Organization of Sound (MIT, Cambridge, MA). *Psychology of Music*, Vol. 19, No. 1 (1991) pp. 77-82.
- Broca, P. (1865). Sur le siège de la faculté du langage articulé. *Bulletins de la Société d'Anthropologie*, 6, 377–393.
- Bronkhorst, Adelbert W. (2000). The Cocktail Party Phenomenon: A Review on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica united with Acustica* 86: 117–128.
- Carlyon RP, Moore BC. (1984). Intensity discrimination: a severe departure from Weber's law. *J Acoust Soc Am* 76:1369-1376.
- Caspary DM, Rupert AL, Moushegian G. (1977). Neuronal coding of vowel sounds in the cochlear nuclei. *Exp Neurol* 54:414-431.
- Chandrasekaran B, Kraus N. (2009). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47:236-246.
- Chen GD, Nuding SC, Narayan S, Sinex DG. (1996). Responses of single neurons in the chinchilla inferior colliculus to consonant-vowel syllables differing in voice onset time. *Audit. Neurosci.* 3, 179-198.
- Cherry, E. Colin (1953-09). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of Acoustic Society of America* 25 (5): 975-979.
- Clarey JC, Paolini AG, Grayden DB, Burkitt AN, Clark GM. (2004). Ventral cochlear nucleus coding of voice onset time in naturally spoken syllables. *Hear Res* 190:37-59.
- Collet, L., (2007). "Phonetic and lexical interferences in informational masking during speech-in-speech comprehension", *Speech Communication*, 49:12, pp. 905-916
- Collet, L., Levy, V., Veuillet, E., Truy, E., & Morgan, A. (1993). Click-evoked otoacoustic emissions and hearing threshold in sensorineural hearing loss. *Ear & Hearing* 14, 141-143.

- Collet, L., Veuillet, E., Moulin, A., Morlet, T., Giraud, A. L., Micheyl, C., et al. (1994). Contralateral auditory stimulation and otoacoustic emissions: A review of basic data in humans. *British Journal of Audiology* 28, 213-218.
- Cunningham J, Nicol T, King C, Zecker SG, Kraus N. (2002). Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res* 169:97-111.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 112:758-767.
- Daly DM, Roeser RJ, Moushegian G. (1976). The frequency-following response in subjects with profound unilateral hearing loss. *Electroencephalogr Clin Neurophysiol* 40:132-142.
- Dau T. (2003). The importance of cochlear processing for the formation of auditory brainstem and frequency following responses. *J Acoust Soc Am* 113:936-950.
- Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hear Res* 229, 132-147.
- Delattre P.. (1955). Les attributs acoustiques de la nasalité vocalique et consonantique. *Studia Linguistica* 8, 103-110.
- Delattre P.. (1951). The physiological interpretation of sound spectrograms, *Publication of the modern language assoc. of Amer*64, 864-876.
- Delgutte B, Kiang NY. (1984). Speech coding in the auditory nerve: I. Vowel-like sounds. *J Acoust Soc Am* 75:866-878.
- Delgutte B, Kiang NY. (1984). Speech coding in the auditory nerve: III. Voiceless fricative consonants. *J Acoust Soc Am* 75:887-896.
- Delgutte B, Kiang NY. (1984). Speech coding in the auditory nerve: IV. Sounds with consonant-like dynamic characteristics. *J Acoust Soc Am* 75:897-907.
- Delgutte B, Kiang NY. (1984). Speech coding in the auditory nerve: V. Vowels in background noise. *J Acoust Soc Am* 75:908-918.
- Delgutte B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J Acoust Soc Am* 68:843-857.
- Delgutte B. (1984). Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J Acoust Soc Am* 75:879-886.
- Delgutte, B., Hammond, B.M. & Cariani, P.A. (1998). Neural coding of the temporal envelope of speech: relation to modulation transfer functions. In *Psychophysical and physiological advances in hearing* (eds A.R. Palmer, A. Rees, A.Q. Summerfield & R. Meddis), pp. 595-603. London, UK: Whurr Publishers.
- Dolan, D. F., & Nuttal, A. L. (1988). Masked cochlear whole-nerve response intensity functions altered by electrical stimulation of the crossed olivocochlear bundle. *J Acoust Soc Am* 83, 1081-1086.
- Dos Santos, C. (2007). *Développement phonologique en français langue maternelle. Une étude de cas*. Université Lumière Lyon 2, Lyon, France.

- Efron R. (1963). Temporal Perception, Aphasia and Déjà Vu. *Brain* 86:403-424.
- Eggermont JJ, Ponton CW. (2002). The neurophysiology of auditory perception: from single units to evoked potentials. *Audiol Neurotol* 7:71-99.
- Eggermont JJ. (2001). Between sound and perception: reviewing the search for a neural code. *Hear Res* 157:1-42.
- Eggermont JJ. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *J Acoust Soc Am* 98:911-920.
- Ehret G, (1997). The auditory midbrain, a "shunting yard" of acoustical information processing. In: Ehret G and Romand R (Eds). The central auditory system. Oxford University Press, New York, pp. 259-316.
- Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.
- Ettlinger G. (1990). "Object vision" and "spatial vision": the neuropsychological evidence for the distinction. *Cortex* 26:319-341.
- Faingold CL, Caspary DM. (1979). Frequency-following responses in primary auditory and reticular formation structures. *Electroencephalogr Clin Neurophysiol* 47:12-20.
- Fant, G. (1960). Acoustic Theory of Speech Production, Mouton
- Fiez JA, Raichle ME, Balota DA, Tallal P, Petersen SE. (1996). PET activation of posterior temporal regions during auditory word presentation and verb generation. *Cereb Cortex* 6:1-10.
- Fletcher, H., & Munson, W.A. (1933). Loudness, its definition, measurement and calculation. *J Acoust Soc Am* 5, 82-108.
- Friederici AD, Alter K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain Lang* 89:267-276.
- Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6:2363-2367.
- Galbraith GC, Bhuta SM, Choate AK, Kitahara JM, Mullen TA, Jr. (1998). Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9:1889-1893.
- Galbraith GC, Threadgill MR, Hemsley Jet al. (2000). Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett* 292:123-127.
- Gardi J, Salamy A, Mendelson T. (1979). Scalp-recorded frequency-following responses in neonates. *Audiology* 18:494-506.
- Gerken GM, Moushegian G, Stillman RD, Rupert AL. (1975). Human frequency-following responses to monaural and binaural stimuli. *Electroencephalogr Clin Neurophysiol* 38:379-386.
- Geschwind N. (1970). The organization of language and the brain. *Science* 170:940-944.
- Giard, M.-H., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain Research*, 633, 353-356.

- Giraud K, Trebuchon-DaFonseca A, Demonet JF, Habib M, Liegeois-Chauvel C. (2008). Asymmetry of voice onset time-processing in adult developmental dyslexics. *Clin Neurophysiol* 119:1652-1663.
- Giraud, A. L., Garnier, S., Micheyl, C., Lina, G., Chays, A., & Chéry-Croze, S. (1997). Auditory efferents involved in speech-in-noise intelligibility. *NeuroReport*, 8, 1779-1783.
- Greenberg S, Ainsworth AW, Fay RR, Popper AN (Eds.), (2004). *New-York* 2, 70, 166.
- Greenberg S, Marsh JT, Brown WS, Smith JC. (1987). Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear Res* 25:91-114.
- Griffiths TD. (2003). Functional imaging of pitch analysis. *Ann N Y Acad Sci* 999:40-49.
- Grill-Spector K, Henson R, Martin A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14-23.
- Grimault N., Micheyl C., Carlyon R.P., Artaud P., Collet L. (2000). Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency . *J. Acoust. Soc. Am.*108, 263-271.
- Guinan Jr., J. J. (1996). The physiology of olivocochlear efferents. In P. J. Dallos, A. N. Popper & R. R. Fay (Eds.), *New York: Springer-Verlag. The Cochlea* (pp. 435-502)
- Habib M. et Robichon F. (1998). Dominance hémisphérique. *Encycl Méd Chir*. Elsevier, Paris; *Neurologie* 17-022-C-10, p.8.
- Hall J. W.. (1979). Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, vol. 205, no. 4412, pages 1297–1299, 35, 88.
- Hauser MD, Fowler CA.(1992). Fundamental frequency declination is not unique to human speech: evidence from nonhuman primates. *J Acoust Soc Am* 91:363-369.
- Hickok G, Poeppel D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92:67-99.
- Hickok G, Poeppel D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci* 8:393-402.
- Hienz, R., Stiles, P., & May, B. (1998). Effects of bilateral olivocochlear lesions on vowel formant discrimination in cats. *Hear Res* 116, 10-20.
- Hoen, Grataloup, Grimault, Perrin, Pellegrino, Meunier, Collet (2005). Who hides behind the masks at the cocktail party? Behavioral assessment of different types of interferences occurring during speech-in-noise comprehension." *Interspeech*.
- Hornickel J, Skoe E, Kraus N. (2008). Subcortical Laterality of Speech Encoding. *Audiol Neurotol* 14:198-207.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA* 4;106(31):13022-7.
- Irvine DRF (1992) Physiology of the auditory brainstem. In: *Mammalian auditory pathway: neurophysiology*, (Popper A, Fay RR, eds), *New York: Springer Verlag*. pp 153–231.

- Javel E. (1981). Suppression of auditory nerve responses I: temporal analysis, intensity effects and suppression contours. *J Acoust Soc Am* 69:1735-1745.
- Jewett DL, Romano MN, Williston JS. (1970). Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 167:1517-1518.
- Johkura K, Matsumoto S, Hasegawa O, Kuroiwa Y. (1998). Defective auditory recognition after small hemorrhage in the inferior colliculi. *J Neurol Sci* 161:91-96.
- John MS, Dimitrijevic A, Picton TW. (2002). Auditory steady-state responses to exponential modulation envelopes. *Ear Hear* 23:106-117.
- Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N. (2008). Brainstem encoding of voiced consonant--vowel stop syllables. *Clin Neurophysiol* 119:2623-2635.
- Johnson KL, Nicol T, Zecker SG, Kraus N. (2008). Developmental plasticity in the human auditory brainstem. *J Neurosci* 28:4000-4007.
- Johnson KL, Nicol TG, Kraus N. (2005). Brain stem response to speech: a biological marker of auditory processing. *Ear Hear* 26:424-434.
- Johnsrude IS, Penhune VB, Zatorre RJ. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123 (Pt 1):155-163.
- Joris PX, Schreiner CE, Rees A. (2004). Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541-577.
- Junius D, Dau T. (2005). Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses. *Hear Res* 205:53-67.
- Kaas, J.H. and Hackett, T.A. (1999) 'What' and 'where' processing in auditory cortex (commentary). *Nat. Neurosci* 2, 1045–1047
- Kawase, T., & Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. I. Enhancement of compound action potentials to masked tones. *Journal of Neurophysiology* 70, 2519-2532.
- Kawase, T., Delgutte, B., & Liberman, M. C. (1993). Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones. *Journal of Neurophysiology*, 70, 2533-2549.
- Keilson SE, Richards VM, Wyman BT, Young ED. (1997). The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J Acoust Soc Am* 102:1056-1071.
- Keuroghlian AS, Knudsen EI. (2007). Adaptive auditory plasticity in developing and adult animals. *Prog Neurobiol* 82:109-121.
- Khalfa, S., Bougeard, R., Morand, N., Veuillet, E., Isnard, J., Guenot, M., et al. (2001). Evidence of peripheral auditory activity modulation by the auditory cortex in humans. *Neuroscience*, 104, 347-358.
- Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., et al. (2001). Peripheral auditory asymmetry in infantile autism *European Journal of Neuroscience* 13, 628-632.
- Kimura RS, Schuknecht HF, Sando I. (1964). Fine Morphology of the Sensory Cells in the Organ of Corti of Man. *Acta Otolaryngol* 58:390-408.

- King C, Warrier CM, Hayes E, Kraus N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett* 319:111-115.
- Kral A, Eggermont JJ. (2000). What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. *Brain Res Rev* 2007;56:259-269.
- Kraus N, Bradlow AR, Cheatham MA et al. Consequences of neural asynchrony: a case of auditory neuropathy. *J Assoc Res Otolaryngol* 1:33-45.
- Kraus N, McGee TJ, Carrell TD, Zecker SG, Nicol TG, Koch DB. (1996). Auditory neurophysiologic responses and discrimination deficits in children with learning problems. *Science* 273:971-973.
- Kraus N, McGee TJ, Koch DB. (1998). Speech sound representation, perception, and plasticity: a neurophysiologic perspective. *Audiol Neurotol* 3:168-182.
- Kraus N, Nicol T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci* 28:176-181.
- Kraus N, Skoe E, Parbery-Clark A, Ashley R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann N Y Acad Sci* 1169:543-557.
- Krishnan A, Parkinson J. (2000). Human frequency-following response: representation of tonal sweeps. *Audiol Neurotol* 5:312-321.
- Krishnan A, Xu Y, Gandour J, Cariani P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25:161-168.
- Krishnan A, Xu Y, Gandour JT, Cariani PA. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 189:1-12.
- Krishnan A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 166:192-201.
- Krizman JL, Skoe E, Kraus N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiol Neurotol* 15:332-342.
- Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. (2007). Hierarchical processing of auditory objects in humans. *PLoS Comput Biol* 3:e100.
- Kumar, U., & Vanaja, C. (2004). Functioning of olivocochlear bundle and speech perception in noise. *Ear and Hearing*, 25, 142-146.
- Kurtzberg D : (1989). Cortical event-related potential assessment of auditory and systems function. *Semin Hear* 10 :252-261.
- Lee KM, Skoe E, Kraus N, Ashley R. (2009). Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 29:5832-5840.
- Liberman A.M., Cooper E.S., Shankweiler D.P. and Studdert-Kennedy M..(1967). Perception of the speech code. *Psychol. Rev.* 74:431-461.
- Liberman MC. (1982). The cochlear frequency map for the cat: labeling auditory-nerve fibers of known characteristic frequency. *J Acoust Soc Am* 72:1441-1449.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.

- Liegeois-Chauvel C, de Graaf JB, Laguitton V, Chauvel P. (1999). Specialization of left auditory cortex for speech perception in man depends on temporal coding. *Cereb Cortex* 9:484-496.
- Liegeois-Chauvel C, Giraud K, Badier JM, Marquis P, Chauvel P. (2001). Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of the human auditory cortex. *Ann N Y Acad Sci* 930:117-132.
- Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P. (2004). Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14:731-740.
- Lisker, L., & Abramson, A. S. (1964). A cross-linguistic study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.
- Liu, L.-F., Palmer, A. R., & Wallace, M. N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *Journal of Neurophysiology* 95, 1926-1935.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B.C.J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci. USA*, 103, 18866-18869.
- Lorenzi, C., Husson, M., Ardoit, M. & Debruille, X. (2006). Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception. *Int. J. Audiol.*, 45, 487-495.
- Luo F, Wang Q, Kashani A, Yan J. (2008). Corticofugal modulation of initial sound processing in the brain. *J Neurosci* 28:11615-11621.
- Maison, S., Durrant, J., Gallineau, C., Micheyl, C., & Collet, L. (2001). Delay and temporal integration in medial olivocochlear bundle activation in humans. *Ear and Hearing*, 22, 65-74.
- Marsh JT, Brown WS, Smith JC. (1975). Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans. *Electroencephalogr Clin Neurophysiol* 38:113-119.
- Marsh JT, Worden FG, Smith JC. (1970). Auditory frequency-following response: neural or artifact? *Science* 169:1222-1223.
- May, B., & McQuone, S. (1995). Effects of bilateral olivocochlear lesions on pure-tone intensity discrimination in cats. *Audit. Neurosci.* 1, 385-400.
- Menard M, Gallego S, Berger-Vachon C, Collet L, Thai-Van H. (2008). Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects. *Hear Res* 235:105-113.
- Micheyl, C., & Collet, L. (1996). Involvement of the olivocochlear bundle in the detection of tones in noise. *J Acoust Soc Am*, 99, 1604-1610.
- Micheyl, C., Khalfa, S., Perrot, X., & Collet, L. (1997). Difference in cochlear efferent activity between musicians and non-musicians. *NeuroReport*, 8, 1047-1050.
- Micheyl, C., Perrot, X., & Collet, L. (1997). Relationship between auditory intensity discrimination in noise and olivocochlear efferent system activity in humans. *Behavioural Neuroscience*, 111, 801-807.

- Miller MI, Sachs MB. (1983). Representation of stop consonants in the discharge patterns of auditory-nerve fibers. *J Acoust Soc Am* 74:502-517.
- Moller AR, Sekiya T, Sen CN. (1989). Responses from dorsal column nuclei (DCN) in the monkey to stimulation of upper and lower limbs and spinal cord. *Electroencephalogr Clin Neurophysiol* 73:353-361.
- Moller AR. Neural mechanisms of BAEP. (1999). *Electroencephalogr Clin Neurophysiol Suppl* 49:27-35.
- Moore B.C.J. (1989). An introduction to the psychology of hearing, Academic Press.
- Moore B.C.J. (1995). Perceptual consequences of cochlear damage, Oxford: University Press.
- Moore, B.C.J. (1973). Frequency difference limens for short-duration tones, *J. Acoust. Soc. Am.* 54, 610.
- Moushegian G, Rupert AL, Stillman RD. Laboratory note. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr Clin Neurophysiol* 5:665-667.
- Musacchia G, Sams M, Skoe E, Kraus N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104:15894-15898.
- Musacchia G, Strait D, Kraus N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 241:34-42.
- Niyogi, P., & Ramesh, P. (2003). The voicing feature for stop consonants: Recognition experiments with continuously spoken alphabets. *Speech Communication*, 41, 349-367.
- Oldfield, R.C. (1971). The assessment and analysis of handedness : The Edinburgh Inventory. *Neuropsychologia*, 9, 97-113.
- Oliver D, Huerta M, (1992). Inferior and superior colliculi. In D.B. Webster, AN Poepper & RR Fay (Eds). The mammalian auditory pathway : Neuroanatomy. New York: Springer-Verlag. (pp. 169-221).
- Palmer A. et Shamma S. (2004). Physiological representations of speech. In: Springer handbook of auditory research *speech processing in the auditory system.*, (vol. 18). p 150–200.
- Parbery-Clark A, Skoe E, Kraus N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 29:14100-14107.
- Parbery-Clark A, Skoe E, Lam C, Kraus N. (2009). Musician enhancement for speech-in-noise. *Ear Hear* 30:653-661.
- Perrot X, Ryvlin P, Isnard Jet al. (2006). Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb Cortex* 16:941-948.
- Perrot, X. (2003). *Cortex auditif et système efférent olivocochléaire médian : influence d'une stimulation électrique corticale sur les otoémissions acoustiques et leur*

- atténuation controlatérale*. Unpublished Thèse de Doctorat en Médecine, Faculté de Médecine Lyon - Nord,
- Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., et al. (2006). Evidence of corticofugal modulation of peripheral auditory activity in humans. *Cerebral Cortex* 16, 941-948.
- Picton TW, Hillyard SA, Krausz HI, Galambos R. (1974). Human auditory evoked potentials. I. Evaluation of components. *Electroencephalogr Clin Neurophysiol* 36:179-190.
- Picton TW, John MS, Dimitrijevic A, Purcell D. (2003). Human auditory steady-state responses. *Int J Audiol* 42:177-219.
- Poeppel D, Hickok G. (2004). Towards a new functional anatomy of language. *Cognition* 92:1-12.
- Poeppel D, Idsardi WJ, van Wassenhove V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 363:1071-1086.
- Poeppel D, Idsardi WJ, van Wassenhove V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philos Trans R Soc Lond B Biol Sci* 363:1071-1086.
- Purcell DW, John SM, Schneider BA, Picton TW. (2004). Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am* 116:3581-3593.
- Rajan, R. (2000). Centrifugal Pathways Protect Hearing Sensitivity at the Cochlea in Noisy Environments That Exacerbate the Damage Induced by Loud Sound. *The Journal of Neuroscience*, 20, 6684-6693.
- Rance G, Dowell RC, Rickards FW, Beer DE, Clark GM. (1998). Steady-state evoked potential and behavioral hearing thresholds in a group of children with absent click-evoked auditory brain stem response. *Ear Hear* 19:48-61.
- Rauschecker, J.P. (1998) Cortical processing of complex sounds. *Curr. Opin. Neurobiol.* 8, 516-521
- Recanzone GH, Engle JR, Juarez-Salinas DL.(in press) Spatial and temporal processing of single auditory cortical neurons and populations of neurons in the macaque monkey. *Hear Res*.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131-1136.
- Rose, J.E., Brugge, J.F., Anderson, D.J. and Hind, J.E. (1968). Patterns of activity in single auditory nerve fibers of the squirrel monkey, In A.V.S. de Reuck & J. Knight (Eds).: *Hearing mechanisms in vertebrates*. London: Churchill, 144.
- Rosen S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* 363:367-373.
- Rosenzweig, M. R. (1951). Representations of the two ears at the auditory cortex. *American Journal of Physiology*, 167, 147-158.
- Ruel, J., Nouvian, R., Gervais d'Aldin, C., Pujol, R., Eybalin, M., & Puel, J.-L. (2001). Dopamine inhibition of auditory nerve activity in the adult mammalian cochlea. *European Journal of Neuroscience*, 14, 977-986.

- Russo N, Nicol T, Musacchia G, Kraus N. (2004). Brainstem responses to speech syllables. *Clin Neurophysiol* 115:2021-2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. (2005). Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 156:95-103.
- Russo NM, Skoe E, Trommer B et al. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol* 119:1720-1731.
- Sachs MB, Young ED. (1980). Effects of nonlinearities on speech encoding in the auditory nerve. *J Acoust Soc Am* 68:858-875.
- Scott SK, Johnsrude IS. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26:100-107.
- Scott SK, Wise RJ. (2004). The functional neuroanatomy of prelexical processing in speech perception. *Cognition* 92:13-45.
- Serniclaes W, Sprenger-Charolles L, Carre R, Demonet JF. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *J Speech Lang Hear Res* 44:384-399.
- Serniclaes, W. (1987). Étude expérimentale de la perception du trait de voisement des occlusives du français. Université Libre de Bruxelles, Bruxelles, Belgique.
- Shinn-Cunningham, B.G. (2009). I want to party, but my hearing aids won't let me! *Hearing J.* 62, 10-13.
- Sinex DG, McDonald LP, Mott JB. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *J Acoust Soc Am* 90:2441-2449.
- Sinex DG, McDonald LP, Mott JB. (1991). Neural correlates of nonmonotonic temporal acuity for voice onset time. *J Acoust Soc Am* 90:2441-2449.
- Sinex DG, McDonald LP. (1988). Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *J Acoust Soc Am* 83:1817-1827.
- Sinex DG, McDonald LP. (1988). Average discharge rate representation of voice onset time in the chinchilla auditory nerve. *J Acoust Soc Am* 83:1817-1827.
- Sinex DG, McDonald LP. (1989). Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve. *J Acoust Soc Am* 85:1995-2004.
- Skoe E, Kraus N. (2010). Auditory Brain Stem Response to complex sounds: a tutorial. *Ear & Hearing*, vol. 31, NO.3.
- Smith JC, Marsh JT, Brown WS. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 39:465-472.
- Snoeren, N. D., Hallé, P. A., & Segui, J. (2006). A voice for the voiceless: Production and perception of assimilated stops in French. *Journal of Phonetics*, 34, 241-268.
- Sohmer H, Pratt H, Kinarti R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42:656-664.
- Sohmer H, Pratt H, Kinarti R. (1977). Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42:656-664.

- Song JH, Banai K, Russo NM, Kraus N. (2006). On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurotol* 11:233-241.
- Song JH, Skoe E, Wong PC, Kraus N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20:1892-1902.
- Steinschneider M, Fishman YI, Arezzo JC. (2008). Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. *Cereb Cortex* 18:610-625.
- Steinschneider M, Reser D, Schroeder CE, Arezzo JC. (1995). Tonotopic organization of responses reflecting stop consonant place of articulation in primary auditory cortex (A1) of the monkey. *Brain Res* 674:147-152.
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG, Jr. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: temporal response patterns. *Brain Lang* 8:326-340.
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG, Jr. (1994). Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr Clin Neurophysiol* 92:30-43.
- Steinschneider M, Volkov IO, Fishman YI, Oya H, Arezzo JC, Howard MA, 3rd. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb Cortex* 15:170-186.
- Steinschneider M, Volkov IO, Noh MD, Garell PC, Howard MA, 3rd. (1999). Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex. *J Neurophysiol* 82:2346-2357.
- Straight DL, Kraus N, Parbery-Clark A, Ashley R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res* 261:22-29.
- Straight DL, Kraus N, Skoe E, Ashley R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 29:661-668.
- Suga N. (2008). Role of corticofugal feedback in hearing. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 194:169-183.
- Suga, N., Xiao, Z., Ma, X., & Ji, W. (2002). Plasticity and Corticofugal Modulation for Hearing in Adult Animals. *Neuron*, 36, 9-18.
- Sun, X., Xia, Q., Lai, C. H., Shum, D. K., Chan, Y. S., & He, J. (2007). Corticofugal modulation of acoustically induced Fos expression in the rat auditory pathway. *Journal of Comparative Neurology*, 501, 509–525.
- Sussman E, Steinschneider M, Gumenyuk V, Grushko J, Lawson K. (2008). The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hear Res* 236:61-79.
- Suzuki T, Hirabayashi M. (1987). Age-related morphological changes in auditory middle-latency response. *Audiology* 26:312-320.
- Tallal P, Miller S, Fitch RH. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Ann N Y Acad Sci* 682:27-47.

- Tremblay K, Kraus N, Carrell TD, McGee T. (1997). Central auditory system plasticity: generalization to novel stimuli following listening training. *J Acoust Soc Am* 102:3762-3773.
- Tzounopoulos T, Kraus N. (2009). Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62:463-469.
- Ulanovsky N, Las L, Nelken I. (2003). Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391-398.
- Université Claude Bernard - Lyon 1, Lyon.
- Van Tasell DJ, Soli SD, Kirby VM, Widin GP. (1987). Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am* 82:1152-1161.
- Veillet E, Magnan A, Ecalle J, Thai-Van H, Collet L. (2007). Auditory processing disorder in children with reading disabilities: effect of audiovisual training. *Brain* 130:2915-2928.
- Veillet, E., Bazin, F., & Collet, L. (1999). Objective evidence of peripheral auditory disorders in learning-impaired children. *Journal of Audiological Medicine*, 8, 18-29.
- Wever EG, Bray CW. (1930). Action Currents in the Auditory Nerve in Response to Acoustical Stimulation. *Proc Natl Acad Sci USA* 16:344-350.
- Wible B, Nicol T, Kraus N. (2004). Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 67:299-317.
- Wible B, Nicol T, Kraus N. (2005). Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 128:417-423.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10:420-422.
- Worden FG, Marsh JT. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol* 25:42-52.
- Zatorre RJ, Belin P, Penhune VB. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* 6:37-46.
- Zatorre RJ, Belin P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946-953.
- Zatorre RJ, Gandour JT. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363:1087-1104.

VI. Communications à des congrès nationaux avec actes

Jeanvoine A, Richard C., Moulin A., Berger-Vachon C, Thai-Van H. (2010) Détection automatique d'indices pertinents dans les réponses électrophysiologiques de type PEA recueillis en réponse à un son de parole. 10ème Congrès Français d'Acoustique, Lyon, Avril 2010.

Jeanvoine, A., Richard, C. Moulin, A. Perrin, E., Laboissière R., Thai-Van, H. Berger-vachon, C. (2010) Détection automatique des indices caractéristiques d'un Potentiel Evoqué Auditif en réponse à un son de parole (PEASP) par l'utilisation de la méthode de détection de forme. Handicap 2010 (IFRATH, juin 2010, Paris).

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

A. Jeanvoine^{1, 2}, C. Richard^{1, 3}, A. Moulin^{1, 2}, E. Perrin^{2, 4}, R. Laboissière^{2, 5}, Thai-Van^{1, 2}, C. Berger-vachon^{1, 2}

1 : Laboratoire Neurosciences Sensorielles, Comportement et Cognition, UMR 5020, H.E.H., 5 place d'Arsonval,

69437 Lyon cedex 03

2 : Université Claude Bernard Lyon1, 43 Bd du 11 Novembre 1918, 69100 Villeurbanne

3 : Service d'ORL et de chirurgie cervico-faciale, CHU de Saint Etienne, Hôpital Nord, 42055 Saint Etienne cedex 02

4 : CREATIS-LRMN, CNRS INSA UCBL UMR 5220 INSERM U630, bâtiment 308, CPE Lyon, 43 Bd du 11 Novembre 1918, 69622 Villeurbanne cedex

5 : INSERM U864, Espace et Action, 16 Avenue Doyen Lépine, 69676 Bron cedex

arnaudjeanvoine@yahoo.fr

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

permet de déterminer les points de plus forte corrélation entre le modèle et la réponse, identifiant par là même, les indices au sein de la réponse, de manière objective.

I. Introduction

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

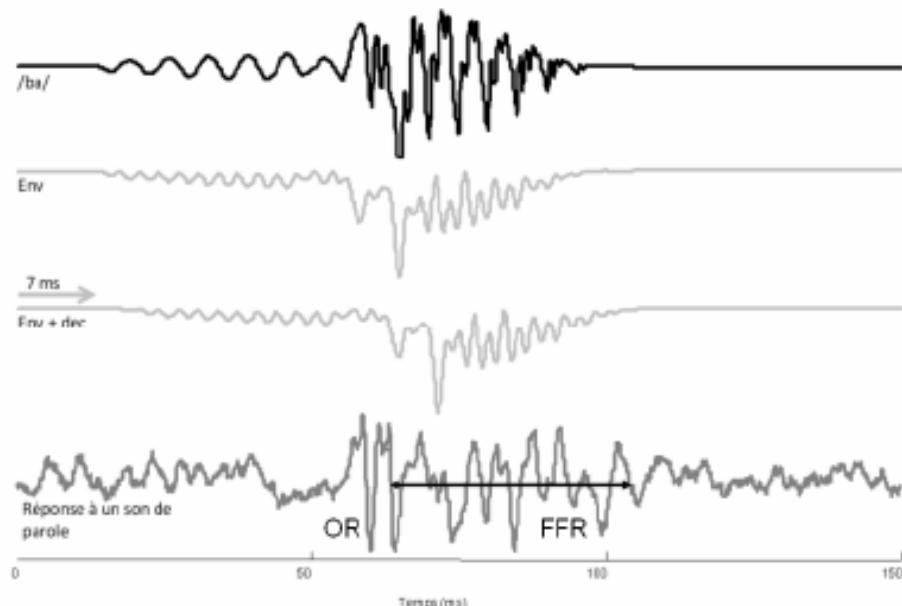


Figure 1 : Potentiel évoqué auditif en réponse à un son de parole.

I. Matériel ET Méthode

A. Méthode de recueil de PEASP

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

Une chaîne de stimulation composée d'un ordinateur portable et d'une carte son externe, dont l'un des canaux est, après amplification, connecté aux écouteurs de type insert, insérés au niveau du conduit auditif externe du sujet, un autre canal étant connecté au trigger.

Une chaîne de recueil, composée de 3 électrodes de surface, positionnées sur le scalp du sujet, au niveau du vertex (électrode « + »), du front (masse), et de chaque mastoïde (électrode « - ») (fig. 2). La mesure du signal, de l'ordre du micro-volt, s'effectue par la différence de potentiel entre l'électrode « + » et l'électrode « - », par rapport à la masse commune de la tête. Le signal des électrodes est amplifié (boîte tétière) avant d'être recueilli par un appareil couramment utilisé pour le recueil des PEAP : le Centor USB, qui assure le moyennage des réponses, synchronisées sur le stimulus.

Du fait de la longueur du stimulus (>50 ms), de son occurrence simultanée au recueil, et de sa similitude avec la réponse de type PEAPSP, (fig. 1), le recueil nécessite une isolation dans une cage de Faraday stricte reliée à la masse, ainsi que l'utilisation d'écouteurs de type d'inserts plutôt que de type casque (Akhoun et coll.), permettant d'isoler le plus possible les électrodes de la transduction électro-acoustique des écouteurs. Un moyennage de plus de 2000 traces en polarité alternée est habituellement nécessaire, avec une intensité de stimulation variant de 50 à 80 dB HL selon les équipes (richard et al.). La fréquence d'échantillonnage est celle maximale admise par le Centor USB (13,3 kHz).

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

B. Caractéristiques du PEAPSP

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

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

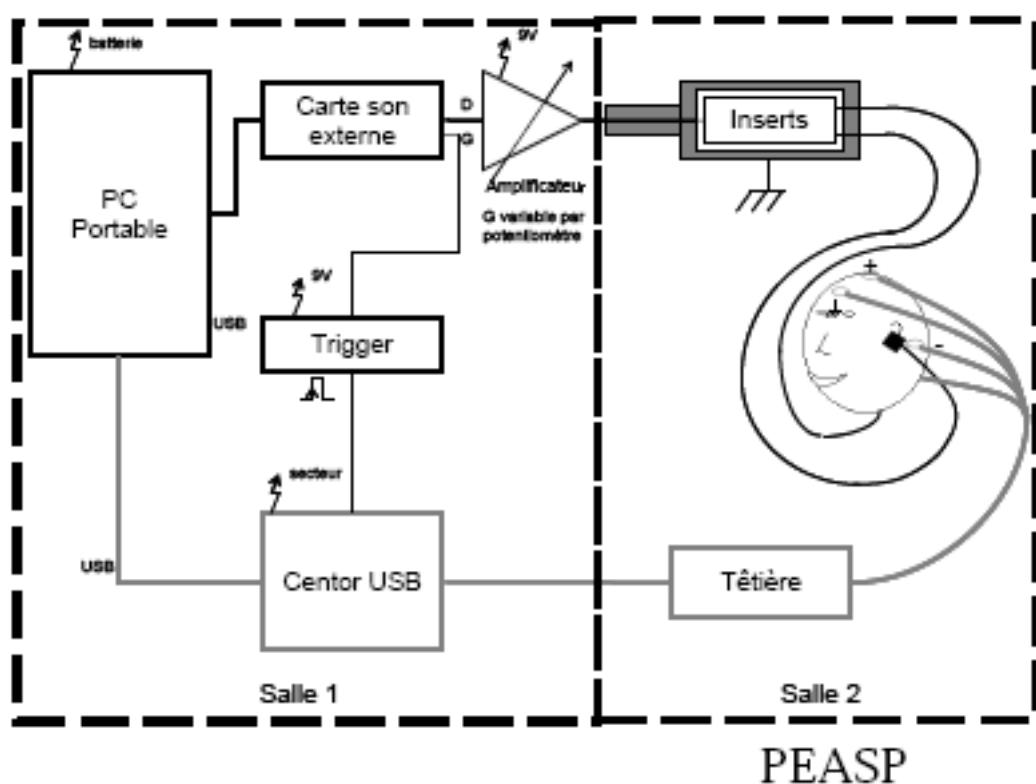


Figure 2 : Chaine de recueil du PEASP

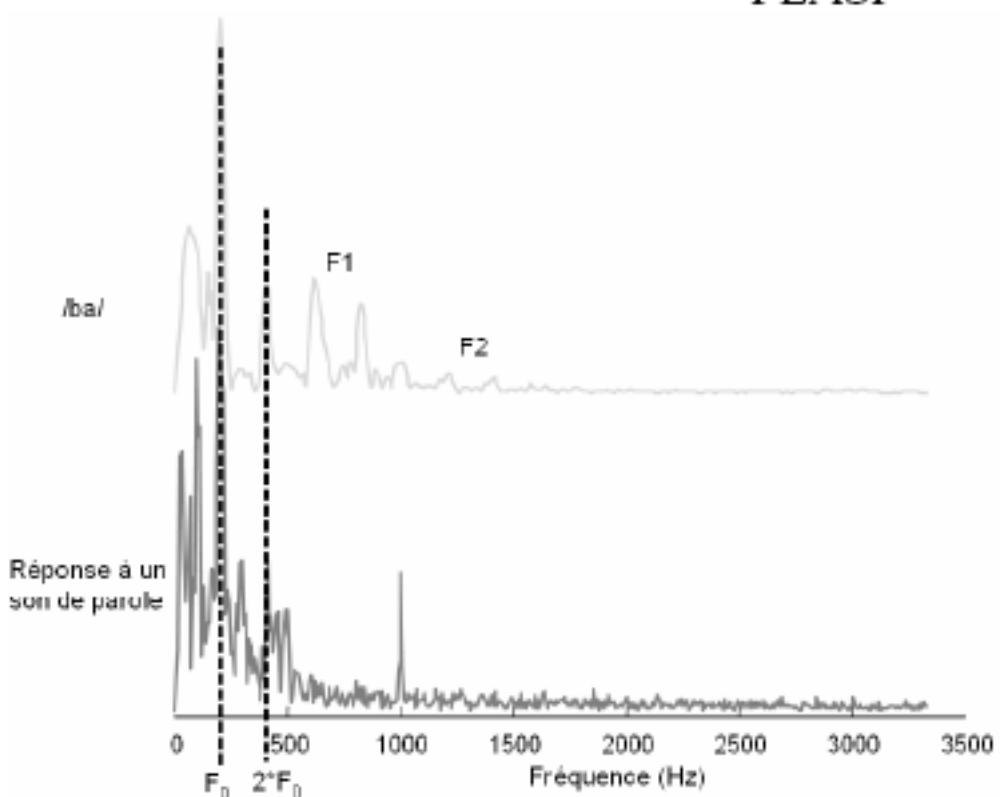


Figure 3 : Spectre de la réponse et du stimulus /ba/.

C. Modélisation d'un PEASPType.

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

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

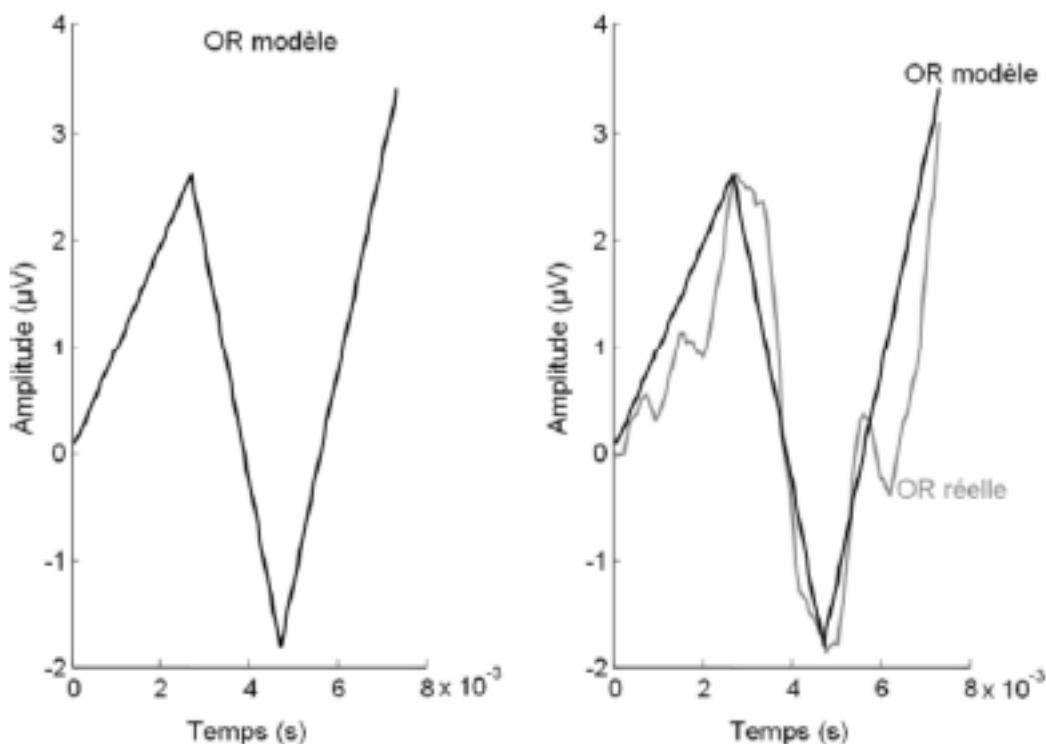


Figure 4 : Modèle de l'OR (gauche) et modèle + réelle (droite)

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

(2 cycles), permettant une plus grande précision temporelle, mais au détriment d'une plus grande sensibilité au bruit. Le modèle sinusoïdal de la FFR est pondéré par une fenêtre de hanning, pour éviter, lors de la corrélation, les effets de bord.

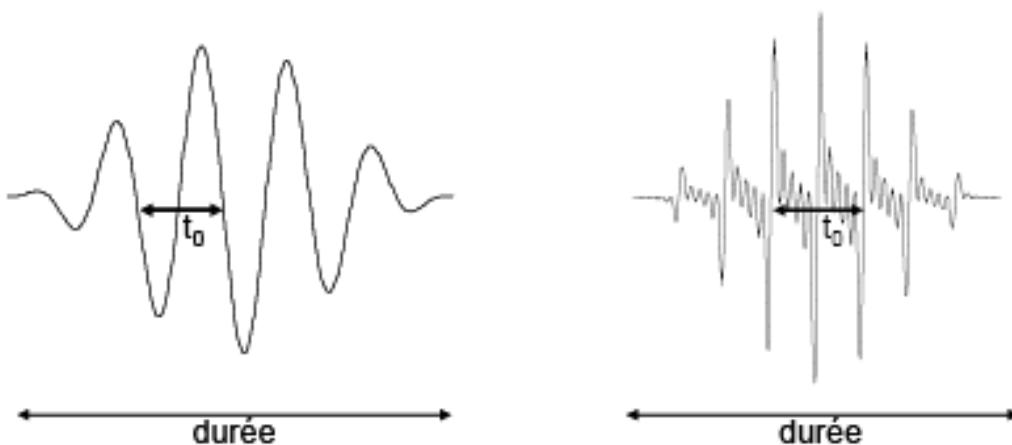


Figure 5 : Modèle sinusoïdal de FFR, à gauche avec une seule harmonique et à droite avec quatre harmoniques, correspondant à la FFR présentée fig. 1 et 3.

D. Méthode de détection

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

III. Résultats et discussions

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

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

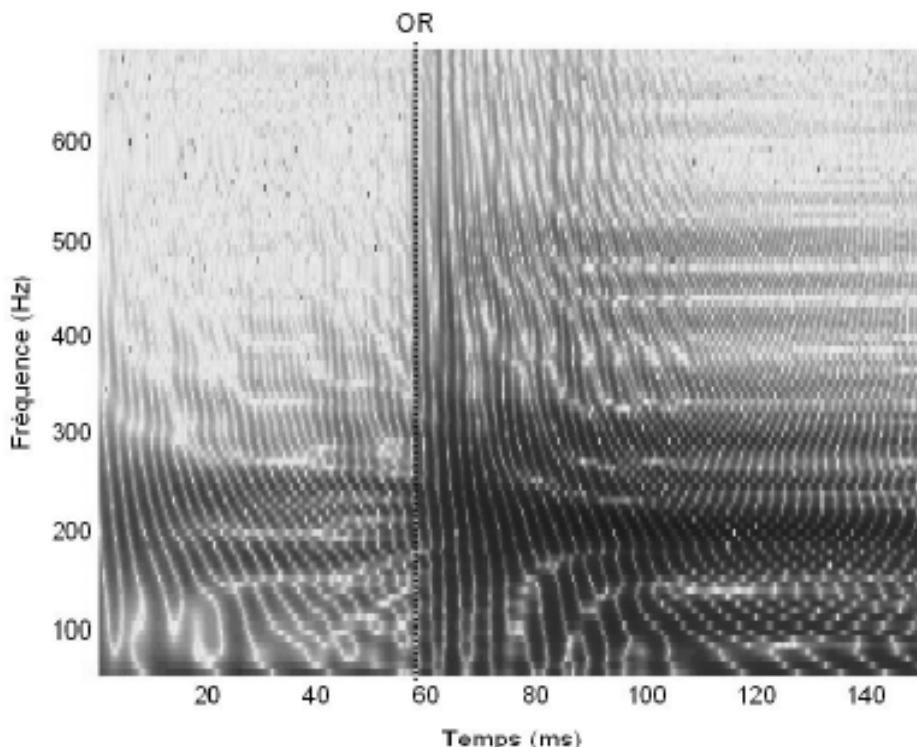


Figure 6 : Détection de l'OR

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

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

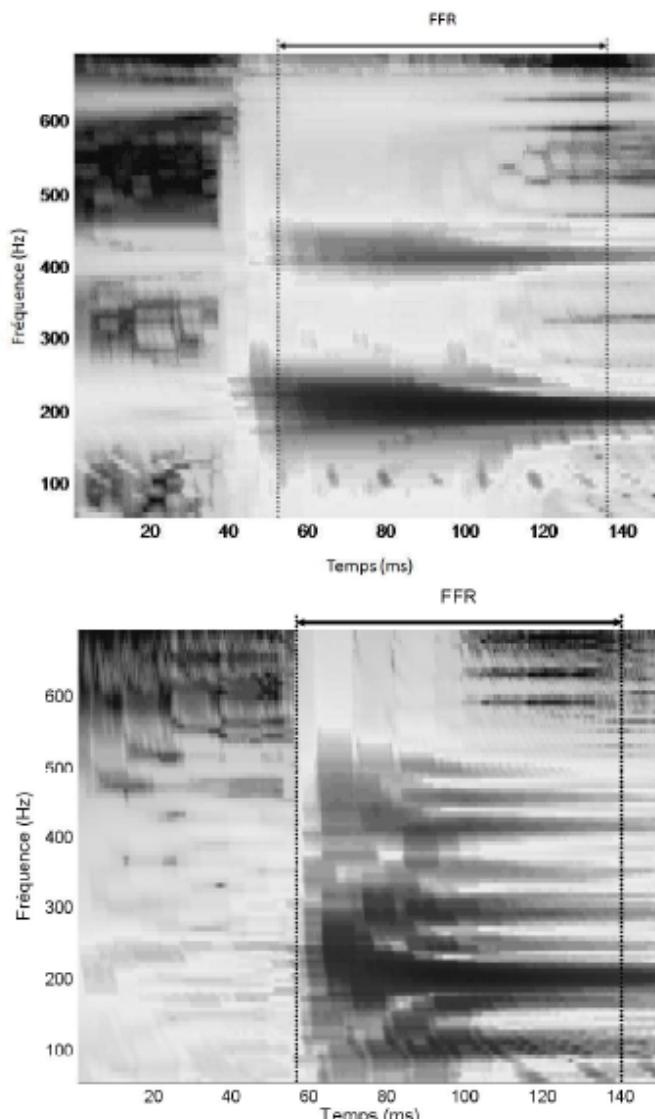


Figure 7 : Détection de la FFR avec 7 oscillations pour le graphique haut et 2 oscillations pour le graphique du bas

IV. CONCLUSION

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

V. Bibliographie

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

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

- [3] N. Russo, T.G. Nicol, G. Musacchia, N. Kraus. «Brainstem responses to speech syllables». *Clinical Neurophysiol*, 2004; 115: 2021_2030
- [4] C. Richard, A. Moulin, A. Jeanvoine, I. Akhoun, E. Veuillet, L. Collet, H. Thai-Van. «Exploration électrophysiologique des voies auditives sous-corticales chez l'Humain : du clic au son de parole». (communication personnelle). *Neurophysiologie Clinique*
- [5] K.L. Johnson, T.G. Nicol, N. Kraus. «Brain stem response to speech: a biological marker of auditory processing». *Ear Hear*, 2005; 26: 424_434.
- [6] E. Skoe, N. Kraus. «Auditory brain stem response to complex sound : a tutorial». *Ear Hear*, 2010; 31: 1_23
- [7] I. Akhoun, A. Moulin, A. Jeanvoine, M. Ménard, F. Buret, C. Vollaire, R. Scorratti, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. «Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status an experimental parametric study», *Journal of Neuroscience Methods*, 2008; 175: 196_205.
- [8] J.T. Marsh, F.G. Worden, J.C. Smith. «Auditory frequency-following response: neural or artefact?». *Science*, 1970; 169: 1222_1223
- [9] I. Akhoun , S. Gallégo, A. Moulin, M. Ménard, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. «The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults». *Clinical Neurophysiol*, 2008; 119: 922_933.
- [10] S.J. Aiken, T.W. Picton. «Envelope following responses to natural vowels». *Audiol Neurotol*, 2006; 11: 213_232
- [11] D. A. Abrams, T. Nicol, S.G. Zecker, N. Kraus. «Auditory Brainstem Timing Predicts Cerebral Asymmetry for Speech». *Journal of Neuroscience*, 2006; 26(43): 11131_11137
- [12] J. Hornikel, E. Skoe, N. Kraus. «Subcortical laterality of speech», *Audiol Neurotol*, 2009; 14: 198_207.

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

10^{ème} Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

A. Jeanvoine^a, C. Richard^{a,b}, A. Moulin^a, C. Berger-Vachon^a, et H. Thai-van^a

a : Laboratoire Neurosciences Sensorielles, Comportement et Cognition, CNRS UMR 5020, HEH, 5 place d'Arsonval, 69437, Lyon cedex 03

b : Service d'ORL et de chirurgie servico-faciale, CHU de Saint Etienne, Hopital Nord, 42055 Saint Etienne cedex 02

arnaudjeanvoine@yahoo.fr

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

ces indices pertinents de manière objective, afin de pouvoir caractériser la variabilité inter-individuelle de l'aspect des Speech ABR.

1 Introduction

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

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

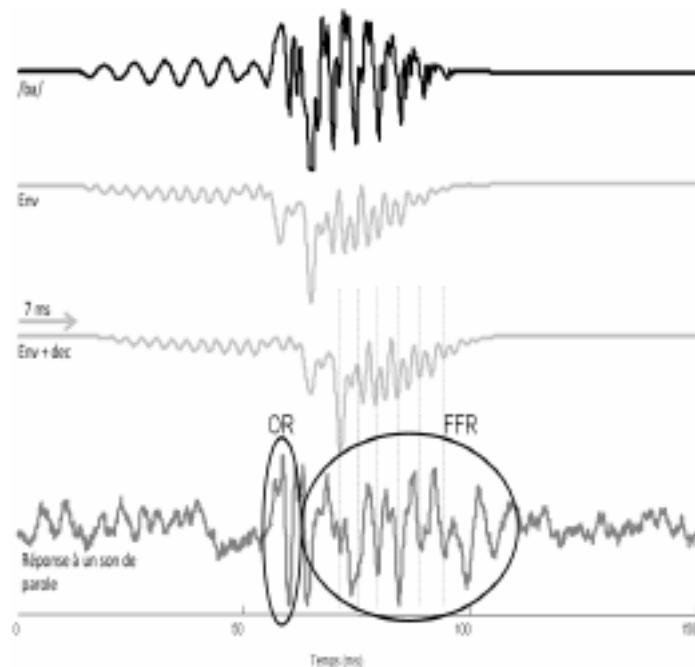


Figure 1a : Potentiel évoqué auditif en réponse à un son de parole (domaine temporel).

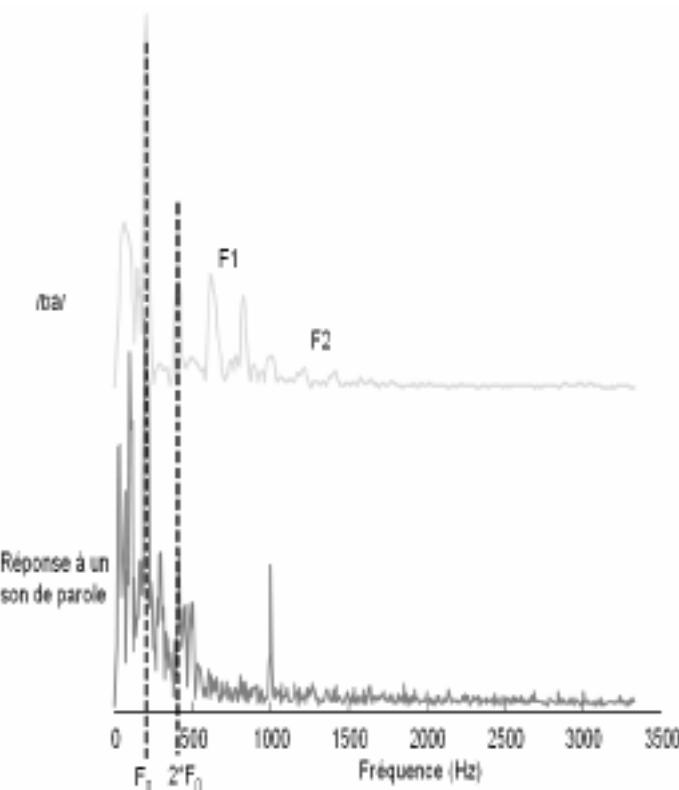


Figure 1b : Potentiel Evoqué Auditif en réponse à un son de parole (domaine fréquentiel).

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

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

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

2 Matériel et méthodes

2.1 Méthode de recueil du Speech ABR

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

- Une chaîne de stimulation composée d'un ordinateur portable et d'une carte son externe, dont l'un des canaux est, après amplification, connecté aux écouteurs de type insert, insérés au niveau du conduit auditif externe du sujet, un autre canal étant connecté au trigger.
- Une chaîne de recueil, composée de 3 électrodes de surface, positionnées sur le scalp du sujet, au niveau du vertex (électrode « + »), du front (masse), et de chaque

mastoïde (électrode « - »). La mesure du signal, de l'ordre du micro-volt, s'effectue par la différence de potentiel entre l'électrode « + » et l'électrode « - », par rapport à la référence « masse » de la tête. Le signal des électrodes est amplifié (boîtier Têteière) puis numérisé avant d'être recueilli par un appareil couramment utilisé en clinique pour le recueil des PEAP : le Centor USB° (Racia Alvar°), qui assure le moyennage des réponses, synchronisées sur le stimulus à l'aide du trigger.

• Du fait de la longueur du stimulus (>50 ms), de son occurrence simultanée au recueil, et de sa similitude avec la réponse Speech ABR, (figure 1a), le recueil nécessite une isolation de la transduction électrique/acoustique dans une cage de Faraday reliée à la masse du système, ainsi que l'utilisation d'écouteurs de type d'inserts plutôt que de type casque classique, permettant d'isoler le plus possible les électrodes des écouteurs [3,4]. Un moyennage de plus de 2000 traces en polarité alternée est habituellement nécessaire, avec une intensité de stimulation variant de 50 à 80 dB HL selon les équipes [5,6] pour obtenir un recueil avec un rapport signal sur bruit suffisant. La fréquence d'échantillonnage du recueil est celle maximale admise par le Centor USB (13,3 kHz).

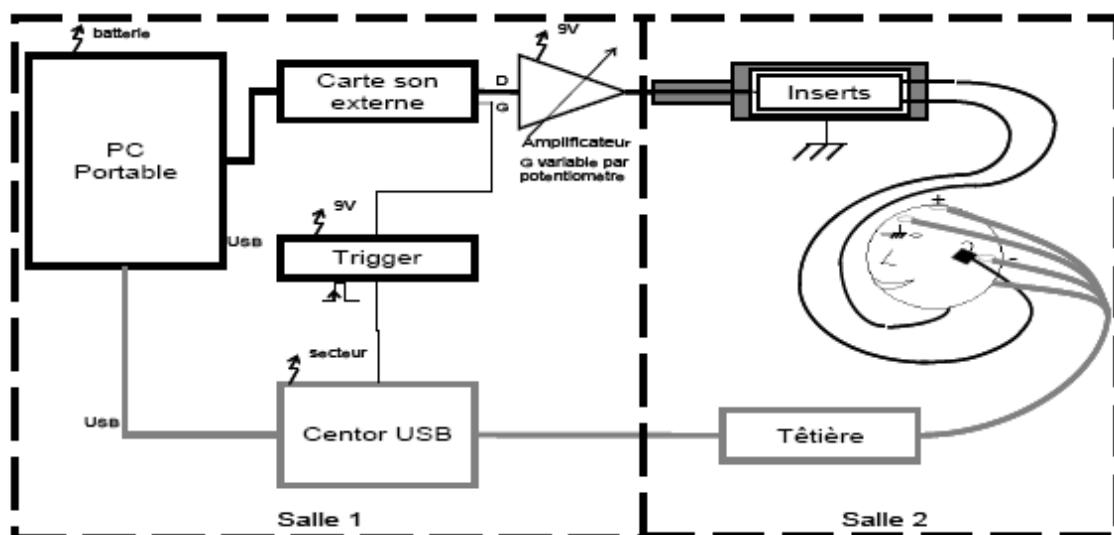


Figure 2 : Chaîne de stimulation et de recueil d'un Speech ABR

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

2.2 Caractéristiques du Speech ABR

Le premier indice pertinent du Speech ABR est une réponse impulsionnelle ressemblant fortement à « l'onde V » des Potentiels Auditifs Précoce classiques (ou OR). Elle est facilement décelable car l'amplitude est largement supérieure à celle du bruit moyen de la trace. Elle est suivie de la réponse soutenue en fréquence (FFR) [5]. La figure 1a montre les traces temporelles du stimulus /ba/, de son enveloppe [7], de cette même enveloppe décalée de 7 ms, et de la réponse électrophysiologique du stimulus [1,5]. Ce décalage d'une dizaine de milliseconde (dépendant du sujet) donne la latence de la FFR, elle est obtenue en corrélant l'enveloppe inversée du stimulus [1], obtenue après un filtrage

passe-bas avec une fréquence de coupure de 500 Hz d'ordre un, avec la réponse de type FFR [8,9,10]. Les caractéristiques des indices sont répertoriées dans le tableau 1.

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

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

Tableau 1 : Récapitulatif des caractéristiques des indices pertinents.

2.3 Utilisation de méthodes mathématiques pour la détection des indices pertinents

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

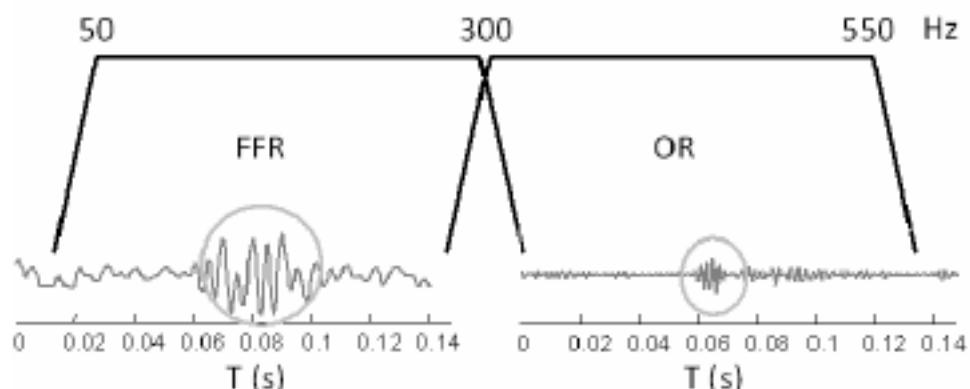


Figure 3 : Séparation de l'OR de la FFR fréquentiellement, la trace temporelle en bas montre les différences temporelles pour l'OR et la FFR.

2.3.1 Méthode de détection de l'OR

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

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

La dérivée permet, à chaque instant de la trace, de connaître le coefficient directeur de l'information temporelle haute fréquence, un coefficient directeur grand montre une

impulsion et donc une montée d'énergie très courte. Il suffit de prendre l'instant où le coefficient directeur est le plus grand négativement car la pente recherchée est décroissante. La partie correspondante est alors isolée par un encadrement de 10 ms centré sur le coefficient directeur trouvé.

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

2.3.2 Méthode de détection de la FFR

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

3 Résultats

3.1 Détection de l'OR

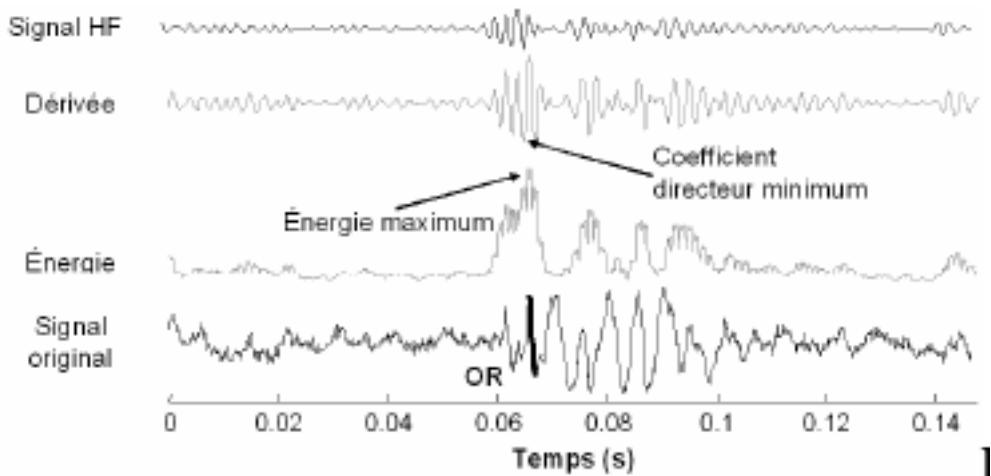


Figure 4 : Détection de l'OR (en gras sur la courbe du bas)

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

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

3.2 Détection de la FFR

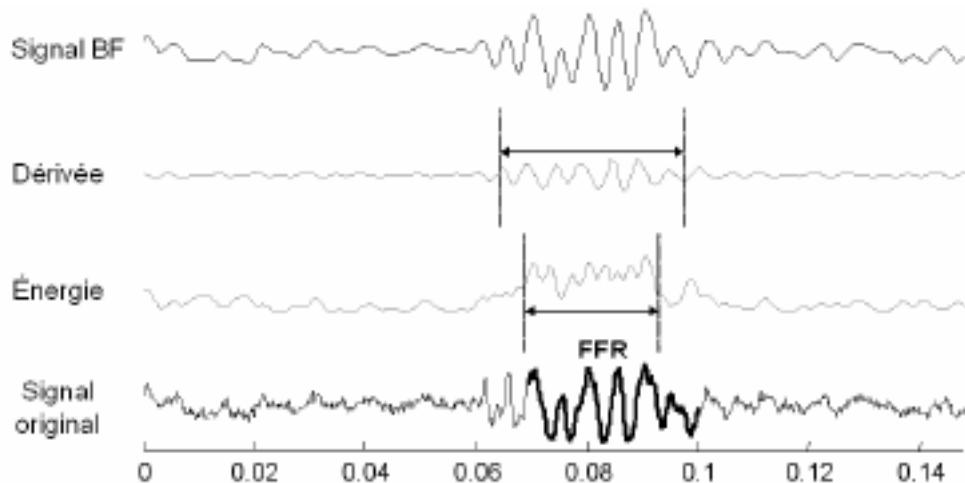


Figure 5 : Détection de la FFR (en gras sur la courbe du bas)

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

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

4 Discussion

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

5 Conclusion

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

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

Références

- [1] I. Akhoun, S. Gallégo, A. Moulin, M. Ménard, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. "The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults", *Clinical Neurophysiol*, 2008; 119: 922_933.
- [2] K.L. Johnson, T.G. Nicol, N. Kraus. "Brain stem response to speech: a biological marker of auditory processing". *Ear Hear*, 2005; 26: 424_434.
- [3] I. Akhoun, A. Moulin, A. Jeanvoine, M. Ménard, F. Buret, C. Vollaire, R. Scorretti, E. Veuillet, C. Berger-Vachon, L. Collet, H. Thai-Van. "Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status an experimental parametric study", *Journal of Neuroscience Methods*, 2008; 175: 196_205.
- [4] J.T. Marsh, F.G. Worden, J.C. Smith. "Auditory frequency-following response: neural or artefact?", *Science*, 1970; 169: 1222_1223
- [5] E. Skoe, N. Kraus. "Auditory brain stem response to complex sound : a tutorial". *Ear Hear*, 2010; 31: 1-23
- [6] C. Richard, A. Moulin, A. Jeanvoine, I. Akhoun, E. Veuillet, L. Collet, H. Thai-Van. "Exploration électrophysiologique des voies auditives sous-corticales chez l'Humain : du clic au son de parole". (communication personnelle). *Neurophysiologie Clinique*
- [7] J.W. Hall. "Auditory brainstem frequency following responses to waveform envelope periodicity". *Science*, 1979; 205: 1297_1299
- [8] N. Russo, T.G. Nicol, G. Musacchia, N. Kraus. "Brainstem responses to speech syllables". *Clinical Neurophysiol*, 2004; 115: 2021_2030
- [9] S.J. Aiken, T.W. Picton. "Envelope following responses to natural vowels". *Audiol Neurotol*, 2006; 11: 213_232
- [10] Aiken SJ, Picton TW. "Envelope and spectral frequency-following responses to vowel sounds". *Hear Res* 2008;245:35-47.

VII. Communications à des congrès nationaux

Akhoun I., Gallégo S., Richard C., Veuillet E., Thai-Van H., Collet L.. Réponse du tronc cérébral à un son de parole (Speech ABR). Corrélation avec les données psychométriques. 112e Congrès de la Société Française d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou. Paris Octobre 2007.

Richard C., Akhoun I., Veuillet E., Collet L., Thai-Van H.. "Réponse du tronc cérébral à un son de parole (speech ABR) dans différentes conditions de bruit.". 113e Congrès de la Société Française d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou. Paris Octobre 2008.

Richard C., Jeanvoine A., Moulin A., Veuillet E., Collet L., Thai Van H.. L'enveloppe d'un signal de parole peut-elle suffire à l'obtention d'un speech ABR ? 114e Congrès de la Société Française d'Oto-Rhino-Laryngologie et de Chirurgie de la Face et du Cou. Paris Octobre 2009.

Richard C., Jeanvoine A., Moulin A., Gallego S., Veuillet E., Collet L., Thai Van H (2009) L'enveloppe d'un signal de parole peut-elle suffire à l'obtention d'un speech ABR ? 116ème Congrès de la Société Française d'O.R.L. et de Chirurgie de la Face et du Cou. Paris. Octobre 2009.

VIII. Publications acceptées et en préparation

*** D'après les résultats de l'expérience 1**

Approche empirique pour déterminer les générateurs neurophysiologiques sous-jacents des potentiels évoqués auditifs engendrés par des sons de parole (Speech ABR).

Akhoun I, Berger-Vachon C., Collet L. (2010),

IRBM

*** Revue de littérature sur les PEAPSP.**

Exploration of the auditory system in humans: from click to speech auditory brainstem responses

Céline Richard, Arnaud Jeanvoine, Evelyne Veuillet, Annie Moulin, Hung Thai-Van
Clinical Neurophysiology (Neurophysiologie Clinique).

Accepté pour publication le 10/08/2010.

*** D'après les résultats de l'expérience 5 (Manuscrit en préparation).**

*** D'après les résultats de l'expérience 6:**

Characteristics of Speech ABR obtained at low stimulation levels.

(Manuscrit en préparation).

[richard_c_annexes_article.pdf](#)

EXPLORATION ELECTROPHYSIOLOGIQUE DES VOIES AUDITIVES SOUS-CORTICALES CHEZ L'HUMAIN:

DU CLIC AU SON DE PAROLE

Céline Richard^{1,2}, Arnaud Jeanvoine¹, Evelyne Veuillet¹,

Annie Moulin¹, Hung Thai-Van¹

EXPLORATION OF THE AUDITORY SYSTEM IN HUMANS:

FROM CLICK TO SPEECH AUDITORY BRAINSTEM RESPONSES

1 Laboratoire Neurosciences Sensorielles Comportement Cognition UMR CNRS 5020, Université de Lyon, Université Lyon 1, Hôpital Edouard Herriot, 5 place d'Arsonval - 69437 LYON cedex 03

2 Service d'ORL et de Chirurgie cervico-faciale, CHU de Saint-Etienne, Hôpital Nord - 42055 St-Etienne cedex 02

Auteur correspondant: Céline Richard

Tel: 04 72 11 05 26

Courriel: celine.richard@chu-st-etienne.fr

celrichdoc@hotmail.com

EXPLORATION ELECTROPHYSIOLOGIQUE DES VOIES AUDITIVES SOUS-CORTICALES CHEZ L'HUMAIN: DU CLIC AU SON DE PAROLE

Céline Richard, Arnaud Jeanvoine, Evelyne Veuillet,

Annie Moulin, Hung Thai-Van

EXPLORATION OF THE AUDITORY SYSTEM IN HUMANS:

FROM CLICK TO SPEECH ABR

Résumé:

La compréhension et l'évaluation des soubassements neurophysiologiques de la perception de la parole au niveau sous-cortical sont au centre des préoccupations de plus en plus de chercheurs, avec des applications cliniques potentielles dans les troubles du langage.

Cet article présente l'évolution des techniques permettant l'évaluation du traitement neuronal de la parole au niveau du tronc cérébral: d'abord les potentiels évoqués auditifs (PEA) en réponse à des clics, ensuite découverte de la réponse soutenue à la fréquence (RSF), enfin PEA en réponse à des sons de parole (PEASP) ou « Speech Auditory Brainstem Response ».

Parce qu'ils permettent l'étude des mécanismes sous-corticaux du traitement des sons complexes, les PEASP apportent une vision objective, non invasive et en temps réel des capacités d'encodage neuronal de chaque individu au niveau du tronc cérébral. Leurs caractéristiques devraient leur permettre d'être utilisés comme marqueurs biologiques d'un déficit de l'encodage sous-cortical de la parole, ainsi que de la rééducation de ce type de déficit chez des personnes présentant des troubles de l'apprentissage et/ou du traitement de l'information auditive.

Mots clefs: Potentiels évoqués auditifs en réponse à un son de parole, tronc cérébral, marqueur biologique, information auditive, langage, réhabilitation auditive, potentiels évoqués auditifs (PEA).

Abstract:

There is a growing and unprecedented interest in the objective evaluation of the subcortical processes that are involved in speech perception, with potential clinical applications in speech and language impairments.

Here, we review the studies illustrating the development of electrophysiological methods for assessing speech encoding in the human brainstem: from the pioneer recordings of click-evoked auditory brainstem responses (ABR), via studies of frequency-following responses (FFR) to the most recent measurements of Speech Auditory Brainstem Response (SABR) or ABR in response to speech sounds.

Recent research on SABR has provided new insights in the understanding of subcortical auditory processing mechanisms. The SABR test is an objective and non invasive tool for assessing individual capacity of speech encoding in the brainstem. SABR characteristics are potentially useful both as a diagnosis tool of speech encoding deficits and as an assessment tool of the efficacy of rehabilitation programs in patients with learning and/or auditory processing disorders.

Key words: Speech Auditory Brainstem Response (SABR), brainstem, biological marker, auditory information, auditory rehabilitation, Auditory Brainstem Responses (ABR)

Introduction

Avant même l'interprétation du message auditif au niveau cortical, la perception de la parole implique tout un ensemble de traitements d'un signal acoustique complexe et extrêmement variable, composé d'informations spectrales et temporelles, à différents niveaux du système auditif [36]. Ainsi, l'information auditive est transmise de manière ascendante au cortex auditif primaire (circonvolutions de Heschl situées au niveau temporal) après différents traitements à chaque relais, du nerf auditif jusqu'au cortex. Le premier relais est effectué au niveau des noyaux cochléaires ventraux et dorsaux à partir desquels un important contingent de fibres croisent la ligne médiane (au niveau des corps trapézoïdes), un second relais est effectué au niveau des complexes olivaires supérieurs (tronc cérébral), ensuite dans les lemnisques latéraux (au niveau de la commissure de Probst), des colliculi inférieurs (mésencéphale) et enfin au niveau des corps genouillés médians (au niveau thalamique). Des décussations s'effectuent au niveau des corps trapézoïdes et à des niveaux plus élevés (noyaux du lemnisque latéral, tubercules quadrijumeaux postérieurs), de telle sorte que la majorité des fibres issues d'une oreille ont croisé la ligne médiane lors de l'arrivée des influx au niveau thalamique [9, 25]. Parallèlement à ces voies afférentes, un système corticofuge efférent auditif module l'information auditive par de multiples boucles de rétrocontrôle, les plus courtes réalisées par les projections cortico-thalamiques et les plus longues constituées par les voies efférentes connectées aux cellules des complexes olivaires supérieurs (tronc cérébral) ainsi qu'aux cellules ciliées externes de l'organe auditif périphérique [54, 62, 76]. Ainsi, l'exploration du traitement des sons de parole par le tronc cérébral et la compréhension de ses mécanismes fourniraient un aperçu des caractéristiques du message auditif tel qu'il est envoyé au cortex.

Différentes approches sont utilisées afin de comprendre les mécanismes physiologiques auditifs intervenant dans la perception de la parole. La première, utilisée en pratique clinique courante, correspond aux potentiels évoqués auditifs (PEA) et teste l'intégrité des voies auditives, de la cochlée jusqu'à la portion moyenne du tronc cérébral. L'enregistrement des PEA fait appel à une technique non invasive et sensible de recueil de l'activité électroencéphalographique en réponse à des stimuli sonores simples tels des clics [7, 82]. Les études neurophysiologiques ont permis de mettre en évidence l'existence d'un traitement pré-cortical de l'information auditive au niveau du tronc cérébral, où l'information temporelle et spectrale contenue dans la parole serait encodée [36, 69]. Cependant, si les PEA en réponse à des clics nous permettent d'objectiver l'intégrité des voies auditives, les caractéristiques de l'encodage spectral et temporel de la parole font appel à des stimuli plus écologiques et de complexité bien supérieure à celle d'un clic, à savoir des stimuli langagiers donnant lieu à des potentiels évoqués auditifs (PEA) spécifiques: les PEA en réponse à un son de parole (PEASP) ou « speech auditory brainstem responses ». Cette approche permettrait de mettre en évidence certains troubles de perception de la parole dus à des dysfonctionnements de l'encodage des sons langagiers au niveau du tronc cérébral, ceci de manière non invasive, fiable et sensible [6, 69, 76].

Cet article s'attachera à une revue sommaire des différents types d'enregistrements neurophysiologiques auditifs de courte latence [PEA en réponse à des clics, réponses soutenues à la fréquence (RSF), PEA stationnaires multiples (PEASM)] ainsi que des caractéristiques acoustiques des sons de la parole, avant d'aborder l'étude de l'encodage des sons langagiers au niveau sous-cortical au moyen des PEASP.

Du clic au son de parole...

Historiquement, les premiers enregistrements électrophysiologiques du système auditif datent des années 1930 [77]. Toutefois, la mise en évidence du PEA en réponse à des

clics en tant que mesure objective de la fonctionnalité des voies auditives ne débutera que 40 ans plus tard [32]. Elle sera suivie par les premiers enregistrements d'une réponse soutenue à la fréquence (RSF ou « Frequency Following Response »), in situ chez l'animal [81] et en champ lointain chez l'homme [50]. L'utilisation des PEA en pratique clinique débuta dès les années 1980, avec une répartition en différents sous-groupes en fonction de leurs latences d'apparition par rapport au stimulus : les PEA de latences précoces ou PEAP (recueillis dans les 10 premières millisecondes), les PEA de latences moyennes (entre 10 et 50 millisecondes) et les PEA corticaux (temps de latence supérieurs à 50 millisecondes). D'autres stimuli sont utilisés pour acquérir les PEA de latences moyennes et tardifs : il s'agit des bouffées tonales ou « tones bursts », correspondant à des sons purs émis pendant des durées de quelques millisecondes (20ms, 500Hz) [25]. Toutefois les PEA du tronc cérébral, obtenus en réponse à des clics, présentent des limites. Ils ne permettent d'explorer ni l'ensemble de la gamme fréquentielle représentative du spectre auditif humain, ni l'encodage des informations temporelles et spectrales contenues dans des stimuli écologiquement valides, comme la parole [73].

La RSF est un phénomène naturel reflétant chez l'humain la synchronisation de l'activité des neurones auditifs sur la fréquence d'un stimulus extérieur. D'un point de vue physiologique, la RSF est un potentiel évoqué représentant l'activité globale de verrouillage de phase dans les cellules du tronc cérébral, entre le noyau cochléaire et le corps genouillé médian du thalamus [12, 17, 27, 55]. Au niveau du colliculus inférieur, l'activité de verrouillage de phase est retrouvée jusqu'à 1000 Hz environ [12, 17, 19, 24, 27, 43, 55, 57] et est composée de pics calqués sur la périodicité d'un stimulus complexe contenant des harmoniques de la fréquence fondamentale (F0) [44, 51, 70, 81]. L'activité de verrouillage de phase contenue dans la RSF permet donc l'encodage des fluctuations fréquentielles rapides du son de parole. La RSF neuronale (de latence supérieure à 6 ms) obtenue en enregistrant la différence de potentiel entre des électrodes placées sur le lobe de l'oreille et sur le scalp, est à différencier du potentiel microphonique cochléaire (de faible latence), que l'on peut faire disparaître en faisant la moyenne des réponses obtenues à des paires de stimuli en opposition de phase [71]. L'information contenue dans la RSF neuronale serait transmise le long de fibres provenant du tour apical de la cochlée et, après 3 à 5 synapses, atteindrait le colliculus inférieur où la RSF neuronale est générée [13, 51, 55, 71]. La RSF reproduit une oscillation à une fréquence qui s'apparente d'un point de vue perceptif à la sensation de hauteur tonale perçue [24]. Il est intéressant de noter qu'un son pur modulé en amplitude ou en fréquence (modulation inférieure à 100 Hz) peut générer une réponse électrophysiologique s'apparentant à une RSF : les PEAS stationnaires multiples (PEASM) ou « auditory steady-state response ».

Les PEASM sont générés à travers le système nerveux auditif par un stimulus continu dans le temps et sont représentés par un potentiel évoqué neuronal qui suit l'enveloppe d'un stimulus complexe périodique avec une activité de « verrouillage de phase ». Pour l'étude des PEASM, plusieurs fréquences porteuses (250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz, 8000Hz) sont utilisées comme stimuli, chacune modulée en amplitude à une fréquence différente (90Hz [64] ou 80Hz [52], [33]). La réponse est enregistrée pour chaque fréquence de modulation d'amplitude associée à une fréquence porteuse précise. Cette étude peut être effectuée sur plusieurs fréquences porteuses simultanément et ceci de manière bilatérale, en conduction osseuse ou aérienne. L'analyse des fréquences recueillies est effectuée de manière individuelle afin de déterminer la présence d'une réponse à chaque fréquence audiométrique. Il a été par ailleurs mis en évidence, une corrélation entre l'amplitude des PEASM et la perception subjective d'intensité sonore ou sonie [52]. Les PEASM s'inscrivent en complément des PEA, notamment dans la prédiction des seuils auditifs sur l'ensemble

de la gamme fréquentielle chez le nouveau-né et l'enfant sédaté ou endormi [61]. De plus, les PEASM couvrent des champs d'investigation vastes tels que l'étude de l'encodage du fondamental des voyelles, [2, 3], l'objectivation de la fonction de sonie [52], ou encore la dyssynchronisation auditive dans le cadre des neuropathies auditives [64].

Cependant, bien que les PEASM permettent une étude plus approfondie de l'audition d'un sujet, le traitement sous-cortical de l'encodage temporel de la parole n'est pas analysable et les stimuli utilisés restent peu écologiques. Or, avec la mise en évidence de troubles du traitement de la structure fine de l'information temporelle, à l'origine de troubles auditifs en particulier en environnement bruyant [48], il apparaît nécessaire de disposer d'un outil diagnostique permettant une étude objective de cet encodage temporel: les PEASP répondent à cette nécessité [36, 63, 78].

Analyse des sons par le tronc cérébral

Caractéristiques d'un son de parole

Un son de parole est codé à la fois dans sa dimension fréquentielle et sa dimension temporelle. L'analyse spectrale des sons de parole permet la description de certains éléments acoustiques tels que leur fréquence fondamentale et leur structure harmonique. La fréquence fondamentale est un composant de basse fréquence de la parole résultant de la vibration des cordes vocales et permettant la perception de la hauteur tonale de la voix d'un individu. La structure harmonique est un phénomène acoustique dont les composants (les harmoniques) sont des multiples de la fréquence fondamentale. L'interaction entre la fréquence de vibration des cordes vocales et les résonances du tractus vocal d'un individu privilégie certains harmoniques proches de la fréquence de résonance du tractus vocal : les formants. Les deux premiers formants jouent d'ailleurs un rôle majeur dans la perception des voyelles [1]. L'utilisation de stimuli contenant des voyelles pour l'étude des sons de parole est indispensable si l'on veut se rapprocher de stimuli naturels.

Les caractéristiques acoustiques temporelles jouent un rôle important dans la perception de la parole [49, 68]. Elles interviennent en complément des informations fréquentielles (fréquence fondamentale et formants). On peut distinguer trois types d'information temporelle en fonction de la gamme de fréquence considérée :

L'enveloppe temporelle (« contour du son ») véhicule les informations relatives au rythme global de la phrase, la délimitation des syllabes et concerne les variations temporelles de moins de 50 Hz (> 20 ms).

La périodicité reflète la prosodie (intonation de la voix), ainsi que le voisement (genre du locuteur, hauteur de la voix) et se situe entre 50 et 500 Hz.

Enfin, la structure fine temporelle (SFT) se situe au-delà de 500 Hz (durées < 2ms). Celle-ci jouerait un rôle aussi bien dans la perception de la hauteur tonale [53] que dans la perception de la parole. En effet, chez le normo-entendant, l'élimination des indices de structure fine entraîne une dégradation de l'intelligibilité de la parole dans un bruit fluctuant [23]. De plus, certains patients avec une perte auditive neurosensorielle présentent une capacité réduite à exploiter ces indices de structure fine temporelle, et donc davantage de difficultés de perception de la parole dans le bruit. En effet, ces indices de SFT participeraient notamment à l'exploitation des creux sonores présents dans un bruit modulé. Cette capacité d'utilisation des creux sonores serait nécessaire à la détection de signaux langagiers et non langagiers dans les situations adverses [23, 49, 53].

Chacun de ces types d'information a son importance vis-à-vis de la compréhension de la parole. Par exemple, la perception catégorielle permettant de distinguer les phonèmes

p/b est basée sur la perception du délai du voisement, long de quelques dizaines de millisecondes en Français. Ce délai du voisement correspond au temps entre le relâchement de l'obstruction dans la production de la consonne et la mise en vibration des cordes vocales. Il est donc important de disposer d'outils cliniques (méthodes objectives et subjectives) permettant d'évaluer l'acuité auditive par rapport à ces caractéristiques temporelles.

Encodage des sons par le tronc cérébral

L'encodage des informations auditives par le tronc cérébral fait appel à la notion de « codage neuronal temporel », caractérisé par la synchronisation des décharges des cellules du système auditif sur la périodicité du stimulus acoustique : c'est le principe de verrouillage de phase ou « phase-locking » [6, 69]. Ce verrouillage de phase se répercute tout le long des voies auditives afférentes en se limitant à des fréquences de plus en plus basses au fur et à mesure que le message nerveux est intégré : il intervient jusqu'à des fréquences de stimulation de 4 à 5 kHz au niveau du nerf auditif, et pour des fréquences inférieures à 80 Hz dans le cortex auditif [59]. Cette activité de verrouillage de phase, présente dans les neurones du tronc cérébral, joue un rôle prépondérant dans l'encodage neuronal des caractéristiques temporelles et spectrales d'un son de parole [65].

Les potentiels évoqués auditifs en réponse à un son de parole (PEASP)

Les PEASP (ou « speech auditory brainstem responses ») sont des PEA en réponse à un son de parole et reflètent l'activation de générateurs neuronaux le long du tractus auditif ascendant, donnant une image fonctionnelle des mécanismes physiologiques impliqués dans le traitement des signaux complexes [13, 19, 24,].

Méthode d'acquisition des PEASP

Les stimuli les plus couramment utilisés sont des syllabes (par exemple /da/ [36, 69]; /ba/ [4]) contenant en moyenne 5 formants, dont la durée varie de 40 millisecondes à plusieurs centaines de millisecondes. Ces stimuli sont le plus souvent présentés de manière monaurale à droite avec des intensités variant de 50 dB SL [12] à 90,3 dB SPL [30]. Le son est délivré via des inserts afin d'éviter l'artefact électromagnétique.

Les réponses EEG induites par les stimuli sont enregistrées à l'aide d'électrodes cutanées placées sur le vertex (électrode positive), le front (masse) et chaque mastoïde (électrodes négatives) [5, 21, 36, 42, 69]. Le moyennage (nombre de répétitions du stimulus) porte sur un nombre variable - de 999 [22, 55] à 6000 [69, 78] - de répétitions collectées pour chaque polarité du stimulus, avec une fenêtre d'enregistrement variant de 52ms [78] à plus de 340 ms en fonction des stimuli utilisés [74] (Tableaux 1 et 1bis). Les montages utilisés pour le recueil sont sensiblement identiques et s'apparentent à celui représenté sur la Figure 1.

En quoi consiste un PEASP-type ?

Les PEA en réponse à une syllabe peuvent être divisés en différentes parties (Figure 2a, b) :

La première partie est représentée par la réponse impulsionale ou « Onset Response », précédée par une phase d'enregistrement pré-stimulus (appelée « pré-moyennage »), destinée à l'évaluation du bruit électrophysiologique ambiant. La réponse impulsionale (ondes V-A) est une réponse transitoire, encodant les variations temporelles rapides inhérentes à la consonne. Elle est similaire au PEA généré par un clic et apparaît approximativement 7 ms après le début du stimulus.

La seconde partie, correspondant à la FFR, apparaît en réponse aux harmoniques et à la structure périodique de la voyelle à la fréquence de la source (voisement) ; elle présente une activité soutenue, la périodicité des pics de cette phase correspondant à l'information fréquentielle contenue dans le stimulus [50, 65]. La période entre les pics C, D, E, F de la FFR (activité de verrouillage de phase) correspond à la fréquence fondamentale du stimulus (F0), tandis que les pics compris entre les ondes C, D, E, F représentent l'activité de verrouillage de phase à la fréquence du premier formant F1 [6]. Une dernière onde apparaît ensuite : l'onde O qui correspond à l'arrêt du son, et temporellement à la fin du stimulus [36].

Selon les équipes et les protocoles, les méthodes d'analyse varient et sont répertoriées dans les Tableaux 1 et 1bis.

La mesure des latences d'apparition des différentes parties de la réponse (réponse impulsionnelle et réponse soutenue à la fréquence) est effectuée par une intercorrélation entre le stimulus et la réponse. Enfin, l'énergie du signal est habituellement calculée à l'aide de la racine quadratique moyenne (RMS) de la réponse soutenue à la fréquence et le rapport signal-bruit peut s'estimer en calculant le bruit sur plusieurs millisecondes d'enregistrement pré-stimulus. Les critères de validité d'une réponse peuvent être établis en utilisant par exemple un test statistique de comparaison de variances (F-test), basé sur le F-ratio entre l'énergie calculée sur les fréquences correspondant aux fréquences du stimulus, et l'énergie calculée sur les fréquences voisines [15, 61, 83].

PEASP : artefact ou réponse neurophysiologique ?

Dans le cas des PEASP, la simultanéité entre la stimulation et l'enregistrement, d'une part, la forte similitude entre la forme du stimulus et le pattern du signal recueilli, d'autre part, font que le risque d'artefact électromagnétique est particulièrement important. Cet artefact, résultant de la superposition du signal neuronal (RSF) et du couplage électromagnétique des transducteurs, est lié aux problèmes de compatibilité électromagnétique entre la chaîne de stimulation (en majeure partie les transducteurs acoustiques) et les électrodes de recueil. En effet, la précision, souvent spectaculaire, de la synchronisation entre la RSF et la voyelle, pose la question de l'authenticité de cette réponse neuronale [71] qui pourrait n'être que le reflet du couplage électromagnétique entre la chaîne de stimulation et les électrodes de recueil. Cet artefact n'existe pas avec les PEA en réponse à des clics, car le PEA n'est pas enregistré de manière concomitante au stimulus. Pour les PEASM, dont la réponse est concomitante au stimulus, l'artefact peut être éliminé grâce aux différentes fréquences de modulation [5]. Les PEASP sont, quant à eux, difficiles à distinguer de l'artefact, les premiers pics fréquentiels du stimulus correspondant à ceux de la réponse soutenue à la fréquence. De plus, dans le cas des PEASP, la polarité alternée (stimulus envoyé alternativement de manière inversée et non inversée [24, 36]) à elle seule ne suffit pas à annuler les réponses artefactuelles. L'élimination des réponses artefactuelles doit être effectuée dans un premier temps en amont, par l'isolation de la chaîne de recueil. En effet, les transducteurs les moins susceptibles de donner lieu à du couplage électromagnétique sont les écouteurs intra-auriculaires (Figure 3) dont la totalité de l'habitacle (hormis les tubes d'air) ainsi que les fils sont encastrés dans une cage de Faraday, reliée à la masse électrique (unique) du dispositif expérimental [5]. En effet, tous les autres transducteurs risquent d'engendrer des réponses contaminées pour les intensités de stimulation habituellement utilisées (Figure 3). Les casques habituellement utilisés en audiologie (TDH 39) sont donc à proscrire dans l'enregistrement des PEASP. Enfin, la validation de la nature neurophysiologique des réponses peut se faire en vérifiant l'absence de réponse dans un modèle inerte et l'absence de réponse chez des patients présentant une cophose (surdité complète) unilatérale [5].

Enfin, l'observation d'un temps de latence de 7 à 10 ms entre le stimulus et la réponse visible sur les PEASP est en faveur du caractère neurophysiologique de celle-ci.

Où est générée la réponse ?

Pour une utilisation clinique, outre la détection objective de certaines pathologies, il est pertinent de poser la question de la localisation des générateurs des PEASP le long des voies auditives. Les neurones du système auditif ont des caractéristiques fonctionnelles spécifiques, telles que des taux de décharges spontanées, des fréquences de réponse et des types de décharges (transitoires, périodiques synchronisées, périodiques mais à une périodicité intrinsèque). De plus, chaque filtre auditif répond avec un décalage temporel (en raison du délai de propagation de la vibration de la membrane basilaire) et chaque noyau cellulaire répond également au son avec des latences progressives d'une milliseconde en sortie de la cochlée jusqu'à une dizaine de millisecondes dans le colliculus inférieur. A priori, aucune réponse identifiable ne devrait parvenir aux électrodes sur le scalp, ou alors noyée dans le bruit de fond résultant de l'ensemble de l'activité neuronale. Toutefois, les phénomènes électriques expliquant la propagation du message nerveux jusqu'au scalp ont récemment été décrits dans le cas des PEA en réponse à des clics, et peuvent être étendus aux PEASP. En effet, d'un point de vue physique, le champ électromagnétique composite accompagnant le signal électrophysiologique peut être transmis jusqu'aux électrodes de scalp lorsque le signal change de milieu biologique (tissu neuronal, liquide céphalo-rachidien, os, peau) et lorsque suffisamment de neurones déchargent de façon synchrone [16]. Les neurones thalamiques, à la différence de ceux du colliculus inférieur, ne semblent pas impliqués dans la genèse des PEASP. En revanche, les caractéristiques de propagation de l'information auditive jusqu'au cortex peuvent s'expliquer par l'existence d'un mode de codage neuronal spécifique dénommé « fonction de transfert du taux de modulation » (rMTF) des cellules du colliculus inférieur. Ce mode repose sur le codage de l'information de modulation de fréquence par des variations du taux de décharge proportionnelles à cette fréquence modulante [38]. A contrario, les noyaux cellulaires de la périphérie cochléaire utilisent uniquement un type de codage temporel-MTF (fonction de transfert de modulation temporelle) : les cellules déchargent chacune selon leurs caractéristiques spécifiques, mais avec une période correspondant à celle du stimulus. La variation périodique de courant synchronisé sur la périodicité de la modulante (F0) présente dans le codage rMTF, semble donc plus à même de résulter en une différence de potentiel enregistrable sur le scalp.

Les PEASP apporteraient donc un aperçu précis du fonctionnement du colliculus inférieur. Ceci serait en particulier très utile au diagnostic différentiel des surdités centrales ou des neuropathies auditives [41].

Intérêts neurophysiologiques des potentiels évoqués auditifs en réponse à un son de parole

Encodage spatial et temporel de l'information auditive

Quelles que soient leurs caractéristiques (voyelles constantes, présence d'un accent de langue, voyelles apériodiques - chuchotées -, voyelles fricatives sans voisement [13, 14]), les stimuli langagiers sont encodés de manière fidèle au niveau du tronc cérébral, l'activité de verrouillage de phase reflétée dans la FFR ayant été démontrée aussi bien pour des sons purs [55], des stimuli complexes [24] que pour des stimuli variant dans le temps [42, 46].

L'étude des PEA en réponse à un son de parole a permis de caractériser de manière plus précise ce codage temporel et spatial de l'information: à faible intensité, la réponse temporelle à chaque stimulus est maximale dans les fibres nerveuses auditives ayant une

fréquence caractéristique proche de celle du stimulus, suggérant une spécificité spatiale de la réponse. Ainsi, en augmentant l'intensité de stimulation, il existerait une diffusion spatiale jusqu'aux unités de hautes fréquences pour le codage des formants [45]. Une étude récente a permis de mettre en évidence un effet de l'intensité du son de parole sur le temps de latence du signal PEASP. En effet la diminution de l'intensité de stimulation (entre 60 et 0 dB SL) entraîne un allongement de temps de latence des PEASP : de 7 à 10 ms pour la réponse impulsionnelle et de 15.6 à 23 ms pour la RSF. Toutefois, l'allongement du temps de latence de la réponse impulsionnelle (-0.6 ms pour 10 dB) est moins important que celui observé pour la RSF (-1.4 ms pour 10 dB), suggérant l'existence de deux voies séparées de traitement de l'information au niveau du tronc cérébral, avec des générateurs différents pour la réponse impulsionnelle et la RSF [4].

Le nerf auditif et les fibres du noyau cochléaire augmentent leur verrouillage de phase (aux harmoniques des formants) mais il apparaît que des populations neuronales différentes encoderaient les premier et deuxième formants [34] qui seraient suffisants à l'identification de la plupart des voyelles [45]. Un « seuil » d'encodage spatial et de verrouillage de phase de l'information auditive dans les situations de la vie courante pourrait donc être obtenu pour chaque sujet sur l'ensemble de la gamme fréquentielle [4]. Toutefois, l'activité de verrouillage de phase serait limitée à la fréquence de coupure au niveau du colliculus inférieur, qui est d'environ 1000 Hz [8, 20, 24].

Un même phonème peut revêtir, selon le contexte, une multitude de patterns acoustiques et le phénomène de co-articulation est prégnant surtout pour les consonnes, les voyelles montrant un pattern acoustique plus stable [47]. D'un point de vue électrophysiologique, l'enregistrement des réponses neuronales dans le bruit à différents sites sous-corticaux et corticaux chez l'animal (cobaye) montre un encodage sous-cortical plus robuste de la voyelle dans le bruit, les voyelles ayant une périodicité plus longue et plus forte que les consonnes [10, 65]. En condition de bruit, la transition formantique et les voyelles (RSF) restent encodées au niveau du tronc cérébral, alors que la réponse impulsionnelle est très réduite ou éliminée [26]. Les problèmes perceptifs d'identification de la consonne dans le bruit peuvent donc être attribués à une diminution de la synchronisation neuronale de la réponse impulsionnelle, alors que la préservation de la RSF permettrait la perception de la voyelle [65]. Cette résistance de la RSF dans le bruit par rapport à la réponse impulsionnelle laisse supposer une certaine différence dans leur processus d'encodage au niveau du tronc cérébral. De plus, l'analyse des PEASP montre que l'encodage de la fréquence fondamentale (qui est importante pour identifier le locuteur, la hauteur tonale -pitch - et percevoir la prosodie) persiste dans le bruit, alors que l'Onset-Response et le premier formant (essentiel à la perception phonémique) se dégradent [65].

La perception de la prosodie (intonation, stress), liée à F0, peut être déficiente dans certaines pathologies comme l'autisme. En effet, Russo et al. [67] ont mis en évidence, sur des PEASP recueillis dans un groupe de 21 enfants présentant un syndrome autistique, une diminution de l'encodage de F0, ce qui pourrait être lié à une moins bonne perception de significations linguistiques (par exemple: affirmation ou question) et de l'émotion véhiculée par la parole.

Pour certains, la détection de la hauteur tonale par le tronc cérébral serait expérience-dépendante, malléable et liée au traitement d'autre facteurs cognitifs comme le langage et la musique : les musiciens encoderaient F0 de manière plus forte et avec un verrouillage de phase plus résistant en comparaison à des sujets non musiciens [67, 72, 80]. En effet, la pratique et l'écoute régulière de la musique feraient notamment intervenir des réseaux cognitifs supérieurs exerçant un effet modulateur corticofuge

(modulation top-down) ayant pour conséquence une amélioration des capacités d'encodage au niveau sous-cortical. Cette intervention corticofuge expliquerait en partie, sur le plan neurophysiologique, les facilités d'apprentissage du langage présentes chez les musiciens [46, 80]. Enfin, l'asymétrie hémisphérique fonctionnelle liée aux traitements du langage a été mise en évidence également sur les réponses de type PEASP [30]. En effet, l'existence d'une asymétrie fonctionnelle du traitement du langage est bien établie au niveau cortical [26] : l'hémisphère gauche serait spécialisé dans le traitement de la parole, de par les caractéristiques acoustiques de celle-ci (événements acoustiques rapides associés aux transitions temporelles courtes, alors que la musique, dont les événements temporeaux et spectraux sont plus soutenus et associés à des transitions plus lentes, serait traitée au niveau de l'hémisphère droit [30]). Au niveau sous-cortical, des études électrophysiologiques ont montré des réponses d'apparition plus rapide et comportant des harmoniques correspondant au premier formant du stimulus plus robustes, lors d'une stimulation langagière monaurale droite par rapport à une stimulation controlatérale [30, 75]. Cette asymétrie périphérique a été également mise en évidence au niveau cochléaire [39] et pourrait être liée au rétro-contrôle cortico-sous-cortical mis en évidence chez l'humain sur les oto-émissions acoustiques [60]. Ainsi, l'hypothèse d'une asymétrie périphérique présente dans les PEASP apparaît tout à fait plausible.

Applications cliniques potentielles des PEASP

L'intérêt des PEASP réside dans les nombreuses informations théoriques et cliniques que ceux-ci nous donnent sur le traitement des stimuli auditifs complexes au niveau du tronc cérébral [36, 65]. Les PEASP peuvent ainsi être utilisés comme marqueurs biologiques d'une déficience d'encodage des sons au niveau du tronc cérébral [37, 40, 78, 79]. Ainsi, le recueil des PEASP dans le bruit permet d'objectiver la dégradation de l'encodage des sons de parole, avec une moins bonne résistance de la consonne par rapport à la voyelle, chez des individus normaux [47] et par extension chez des personnes présentant des troubles de l'intelligibilité.

Les difficultés de traitement du signal au niveau central observées chez des enfants présentant des troubles des apprentissages, se traduisent sur les PEASP par une diminution de la réponse RSF, une amplitude plus faible et une durée allongée de la réponse impulsionale (complexe V-A) [78]. Ceci suggère qu'une représentation appauvrie des composants-clés du signal de parole au niveau du tronc cérébral, contribuerait aux difficultés du traitement du même signal à un niveau plus central [78]. Différents groupes peuvent ainsi être individualisés au sein de ces mêmes enfants, selon qu'une corrélation est retrouvée ou non entre les mesures effectuées au niveau de leur tronc cérébral (PEASP) et celles effectuées au niveau cortical. Lorsque cette corrélation existe, les PEASP sont un outil précieux pour le clinicien, puisqu'ils permettent de juger de l'efficacité des programmes de réhabilitation [40, 66, 79]. L'absence d'anomalie des PEASP chez certains enfants souffrant de troubles des apprentissages a été interprétée comme la conséquente modulation corticofuge du traitement de la parole par le tronc cérébral [1, 72, 73]. Un système de mesure dédié à l'enregistrement des réponses du tronc cérébral à des sons complexes (BioMAP: Biological Marker of Auditory Processing) a été proposé par Kraus et collègues afin d'identifier, chez les enfants, les troubles des apprentissages accessibles à une rééducation [28]. De plus, les PEASP ont également mis en évidence, chez les sujets autistes, des troubles de perception de la prosodie au niveau sous-cortical [67].

Conclusions

Les PEA en réponse à un son de parole (PEASP) semblent être une technique prometteuse d'exploration audiologique de l'encodage de la parole par le tronc cérébral [36,

63, 78]. L'exploration de l'encodage des stimuli acoustiques complexes par le tronc cérébral est une étape essentielle à une meilleure compréhension et réhabilitation des processus à la base de l'audition normale et pathologique chez l'humain.

Références

- [1]. Abrams DA, Nicol T, Zecker SG, Kraus N. Auditory brainstem timing predicts cerebral asymmetry for speech. *J Neurosci* 2006;26:11131-11137.
- [2]. Aiken SJ, Picton TW. Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 2008;245:35-47.
- [3]. Aiken SJ, Picton TW. Envelope following responses to natural vowels. *Audiol Neurotol* 2006;11:213-232.
- [4]. Akhoun I, Gallego S, Moulin A et al. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol* 2008;119:922-933.
- [5]. Akhoun I, Moulin A, Jeanvoine A et al. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study. *J Neurosci Methods* 2008;175:196-205.
- [6]. Banai K, Abrams D, Kraus N. Sensory-based learning disability: Insights from brainstem processing of speech sounds. *Int J Audiol* 2007;46:524-532.
- [7]. Banai K, Nicol T, Zecker SG, Kraus N. Brainstem timing: implications for cortical processing and literacy. *J Neurosci* 2005;25:9850-9857.
- [8]. Batra R, Kuwada S, Maher VL. The frequency-following response to continuous tones in humans. *Hear Res* 1986;21:167-177.
- [9]. Biacabe B, Mom T, Avan P, Bonfils P. Anatomie fonctionnelle des voies auditives. *Encycl Méd Chir* (Elsevier, Paris), Oto-rhino-laryngologie, 20-022-A-10, 1999, 7p.
- [10]. Cunningham J, Nicol T, King C, Zecker SG, Kraus N. Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex. *Hear Res* 2002;169:97-111.
- [11]. Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N. Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 2001;112:758-767.
- [12]. Daly DM, Roeser RJ, Moushegian G. The frequency-following response in subjects with profound unilateral hearing loss. *Electroencephalogr Clin Neurophysiol* 1976;40:132-142.
- [13]. Delgutte B. Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *J Acoust Soc Am* 1980;68:843-857.
- [14]. Delgutte B. Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J Acoust Soc Am* 1984;75:879-886.
- [15]. Dobie RA, Wilson MJ. A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids. *J Acoust Soc Am* 1996;100:2236-2246.
- [16]. Eggermont J. J. Auditory evoked potentials. principles and clinical applications., chapitre Electric and magnetic fields of synchronous neural activity: Peripheral and central origin of auditory evoked potentials., pages 1–30. Lippincott, Williams et Wilkins. Burkard

R, Don M, Eggermont JJ (Eds.), Philadelphia, Baltimore, New-York, London, Buenos-Aires, Hong-Kong, Sydney et Tokyo, 2007. 21, 36, 68, 90.

- [17]. Faingold CL, Caspary DM. Frequency-following responses in primary auditory and reticular formation structures. *Electroencephalogr Clin Neurophysiol* 1979;47:12-20.
- [18]. Galbraith GC, Amaya EM, de Rivera JMet al. Brain stem evoked response to forward and reversed speech in humans. *Neuroreport* 2004;15:2057-2060.
- [19]. Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 1995;6:2363-2367.
- [20]. Galbraith GC, Bhuta SM, Choate AK, Kitahara JM, Mullen TA, Jr. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 1998;9:1889-1893.
- [21]. Galbraith GC, Threadgill MR, Hemsley Jet al. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett* 2000;292:123-127.
- [22]. Gerken GM, Moushegian G, Stillman RD, Rupert AL. Human frequency-following responses to monaural and binaural stimuli. *Electroencephalogr Clin Neurophysiol* 1975;38:379-386.
- [23]. Gnansia D, Jourdes V, Lorenzi C. Effect of masker modulation depth on speech masking release. *Hear Res* 2008;239:60-68.
- [24]. Greenberg S, Marsh JT, Brown WS, Smith JC. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear Res* 1987;25:91-114.
- [25]. Guérit J.-M. In: *Les potentiels évoqués*. Editions Masson, Paris, 1993:92-103.
- [26]. Habib M. et Robichon F. Dominance hémisphérique. *Encycl Méd Chir*. Elsevier, Paris; Neurologie, 17-022-C-10, 1998, p.8.
- [27]. Hall J. W.. Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, vol. 205, no. 4412, pages 1297-1299, Sep 1979. 35, 88.
- [28]. Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N. Neural plasticity following auditory training in children with learning problems. *Clin Neurophysiol* 2003;114:673-684.
- [29]. Hoormann J, Falkenstein M, Hohnsbein J. Effect of selective attention on the latency of human frequency-following potentials. *Neuroreport* 1994;5:1609-1612.
- [30]. Hornickel J, Skoe E, Kraus N. Subcortical Laterality of Speech Encoding. *Audiol Neurotol* 2008;14:198-207.
- [31]. Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA*. 2009 Aug 4;106(31):13022-7.
- [32]. Jewett DL, Romano MN, Williston JS. Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 1970;167:1517-1518.
- [33]. John MS, Dimitrijevic A, Picton TW. Auditory steady-state responses to exponential modulation envelopes. *Ear Hear* 2002;23:106-117.
- [34]. Johnson KL, Nicol T, Zecker SG, Bradlow AR, Skoe E, Kraus N. Brainstem encoding of voiced consonant-vowel stop syllables. *Clin Neurophysiol* 2008;119:2623-2635.

- [35]. Johnson KL, Nicol T, Zecker SG, Kraus N. Developmental plasticity in the human auditory brainstem. *J Neurosci* 2008;28:4000-4007.
- [36]. Johnson KL, Nicol TG, Kraus N. Brain stem response to speech: a biological marker of auditory processing. *Ear Hear* 2005;26:424-434.
- [37]. Johnson KL, Nicol TG, Zecker SG, Kraus N. Auditory brainstem correlates of perceptual timing deficits. *J Cogn Neurosci* 2007;19:376-385.
- [38]. Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. *Physiol Rev* 2004;84:541-577.
- [39]. Khalifa S, Veuillet E, Collet L. Influence of handedness on peripheral auditory asymmetry. *Eur J Neurosci*. 1998 Aug;10(8):2731-7.
- [40]. King C, Warrier CM, Hayes E, Kraus N. Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett* 2002;319:111-115.
- [41]. Kraus N, Bradlow AR, Cheatham MA et al. Consequences of neural asynchrony: a case of auditory neuropathy. *J Assoc Res Otolaryngol* 2000;1:33-45.
- [42]. Krishnan A, Parkinson J. Human frequency-following response: representation of tonal sweeps. *Audiol Neurotol* 2000;5:312-321.
- [43]. Krishnan A, Xu Y, Gandour J, Cariani P. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 2005;25:161-168.
- [44]. Krishnan A, Xu Y, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004;189:1-12.
- [45]. Krishnan A. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 2002;166:192-201.
- [46]. Lee KM, Skoe E, Kraus N, Ashley R. Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 2009;29:5832-5840.
- [47]. Liberman A.M., Cooper E.S., Shankweiler D.P. and Studdert-Kennedy M.. Perception of the speech code. *Psychol. Rev.* 1967;74:431-461.
- [48]. Lorenzi C, Debruille L, Garnier S, Fleuriot P, Moore BC. Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *J Acoust Soc Am* 2009;125:27-30.
- [49]. Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA* 2006;103:18866-18869.
- [50]. Marsh JT, Brown WS, Smith JC. Far-field recorded frequency-following responses: correlates of low pitch auditory perception in humans. *Electroencephalogr Clin Neurophysiol* 1975;38:113-119.
- [51]. Marsh JT, Worden FG, Smith JC. Auditory frequency-following response: neural or artifact? *Science* 1970;169:1222-1223.
- [52]. Menard M, Gallego S, Berger-Vachon C, Collet L, Thai-Van H. Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects. *Hear Res* 2008;235:105-113.

- [53]. Moore BC. 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* 2008;9:399-406.
- [54]. Morlet T, Ferber-Viart C, Putet G, Sevin F, Duclaux R. Auditory screening in high-risk pre-term and full-term neonates using transient evoked otoacoustic emissions and brainstem auditory evoked potentials. *Int J Pediatr Otorhinolaryngol* 1998;45:31-40.
- [55]. Moushegian G, Rupert AL, Stillman RD. Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr Clin Neurophysiol* 1973;35:665-667.
- [56]. Musacchia G, Sams M, Nicol T, Kraus N. Seeing speech affects acoustic information processing in the human brainstem. *Exp Brain Res* 2006;168:1-10.
- [57]. Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci U S A* 2007;104:15894-15898.
- [58]. Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 2008;241:34-42.
- [59]. Palmer A. et Shamma S. Physiological representations of speech. In: Springer handbook of auditory research (vol. 18). speech processing in the auditory system., p 150–200. Springer. Greenberg S, Ainsworth AW, Fay RR, Popper AN (Eds.), New-York, 2004. 2, 70, 166.
- [60]. Perrot X, Ryvlin P, Isnard Jet al. Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb Cortex* 2006;16:941-948.
- [61]. Picton TW, John MS, Dimitrijevic A, Purcell D. Human auditory steady-state responses. *Int J Audiol* 2003;42:177-219.
- [62]. Pratt H., Aminoff M., Nuwer M.R., Starr A. Potentiels évoqués auditifs de courte latence. Dans: Guide pratique de neurophysiologie clinique. Recommandations de la Fédération internationale de neurophysiologie clinique. scientifiques et médicales Elsevier SAS. ; 2002.p.95-110.
- [63]. Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am* 2004;116:3581-3593.
- [64]. Rance G, Dowell RC, Rickards FW, Beer DE, Clark GM. Steady-state evoked potential and behavioral hearing thresholds in a group of children with absent click-evoked auditory brain stem response. *Ear Hear* 1998;19:48-61.
- [65]. Russo N, Nicol T, Musacchia G, Kraus N. Brainstem responses to speech syllables. *Clin Neurophysiol* 2004;115:2021-2030.
- [66]. Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 2005;156:95-103.
- [67]. Russo NM, Skoe E, Trommer Bet al. Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol* 2008;119:1720-1731.
- [68]. Shannon RV. The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol* 2002;11:124-127.

- [79]. Skoe E, Kraus N. Auditory Brain Stem Response to complex sounds: a tutorial. *Ear & Hearing*, vol. 31, NO.3, in press.
- [70]. Smith JC, Marsh JT, Brown WS. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 1975;39:465-472.
- [71]. Sohmer H, Pratt H, Kinarti R. Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 1977;42:656-664.
- [72]. Song JH, Banai K, Kraus N. Brainstem timing deficits in children with learning impairment may result from corticofugal origins. *Audiol Neurotol* 2008;13:335-344.
- [73]. Song JH, Banai K, Russo NM, Kraus N. On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurotol* 2006;11:233-241.
- [74]. Song JH, Skoe E, Wong PC, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 2008;20:1892-1902.
- [75]. Strait DL, Kraus N, Skoe E, Ashley R. Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur J Neurosci* 2009;29:661-668.
- [76]. Suga N. Role of corticofugal feedback in hearing. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol*. 2008 Feb;194(2):169-83.
- [77]. Wever EG, Bray CW. Action Currents in the Auditory Nerve in Response to Acoustical Stimulation. *Proc Natl Acad Sci U S A* 1930;16:344-350.
- [78]. Wible B, Nicol T, Kraus N. Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 2004;67:299-317.
- [79]. Wible B, Nicol T, Kraus N. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 2005;128:417-423.
- [80]. Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 2007;10:420-422.
- [81]. Worden FG, Marsh JT. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr Clin Neurophysiol* 1968;25:42-52.
- [82]. Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C. Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci U S A* 2005;102:14110-14115.
- [83]. Zurek PM. Detectability of transient and sinusoidal otoacoustic emissions. *Ear Hear* 1992;13:307-310.

Légendes des tableaux et figures

Tableau 1 et 1 bis

Revue méthodologique des PEASP. Détail des paramètres de stimulation et de recueil de réponses de type PEASP (ou speech ABR) chez l'humain, utilisés par les principales équipes:

Tableau 1

Colonne 1: Numérotation signalée entre crochets, se rapportant à la référence (Ref) de l'article. Classement chronologique. Colonne 2 : Sujets : caractéristiques et

nombre : A=adulte, E=enfants, ASD= enfant présentant un syndrome autistique, mu=musicien, mua=musicien amateur, Na=sans problème d'apprentissage, NE= normo-entendant, nmu= non musicien, Pa= problèmes d'apprentissage, SNu = surdité neurosensorielle unilatérale. Colonne 3: Placement des électrodes : C= cervical, C7= 7 ième vertèbre cervicale, cae= conduit auditif externe, cz= vertex, f= front, lob d= lobe droit, lob g=lobe gauche, (m)= masse, ma= mastoïde. Colonne 4: Caractéristiques du filtre (F) utilisé : oct=octave,

Tableau 1 bis:

Colonne 1 : Numérotation signalée entre crochets, se rapportant à la référence (Ref) de l'article. Colonne 2 : Caractéristiques du stimulus audiologique utilisé. - Type de stimulus utilisé: F0= fréquence fondamentale; -Type de présentation du stimulus au sujet: B=binaural, BB= bruit blanc, cae= conduit auditif externe, d= droit, g= gauche, I= image, M= monaural, OD= oreille droite, OG= oreille gauche, PA= polarité alternée, S=son à 40 dB SPL, Vi= vidéo. -Fenêtre d'enregistrement (FE): ms= milliseconde, -Durée du stimulus (D) utilisé en millisecondes, -Nombre de réponses moyennées (Mn) pour chaque enregistrement, -Fréquence de récurrence (FR) de la stimulation par seconde, -Intensité de la stimulation envoyée au sujet: □ = dB HL (hearing level), □ = dB SL (sensitive level), □ = dB SPL (sound pressure level).

Figures:

Figure 1: Système de recueil des PEASP (PEA en réponse à un son de parole).

L'ensemble des enregistrements est effectué dans une cage de Faraday reliée à la masse. Le stimulus est envoyé par l'ordinateur portable et la carte son externe via les inserts jusqu'au sujet dont l'activité électroencéphalographique est recueillie par des électrodes cutanées. Les électrodes cutanées sont placées sur le vertex (électrode positive), le front (masse) et au niveau de chaque mastoïde. L'enregistrement de cette activité est réalisé par un système de recueil de potentiels évoqués auditifs précoces (ici le « centor USB [®] »), avec un déclencheur (trigger) externe permettant la synchronisation du recueil sur le stimulus.

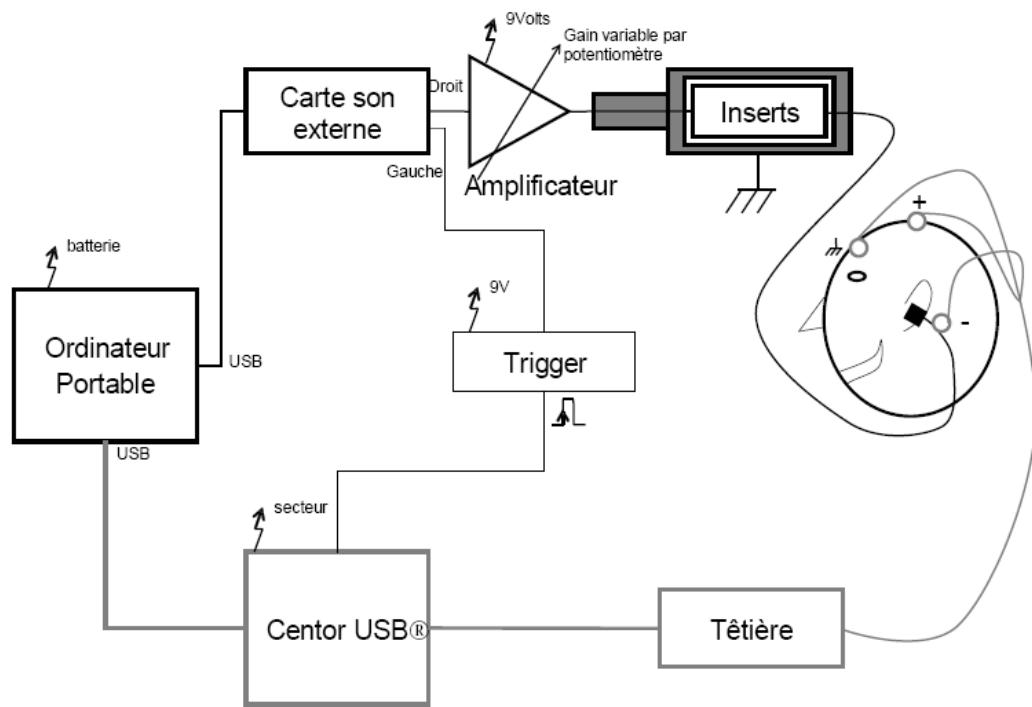
Figure 2: PEASP en réponse à un /ba/:

a: Représentations temporelles du stimulus /ba/ et du PEASP obtenu en réponse à ce stimulus: □ Représentation temporelle de l'enveloppe d'un stimulus /ba/ de 80 millisecondes ; □ Représentation de l'enveloppe retournée du même stimulus /ba/ ; ▲ L'enveloppe retournée est décalée de 9 millisecondes, ceci permettant d'observer une synchronicité avec la trace PEASP ; ■ Représentation temporelle de la réponse PEASP. Cette réponse peut être divisée en différentes parties (Figure 3 a, b): - la première partie est représentée par « l'Onset Response » (ondes V-A), réponse transitoire, encodant les variations temporelles rapides inhérentes à la consonne, similaire à la réponse observée à un click et apparaissant approximativement entre 7 et 9 ms après le début du stimulus. -la seconde partie: réponse soutenue à la fréquence, ou « Frequency Following Response » (FFR). La périodicité des pics de cette phase correspond à l'information fréquentielle contenue dans le stimulus [65]. La période entre les pics C, D, E, F de la FFR (activité de phase-locking) correspond à la fréquence fondamentale du stimulus (F0), tandis que les pics entre les ondes C, D, E, F représentent l'activité de calage de phase à la fréquence du premier formant F1 [6]. Une dernière onde apparaît ensuite: l'onde O qui correspond à l'arrêt du son, et temporellement à l'offset du stimulus [36].

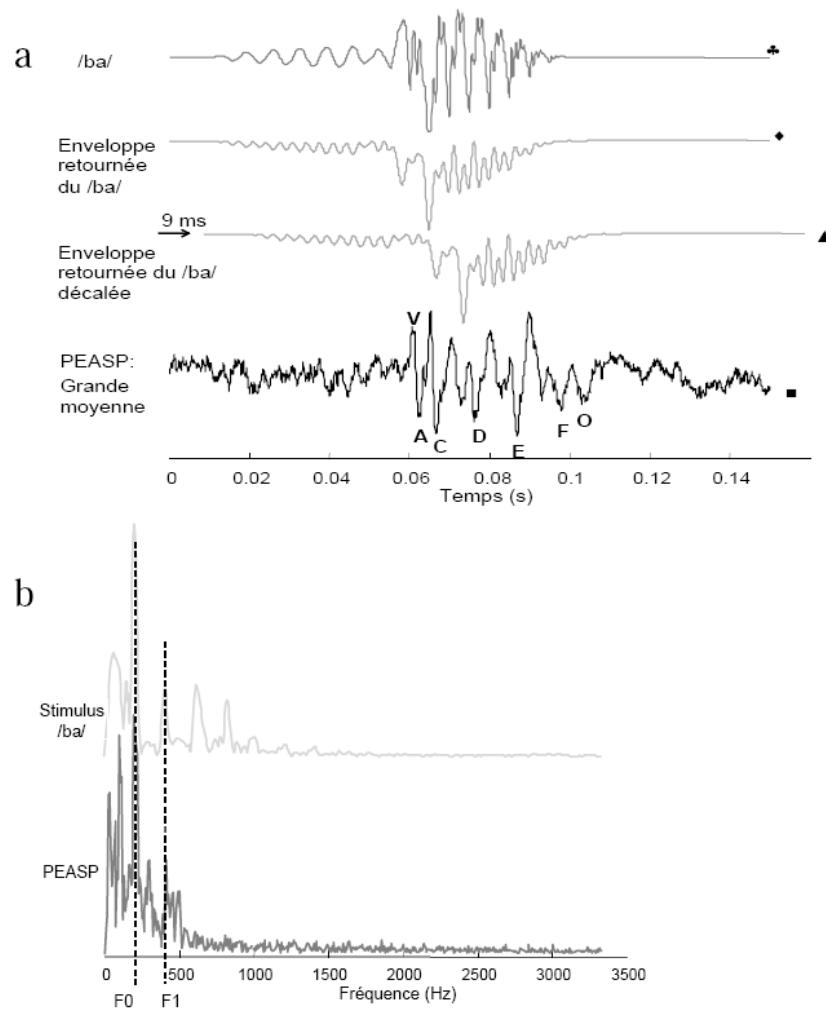
b: Représentation spectrale d'une réponse PEASP en réponse à un stimulus /ba/: Le stimulus /ba/ de 80 millisecondes est représenté en gris pâle. La réponse spectrale du PEASP est représentée en gris foncé. Cette représentation en fonction de la fréquence permet d'observer une corrélation spectrale au niveau de deux pics F0 et F1. Ces deux pics correspondent à la fréquence fondamentale et à la fréquence du premier formant de ce même stimulus.

Figure 3: Modélisation de l'effet des différents transducteurs sur l'artefact électromagnétique.

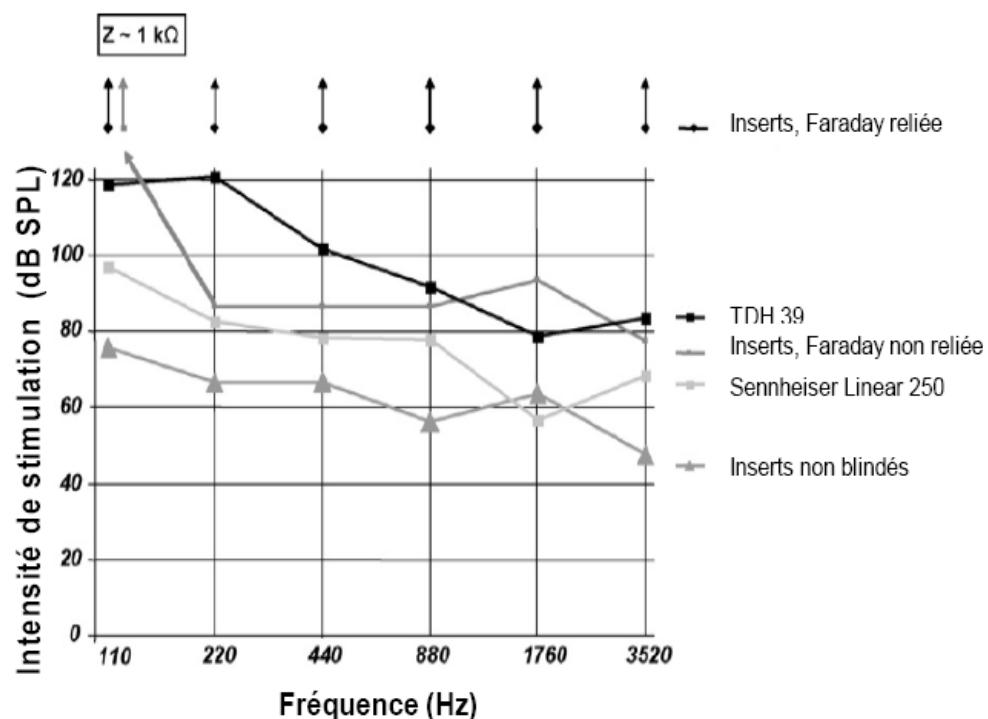
Seuil d'intensité (en dB SPL), en fonction de la fréquence (Hz), pour lequel un artefact électromagnétique est apparu, dans des conditions réelles ($Z=1\text{ k}\Omega$), pour différents types de transducteurs. La condition expérimentale « inserts non blindés » présente le plus d'artefacts quelle que soit la fréquence observée, alors que l'utilisation de ces mêmes inserts reliés à une cage de Faraday permet une diminution de l'artefact qui s'accentue de manière significative lorsque la cage de Faraday est reliée à la masse. Concernant l'utilisation des casques, le casque TDH 39 présente moins d'artefact que le Sennheiser Linear 250. Ainsi la condition expérimentale présentant le moins d'artefact est obtenue lors de l'utilisation d'inserts reliés à une cage de Faraday, elle-même reliée à la masse.



Richard et al., Figure 1



Richard et al., Figure 2 a, b



Richard et al., Figure 3

Tableau 1 :

Ref	SUJETS		PLACEMENT DES ELECTRODES	F
	TYPE	NOMBRE		
[12]	A	5	cz, lob (d+g), jambe(m)	0,2 - 3500 Hz ; 6dB/oct
[22]	A	5	cz, lob (d+g), jambe gauche (m)	400 - 4000 Hz ; 6dB/oct
[17]	A	NE-8 ; SNu-8	cz, lob (d+g), rotule g (m)	-
[8]	A+E	7	cz, lob (d+g)	3 - 10kHz ; 12dB/oct
[24]	A+E	12	cz, C postérieur, C latéral (m)	150-2800 Hz
[29]	A	11	cz, ma (d+g), f (m)	100-600 Hz
[19]	A	2	cz, lob (d+g), ma (d+g), f (m)	30-3000 Hz ; 6dB/oct
[20]	A	9	cz, lob (d+g), ma (d+g), f (m)	300-3000 Hz ; 6dB/oct
[42]	A	8	f, C7, ma g (m)	100-3000 Hz ; 6dB/oct
[21]	A	8	cae d et g, cz, f (m)	30-3000 Hz
[45]	A	8	f, C7, ma g (m)	100-3000 Hz ; 6dB/oct
[40]	E	Na-33 ; Pa-54	cz, lob droit, f (m)	100-2000 Hz
[78]	E	Na-9 ; Pa-11	cz, ma d, f (m)	100-2000Hz
[18]	A	11	cz, ma (d+g), f (m)	30-3000 Hz ; 6dB/oct
[65]	E	Na-5 ; Pa-14	cz, lob d, f (m)	100-2000Hz
[44]	A	13	cz, C7, f (m)	100-3000 Hz ; 6dB/oct
[79]	E	Na-9 ; Pa-11	cz, ma d, nez, f (m)	100-2000Hz
[7]	E	Na-46 ; Pa-74	cz, lob d, f (m)	100-2000Hz ; 6dB/oct
[66]	E	Na-10 ; Pa-9	cz, lob d, f (m)	100-2000Hz
[1]	E	Na-37 ; Pa-30	cz, lob d, f (m)	-
[73]	E	Na-115 ; Pa-119	cz, lob d, f (m)	100-2000Hz ; 6dB/oct
[56]	A	10	cz, nez, f (m)	75-2000HZ
[57]	A	nmu-13 ; mu-16	cz, lob, f (m)	-
[37]	E	Na-24 ; Pa-46	cz, lob d, f (m)	100-2000Hz
[80]	A	nmu-10 ; mu-10	cz, lob d, f (m)	-
[67]	E	Na-21 ; ASD-21	cz, lob, f (m)	80-1000Hz ; 12dB/oct
[35]	E	104	cz, lob d, f (m)	100-2000Hz
[34]	E	22	cz, lob ipsilatéral, f (m)	70-2000Hz
[74]	A	23	cz, lob d, f (m)	80-1000Hz ; 12dB/oct
[58]	A	nmu-12 ; mu-14	cz, lob d, f (m)	0,5-2000Hz
[72]	E	Na-90 ; Pa-93	cz, ma d, f (m)	100-2000Hz ; 6dB/oct
[4]	A	23	cz, ma (d+g), f (m)	2-4000Hz
[5]	A	NE-6 SNu-6	cz, ma (d+g), f (m)	2- 4000Hz
[46]	A	nmu-11 ; mu-10 ; mu-5	cz, lob (d+g), f (m)	20-2000Hz ; 12dB/oct
[28]	E	Na-7 ; Pa-42	cz, lob d, f (m)	100-2000Hz
[11]	E	Na-9 ; Pa-9	cz, lob d, f (m)	100-2000Hz
[31]	A	12	cz, lob (d+g), f (m)	100-2000Hz ; 12dB/oct

Tableau 1 :

STIMULUS AUDIOLOGIQUE							
	TYPE	TYPE DE PRÉSENTATION	FE (ms)	D (ms)	M n	FR(/s)	INTENSITE
[12]	sinusoïdes	B	-	18	999	4	70♦
[22]	tone bursts	B/ M d/ M g	62,5	16	999	1	60-80♦
[17]	tone bursts	B/ M d/ M g	180	14	1024	500	50 à 60♦
[8]	tone bursts/sons continus	M	-	14	1500-5000	-	95♦
[24]	signal composite	B	65,5	40	1000	9-15	53-70♦
[29]	tone bursts	M OG, OD + BB 65 dB SPL	55	52	3000	-	85♦
[19]	mots/sons purs	M OD	< 640	-	1000	2	77♦
[20]	/e/ (170Hz), /a/ (230Hz)	stimulus différent OD/OG	750	320	1000	-	76-84♣
[42]	tone bursts	M OD	100	80	3000	6,13	65-95♣
[21]	10 sons purs	M OG	105	75	2000	-	70♣
[45]	/u/, /i/, /a/	M, PA	80	60	2000	-	55-65-75-85♣
[40]	/da/	M OD, PA	100	40	3000	11,1	68-80♣
[78]	/da/	M OD, PA, Vi: I+S	80	40	6000	-	80♦
[18]	"chicken pot pie"	M OD	1100	880	1000	-	70♣
[65]	/da/	M OD, PA, Vi: I+S	70	40	3*1000	-	80♦
[44]	4 sons (mandarin, même F0)	M OD, PA	260	250	2000	3,13	60♣
[79]	/da/	M OD, PA, Vi: I+S	52	40	6000	-	80♦
[7]	/da/	M OD, PA, Vi: I+S	70	40	3*1000	-	80♦
[66]	/da/	M OD, PA, Vi: I+S	70	40	6000	-	80♦
[1]	/da/	M OD, PA, Vi: I+S	-	40	6000	-	80♦
[73]	/da/	M OD, PA, Vi: I+S	70	40	3*1000	-	80,3♦
[56]	/da/	B, PA	-	100	3*1000	-	84♦
[57]	/da/, sol de violoncelle	M OD, PA, cibles	-	350	3*2400	-	-
[37]	/da/	M OD, PA, Vi: I+S	70	40	3*2000	-	80♦
[80]	3 stimuli mandarin (différentes F0)	M OD, PA, Vi: I+S	-	-	-	-	-
[67]	/ya/ (interrogatif vs affirmatif)	M OD, PA, Vi: I+S	300	230	4800	-	60♦
[35]	/da/	M OD, PA, Vi: I+S	74,67	40	3*2000	-	80♦
[34]	/ga/ /da/ /ba/	M OD, PA, Vi: I+S	230	170	4000	4,35	83♦
[74]	3 sons mandarins (/ni/)	M OD, PA, Vi: I+S	340ms	278,5	2*1200	-	70♦
[58]	/da/	B, PA	-	350	12*600	-	84♦
[72]	/da/	M OD, PA, Vi: I+S	70	40	3*2000	-	80♦
[4]	/ba/ (intensité croissante)	M OD, PA	-	60	3000	11,1	0 à 60♣
[5]	Sons purs	Serum/melon	-	360	3000	-	65♦
[46]	accords musicaux	B, PA, Vi: I sous titrée	-	400	2*3000	-	70♦
[28]	/da/	M OD, PA, Vi: I+S	60	40	3*1000	-	80♦
[11]	/da/	M OD, PA, Vi: I+S	110	40	4000	-	80♦
[31]	/da/	M OD puis OG, PA, Vi: I	74,67	40	3*2000	-	90,3♦

Tableau 1 bis :

Characteristics of Speech ABR obtained at low stimulation levels.

C. Richard*, A. Jeanvoine*, H. Thai-Van*, A. Moulin*.

*Manuscrit en préparation.

Abstract

Speech auditory brainstem responses (SABR) are a very promising tool for investigating a wide range of speech perception and language impairments, as it represents a unique objective, non-invasive means of auditory brainstem timing exploration in humans. However, such SABR responses are very sensitive to parameters changes. The present study analysed systematic changes in both onset responses (OR) and Frequency Following Responses of SABR and click evoked ABR (CABR) as a function of stimulation intensity. The latency intensity functions showed a variation of 0.6 ms per 10 dB for CABR, more than 1 ms per 10 dB for OR, and between 0.7 and 1 ms per 10 dB for the different peaks of the FFR. The threshold allowing to record SABR in normally hearing subjects was 20 dB SL. At moderate intensities (20 to 30 dB SL), the representation of SABR was bilateral (with

a monaural stimulation), and the SABR response increased its amplitude linearly with the stimulation intensity for the stimulated side. The neural synchronisation on the fundamental frequency of the BA stimulus (F0) showed a non-monotonic behaviour as a function of stimulation intensity.

Introduction

Speech auditory brainstem responses (SABR) are a promising tool for specific language impairments (King et al., 2002; Song et al., 2008), speech in noise perception deficits (Anderson et al., 2010) or monitoring the effects of auditory training (Russo et al., 2005; Song et al., 2008). They are neurophysiological responses to natural syllables (such as /da/ (Skoe et al., 2010), or /ba/ (Akhoun et al., 2009), that are composed of two main parts: the onset response (waves V-A), and a sustained event: the frequency following response (FFR). The onset response is often likened to wave V of the click evoked ABR, and corresponds to the beginning portion of the consonant, whereas, the FFR is synchronized to the periodicity of the sound reflecting the neural phase-locking (Chandrasekaran et al., 2010). SABR represent therefore an objective, non-invasive technique of auditory temporal coding exploration in the brainstem in humans.

The majority of studies dealing with SABR use quite high stimulus intensities (60 to 85 dB SPL). This ensures good signal to noise ratio in the recorded SABR, and clear individual responses. In addition, in normally hearing subjects, such levels are high enough above the hearing thresholds, so that the inherent variability in hearing thresholds (within the normal hearing range) is unlikely to affect the SABR. On the other hand, it is possible that lower stimulus intensities might be more sensitive to minute changes in brainstem processing, and might be more relevant to day to day speech perception difficulties, that occur for moderate intensities. In addition, recording SABR in subjects with hearing losses can be a problem: increasing the stimuli levels can lead to artefactual responses. Indeed, due to the strong similarity between the SABR waveform, and the acoustic stimulus pattern, high levels (above 100 dB SPL) can lead to electromagnetic leakage and artefactual responses, even when using shielded insert earphones (Akhoun et al., 2008). In addition, for moderate hearing losses, 80 dB SPL stimulus levels can be as low as 30 dB SL. Characterising SABR pattern as a function of stimulus intensity seems a prerequisite before attempting to evaluate Speech Processing in hearing impaired individuals, if only to determine the minimum threshold likely to yield a speech ABR response. Although in most studies, normal-hearing subjects are defined as having a hearing threshold of 20 dB HL or less at octave frequencies between 250 and 8 kHz, a lot of differences, within this hearing threshold range, can be observed in more sensitive measures of cochlear function, such as otoacoustic emissions. Dhar et al. (2009) tried to correlate cochlear status, in normally hearing subjects, with speech ABR characteristics : they observed a correlation between the SABR harmonics (especially f1) and the strength of distortion acoustic otoacoustic emissions (DPOAEs), with greater harmonics in ears showing high levels DPOAEs, i.e. healthier ears. Even in normally hearing subjects, small differences in cochlear status can account with variations in the speech ABR response. It is therefore very likely, that speech ABR would be very sensitive to changes in stimulation intensity. The latencies of the FFR and OR as a function of stimulus intensity show a different relation, accounting for differences in generation mechanisms (Akhoun et al. 2008). Although the OR is often likened to the peak V of CABR, several differences between both responses have been pointed out. First, as the OR is elicited by more ecological stimuli (such as syllables) rather than a click, one can hypothesise that the OR would be more susceptible to minute differences in auditory processing capabilities than wave V of the ABR. This has been confirmed by several studies, pointing out latency differences in OR

in groups of children with learning disabilities, without any differences observed in wave V. Furthermore, a weak correlation has been observed between wave V latency and OR latency (Song et al., 2006). The aim of the present study is to systematically characterize speech ABR in response to a french /ba/ (both OR and FFR), as a function of stimulation intensity, and to compare it with click evoked ABR behavior in the same subjects.

Materials and methods

1. Subjects:

Ten native-French-speaking adults, ranging in age from 18 to 30 years, participated in the study. Audiometric pure-tone thresholds were not greater than 15 dB HL (Audiometer AC 40®) at half-octave frequencies from 250 Hz to 8000 Hz. All subjects were free of any otologic disease. Peripheral right laterality of each participant was assessed by the Edinburgh's test. This study was carried out in accordance with the French Huriet act pertaining to biomedical research, with the agreement of the local Ethics Committee on Biomedical Research of South-East region, France (CCPPRB # 05-024). Informed written consent was obtained for each participant.

2. Brainstem recording and stimuli:

2.a. Click Auditory Brainstem Responses (ABR):

The stimulus was a gated train of biphasic 80 μ s click, presented at a rate of 20 pulses per second. Stimulation was monaural on the right ear, with intensities ranging from 0 to 60 dB HL in a random order, and 2000 clicks were averaged for each trace. Stimulus presentation and recordings were performed in a sound proof room, using a Centor° USB system (Racia-Alvar®) ABR recording system. Click evoked ABR were recorded using four twined silver electrodes placed between vertex (+), left and right mastoid, with forehead as ground (Galbraith et al., 2000). Electrode impedances were maintained below 5 k Ω .

2.b. Speech Auditory Brainstem Responses:

Speech auditory brainstem responses were collected during the same session as the CABR, using the same setup as for the CABR recording, and the same electrode placement.

The stimulus was a 100 ms long natural /ba/syllable, with a voice onset time of 55 ms for the /b/. Individual thresholds for the BA stimulus were obtained from each subject using a transformed up/down procedure, with the psychoacoustics toolbox developed by M. Grassi (Grassi et al., 2009), using the same setup as for the SABR stimulation. The stimuli intensities ranged from 0 dB SL up to 50 dB SL by 10 dB steps, with an additional recording at 46 dB SL. All the different intensities were presented at random. For all subjects, stimuli were presented by a PC-based stimulus delivery system that output the signal through an external sound card (Roland®, UA4FX) which permits to control the timing and the intensity of stimulus delivery and also triggered the PC-based evoked potentials averaging system. Then, stimuli were delivered, with alternating polarity, monaurally through insert earphones (Etymotic Research ER-3A) to the right ear of the subject. The repetition rate used was 5Hz, and ten ms of silence was recorded before the beginning of the sound in order to assess the ambient noise level. The order of test conditions was randomized both within and across subjects.

Two thousand and four hundred epochs (alternating polarity) were averaged, with a 150 ms analysis window (sampling frequency =13.3kHz). Speech ABRs were filtered online by an 80–3200 Hz bandpass Finite Impulse Response filter (order 2). Each epoch with amplitude greater than 75 μ V was discarded. The ambient noise was assessed by

calculating root-means square (RMS) for the whole signal divided by RMS obtained for the silence period before averaging.

3. Data analysis:

The latencies of waves III and V of the CABR were visually obtained by two independent observers. Similarly, the characteristic peaks of the response to /ba/ (Onset response and the different peaks of the FFR) were manually identified first on the grand-average of traces and then on each individual traces, by two independent observers, both blind to the condition of presentation.

Analysis was carried out on both grand-average curves (algebraic summation of the individual data on alternate polarity) and individual traces. RMS values were obtained for the FFR part, in order to assess the influence of stimulation intensity. Fast Fourier Transform was calculated on the FFR part, on a 30 ms window, in order to assess the amplitude of the fundamental frequency representation on the FFR (F0). The FFR was carried out on the full traces using a sliding window, with 1 ms lag, in order to characterize the synchronisation of F0 as a function of the stimulus intensity, and as a function of time. In addition, to see if there was a difference between the beginning of the FFR and the end, we analyzed the FFR in two 15ms bins, termed the first part and end part respectively.

Results

1. Latencies:

Fig. 1a,b shows the stimulus used and all the traces obtained for stimulus intensities ranging from 50 dB SL down to 0 dB SL. The onset response, and the different waves of the FFR have been labeled.

The OR latencies increased significantly from 6.3 ms at 50 dB SL to 9 ms at 30 dB SL, as did the cABR wave V, rising from 7.3 ms at 50 dB SL to 8.85 at 30 dB SL. Fig. 2 compares cABR latencies with SABR OR latencies and SABR FFR latencies in the same subjects, as a function of stimulus intensity in dB SL.

The slope of onset response as a function of stimulus intensity was calculated for each subject for intensities ranging from 30 dB SL to 50 dB SL. The average slope was -1.03 ms for 10 dB increase in intensity (sem=0.48). This was performed on 9 subjects (subject 7 was eliminated because of no visible onset response at 30 and 40 dB SL). For the same subjects, slopes of CABR functions were calculated, with an average of 0.57 ms increase (sem=0.04) per 10 dB decrease (fig. 2). No significant differences were obtained between the speechABR onset response slope and cABR slope with wave V and wave III, due to significantly greater variability in SABR latencies versus CABR latencies. No significant correlations were obtained between speechABR OR slope and cABR slope as function of stimulation intensity.

FFR latencies were obtained for each peak, labeled from B to J, and decreased significantly with increasing stimulus intensity. The slopes tended to get lower for the later waves: it ranged from 1.24 ms decrease per 10 dB increase (wave C) to 0.73 ms decrease per 10 dB increase (wave G), but no significant difference has been found.

2. FFR characteristics analysis

2a. RMS values of the FFR

FFR amplitude (RMS values) increased significantly with intensity, especially for intensities above 30 dB SL. A linear model was fitted, and revealed a highly significant effect of the intensity ($F(1,140)=51.8$, $p<0.0001$), with an interaction with the channel

($F(1,140)=7.1$, $p<0.01$). Similar highly significant effects of stimulation intensity on the FFR RMS amplitude was observed for the first and last part of the FFR. The difference between the right and left channels increased with stimulation intensity (anova, $F(1,114)=19.9$, $p<0.0001$). (fig. 3). However, this difference didn't change significantly whether only the first part of the FFR or the last part of the FFR was considered. The right channel revealed significantly greater FFR for stimulation intensities above 40 dB SL, whereas left channel showed a constant FFR across intensities from 30 dB SL.

2b. Fundamental frequency of the FFR.

Spectral analysis of the FFR showed a clear F0 peak at 205 Hz, that increased significantly with stimulation intensity $F(1,114)=49.7$, $p<0.0001$, with a significant interaction with the channel involved ($F(1,114)=13.3$, $p<0.0005$) (fig. 4). The right channel showed an almost linear increase of f0 with stimulation intensity, whereas left channel showed a fairly constant level from 20 dB SL. There was a very strong interaction between intensity and channel ($F(1,236)=18.5$, $p<0.0001$).

A significant differential effect of stimulation intensity on F0 was revealed between the first and the last part of the FFR : $F(1,236)=5.4$, $p<0.03$) with the end part being significantly greater than the beginning part (fig. 5), especially for mid-intensity range (20 to 30 dB SL).

2c. Time/Frequency structure of the FFR

As expected, spectrograms of the full traces revealed the greatest amplitude for a frequency band of 200-232 Hz (corresponding to the fundamental frequency F0 of the /BA/ stimulus), and for a time slice corresponding to the FFR. Analysis of the frequency band over time, for different stimulation intensities, showed different behavior depending on intensity. (fig. 6a). The maximum amplitude value over the frequency band increased significantly with intensity ($F(1,140)=44.5$, $p<0.0001$), especially for intensities above 40 dB SL, with a significant interaction with the channels: the left channel showed a fairly constant maximum value over the intensities over 20 dB SL (fig. 6b). The surface under the curve revealed similar results (fig. 6c). The Q value, calculated as the ratio of the maximum value over the duration of the FFR, showed a significant increase with intensity, suggesting that the duration of the frequency band corresponding to F0 doesn't increase with intensity: the amplitude of the frequency band alone increases.

Discussion

Grand-average data show a clear SABR for stimulus levels as low as 20 dB SL, but the occurrence of individual responses starts usually at 40 dB SL. Contrary to expectation, the left channel showed a remaining SABR response at 10 dB SL and 0 dB SL, whereas the same levels didn't show any obvious response on the right channel. Similarly, analysis of energy of the FFR as a function of stimulus levels, showed an almost linear increase of the FFR amplitude with stimulus levels for the right channel. At moderate stimulus levels (20 to 30 dB SL), both right and left channels showed a similar response, whereas the stimuli were strictly monaural. As stimulus intensity went beyond 30 dB SL, the FFR amplitude recorded with the right channel continued to increase, whereas the FFR recorded with the left channel remained constant, ruling out the effect of a transcranial transfer of the stimuli that could occur for 50 dB SL. Besides, the levels of the stimuli used (maximum of 50 dB SL), and the fact that we used insert ear-phones, allows to rule out the hypothesis of a transcranial transfer from the right to the left ear. Frequency analysis of the FFR, and the amplitude of F0, showed a similar pattern, with a bilateral representation of F0 for moderate stimulus levels (20 to 30 dB SL), and a mostly unilateral representation of F0 for higher levels. It seems that auditory temporal processing is optimized at moderate stimulus levels, by recruiting both left

and right auditory brainstem pathways even with a strictly monaural stimulation. For higher levels, such as those used in most SABR studies, the side of the stimulation is privileged, whereas the contralateral side shows some constant response, which is still greater than the response obtained at 0 dB SL.

Temporal analysis of the FFR showed a different behavior of the FFR with stimulus intensity, whether the first part or the last part of the FFR is considered: for moderate stimulus levels (20 to 30 dB SL), the neural synchrony on the fundamental frequency of the /BA/ is weaker on the first part than on the last part of the FFR, where at greater stimulus levels, no difference is obtained. It is possible that, at moderate stimulus levels, the auditory pathways require some more time to fully synchronize on F0. Investigating in a more precise manner, the timing of the F0 synchronization along the FFR, at moderate stimulus level, could perhaps yield some explanation on differences in brainstem auditory timing deficits in specific language impairments. The pattern of the time/frequency analysis of the encoding of the frequency band encompassing F0, showed a non-monotonic behavior with stimulus intensity, with 3 parts: a first part, with very little response (from 0 to 10 dB SL), a part with moderate response (20 to 40 dB SL), and a final part with higher responses (46 and 50 dB SL). Moderate stimulus levels (20 and 30 dB SL) showed greater F0 amplitude at mid-FFR, whereas, as stimulation intensity increases, the pattern widens towards both earlier and later parts of the FFR. This extension of the synchronization along the FFR with stimulus levels, is accompanied by an increase in the strength of synchronization, so that the selectivity, in time, of the synchronization, is greater at high stimulus levels.

Due to those different thresholds, (i.e. 20 dB SL to obtain a SABR response, 46 dB SL to obtain a “fully formed” SABR), individual thresholds will have to be taken into account, especially when investigating subjects with moderate hearing losses. Furthermore, as a link between sensitive measures of the cochlea status (by means of otoacoustic emissions) and SABR has been shown (Dhar et al., 2009), it is likely that cochlea status has a substantial influence on SABR recorded with moderate stimulus levels.

In addition to different patterns of FFR according to stimulus levels, the present study shows a definite difference between cABR wave V and OR modifications as a function of stimulation intensity, with a variation of more than 1 ms per 10 dB for the OR, versus only 0.57 for the cABR. The different peaks of the FFR show different latencies, with a slope that tends to decrease for the latest peaks. As pointed out by other several studies for a fixed stimulation level, the SABR onset response cannot be considered like a CABR wave V, and is more prone to changes with stimulus intensity than CABR. This is another argument for carefully monitoring at which stimulus level SABR should be recorded. SABR seems to be generally more sensitive to parameters changes than CABR: for instance, the stimulus rate is known to affect more SABR than CABR (Krizman et al., 2010). Although SABR is a very promising technique for exploration of brainstem auditory timing, much work is needed for carefully validating a set of parameters, across different languages.

References:

- Akhoun I, Gallego S, Moulin A et al (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin Neurophysiol*;119:922-933.
- Akhoun I, Moulin A, Jeanvoine A et al. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: an experimental parametric study (2008). *J Neurosci Methods*;175:196-205.

- Anderson S, Skoe E, Chandrasekaran B, Kraus N. Neural timing is linked to speech perception in noise (2010). *J Neurosci*;30:4922-4926.
- Chandrasekaran B, Kraus N. The scalp-recorded brainstem response to speech: neural origins and plasticity (2010). *Psychophysiology*;47:236-246.
- Galbraith GC, Threadgill MR, Hemsley Jet al. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci Lett* 2000;292:123-127.
- Grassi M, Sorenzo A, Borgo M (2010) Psychoacoustics: a comprehensive MATLAB toolbox for auditory testing. *Behavior Research Methods*.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proc Natl Acad Sci USA*. Aug 4;106(31):13022-7.
- King C, Warrier CM, Hayes E, Kraus N, (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neurosci Lett*;319:111-115.
- Kraus N, McGee TJ, Koch DB. Speech sound representation, perception, and plasticity: a neurophysiologic perspective. *Audiol Neurotol* 1998;3:168-182.
- Krizman J, Skoe E, Kraus N (2010) Stimulus Rate and Subcortical Auditory Processing of Speech. *Audiology and Neurotology*, 15, 332-342.
- Moushegian G, Rupert AL, Stillman RD. Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr Clin Neurophysiol* 1973;35:665-667.
- Russo N, Nicol T, Musacchia G, Kraus N, (2004). Brainstem responses to speech syllables. *Clin Neurophysiol*;115:2021-2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N. (2005) Auditory training improves neural timing in the human brainstem. *Behav Brain Res.* ; 156(1):95-103.
- Russo NM, Skoe E, Trommer Bet al, (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clin Neurophysiol*;119:1720-1731.
- Sinnott JM, Adams FS. Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *J Acoust Soc Am* 1987;82:1539-1547.
- Skoe E, Kraus N, (2010). Auditory Brain Stem Response to complex sounds: a tutorial. *Ear & Hearing*, vol. 31, NO.3.
- Song JH, Banai K, Russo NM, Kraus N, (2006). On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurotol*;11:233-241.
- Song JH, Skoe E, Wong PC, Kraus N, (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci*;20:1892-1902.
- Steinschneider M, Schroeder CE, Arezzo JC, Vaughan HG, Jr. Speech-evoked activity in primary auditory cortex: effects of voice onset time. *Electroencephalogr Clin Neurophysiol* 1994;92:30-43.

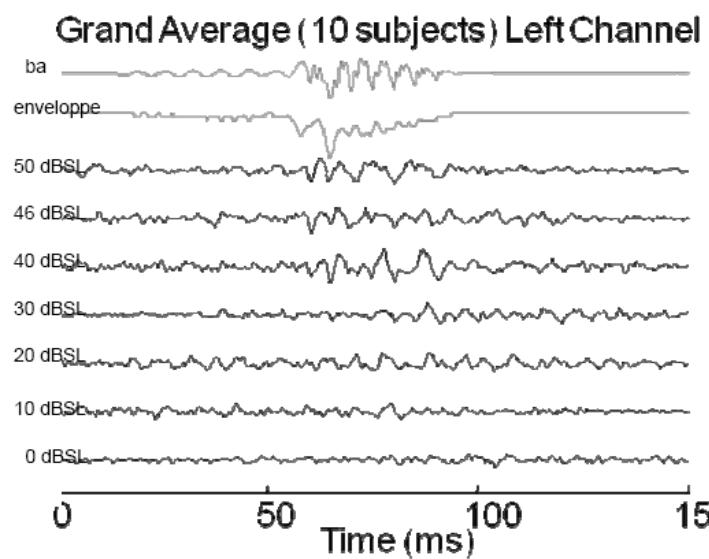
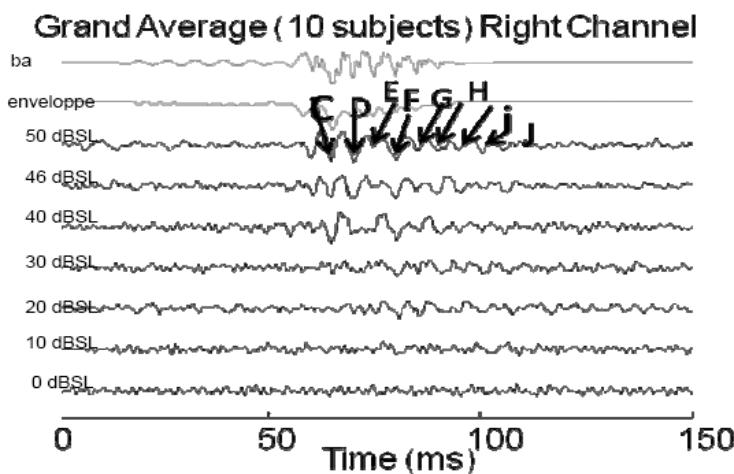


Fig. 1a and b: Grand-average traces of speech ABR recorded in 10 normally hearing subjects. The two top traces represent the /ba/ stimulus and its envelope. The SABR are ordered by stimulation intensities, from top to bottom. The stimulation was monaural (right ear), and the traces have been recorded on the right channel (fig. 1a) and on the left channel (fig. 1b).

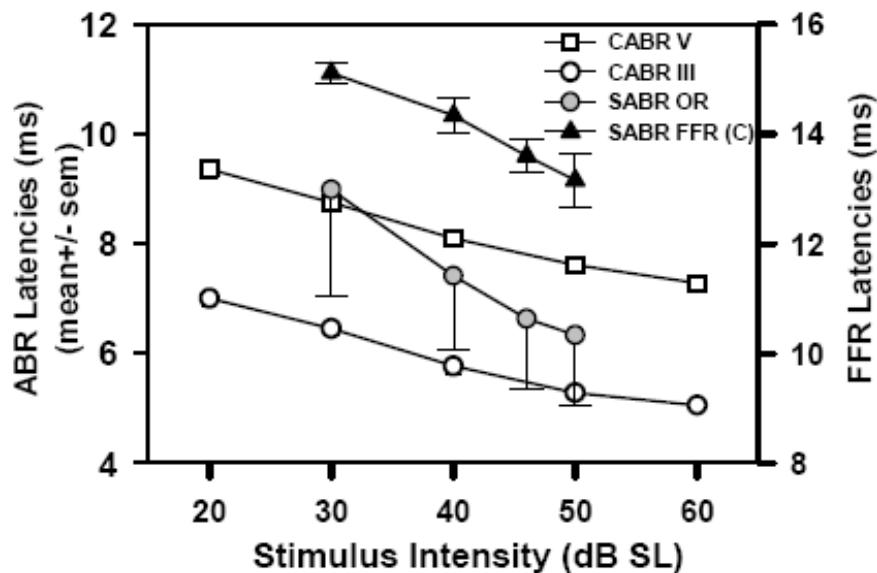


Fig. 2 : Latencies of Auditory brainstem responses (ABR) in ms as a function of stimulation intensity (dB SL). The white symbols show click evoked ABR (CABR) latencies, with squares for wave V and dots for wave III. Grey and black symbols are for speech ABR latencies, for Onset response and wave C of the FFR respectively.

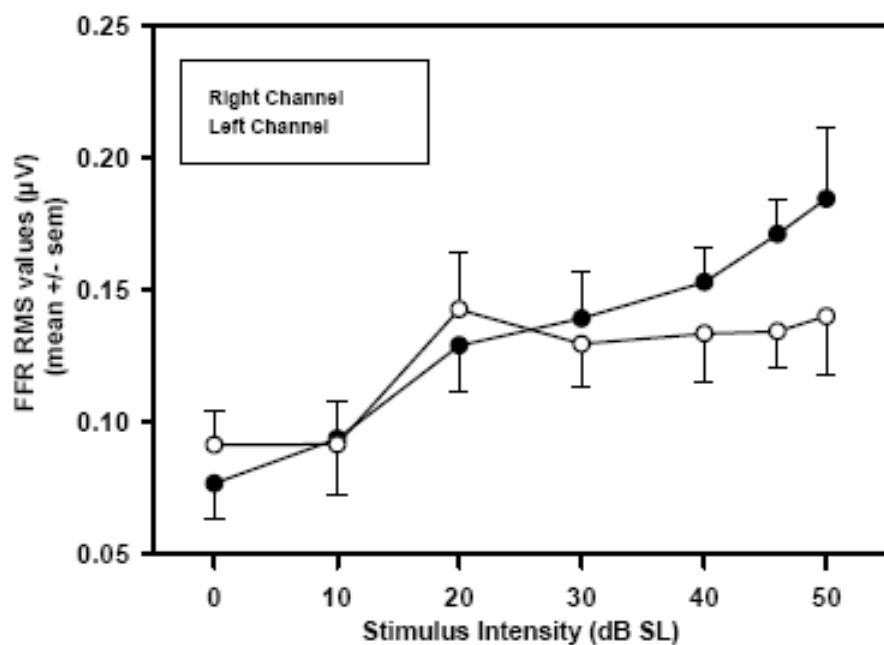


Fig. 3 Root Mean Square value (μ V) of the Frequency Following Response (FFR) as a function of stimulus intensity (dB SL), in response of a stimulus /BA/ of

120 ms duration. The stimulus was applied in the right ear, and the speech ABR was recorded using the right channel (black dots), and the left channel (white dots).

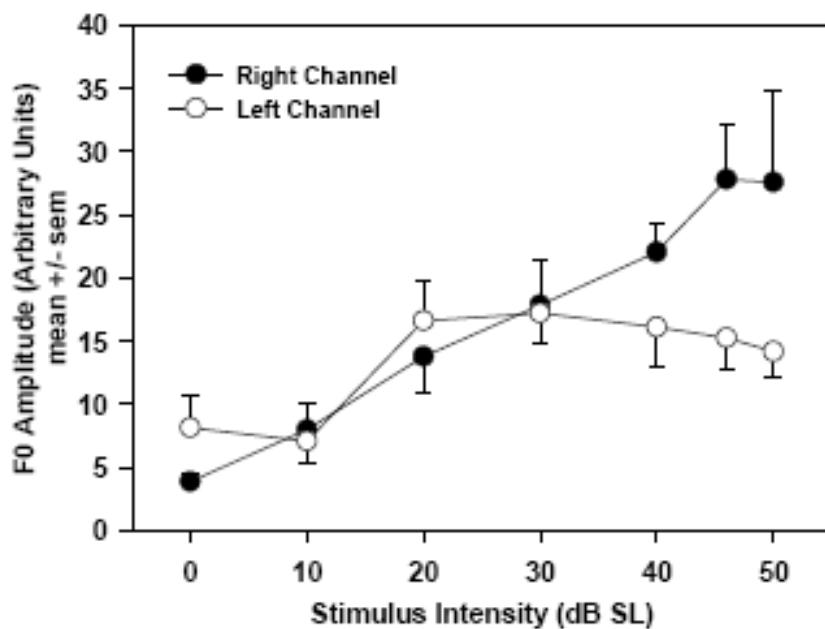


Fig.4: F0 peak amplitude (on the FFT of the FFR), corresponding to the fundamental frequency of the /ba/ stimulus, as a function of stimulus intensity in dB SL. The stimulus was applied in the right ear, and the speech ABR was recorded using the right channel (black dots), and the left channel (white dots).

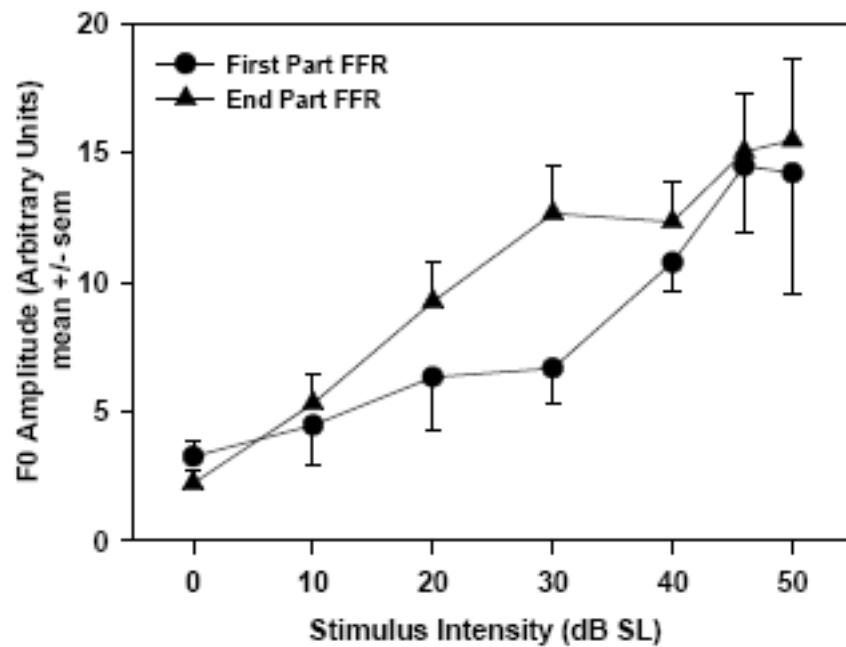


Fig.5: F0 peak amplitude (on the FFT of the FFR), corresponding to the fundamental frequency of the /ba/ stimulus, as a function of stimulus intensity in dB SL. The stimulus was applied monaurally to the right ear. F0 amplitude was calculated on the first part of the FFR (dots) and on the second part of the FFR (triangle), with statistically significant differences for mid-intensity range.

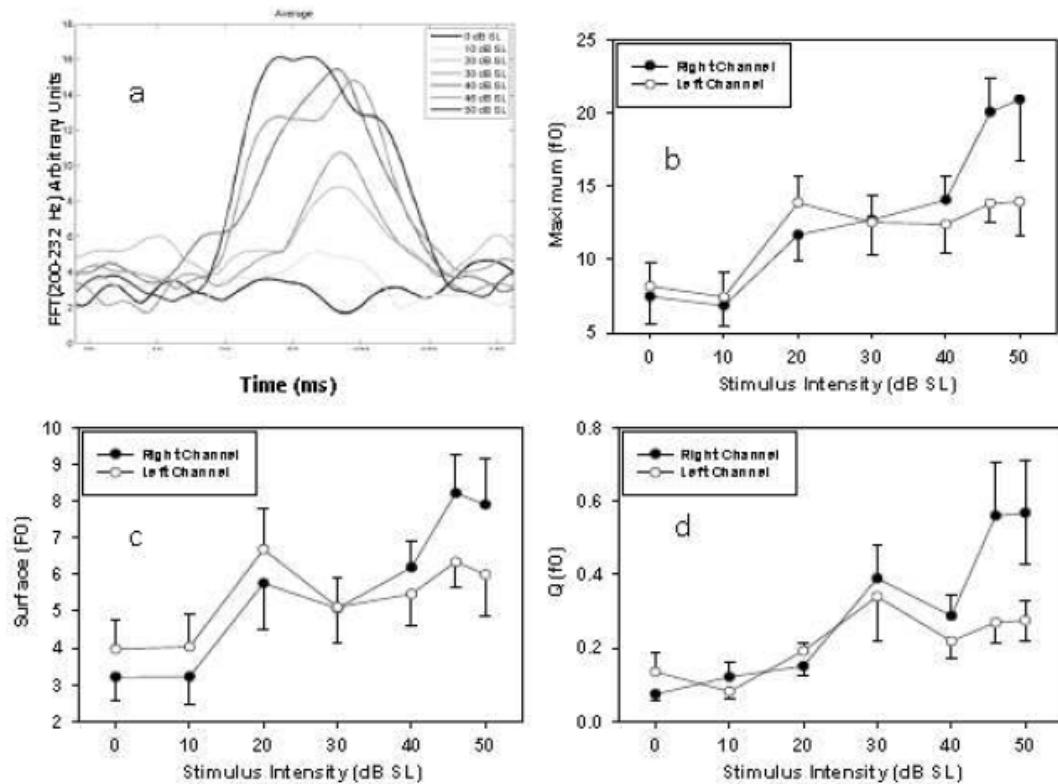


Fig. 6: Panel (a) shows the amplitude of the 200-232 Hz frequency band as a function of time, for the different stimulus intensities. Panel (b) shows the maximum of the function described in a, as a function of stimulus intensity, whereas panel (c) shows the surface under the curves shown in panel (a). Panel (d) shows the “Q values” obtained from the curves in panel (a), and calculated as the ratio of the maximum value divided by the duration over which the frequency band was greater than 50% of the maximum value.