



**HAL**  
open science

# ANALYSE DE L'ENCODAGE DES SIGNAUX DE PAROLE DANS LE TRONC CEREBRAL (SPEECH AUDITORY BRAINSTEM RESPONSE)

Idrick Akhoun

► **To cite this version:**

Idrick Akhoun. ANALYSE DE L'ENCODAGE DES SIGNAUX DE PAROLE DANS LE TRONC CEREBRAL (SPEECH AUDITORY BRAINSTEM RESPONSE). Ingénierie biomédicale. Université Claude Bernard - Lyon I, 2008. Français. NNT: . tel-00568938

**HAL Id: tel-00568938**

**<https://theses.hal.science/tel-00568938>**

Submitted on 24 Feb 2011

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

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

THESE DE L'UNIVERSITE DE LYON - délivrée par  
L'UNIVERSITE CLAUDE BERNARD LYON 1  
ECOLE DOCTORALE ELECTRONIQUE,  
ELECTROTECHNIQUE ET AUTOMATIQUE  
DE LYON

## Diplôme de Doctorat

Arrêté du 7 Août 2006

N° d'ordre : 270 - 2008  
2008

### ANALYSE DE L'ENCODAGE DES SIGNAUX DE PAROLE DANS LE TRONC CEREBRAL (SPEECH AUDITORY BRAINSTEM RESPONSE)

Thèse dirigée par Christian BERGER-VACHON

préparée au Laboratoire Neurosciences. CNRS UMR 5020

soutenue publiquement par Idrick AKHOUN le 3 décembre 2008

#### Jury :

<i>Rapporteurs :</i>	Jean-Luc SCHWARTZ	- CNRS (GIPSA)
	Paul AVAN	- Université de Clermont-Ferrand 1
<i>Président du Jury :</i>	Christian BERGER-VACHON	- LNSCC
<i>Co-encadrant :</i>	Hung THAI-VAN	- LNSCC
<i>Examineurs :</i>	Sandrine CABROL	- Cochlear France
	Eric TRUY	- LNSCC



## Remerciements

Un immense merci au laboratoire neurosciences sensorielles, comportement et cognition, l'UCBL (Christian Berger-Vachon, Lionel Collet) ainsi qu'à Cochlear France (Sandrine Cabrol), et au CIES de Lyon (Mohand Hacid) pour la mise en œuvre de la convention CIFRE et du monitorat.

Merci à Cochlear France, à la SFA (Société Française d'Acoustique), à l'ASA (Acoustical Society of America), ainsi qu'au laboratoire pour m'avoir soutenu aussi généreusement pour que je puisse participer à tous ces congrès, et rendre visite au laboratoire BrainVolts (Kraus et al.) à Chicago.

Merci à tous ceux qui dans l'équipe m'ont aidé dans mon travail. Les collègues impliqués dans les études Speech ABR sont les suivants : Lionel Collet (chef d'équipe), Hung Thai-Van (chef d'équipe), Stéphane Gallégo, Arnaud Jeanvoine (stagiaire L3, M1, M2), Mikaël Dude Ménard, Jeanne Guiraud, Fabien Seldran, Julien Hoarau (office-mates), Annie Moulin, William Darancy (stagiaire M2 GBM, puis Audioprothèse), Céline Richard (stagiaire M2 Sces Co, thésarde), Clémence Pavillon (stagiaire M2 Sces Co), Thomas Bridonneau et Clément Barral (stagiaire dentaire), Sarah Benzidane et Elisabeth Ambadiang (stagiaires M1-GBM), Rémi Algore (stagiaire L3 GBM), Arnaud Norena et Zak Khayli (stagiaire M1 Neurosciences), Evelyne Veuillet, Eric Truy, Johan Poncelet (stagiaire M2 Neurosciences) et Michel Hoen. Merci pour l'expertise extérieure : François Buret, Christian Vollaire, Riccardo Scorretti, Noel Burais et Guy Clerc (Ampère). Frédéric Marmel, Barbara Tillmann, Etienne Gaudrain, Nicolas Grimault, Samuel Garcia, Marc Thevenet, Belkacem Messaoudi, Michel Vigouroux, Bernard Bertrand, Michel Chaput (LNSCC), Nina Kraus (BrainVolts), Arnaud Norena et Yves Cazals (Marseille). Jean-Luc Schwartz et Paul Avan, en tant que rapporteurs, ont été d'une très grande aide pour extraire l'essence de ce travail. En tant que collègues, ils ont prodigué des idées et des conseils très précieux sur les perspectives des Speech ABR. Pour la mise en page [L<sup>A</sup>T<sub>E</sub>X](#) : ce modèle de thèse a été utilisé. Merci à Colette McKay (Manchester) et son équipe pour l'accueil, la confiance retrouvée, et les idées sur ces études.

La famille : Chantal, Abdul-Hacq, Nanou, Maryam, Khalil, Papou, Mémé Marie, l'oncle Kukanya.

Les amis : Jean-André et Alexandra, Brigitte, JL, Fabien, Nico H, Caro Robert, Michel Guillaud, Mouhand, Maryse, Philippe Litaudon, Pierre Breuvert, Didier Derks, Galab et Fanfan, Les étudiants d'audioprothèse, ceux de GBM. Andreas, Jean-Luc Peters, Max Faure, Romu Flav, Rafael Ejido, Val Maertens, Greg Bruyère, Kai-Guo Yue, Pierre Bertrand, Paul, Sharon, Lily et Kailyn Hepburn.

Ce travail est dédié à Alengo Nyamay'Antu, ainsi qu'à Seb et Jean-Louis Hervouet.





# Table des matières

<b>1</b>	<b>Auditory Brainstem Response (ABR) : Contexte scientifique et analyse de l'existant.</b>	<b>1</b>
1.1	Bases neurales des ABR . . . . .	1
1.1.1	Principales étapes du traitement des sons . . . . .	2
1.1.2	Aspects du codage neural. . . . .	6
1.1.3	Modélisation de l'encodage des sons. . . . .	7
1.1.4	Perception psychoacoustique des sons. . . . .	10
1.2	Présentation générale de l'implant cochléaire : . . . . .	14
1.2.1	Approche technique : . . . . .	16
1.2.2	Les processeurs vocaux : . . . . .	18
1.2.3	Le réglage de l'implant . . . . .	19
1.3	ABR : principe et historique . . . . .	21
1.4	ABR : Analyse de l'existant . . . . .	23
1.5	ABR : Avancées et Zones d'ombres : . . . . .	38
1.6	Objectif du travail . . . . .	45
<b>2</b>	<b>Etude expérimentale : authentification du Speech ABR.</b>	<b>47</b>
2.1	Dispositif expérimental de recueil des Speech ABR . . . . .	48
2.1.1	Chaîne de stimulation : envoyer une stimulation acoustique . . . . .	48
2.1.2	Chaîne de recueil : enregistrer l'activité électrophysiologique . . . . .	50
2.1.3	Trigger : synchroniser le recueil sur la stimulation . . . . .	51
2.1.4	Principales méthodes d'analyse des signaux . . . . .	51
2.2	Authentification des Speech ABR 1. Etude paramétrique de quantification de l'artéfact . . . . .	52
2.2.1	Introduction . . . . .	52
2.2.2	Matériel et Méthodes . . . . .	54
2.2.3	Résultats . . . . .	55
2.3	Authentification des Speech ABR 2. Etude comparative chez cophotiques unilatéraux . . . . .	63
2.3.1	Introduction . . . . .	63
2.3.2	Matériel et Méthodes . . . . .	63
2.3.3	Résultats . . . . .	63
2.4	Bilan : précautions à prendre et indices d'authenticité du Speech ABR . . . . .	68
2.5	Article Journal of Neuroscience Methods [Akhoun 2008d] . . . . .	69
<b>3</b>	<b>Influences acoustiques sur les ABR.</b>	<b>81</b>
3.1	Etude pilote : Speech ABR à /ba/ . . . . .	82
3.1.1	Matériel et méthodes . . . . .	82
3.1.2	Résultats : . . . . .	84

3.2	Effets comparés de l'enveloppe et de la structure fine sur le Speech- (et Analogue-) ABR . . . . .	87
3.2.1	Parole vs Analogues-en-enveloppe-temporelle, inharmoniques et spectralement appauvris 1. . . . .	87
3.2.2	Parole vs Analogues-en-enveloppe-temporelle, inharmoniques et spectralement appauvris 2. Variation sur les analogues. . .	104
3.2.3	Parole vs Analogues-en-structure-harmonique, et spectrale- ment riches. . . . .	117
3.3	Effet de la structure harmonique : sons purs vs sons complexes har- moniques. . . . .	120
3.3.1	Matériel et Méthodes . . . . .	120
3.3.2	Résultats . . . . .	120
3.3.3	Discussion . . . . .	122
3.4	Effet de la nature du stimulus et de la hauteur tonale sur le Speech-, Music-, Analogue- ABR . . . . .	123
3.4.1	Objectifs : . . . . .	123
3.4.2	Matériel et méthodes : . . . . .	123
3.4.3	Résultats : . . . . .	127
3.5	Effet du niveau sonore sur le Speech ABR . . . . .	136
3.5.1	Matériel et méthodes : . . . . .	136
3.5.2	Résultats : . . . . .	136
3.6	Bilan : indices acoustiques influençant les ABR . . . . .	139
3.7	Article Clinial Neurophysiology [Akhoun 2008c] . . . . .	140
3.8	Article Proceedings of Meetings on Acoustics [Akhoun 2008a] . . . .	141
<b>4</b>	<b>Discussion générale</b>	<b>161</b>
<b>5</b>	<b>Conclusion générale</b>	<b>171</b>
<b>A</b>	<b>Calibration de la stimulation sonore avec les inserts-earphones</b>	<b>173</b>
<b>B</b>	<b>Modélisation de l'activité neurale par AIM</b>	<b>175</b>
<b>C</b>	<b>Speech ABR et analyse temps-fréquence. Effet jitter.</b>	<b>179</b>
C.1	Annulation artificielle de la FFR (effet jitter) . . . . .	179
C.2	Etude du compromis moyennage / cartes temps fréquences instantanées.	179
	<b>Bibliographie</b>	<b>187</b>

# Auditory Brainstem Response (ABR) : Contexte scientifique et analyse de l'existant.

---

## Sommaire

---

<b>1.1 Bases neurales des ABR</b> . . . . .	<b>1</b>
1.1.1 Principales étapes du traitement des sons . . . . .	2
1.1.2 Aspects du codage neural. . . . .	6
1.1.3 Modélisation de l'encodage des sons. . . . .	7
1.1.4 Perception psychoacoustique des sons. . . . .	10
<b>1.2 Présentation générale de l'implant cochléaire :</b> . . . . .	<b>14</b>
1.2.1 Approche technique : . . . . .	16
1.2.2 Les processeurs vocaux : . . . . .	18
1.2.3 Le réglage de l'implant . . . . .	19
<b>1.3 ABR : principe et historique</b> . . . . .	<b>21</b>
<b>1.4 ABR : Analyse de l'existant</b> . . . . .	<b>23</b>
<b>1.5 ABR : Avancées et Zones d'ombres :</b> . . . . .	<b>38</b>
<b>1.6 Objectif du travail</b> . . . . .	<b>45</b>

---

## 1.1 Bases neurales des ABR

Les mécanismes physiologiques qui entrent en jeu dans le codage d'un son sont globalement très complexes, et ce code neural conserve encore pas mal de zones d'ombres. L'onde acoustique perçue par l'oreille est transformée en message nerveux, qui est traité au fur et à mesure dans le système auditif de la cochlée jusqu'au cortex auditif. Les noyaux successifs du tronc cérébral sont constitués de cellules anatomiquement et fonctionnellement très distinctes : on suppose que chacune joue un rôle très particulier dans le codage des sons (développé ci-dessous et en annexe à partir des 3 revues suivantes : [Joris 2004], [Palmer 2004], [Eggermont 2001]). Dans tout le système auditif, deux aspects principaux du code neural se distinguent : le codage spatial et le code temporel.

### 1.1.1 Principales étapes du traitement des sons

Les 4 grandes étapes de l'encodage d'un son dans le système auditif sont les suivants :

- *Etape 1 : Entrée de la vibration acoustique dans l'oreille.*

Après avoir été amplifiée par (filtrée dans) le conduit auditif externe, l'onde sonore fait vibrer le tympan. Le tympan est relié à la cochlée via la chaîne des osselets qui constitue l'oreille moyenne, et la vibration du tympan se traduit par un mouvement de piston de la platine de l'étrier contre la paroi de la fenêtre ovale de la cochlée, Fig. 1.1. Notons au passage que la vibration passe d'un milieu aérien à un milieu liquidien : l'amplification apportée par les osselets correspond à l'adaptation d'impédance (facteur 20) entre ces deux milieux.

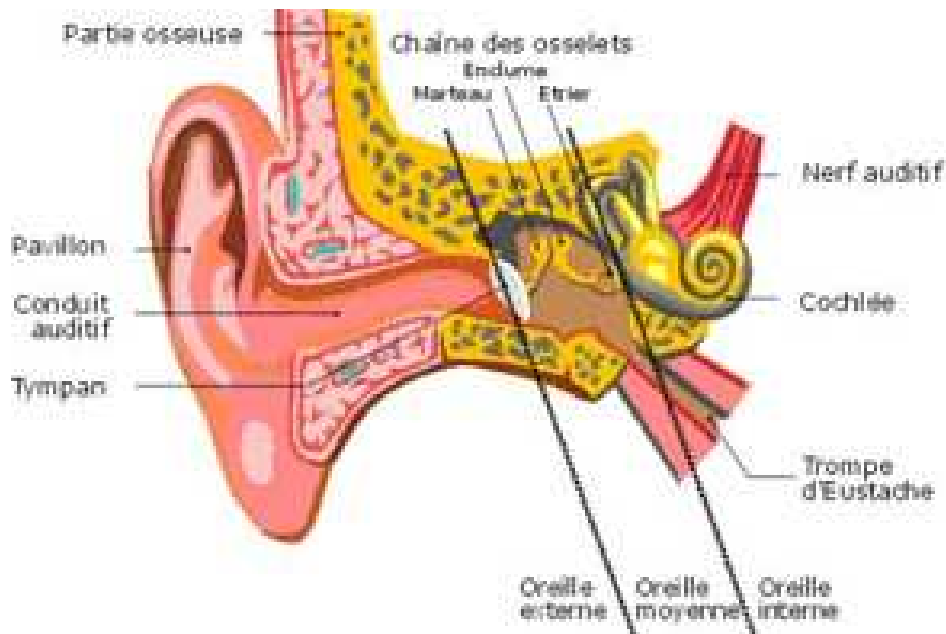


FIGURE 1.1 – **Coupe anatomique des 3 oreilles.** Le son est tout d'abord pré-filtré dans le conduit de l'oreille externe, avant de provoquer une vibration liquidienne dans la cochlée par l'intermédiaire de la chaîne d'osselets, qui procèdent à une adaptation d'impédance entre les deux milieux air-liquide. C'est dans l'oreille interne que le son est transformé en message neural.

- *Etape 2 : Transduction sensorielle : vibration mécanique  $\Rightarrow$  potentiels d'action.*  
La transduction sensorielle transforme la vibration de la membrane basilaire de la cochlée en message nerveux sous forme de potentiels d'action grâce aux cellules ciliées internes (CCI) qui la tapissent (formant l'organe de Corti).

Notons au passage que la transduction cochléaire est asymétrique entre les parties positives et négatives de la vibration de la membrane : elle applique un redressement du signal. Autrement dit, l'oreille est très peu sensible à la phase de la stimulation : le potentiel d'action n'est créé que dans le cas où les CCI penchent vers le stéréocil le plus long [Khanna 1989], comme sur la Fig. 1.2.

Chaque portion de la cochlée code une bande de fréquence correspondant au maximum vibratoire local de la membrane basiliaire. Les fibres nerveuses en aval des cellules ciliées tapissant cette portion de la cochlée véhiculent donc les informations liées à ces bandes de fréquences (aussi appelées filtres auditifs, ou bandes critiques). Ces bancs de filtres ont été décrits grâce à des expériences de masquage psychoacoustiques, et la fonction de transfert de ces filtres Gamma-Tones s'exprime de la façon suivante [Patterson 1995] :  $g(t) = at^{n-1} \exp(-2\pi b \text{ERB}(f_c)t) \cos(2\pi f_c t + \phi)$ , (a, b et n étant des paramètres particuliers,  $f_c$  la fréquence caractéristique centrale du filtre, correspondant à un endroit précis de la cochlée,  $\phi$  le paramètre de phase et t la variable temps). Les  $\text{ERB}(f_c)$  sont des filtres rectangulaires équivalents à l'action des filtres auditifs mesurés psychoacoustiquement à un niveau sonore modéré, et dont la réponse impulsionnelle donne un rectangle centré en  $f_c$  et de largeur de bande suivante :  $\text{ERB}(f_c) = 24.7 + 0.108 * f_c$  en Hz [Glasberg 1990]. La distance entre ces filtres est estimée en Bark. Notons que pour passer de l'échelle linéaire (Hz) à l'échelle Bark, on applique la formule suivante :  $Z(\text{Bark}) = 6.7 * \text{arcsinh}((f(\text{Hz}) - 20)/600)$ . De même, dans l'autre sens :  $f(\text{Hz}) = \sinh(Z(\text{Bark})/6.7) * 600 + 20$ .

- *Etape 3 : Le message nerveux remonte les voies auditives vers le cortex.*  
Le message nerveux ainsi engendré suit les voies auditives afférentes à travers le système nerveux périphérique, puis central. Ce système nerveux auditif est symétrique : il est anatomiquement identique à gauche et à droite. Le système nerveux périphérique est constitué des afférences nerveuses situées en périphérie immédiate de la cochlée. Il comporte deux parties : le ganglion spiral, qui regroupe les axones des CCI, et les fibres nerveuses du nerf auditif. Le système nerveux auditif central est quant à lui constitué d'une série de noyaux successifs interconnectés, dont les projections partent également vers les noyaux contralatéraux, Fig. 1.3, (en amont de l'autre oreille). Les principaux noyaux sont : le noyau cochléaire, le complexe olivaire supérieur, le leménisque latéral, le colliculus inférieur, le corps genouillé médian. Chacun de ces noyaux est composé d'une grande variété de cellules nerveuses très spécialisées, dont les dendrites sont très ramifiées et qui reçoivent des afférences (signaux d'entrée) d'un grand nombre de cellules de l'étage inférieur. Du fait de cette complexité du système nerveux auditif central, il est communément admis que le signal nerveux est pré-traité (mais dans quelle mesure ?) avant même d'atteindre le cortex auditif où le message lié au son sera alors intégré.
- *Etape 4 : Le rôle du cortex dans la prise de conscience du message lié au son.*

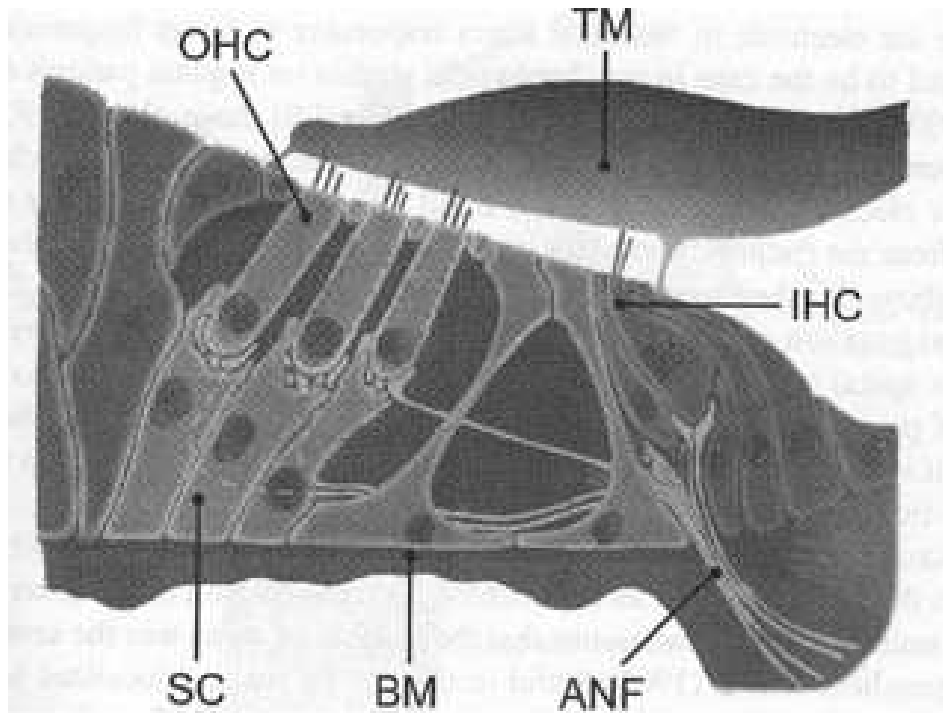


FIGURE 1.2 – **Coupe de la cochlée.**La vibration liquidienne dans la cochlée se traduit par des ondes stationnaires qui font vibrer la membrane basilaire (BM) tandis que la membrane tectoriale (TM) reste immobile. Ce mouvement fait se courber les cellules ciliées internes (IHC), et lorsque celles-ci sont abaissées, elles permettent l'absorption du liquide endolymphatique vers le perilymphatique. Du fait de la composition ionique différente de ces deux milieux, une différence de potentiel se crée, et le message nerveux auditif est ainsi engendré. Le potentiel d'action (potentiel évoqué, comme ce potentiel d'action est issu d'une stimulation sensorielle) se propage ensuite dans le nerf auditif (ANF) vers le tronc cérébral et le cerveau. [Clark 2003]

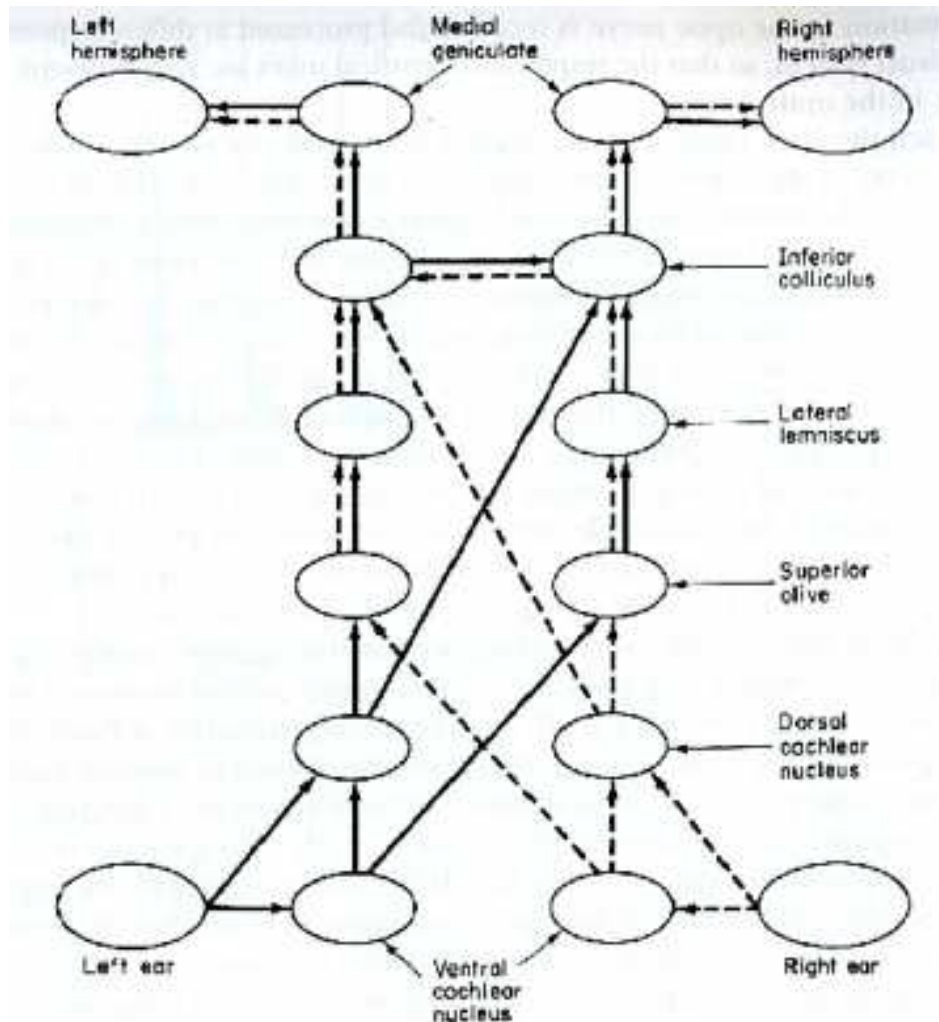


FIGURE 1.3 – **Synoptique des voies auditives, entre le récepteur périphérique et le cerveau.** Les voies auditives exercent une série très complexe de traitement sur l'information auditive, avant que le son ne soit interprété cognitivement dans le cortex. Anatomiquement, les études animales ont permis de connaître dans le détail tous ces noyaux chez l'homme (par extrapolation darwinienne), mais fonctionnellement, beaucoup reste encore à élucider. La difficulté d'accès à ces sites chez l'humain éveillé explique le manque de connaissances à ce sujet. Le Speech ABR est enregistré entre le noyau cochléaire et le thalamus, avec comme site privilégié le colliculus inférieur. D'après [Moore 2005]



Lorsque le message nerveux pré-traité par le système nerveux auditif central arrive dans le cortex auditif, les mécanismes cognitifs entrent en action. Jusqu'alors ce message nerveux n'était que sensoriel, le rôle du cortex est d'apporter cette dimension cognitive. Le cortex va alors trouver un sens à ce qui a été entendu, en faisant appel au besoin aux autres sens (modalité multisensorielle) et à coup sûr à la mémoire, il peut encore choisir de l'ignorer (bruit ambiant).

- *Etape 5 : Le feedback neural.*

Le cortex peut aussi exercer une activité efférente et moduler l'activité du système nerveux central ou périphérique, des fibres efférentes sont directement connectées aux CCE et permettent la modulation de leur activité. Pour les basses intensités, les CCE exercent une action mécanique sur la membrane basilaire de la cochlée et la rendent plus sensibles en pré-incurvant les CCI. Les CCE jouent le rôle d'amplificateur cochléaire. C'est ainsi que les normo-entendants peuvent avoir une dynamique auditive de 120 dB environ. Sans l'action des CCE, les basses intensités sonores ne pourraient pas être perçues entraînant une surdité. De nombreux travaux ont été menés au sein du laboratoire [Khalfa 2001], [Perrot 2006] pour mettre en évidence l'implication de ce système efférent, dont le rôle n'est malgré tout pas encore totalement élucidé.

### 1.1.2 Aspects du codage neural.

Le code neural temporel est lié au phase-locking. Il utilise la synchronisation des neurones pour véhiculer des informations sur la périodicité du stimulus acoustique. Le code neural spatial découle de la tonotopie cochléaire (décomposition fréquentielle du plus aigu au plus grave des sons le long de la membrane basilaire, comme sur le clavier d'un piano); le long des voies auditives, les neurones ont souvent une fréquence caractéristique à laquelle ils déchargent le mieux, et ainsi, le son décomposé en fréquence dans la cochlée est véhiculé dans des canaux (filtres) auditifs passe-bandes. Notons que ces deux codes sont complémentaires.

Le *codage spatial* est régi par le principe de la tonotopie : la cochlée est sensible aux fréquences de plus en plus basses le long de la membrane basilaire, un peu comme le clavier d'un piano 1.4. En d'autres termes, le son perçu à un instant donné est encodé au fur et à mesure du plus aigu au plus grave dans une succession de canaux auditifs partiellement recoupées (1.9) quantifiés ci-dessus par les ERB, et sur des bandes de fréquence centrées sur des fréquences de plus en plus basses. Le grand nombre de cellules ciliées (30000) qui tapissent la membrane basilaire explique la haute sélectivité fréquentielle auditive d'un normo-entendant.

Le *code temporel* intervient en complément du code spatial : 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 (**phase-locking** :



FIGURE 1.4 – **Codage spatial (tonotopique) des sons.** La cochlée a pour propriété d’être organisée tonotopiquement, comme le clavier d’un piano : 1 segment de sa membrane basiliaire est sensible à une fréquence particulière : de 20 kHz à la base quelques centaines de Hz à l’apex. Cette propriété est due à la mécanique cochléaire, et aux phénomènes de résonance, qui sont tributaires de l’épaisseur de la membrane.

Fig. 1.5). 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. Les fréquences limites sont de l’ordre de 4 kHz sur le nerf auditif, à 80 Hz dans le cortex auditif en passant par 1 kHz pour le noyau cochléaire et 300 Hz dans le colliculus inférieur. Au niveau du colliculus inférieur, souvent décrit comme le premier centre d’intégration du message auditif, le code temporel passe du mode « tMTF » au mode « rMTF » (Fig. 1.6), c’est-à-dire que les cellules peuvent se synchroniser sur une fréquence modulante préférentielle (en restant modérément synchronisée sur les autres fréquences modulantes), tandis qu’en rMTF, en plus d’être plus ou moins synchronisée sur la fréquence modulante (tMTF), les cellules répondent de façon plus ou moins importante (en terme de nombre de décharge nerveuses), [Joris 2004]. Notons que le fait que le phase-locking soit limité aux basses fréquences suggère qu’il n’a pas pour but de véhiculer l’information acoustique de façon indépendante (la plupart des informations acoustiques importantes pour la compréhension de la parole, tels que les formants, se situent à des fréquences bien supérieures aux limites de phase-locking dans le tronc cérébral supérieur), mais qu’il doit certainement agir de concert avec le code spatial.

### 1.1.3 Modélisation de l’encodage des sons.

La grande complexité anatomo-fonctionnelle du système auditif dans le tronc cérébral fait que sa description, tant anatomique que fonctionnelle, ne peut être exhaustive. Par conséquent, l’approche modélisante permet de considérer ces cellules par ensembles. Ainsi il est possible de se faire une idée de leur fonctionnement global. A ce titre, ce fonctionnement global est bien illustré par la description du *Periodicity Analysis Network* (PAN) décrit par [Voutsas 2005], dont voici les principales étapes, et dans (ii) *l’Auditory Image Model*, décrit plus bas, au chapitre

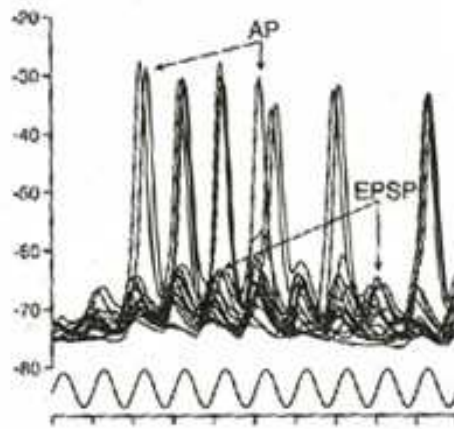


FIGURE 1.5 – Codage temporel : Lien entre un son sinusoïdal (son pur) et les potentiels d'action engendrés (phase-locking). La périodicité du stimulus se retrouve dans la décharge des cellules nerveuses. [Clark 2003]

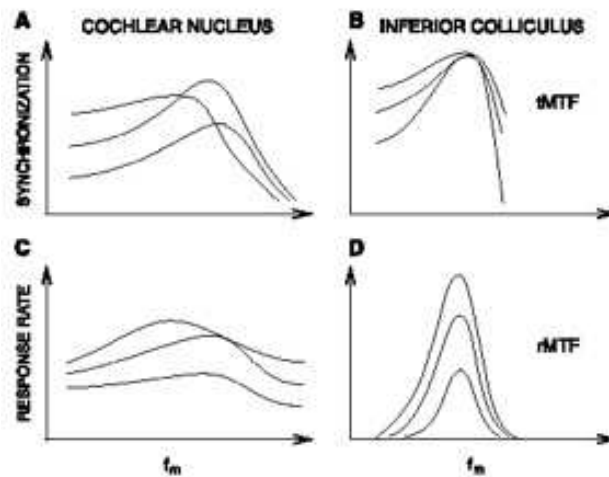


FIGURE 1.6 – Les deux types de codage temporel se synchronisant sur une modulation d'amplitude : tMTF et rMTF. Le phase-locking est présent dans tous les noyaux du système auditif avec plus ou moins d'importance. A partir du colliculus inférieur, les voies auditives convergent, et le codage temporel passe du tMTF (temporal Modulation Transfer Function), qui apparaît depuis le nerf auditif, au rate Modulation Transfer Function). [Joris 2004]

3.2.

**Le Periodicity Analysis Network** [Voutsas 2005] :

Dans ce modèle, les cellules disc-shape du colliculus inférieur ont été décrites comme des détecteurs de coïncidences des spikes de cellules situées en amont du noyau cochléaire. Ces disc-shape fonctionnent comme des autocorrélateurs entre les activités oscillatoires soutenues des cellules de oscillantes de type chopper, et les activités de cellules intégratrices de type pauser qui mettent un certain temps à atteindre leur seuil de déclenchement de spikes. Ces deux types de cellules sont sollicités par des activités transitoires de synchronisation (trigger) des cellules onset, (Fig.1.7).

Le réseau de neurones décrit dans ce modèle, et résumé Fig.1.7, est composé de trois types de cellules : le complexe **intégrateur** est fait de deux cellules inhibitrices flip / flop (qui nous sembleraient inspirées des cellules *bushy cells* du noyau cochléaire antéro-ventral, dont les caractéristiques de décharge à haute spontanéité sont du type *primary-like*), couplées aux cellules intégratrices par elles-mêmes (cellules *fusiformes* du noyau cochléaire dorsal, au profil de décharge *pauser buildup*). Ensuite, on trouve les **oscillatrices** (cellules *stellate cells* du noyau cochléaire postero-ventral, au profil de décharge *chopper*) et les cellules **trigger** à déclenchement (cellules *octopus* ou *multipolar* du noyau cochléaire postero-ventral, au profil de décharge *pure onset* ou resp. *onset chopper*). Ces trois types de cellules, selon la coordination de leur profil de décharge, peuvent engendrer une activité du quatrième type de cellule : les **détecteurs de coïncidences**. Les détecteurs de coïncidences sont stimulés lorsqu'elles reçoivent une afférence simultanée entre les oscillantes (i) et les intégratrices (ii). Ce double-processus parallèle qui donne (i) et (ii) est le suivant.

(i) Les cellules trigger, qui envoient un spike par période modulante  $\tau_m = \frac{1}{F_m}$ , déclenchent un certain nombre  $k$  de décharges à une certaine fréquence  $\tau_k$  propre à la cellule oscillante (partie gauche de la Fig. 1.7).

(ii) En parallèle (partie à droite de la Fig. 1.7), juste avant le déclenchement dû à la cellule trigger, la cellule intégratrice est inhibée par une des deux cellules flip / flop (notée FF2). La décharge de la cellule trigger entraîne une excitation de FF1 qui inhibe FF2, relâchant l'inhibition de celle-ci sur la cellule intégratrice, qui commence à emmagasiner les spikes provenant du nerf auditif (qui décharge le plus à une certaine fréquence caractéristique, d'après la tonotopie cochléaire du code neural spatial), jusqu'à ce que le seuil de la cellule intégratrice soit atteint (au bout de  $n$  entrées). Notons bien la distinction entre la cellule trigger (synchronisées sur  $F_m$ ) et l'afférence des cellules intégratrices (qui déchargent selon  $\tau_c$ ). Les cellules intégratrices déchargent tous les  $n\tau_c$ ,  $\tau_c$  étant la cadence de décharge des fibres nerveuses dont la fréquence caractéristique est  $F_c$ . Toutefois, il nous semble que  $\tau_c \neq \frac{1}{F_c}$  (en raison des limites du phase-locking).

Dans le cas des 4 profils de décharge présentés Fig. 1.7, en dessous de la modulation d'amplitude, on a donc coïncidence de décharge (i) = (ii) pour :  $k.\tau_c + m.\tau_m = n.\tau_c$ . Notons que cette équation de coïncidence est sensible au code temporel (périodicité du stimulus ( $\tau_m$ )) et au code spatial (fréquence caractéristique des fibres nerveuses en amont de la cellule intégratrice). Cet aspect est illustré par la répartition de ces cellules sur un plan défini par les dimensions orthogonales fréquence (tonotopie) et périodicité (phase-locking sur  $F_m$ ). Fig.1.8

Enfin, les ambiguïtés dues aux décharges des cellules coincidentes pour les harmoniques de  $F_m$  sont levées par l'utilisation de cellules inhibitrices du noyau ventral du leménisque latéral.

Ce modèle reproduit bien la complémentarité du code temporel et du code spatial, tel que c'est illustré (Fig.1.8). Quelle que soit la modulation d'amplitude, les cellules trigger se synchroniseront sur  $F_m$  (et déclencheront l'activité de cellules oscillantes qui déchargeront pendant un certain laps de temps à une fréquence quelconque qui leur est propre), sélectionnant une ligne donnée du plan de cellules coincidentes et la porteuse donnée déterminera le filtre auditif qui fera atteindre aux intégratrices le plus rapidement leur seuil, sélectionnant ainsi la colonne correspondante.

#### 1.1.4 Perception psychoacoustique des sons.

Chez l'humain, les opportunités d'exploration neurophysiologiques sont très limitées hormis en mode per-opératoire tel que lors de la pose d'implants du tronc cérébral ou de pose d'électrodes corticales pour calmer les crises chez les patients épileptiques. Cependant, la théorie évolutionniste permet d'extrapoler les résultats obtenus chez l'animal avec ceux attendus chez l'Homme. L'étude de la perception auditive (*psychoacoustique*) est l'approche naturelle (et très complémentaire de la neurophysiologie) qui consiste à tester l'acuité auditive du sujet dans différents types de tâches auditives.

Les caractéristiques acoustiques temporelles jouent un rôle important dans la perception de la parole [Shannon 2002], [Lorenzi 2006]. 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 ( $\geq 20ms$ )
- La **périodicité** reflète l'intonation, l'identité du locuteur, et se situe entre 50 et 500 Hz.
- Enfin, la **structure fine** temporelle se situe au-delà de 500 Hz (durées  $\leq 2ms$ ), et permet la perception, par exemple, des transitions formantiques.

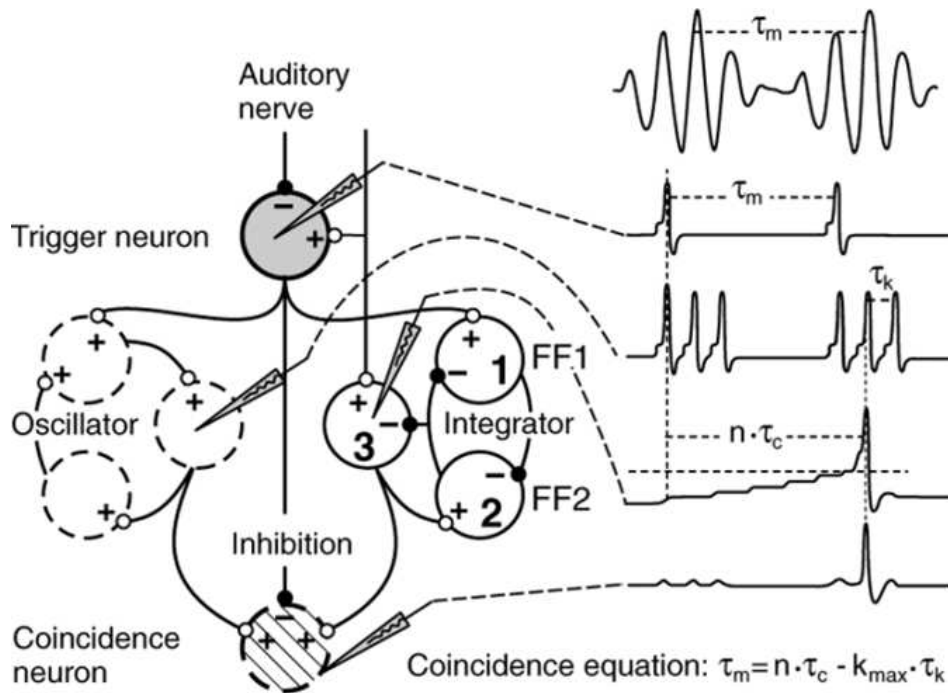


FIGURE 1.7 – Modèle de « **Periodicity Analysis Network** » dans le **Tronc Cérébral**. De l'amont (cochlée) vers l'aval (colliculus inférieur), il fonctionne de la façon suivante. (1) Le son modulé en amplitude (en haut à droite, de période modulante  $\tau_m$ ) engendre une activité de phase-locking dans les cellules du nerf auditif (deuxième histogramme de période de décharge  $\tau_m$ ). (2) Ces spikes vont servir à déclencher (**trigger**) l'activité de 2 différents types de cellules du noyau cochléaire entrent en jeu : les **intégrateurs** à droite, qui n'atteignent leur seuil de décharge que tous les  $n\tau_c$ ; et les **oscillateurs** à gauche dont la seule activité possible est l'oscillation à la période  $\tau_k$ . (3) Au sommet de cette hierarchie, les neurones de **coincidence** du colliculus inférieur déchargent seulement lorsque le signal provenant des oscillateurs et les intégrateurs arrive en phase. Des afférences du lemnisque latéral inhibent en partie les neurones de coincidences. Ce qui mène à l'équation de coincidence, voir texte. [Voutsas 2005]

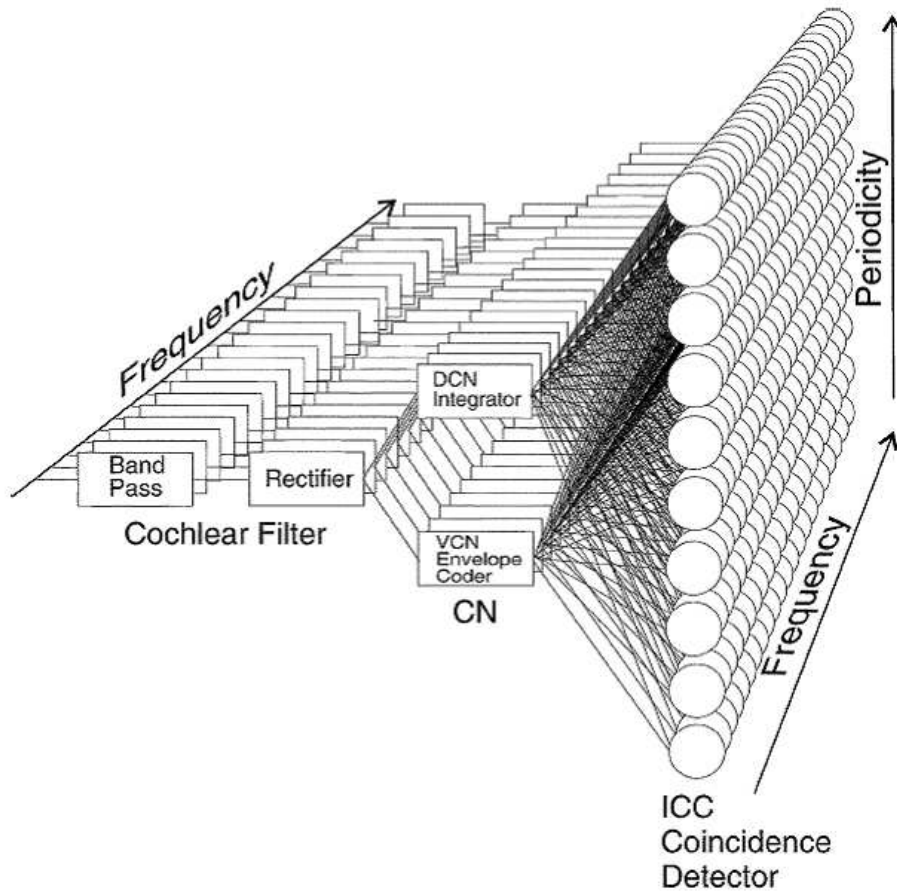


FIGURE 1.8 – Schéma très simplifié du fonctionnement dual temporel/spatial du codage neural de la périodicité entre la cochlée et le colliculus inférieur. Pour un signal avec une modulation d'amplitude donnée, une colonne donnée de cellules coïncidentes sera activée, et une porteuse donnée déterminera la ligne à activer. La flèche des fréquences représente la tonotopie cochléaire. En sortie de la cochlée, des cellules déclenchantes *onset trigger* (non-représentées) déclencheront en parallèle l'activité (i) des cellules oscillantes *chopper* (noyau cochléaire ventral) = codeur d'enveloppe, et (ii) les cellules *pauser* intégratrices (noyau cochléaire dorsal). Les cellules *disc-shape* du colliculus inférieur central seront à l'affût de décharges en synchronisées entre ces 2 types de cellules du noyau cochléaire.[Voutsas 2005]



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 de voisement, long de quelques dizaines de millisecondes. Il est donc important de disposer d'outils cliniques (méthodes objectives et subjectives) tels que les Speech ABR, permettant d'évaluer l'acuité auditive par rapport à ces caractéristiques temporelles.

Afin d'évaluer les capacités de traitement de l'information auditive temporelle, de multiples protocoles de psychoacoustiques ont été validés et s'organisent autour de 3 axes principaux :

- *Détection d'un silence au sein d'un stimulus acoustique continu* (gap-detection) :

Afin d'éviter la perception d'indices spectraux, l'ensemble de la tâche est présentée au sein d'un bruit de masque, le plus souvent bruit blanc à large bande, avec une encoche fréquentielle correspondant à la fréquence du stimulus acoustique utilisé. La durée minimale du silence perçu dépend du type de stimulus utilisé avant et après le silence.

Le seuil minimal de détection est de 2 à 3 ms pour un stimulus acoustique de type bruit à large bande, à une intensité confortable (30 dB SL), et de 3 à 5 ms pour des sons purs. Pour des intensités plus proches du seuil de perception du sujet, la durée du plus petit silence perçu augmente. En revanche, pour des intensités de stimulation largement audibles (30 dB SL et au-delà), la durée minimum perçue est remarquablement constante, que la durée totale de la stimulation varie de 20 à 400 ms, que la position du silence au sein de la stimulation varie, ou que l'intensité de la stimulation varie au-delà de 30 dB SL.

Lorsque des bandes de bruit sont utilisées, plus la bande de bruit est large, plus le silence est facile à détecter, selon la formule : seuil  $s \approx \sqrt{\frac{1}{B}}$ , B représentant la largeur de bande en Hz.

Pour des sons purs, les silences minima perçus sont de l'ordre de 5 ms pour les fréquences de 400 à 2000 Hz, et augmente pour les fréquences inférieures à 400 Hz (18 ms à 100 Hz).

Plutôt que la détection d'un silence, d'autres expériences se sont attachées à mesurer la plus petite différence de durée de silence entre deux sons (gap duration detection). Lorsque les fréquences des sons avant et après le silence sont différentes (détection inter-canaux), le seuil de détection se dégrade, de même lorsque l'intensité des sons se rapproche des seuils de perception des sujets [Grose 2001].

- *Discrimination de durée* : Elle fait référence à la capacité de détecter un changement de durée entre 2 sons. En pratique, deux sons successifs de même puissance mais de longueur différente sont présentés, et le sujet doit identifier le son le plus long. Lorsque la durée du son augmente, la fraction discriminée



$\Delta t/T$  diminue, avec, pour des sons de durées respectives de 10 ms et 1000 ms, des différences respectives de durée de 4 et 60 ms peuvent être perçues.

- *Les fonctions de détection de modulation ou Temporal modulation transfer function (TMTF)*. Elles consistent à mesurer le seuil de détection d'un changement d'amplitude d'un son (porteuse), en fonction de la rapidité des changements, et de l'importance des changements. Le son modulé en amplitude peut être soit une sinusoïde, soit un bruit à bande étroite.

Les seuils de détection de la modulation d'amplitude sont constants pour des modulations inférieures à 50 Hz (20 ms), et diminuent linéairement avec la fréquence de modulation de 3 dB/octave, jusqu'à 1000 Hz (1 ms). Au-delà de 1000 Hz, le seuil de détection de la modulation ne varie plus.

Il y a peu de variations des seuils de détection avec l'intensité du signal, sauf à proximité du seuil de perception où les seuils de détection se dégradent. La sensibilité à la modulation augmente avec la largeur de bande du son modulé, mais varie peu avec la fréquence de la porteuse. Pour des échelles de temps plus importantes, l'étude du traitement de l'information temporelle fait appel à tâches de discrimination de séquences de sons variant selon différents rythmes, des tâches de détection de l'ordre de plusieurs sons successifs, et se rapproche de l'analyse des scènes auditives. La sélectivité fréquentielle et le masquage et bandes critiques.

Le système auditif code l'information spatialement par suite de la décomposition fréquentielle cochléaire. On retrouve en psychoacoustique cette notion de décomposition du son en bancs de filtres qui se recouvrent partiellement (Fig. 9) : les filtres auditifs. Les sorties de chaque filtre auditif correspondent à un canal auditif. Cette idée de décomposition du signal sur la membrane basilaire en bancs de filtres remonte à Helmholtz (1863), et elle a été étayée au début du XX<sup>e</sup>sc. (Mayer, 1894; Weger et Lane, 1924) en montrant qu'un son pur était plus facilement masqué par un son pur de fréquence proche présenté simultanément, mettant ainsi en évidence la nécessité de distinguer (résoudre) les composantes fréquentielles dans des bandes critiques distinctes. Cette notion de bande critique est proche de celle de filtre auditif. Les filtres auditifs sont d'autre part plutôt continus que discrets. On peut donc utiliser cette notion quelle que soit la fréquence envisagée.

## 1.2 Présentation générale de l'implant cochléaire :

On trouve dans la littérature des publications très bien vulgarisées sur l'implant cochléaire [Loizou 1999], [Zeng 2004], [McDermott 2004], [Clark 2006], et un ouvrage de référence [Clark 2003]. De ces ouvrages proviennent les principaux axes d'intérêt repris dans ce chapitre, concernant la réhabilitation auditive.

L'implant cochléaire peut être indiqué pour tous types de surdités neurosensorielles profondes à totales liées à la non-fonctionnalité des cellules ciliées de la cochlée. Toutefois, il est impératif que le système auditif soit fonctionnel afin que le message

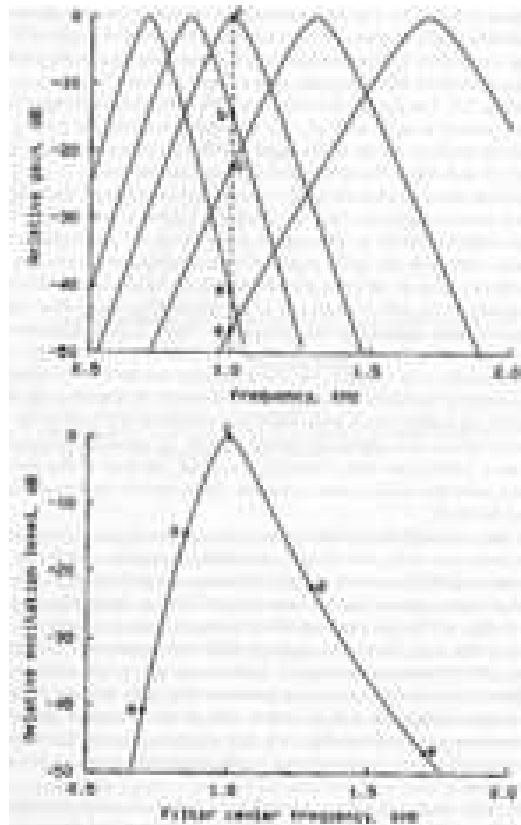


FIGURE 1.9 – **Filtres Auditifs.** En sortie de la cochlée, qui décompose tonotopiquement (en fréquence) les sons, les informations fréquentielles sont réparties en bandes de fréquences : les filtres auditifs. L'auditeur peut par suite tirer partie de la cohérence entre l'information véhiculée par différents filtres pour se focaliser sur une partie du son. Les sorties de ces filtres se chevauchent. Un son pur de 1 kHz n'est pas représenté que dans un filtre auditif, mais dans 5 filtres se chevauchant ici. Ce son pur a plus ou moins d'impact en fonction de la position de l'abscisse 1 kHz dans ces filtres. La réponse la plus grande est celle obtenue pour l'intersection entre la droite verticale  $f = 1\text{kHz}$  et l'enveloppe spectrale du filtre auditif centré sur 1 kHz (schéma du dessus). L'importance de l'excitation produite est reproduite sur le schéma du dessous. [Moore 2005].

nerveux artificiellement induit soit intégré dans le cortex. Il est d'autre part impératif que le patient candidat soit conscient des efforts à fournir pour s'habituer aux sensations auditives que l'implant peut apporter, que ses capacités cognitives puissent lui permettre de s'habituer et que sa rééducation soit envisageable.

### 1.2.1 Approche technique :

L'implant cochléaire est un faisceau d'une vingtaine d'électrodes réparties le long de la cochlée qui stimule directement le nerf auditif avec une séquence d'impulsions électriques. Cette séquence d'impulsions est directement déduite du son capté par le microphone externe par l'intermédiaire d'une stratégie de codage. Cette stratégie de codage est gérée par le processeur de l'implant, et répartit l'information du son sur les électrodes en se basant sur le codage spatial. Toute l'efficacité de l'implant est directement liée à celle de la stratégie de codage, lesquelles ont beaucoup évolué ces dernières années. [Zeng 2004], pour une revue.

Le son est capté par le microphone externe Fig.1.10, il est pré-amplifié avec contrôle automatique de gain, puis il est traité dans le microprocesseur (DSP) de la partie externe (généralement un contour d'oreille) où il est transformé en séquence d'impulsions pour les électrodes correspondantes. Cette information, tout comme l'énergie nécessaire à la stimulation électrique sont transmises par antenne à la partie interne via liaison Radio-Fréquence. Chaque électrode est dévolue à une bande de fréquence précise correspondant plus ou moins à la tonotopie cochléaire : les plus hautes bandes de fréquence stimulent les électrodes basales et les plus basses fréquence les électrodes apicales. Le courant délivré par les électrodes est fixé à partir des seuils auditifs (seuil de perception – T, et de confort – C) propres au patient. L'implant dans son ensemble est alimenté par un bloc accumulateur rechargeable (ou des piles zinc-air, suivant le modèle) stockés dans la partie externe ; l'énergie nécessaire à la stimulation électrique est acheminée vers la partie interne par la liaison radiofréquence.

La stimulation électrique doit opérer le plus localement possible afin de conserver une précision fréquentielle la plus satisfaisante possible, et d'éviter de stimuler des fibres nerveuses qui ne correspondent pas à la bande de fréquence choisie [van den Honert 1987]. Il existe plusieurs façons sensiblement équivalentes de délivrer cette stimulation électrique pour les électrodes. Afin d'éviter toute électrolyse, et que la stimulation ne soit pas invasive, elle doit rester non-galvanique (Shepherd et Clark, 1983) : la somme des charges apportées par les électrodes dans la cochlée doivent s'annuler (stimulation biphasique ou triphasique). Dès lors, on peut fixer la masse électrique en dehors du faisceau d'électrodes (stimulation monopolaire), on peut choisir l'électrode voisine comme masse (stimulation bipolaire), ou encore fixer une électrode précise du faisceau comme masse commune (stimulation à masse commune).

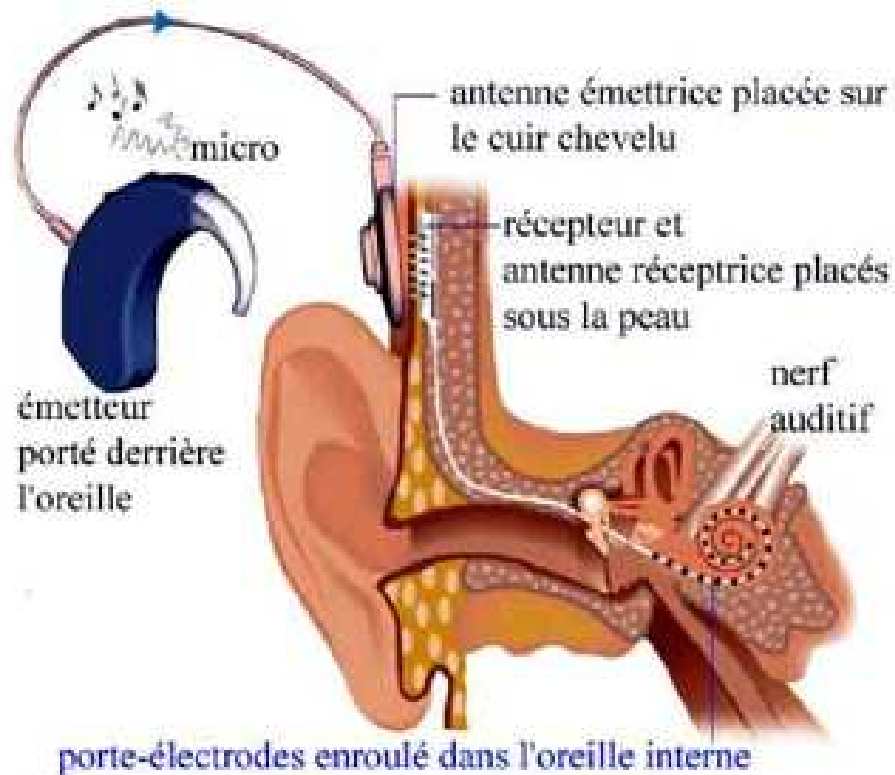


FIGURE 1.10 – **Vue d'ensemble de l'implant cochléaire.** La partie externe est constituée d'un boîtier qui capte les sons par l'intermédiaire du microphone et pratique un traitement du signal pour adapter l'information acoustique en information électrique. La partie interne est mise en place chirurgicalement, et la partie externe qui gère le traitement du signal peut être reprogrammée en fonction de l'évolution de l'audition du patient ou des avancées technologiques (d'après Neurelec)

### 1.2.2 Les processeurs vocaux :

Comment transformer le son capté par le micro en séquence de pulsations électriques sur les électrodes correspondantes? C'est le rôle (fondamental) de la stratégie de codage réalisée dans les processeurs vocaux situés dans la partie externe de l'implant. L'efficacité de l'implant cochléaire a évolué avec les progrès faits dans les stratégies de codage. Les premiers implants cochléaires dans les années 70 ne comportaient qu'une électrode qui était stimulée avec des trains de pulses à basse fréquence fixée sur F0 (Fourcin et al., 1979). En parallèle, des stratégies de codage analogiques pour l'implant multi-électrode Inneraid<sup>®</sup> visait à délivrer un courant ininterrompu et directement fonction des sorties des bancs de filtres correspondant aux électrodes : stratégies **CA** (continuous analog. Eddigton, 1980) et **SAS** (Simultaneous Analog Stimulation. Kessler, 1980).

Aujourd'hui les stratégies de codage sont basées sur la détection de maxima spectraux et exploitent au maximum la cadence de stimulation qui doit être rapide, mais pas trop pour éviter la fatigabilité des fibres nerveuses et pour s'adapter à leur période réfractaire (le nerf auditif a besoin d'une milliseconde environ avant de pouvoir déclencher une nouvelle décharge neurale).

- **stratégies de codage par extraction d'indices vocaux (MPEAK)** Dans les années 80, l'Université de Melbourne a développé l'implant cochléaire multiélectrode, et la stratégie de codage qui fut retenue **MPEAK** (Patrick et al., 1990; Patrick et Clark, 1991) consistait à détecter les caractéristiques propres à la parole dans le son capté par le microphone. Les électrodes étaient stimulées à une cadence correspondant au voisement, ou à 250 pulses par seconde (pps) si aucun voisement n'était détecté. Les 2 électrodes correspondant aux 2 formants étaient stimulées, tout comme les électrodes basales à partir de l'estimation de 3 filtres passe-bande centrés au-delà de F2 (afin de retransmettre certaines consonnes de haute fréquence).
- **Les stratégies de codage par estimation spectrale (SPEAK, CIS, ACE)** Paradoxalement, la meilleure façon de retransmettre une intelligibilité maximale de la parole n'est pas de se focaliser sur les paramètres habituels de la parole (F0, formants), mais de faire une estimation spectrale globale et de sélectionner les maxima, sans forcément qu'ils correspondent aux formants. Ceci est d'autant plus le cas s'il s'agit de sons environnementaux ou de musique.

La stratégie **SPEAK** (Spectral Peak – McDermott et al., 1992) a été conçue à cet effet : le son subit une transformée de Fourier, puis le spectre est décomposé en autant de bandes de fréquences qu'il y a d'électrodes disponibles. Les bandes de fréquences les plus énergétiques entraînent la stimulation des électrodes correspondantes (sans forcément qu'il s'agisse des formants, même pour un son de parole). Le nombre de bandes de fréquence retenues peut être réglé en fonction de la sensibilité du patient (autour de 10 maxima sont en général requis). La stimulation électrique est biphasique, la cadence de sti-

mulation est fixée à 250 pps, et la charge délivrée amplitude \* largeur d'une impulsion) par la stimulation est proportionnelle à l'amplitude de l'enveloppe spectrale estimée dans les bandes de fréquences les plus énergétiques.

Lorsque deux électrodes adjacentes déchargent simultanément, les courants induits dans le nerf auditif peuvent interférer et la qualité de la stimulation en pâtit. La stratégie *CIS* (Continuous Interleaved Sampling – stimulation intercalée en continu. Wilson et al., 1991) vise à limiter cet inconvénient en stimulant les électrodes les unes après les autres. Les 6 bandes de fréquences les plus énergétiques sont sélectionnées, et les 6 électrodes correspondantes sont stimulées : les unes après les autres. La cadence de stimulation est beaucoup plus élevée que dans les stratégies précédentes : 4000 pps. La stratégie *ACE* (Advanced Combination Encoder – Encoder combinateur avancé. Vandali et al., 2000, Cochlear Ltd.) fut développé pour l'implant cochléaire Nucleus (Cochlear Ltd.), et elle combine les caractéristiques de *SPEAK* et *CIS*, en stimulant avec une cadence encore plus élevée. Le son est décomposé en une vingtaine de bandes de fréquences (adapté à l'implant Nucleus qui a 22 électrodes) qui se chevauchent partiellement et dont une dizaine de bandes de fréquence d'énergies maximales sont sélectionnés, comme pour la stratégie *SPEAK*, mais les électrodes sont stimulées les unes après les autres, comme pour la stratégie *CIS*, avec une cadence pouvant aller jusqu'à 14400 pps. Les électrodes stimulent ainsi les fibres nerveuses afférentes du nerf auditif le plus rapidement possible sans dépasser la limite fixée par la période réfractaire du nerf (4000 Hz).

### 1.2.3 Le réglage de l'implant

Les avancées technologiques ont permis de disposer de stratégies de codage, et de techniques d'implantation très performantes. Mais pour le patient, c'est environ un mois après la chirurgie que le « combat » commence. La personne devra se rendre régulièrement dans un centre hospitalier de réglage de l'implant où elle aura deux principaux rendez-vous. L'un avec un orthophoniste pour suivre l'évolution dans l'acquisition du langage et des sons environnementaux, l'autre avec un audioprothésiste afin de régler le processeur de l'implant. Parmi ces paramètres à régler, l'audioprothésiste fixe les nouveaux seuils T et C qui évoluent rapidement les 6 premiers mois après l'implantation. De plus, si les cadences de stimulation, ou le nombre de maxima choisis pour la stratégie de codage étaient jugés trop inconfortables par le patient, l'audioprothésiste pourra revoir ces paramètres à la hausse.

Cette étape de réglage conditionne en grande partie la réussite de l'implantation, et il est important que le personnel ait à disposition des outils performants qui permettent d'avoir une idée précise de la qualité de la perception sonore par le patient. Il se trouve que souvent, la simple description verbale reste très limitée pour décrire sa perception des sons. D'autant plus qu'il s'agit de personnes retrouvant l'audition après une longue période de silence forcé, si ce n'est pas la toute première

fois qu'ils entendent pour les sourds congénitaux. Ce problème d'estimation des seuils, ou de la qualité de la perception, se pose en particulier pour les très jeunes enfants et les patients difficiles à tester.

D'où la nécessité de mesures objectives (explorations fonctionnelles électrophysiologiques) pour assister le régleur (audioprothésiste, ou orthophoniste). Parmi ces mesures objectives, les potentiels évoqués sont les plus couramment utilisés. Les *eCAP* (electrically evoked Compound Action Potential – potentiels évoqués composites engendrés électriquement) sont issu de la stimulation d'une électrode et elle est recueillie sur l'électrode voisine. Les eCAP renseignent sur l'efficacité de la transmission des courants en périphérie de la cochlée, dans le ganglion spiral. Les *PEAPs* (potentiels évoqués auditifs précoces) sont recueillis sur le scalp en réponse à la stimulation d'une électrode. Ils sont engendrés par le système nerveux central au niveau du tronc cérébral : chaque onde (I, III, V) sur le tracé correspond respectivement à la réponse du ganglion spiral, du noyau cochléaire et du colliculus inférieur. La présence de ces ondes traduit la réactivité de ces noyaux neurax [Gallégo 1998] ou [Guiraud 2007]. Mais ces outils sont à développer, et de nouvelles techniques apparaissent, telles que les *ASSR* (Auditory Steady State Response – potentiels évoqués stationnaires et multiples, [Picton 2003a] ou [Ménard 2004]. Les ASSR vont plus loin que les PEAPs et évaluent l'encodage de sons stationnaires (des modulations d'amplitude) également dans le système nerveux central, dans le tronc cérébral supérieur. De la même manière, les *Speech ABR* [Banai 2008] ou [Akhoun 2008d] consistent à suivre l'encodage d'une syllabe dans le tronc cérébral. Mais chez l'implanté, l'activité du système auditif est enregistrée en même temps que la stimulation électrique a lieu, ce qui pose des problèmes d'artefacts encore non-résolus à l'heure actuelle.

Les ABR en réponse à des sons complexes peuvent apporter des perspectives intéressantes dans l'évaluation objective de la perception temporelle des sons, notamment de parole. L'état de l'art dans cette technique, qui sera étudiée en détail tout au long de ce travail, est présentée dans les paragraphes ci-dessous.

## 1.3 ABR : principe et historique

**Principe :** Les potentiels évoqués auditifs (PEA) permettent d'enregistrer l'activité électrique résultant de la propagation et du traitement du message nerveux auditif dans les noyaux successifs des voies auditives, chez l'humain éveillé. L'enregistrement des PEA est très simple à mettre en œuvre et implique une chaîne de stimulation auditive et une chaîne de recueil électro-physiologique provenant de 3 électrodes de contact. 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 biophysique et lorsque les fibres nerveuses véhiculant ce signal sont suffisamment alignées [Eggermont 2007]. S'il est de fait moins précis que les enregistrements faits chez l'animal in situ, les PEA sont un outil de choix, car on peut alors prendre en compte les mécanismes attentionnels et cognitifs absents (ou différents) chez l'animal, a fortiori anesthésié.

Le signal issu des PEA dépend à la fois de la structure acoustique du stimulus, de la durée de l'acquisition ainsi que de la gamme de fréquences des filtres d'acquisition, et enfin (ce qui en fait son intérêt clinique), de l'état fonctionnel des voies auditives du sujet. L'exemple le plus simple est celui d'un normo-entendant stimulé par une impulsion acoustique (clic). Le signal recueilli consiste en une série de pics successifs [Picton 1974], [Melcher 1996] qui traduisent la réponse des noyaux successifs des voies auditives (1ms : ganglion spiral, 3ms : noyau cochléaire, 5ms : leménisque latéral, colliculus inférieur, corps genouillé médian dans le thalamus, au delà de 100ms : cortex auditif). Nous nous limitons ici à la description des PEA subcorticaux, sensés engager des mécanismes auditifs non-influencés par des processus mentaux intégrés tels que l'attention et la cognition (ce qui n'exclut pas des influences réciproques avec les aires corticales, qui eux le sont). Toutefois, l'importance de ce bas-niveau de traitement est justifié par l'obligatoire transit du message auditif afférent en route vers l'intégration du message auditif dans les aires corticales.

L'absence, ou la détérioration d'un des pics du PEA peut donc être relié à un dysfonctionnement du traitement du signal auditif à un endroit donné des voies auditives [Musiek 2007], [Jewett 1970]. Le personnel clinique en audiologie a couramment recours aux PEA (en appoint de la description du patient) pour le diagnostic de la surdité, ou pour évaluer l'efficacité des techniques de réhabilitations auditives comme par exemple les implants cochléaires [Gallégo 1998], [Brown 2003]. Ces PEA sont très utiles pour avoir un aperçu de l'activité auditive transitoire, en réponse au clic ; mais un stimulus acoustique complexe peut considérablement élargir l'horizon d'étude [Banai 2006],[Banai 2007],[Galbraith 2001], pour des revues sur les Speech ABR et [Martin 2008] pour une revue sur l'utilisation de la parole dans les différentes techniques d'enregistrement de PEA. Les parties impulsionnelles du son permettent de solliciter l'activité transitoire (Onset Response) et les parties périodiques du son permettent d'observer l'activité synchronisée sous-jacente (FFR), d'où son potentiel intérêt pour évaluer la qualité du traitement auditif



temporel dans le tronc cérébral.

Le Speech ABR permet donc d'avoir un double-aperçu de l'activité transitoire (Onset Response) et synchronisée (FFR) sous-jacente, d'où son potentiel intérêt pour évaluer la qualité du traitement auditif temporel dans le tronc cérébral. Toutefois, il est fondamental de maîtriser deux aspects de cette technique, dans le but de la valider et d'en permettre l'utilisation clinique. Premièrement, il est nécessaire de pouvoir corrélérer la sensibilité de ces mesures électro-physiologiques et la sensibilité perceptive (psychophysique) des sujets [Purcell 2004]. Deuxièmement, il est crucial de cibler les sites neurophysiologiques d'où provient la réponse. Or, si cette démarche est répandue chez l'animal anesthésié, elle est très lourde à implémenter chez l'humain (éveillé), avec des électrodes implantées dans le cerveau. Or, l'aperçu du traitement de l'information auditive dépend de l'échelle d'étude : depuis l'allure de la décharge de potentiels d'action d'une (ou d'un groupe bien ciblé de) fibre nerveuse jusqu'à l'analyse psychoacoustique de la perception globale d'un son qui engage le système nerveux et cérébral dans son ensemble. D'un côté, l'acuité auditive humaine est relativement bien décrite par la psychoacoustique [Moore 2005] ou par la neuroimagerie corticale [Moore 2008], et de l'autre les caractéristiques de réponse électrophysiologique chez l'animal sont également bien documentées [Joris 2004]. Il existe donc un angle mort dans la connaissance des mécanismes neurophysiologiques en jeu chez l'humain éveillé, où de nombreuses interactions entrent en compte. Ce manque de connaissances affecte entre autres les hypothèses utilisées dans les techniques de modélisations, et limite la compréhension des mécanismes déficients lors de la surdité. Le Speech ABR est donc un moyen d'explorer le traitement auditif, et l'analyse de ce signal électrophysiologique peut renseigner sur ce fonctionnement auditif global.

**Historique :** Les ABR en général se présentent comme un outil audiolologique prometteur et dont les fonctionnalités demeurent à l'heure actuelle relativement mal explorées, ce qui est paradoxal au vu du recul dont nous disposons vis-à-vis des tout premiers enregistrements électrophysiologiques dans le système auditif ([Wever 1930], Fig.1.13). Il est notamment frappant de constater que cet article pionnier posait déjà des questions très actuelles (authenticité de la réponse, opportunités cliniques). Les principales étapes des investigations sur les ABR furent les suivants. (i) La réponse transitoire des clic-ABR fut mise en évidence comme une mesure objective de la fonctionnalité des voies auditives [Jewett 1970]. Ensuite suivirent (ii) les premiers enregistrements de réponse soutenues (Frequency Following Response – FFR) in situ chez l'animal [Worden 1968], puis en champ lointain chez l'homme [Marsh 1975]. Ensuite, Greenberg, Marsh, Brown et Smith ([Smith 1975], [Greenberg 1987]) firent le lien entre la perception de la hauteur tonale et la FFR. Puis, en parallèle, (iii) Picton s'investissait dans l'étude des différentes techniques de potentiels évoqués auditifs qui allaient aboutir sur les nombreux travaux de son équipe sur les Auditory Steady State Responses (ASSR), qui s'apparentent forte-

ment à la FFR. Notons au passage que l'amplitude des potentiels d'action composite du ganglion spiral (electrical compound action potentials - eCAPs) en réponse à une impulsion électrique pourrait être reliée aux seuils perceptifs, bien que ces corrélations ne soient pour l'instant pas satisfaisantes [Thai-Van 2004], [McKay 2005]. Les eCAPs ne sont pas traités dans ce travail de thèse, mais nous pouvons d'ores et déjà imaginer les bénéfices que les ABR et FFR pourraient apporter dans la compréhension des mécanismes liés à la réhabilitation auditive par l'implant cochléaire.

L'utilisation de signaux de parole (Speech ABR, en vue de disposer d'un outil d'audiométrie vocale objectif), et qui sont détaillés ci-dessous date d'une quinzaine d'années [Galbraith 1995a]. Galbraith et ses collaborateurs se sont intéressés principalement aux influences cognitives sur la qualité de la FFR. Krishnan et son équipe se sont intéressés aux influences de la plasticité à long-terme sur la FFR, ainsi que sur des produits de distorsion cochléaire sur la FFR. L'équipe de Kraus dispose d'un grand nombre de résultats (et de revues de littérature) sur l'impact des problèmes d'apprentissage, ainsi que sur la plasticité à court et à long terme sur les speech- et music- ABR.

## 1.4 ABR : Analyse de l'existant

### Utilisation clinique

- *Problèmes d'apprentissages*

L'équipe Kraus, déjà forte de contributions majeures dans l'étude des PEA corticaux dans les pathologies du langage [Kraus 1996], a apporté un grand nombre de connaissances sur l'implication des mécanismes subcorticaux du traitement auditif dans les pathologies liées au traitement du langage et l'autisme, en particulier chez les enfants.

Les travaux de l'équipe dans cette thématique ont été repris dans deux revues de littérature [Banai 2006], [Banai 2007]. Le principe de leur approche est d'enregistrer des Speech ABR engendrés par un son consonne-voyelle (toujours le même /da/) très court, avec un bruit suivi d'une brève voyelle. Ce son a systématiquement engendré un Speech ABR très net (chez les enfants contrôles comme chez ceux diagnostiqués avec des troubles d'apprentissage) avec une Onset Response suivie de 4 ou 5 pics de FFR se terminant par une offset response ([Russo 2004], [Johnson 2005], [Banai 2005], [Wible 2005]).

Leur *rationalnel* est de relier les troubles d'apprentissage à un dysfonctionnement du traitement auditif, en particulier dès le tronc cérébral. Pour eux, la synchronie des réponses neurales détermine la qualité du traitement de l'information auditive. Cette synchronie se traduit dans les Speech ABR sous forme de latence : plus les latences sont longues, moins exacte est la synchronie dans les cellules du tronc cérébral, et par suite, plus dégradé se

### AUDITORY NERVE IMPULSES

By placing an electrode on the cat's auditory nerve near the medulla, with a grounded electrode elsewhere on the body, and leading the action currents through an amplifier to a telephone receiver, the writers have found that sound stimuli applied to the ear of the animal are reproduced in the receiver with great fidelity. Speech is easily understandable. Simple tones, as from tuning forks, are received at frequencies which, so far as the observer can determine by ear, are identical with the original. Frequencies as high as 3,300 cycles per second are audible.

Numerous checks have been used to guard against the possibility of artifact. No response was obtained when the active electrode was placed on any other tissue. After destruction by pithing of the cochlea on the electrode side, the intensity of the response was diminished; after destruction of the cochlea on the other side as well, the response ceased. However, the possibility is still conceivable that these results are due to purely mechanical action of the nerve, which is brought about by mechanical vibrations transmitted from the cochlear structure acting as a special receptor and transmitter.

Further experiments are in progress.

ERNEST GLEN WEVER

CHARLES W. BRAY

PSYCHOLOGICAL LABORATORY,  
PRINCETON UNIVERSITY

FIGURE 1.11 – Première communication sur les speech-ABR. [Wever 1930]

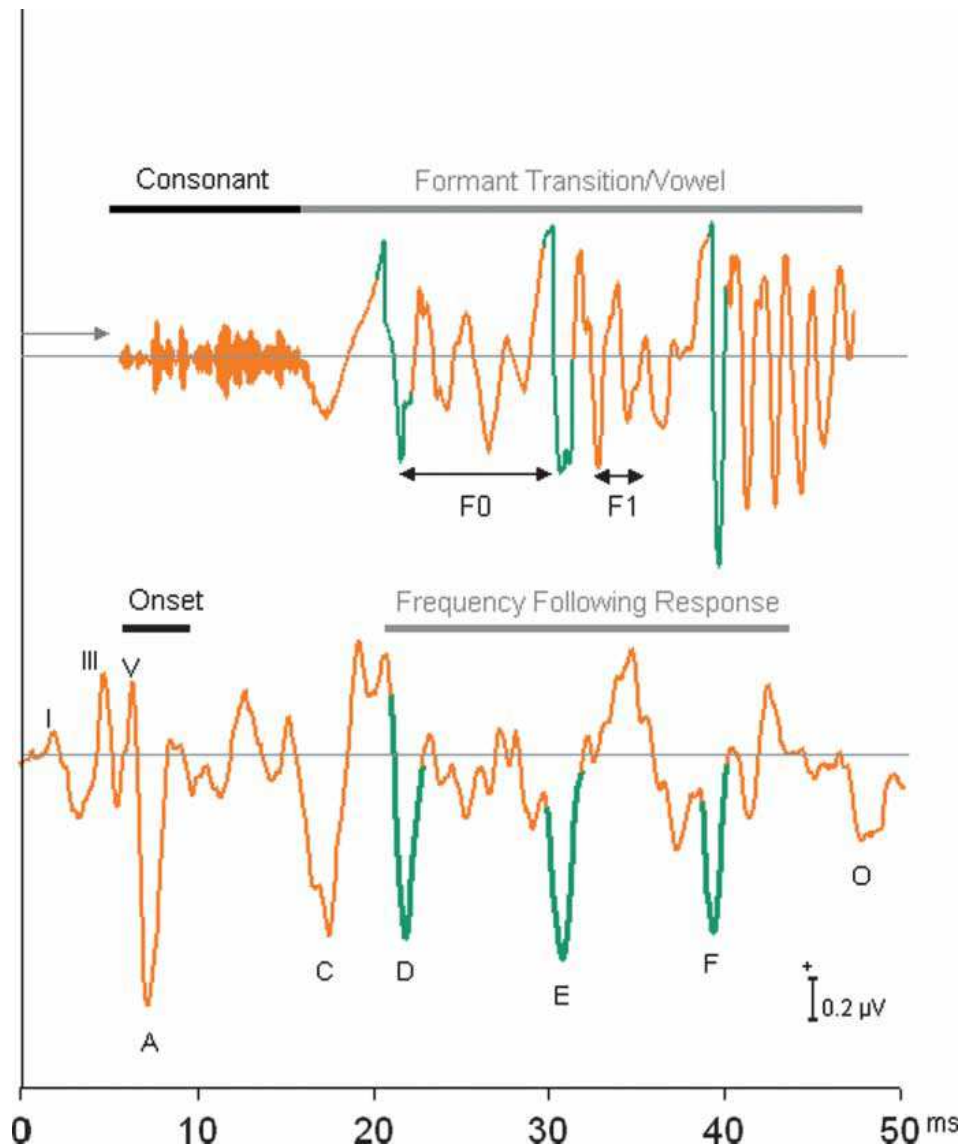


FIGURE 1.12 – Description du Speech ABR, en réponse à /da/. Le stimulus consonne-voyelle est constitué d'une explosion acoustique durant la transition consonne-voyelle qui engendre l'Onset Response. Puis la Frequency Following Response est synchronisée sur la voyelle, comme le montrent les parties en vert, qui se correspondent de près entre le stimulus et la réponse. Avant le stimulus se situe une partie silence durant laquelle la ligne de base de pré-moyennage permet d'avoir une idée du bruit de recueil ambiant. [Banai 2007]

trouve le traitement de l'information auditive.

La maturation corticale se poursuit quelques temps après que celle du tronc cérébral soit achevé. Or, étant donné que le message auditif doit nécessairement être traité dans le tronc cérébral avant d'être intégré (et compris) dans les aires corticales, il se pourrait que les aires corticales reçoivent un message auditif afférent dégradé, ce qui se répercute par une audition, une compréhension (et donc un apprentissage) dégradés.

[Russo 2004] a confirmé par ses résultats que les enfants avec des problèmes d'apprentissages avaient des Speech ABR différents. Premièrement, les latences de l'Onset Response et de la FFR sont plus longues chez ces enfants que chez les enfants contrôles. Les effets du bruit (connus pour dégrader fortement les performances auditives des enfants avec troubles d'apprentissage) se sont manifestés par une dégradation du premier harmonique spectral, et une bonne robustesse de la FFR à F0. [Kraus 2005] en ont déduits que le Speech ABR pourrait être considéré comme l'observation d'un phénomène précurseur à la théorie du *what and where* auditif cortical (le *what* étant la source pour la perception de l'identité du locuteur, véhiculé par F0 et le *where* étant le filtre, ou encore l'information véhiculée, c'est-à-dire la nature de la voyelle).

Le lien entre traitement cortical et subcortical a été établi par [Banai 2005], qui ont montré que 40% des enfants avec des problèmes d'apprentissage et des Speech ABR avec des latences plus longues que les enfants contrôles, avaient des ondes corticales (mismatch negativity - MMN) absentes (contre 15% des enfants contrôles). Un autre lien entre la force de la latéralité de l'activité corticale (observe par cartographie EEG multicanaux) et les latences courtes des Onset Response ou FFR a été mis en évidence par [Abrams 2008]. Plus récemment, les résultats trouvés chez les enfants présentant des troubles d'apprentissage se sont retrouvés chez les enfants autistes [Russo 2008].

- **Audiométrie**

Terry Picton et son équipe ont énormément contribué à faire évoluer les perspectives sur l'utilisation de la FFR (en fait ASSR, car le stimulus est joué en continu) en audiologie clinique. Les ASSR utilisent traditionnellement un stimulus acoustique modulé en amplitude en basse fréquence. L'*hypothèse de départ des ASSR* est que la fréquence porteuse va stimuler spatialement (tonotopiquement) un endroit précis de la cochlée, et que l'activité neurale résultante sera codée temporellement le long des voies auditives, par phase-locking sur  $F_m$ .

Dans ce laboratoire ont été mis au point des outils originaux de traitement du signal ASSR ainsi que pour limiter les artéfacts électromagnétiques ou de calcul [Picton 2004]. L'équipe a mis au point le dispositif expérimental **MASTER**, qui traite les ASSR spectralement en temps réel. Les informations d'amplitude et de phase sont déduites de transformées de Fourier

sur des longues durées d'1s, et détermine automatiquement quand arrêter l'enregistrement (typiquement, dès qu'un pic à la fréquence modulante est deux fois plus grand que le bruit spectral ambiant de mesure).

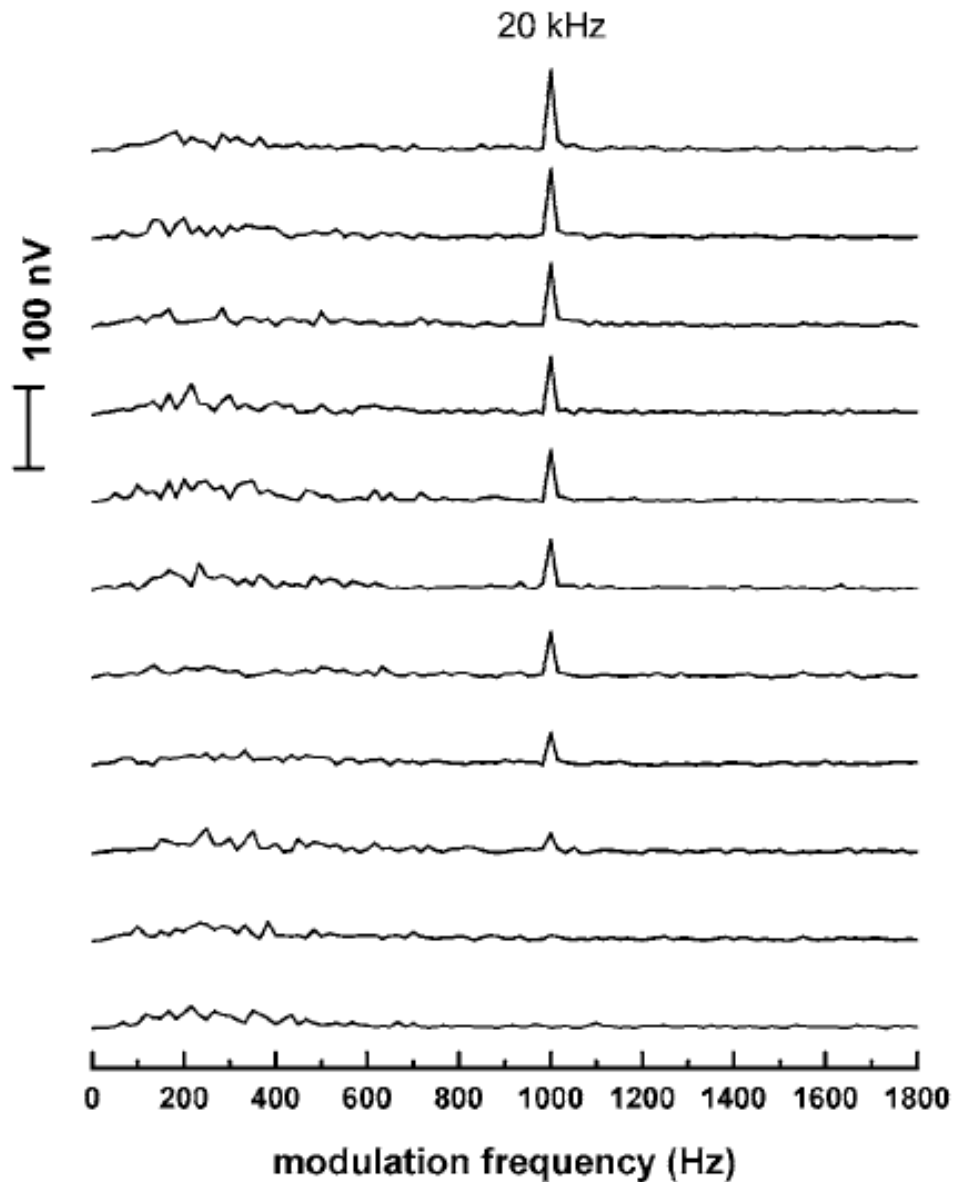


FIGURE 1.13 – Un exemple de l'utilisation d'ASSR pour mesurer l'audition d'un dauphin, avec une fréquence modulante de 1000 Hz, et une fréquence porteuse de 20 kHz, correspondant tonotopiquement au site cochléaire testé [Finneran2007].

- *Surdité*



[Dimitrijevic A 2004] a pu corrélérer les performances d'audiométrie vocale à l'amplitude des ASSR, qu'ils ont trouvé moindre chez les *malentendants*. Les auteurs ont montré qu'il était possible d'enregistrer des ASSR chez les malentendants, et a mis en évidence une augmentation de l'amplitude des ASSR avec une *aide auditive*. Mais il n'est pas précisé comment l'artéfact lié au fonctionnement de la prothèse a été supprimé.

### Influences cognitives

- *Plasticité à long terme*

L'équipe Kraus et celle de Krishnan se sont intéressés à l'effet de l'exposition à un environnement donné sur le long-terme (musique, langue maternelle) sur le Speech ABR ; et l'hypothèse principale des collaborateurs de Kraus rejoint les travaux antérieurs de Krishnan.

[Krishnan 2005] et [Xu 2006] ont comparé les FFR obtenues en réponse à des phonèmes Mandarins, appelés *linguistic-pitch* (/ma/) chez des sujets ayant le Mandarin comme langue maternelle et des natifs anglophones, pour montrer que les Chinois présentaient une FFR mieux synchronisée<sup>1</sup>). Les 4 phonèmes /ma/ diffèrent les uns des autres par la dynamique de leur F0, qui détermine la signification linguistique du mot.

Pour l'équipe Kraus, l'apprentissage de la musique façonne les voies auditives subcorticales, si bien que les musiciens (anglophones) sont plus sensibles aux variations de hauteur tonale (pour des linguistic pitch de Mandarin) par rapport à des sujets contrôles. Ce façonnage des voies auditives se retrouve dans la meilleure synchronisation de la FFR pour les musiciens [Wong 2007] sur des linguistic pitch dont la  $F_0$  est dynamique, Fig.1.15 (voir aussi [Krishnan 2004], Fig.1.14), ou pour des notes de violoncelle, par ailleurs sensés être les plus proches d'une voix humaine [Musacchia 2007].

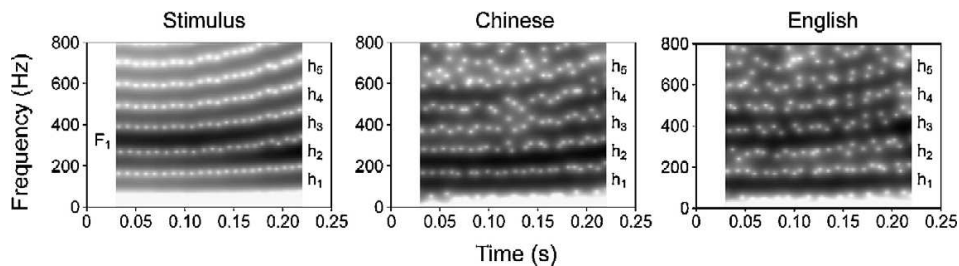


FIGURE 1.14 – Autocorrélogramme des FFR en réponse à des sons de langue tonale. Le stimulus est à gauche, la FFR à droite. [Krishnan 2004]

1. Ils ont au passage introduit la méthode de *pitch-tracking* de la FFR reprise par l'équipe Kraus, issue du logiciel de linguistique Praat ; le *pitch-strength* associé est l'amplitude du pic d'autocorrélation de la FFR par sections de 40 ms

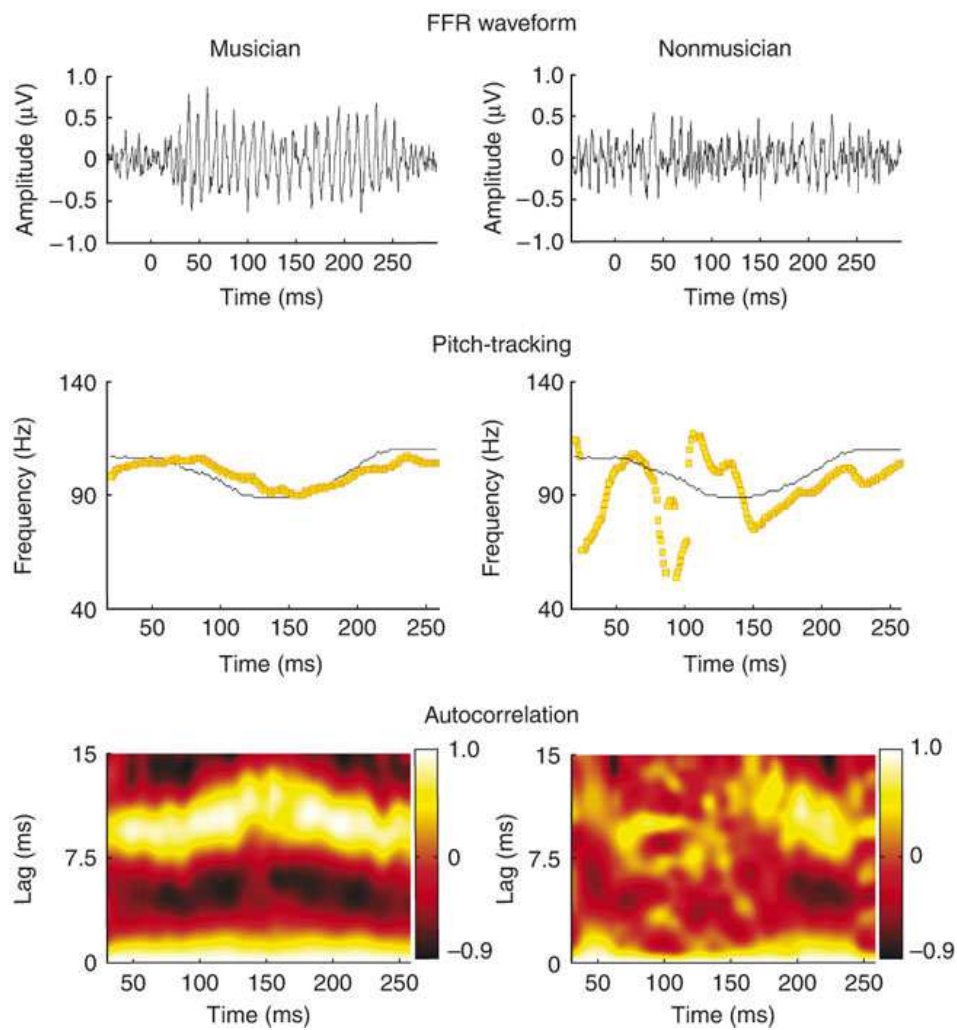


FIGURE 1.15 – **Comparaison des ABR en réponse à un son de langue tonale, chez les musiciens et chez les non-musiciens.** La meilleure synchronisation de la FFR chez les musiciens est ici reliée à un meilleur traitement de la hauteur tonale en général, en conséquence d'une plasticité à long-terme dû à la pratique musicale. [Wong 2007]



- *Plasticité à court et moyen terme*

Dans une communication récente [Song 2008] ont précisé que la force du pitch matching d'une FFR à des linguistic pitch non-usuels pouvait être améliorée par une série d'entraînement. Ce qui suggère un effet de la plasticité d'apprentissage à court-terme. Avant cela, [Russo 2004], [Wible 2005], avaient montré des latences de Speech ABR significativement améliorées par des séances d'entraînement auditif, sous-entendant une certaine forme de plasticité à moyen-terme (sur environ un mois).

- *Attention, Effets du Sommeil*

La technique des ASSR a permis de montrer que l'état de *sommeil* [Picton 2003b] ou d'*attention* [Ross 2004] influençaient l'amplitude des ASSR seulement pour les fréquences modulantes inférieures à 40 Hz. Leur hypothèse est que la fréquence modulante définit le générateur neuro-physiologique sous-jacent de la réponse ASSR est générée, en se basant sur la gamme de fréquence du phase-locking, chap.1.1 : plus les noyaux cellulaires sont intégrés moins la fréquence limite de phase-locking des cellules est élevée. Toutefois, la réciproque semble difficile à admettre : un phase-locking de 40 Hz pourrait tout à fait apparaître en périphérie cochléaire.

Galbraith a montré que l'amplitude des *potentiels corticaux* (avec des sons pur de 230 Hz) étaient plus importants lorsque l'attention était portée sur le son, par rapport à quand il était ignoré. A l'inverse, la FFR était insensible à l'*attention*. Mais il publia aussi le résultat inverse la même année [Galbraith 1993] : en utilisant deux sons purs (200 et 400 Hz) dans une condition dichotique (oreille droite et gauche stimulées différemment), et en considérant séparément la première et la seconde moitié du stimulus, il rapporta une préférence de l'oreille droite, ainsi qu'un effet de l'attention\*fréquence\*oreille sur l'amplitude spectrale (par FFT). Un effet de l'attention sur l'amplitude spectrale de la composante correspondant à F0 fut aussi mentionné avec des voyelles en condition dichotiques [Galbraith 2003].

Lors du moyennage, où un grand nombre de répétitions est nécessaire, le stimulus de parole répétitif a tendance à être entendu comme des mots significatifs, parfois très éloignés du stimulus par lui-même, qui peuvent changer au long de l'expérience (phénomène cognitif de *verbal transformation*). Les FFR ne furent pas différentes dans le cas de verbal transformation, même si un effet groupe\*condition fut identifié en séparant les sujets selon ceux qui faisaient plus ou moins souvent l'expérience de verbal transformation [Galbraith 1997]. De la même façon, une FFR moins bien définie fut obtenue en réponse à des sons de parole diffusés à rebours, comparé à des sons de parole normaux [Galbraith 2004].

- *Effet du contexte*

Est-ce que la sensibilité de la FFR est liée à la nature du stimulus ? Pour approfondir leur approche de l'encodage de la hauteur tonale par la FFR (voir ci-dessous), [Krishnan 2008] ont eu recours à des stimuli ne faisant intervenir que la *dimension temporelle* (et non spatiale) de la hauteur tonale (dit pitch temporel). Il s'agit de bruits blancs récurrents (iterated ripple noise *IRN*) : le principe est de sommer au fur et à mesure une portion de bruit blanc par lui-même, décalé d'une certaine période ( $T_0$ ), et ceci un certain nombre de fois. Au final, le son ainsi produit induit une sensation de hauteur tonale à la fréquence  $\frac{1}{T_0}$ . L'avantage de ce son est de ne produire aucun indice spectral (tonotopique), et de n'être relié qu'au pitch temporel. Dans de récentes études [Swaminathan 2008], ils ont répliqué leurs comparaisons de l'influence de la langue tonale, cette fois-ci en remplaçant les linguistic-pitch par des IRN équivalents (homologues) ayant le même profil de F0. Il en ressort que la FFR engendrée par les IRN sont là-encore mieux synchronisées pour les Chinois que les anglophones. Leur conclusion principale est que la FFR n'est pas sensible au contexte langagier (phonèmes /ma/) ou pas (IRN homologues des /ma/). Ceci rejoindrait les théories issues de la neuroimagerie, qui stipulent que les voies du spécifiques du traitement de la parole ne se déclarent que dans les aires corticales, et qu'auparavant, la FFR n'est sensible qu'aux caractéristiques acoustiques (dynamique de F0) du son [Zatorre 2008].

De façon intéressante, leur équipe et la notre eurent tous deux recours aux théories neurophysiologiques du pitch développées par l'équipe Langner, détaillé dans le Periodicity Analysis Model dans le chap. 1.1 (sans concertations préalables). Ainsi, ils spéculent que les réseaux neuronaux d'encodage de la périodicité Fig.1.7 seraient susceptibles d'être affinés par *exposition à long-terme* à une langue comme le Mandarin qui utilise les caractéristiques tonales pour véhiculer le sens linguistique (par une action sur les effets corticofuges ou sur les réseaux locaux d'excitation / inhibition), comme le suggéreraient les FFR mieux définies chez les sujets Chinois. En remarquant une FFR moins bien définie pour des portions des stimuli ayant une F0 linéaire, ils ont pu suggérer une importance particulière de l'*accélération des variations de F0* (qui se retrouvent par ailleurs dans la transition formantique, indice acoustique crucial dans la perception de la parole) dans le processus neurophysiologique qui aboutit à la FFR, phénomène qu'ils avaient déjà observé en comparant l'effet des profils curvilignes vs. linéaires [Xu 2006].

- *Multimodalité*

[Musacchia 2006] ont recherché une expression de l'effet McGurk dans le Speech ABR (processus cognitif qui fait percevoir /ga/ lorsque /da/ est prononcé et que l'image du locuteur correspond à /ka/), et a rapportée que l'onde delta (sous-onde de l'Onset Response) était décalée dans ce cas, Fig.1.16. Mais précisons déjà que ce résultat très discret fait débat. Toutefois, il est intéres-

sant de constater que les ABR obtenus en réponse à des stimuli purement visuels n'ont donné aucune réponse autre que du bruit ambiant Fig.1.17. En revanche, les auteurs ont plus récemment montré que le meilleur encodage des sons musicaux (violoncelle) et de linguistic pitch par les musiciens était encore accentué en modalité audiovisuelle [Musacchia 2007].

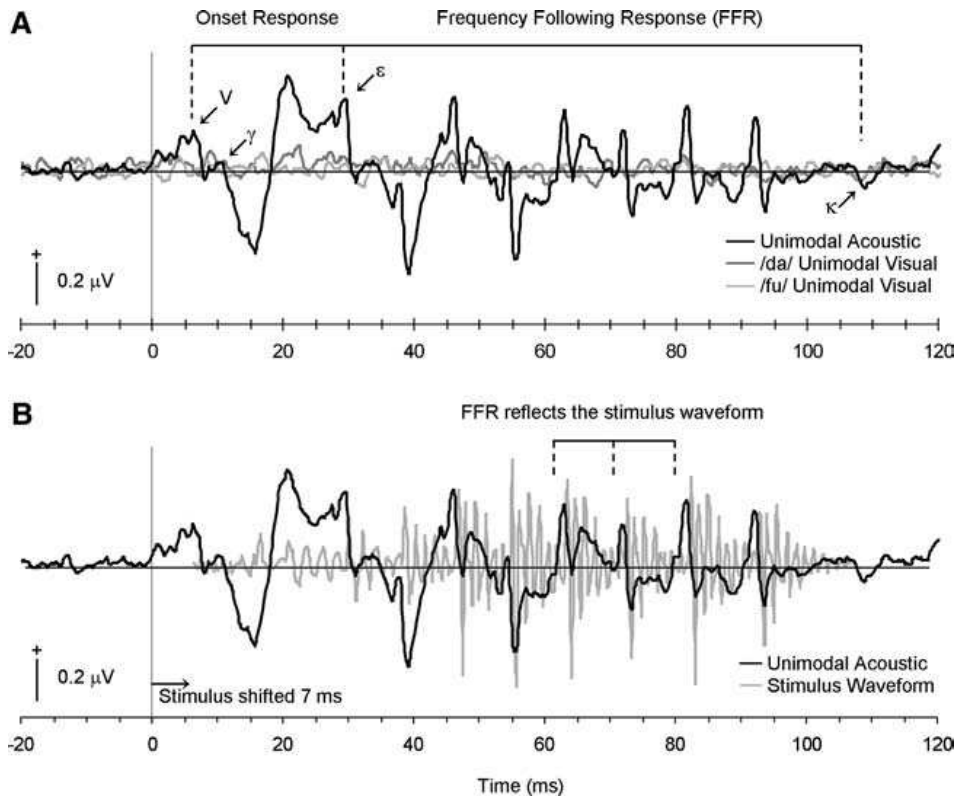


FIGURE 1.16 – Speech ABR en modalité auditive et visuelle. Pas de Speech ABR en modalité visuelle. [Musacchia 2006]

### Corrélats psychoacoustiques

- *Relation avec la hauteur tonale*

[Greenberg 1987] ont effectué un corrélat électrophysiologique de la perception de la hauteur tonale (low-pitch) par la FFR, en montrant que la fréquence synchronisante de la FFR peut être reliée à la hauteur tonale perçue dans la plupart des conditions remarquables (missing fundamental, zone de dominance ; harmoniques résolus, non-résolus, déphasés ou pas ; sons complexes inharmoniques, voir ci-dessous). Dans leur étude, ils proposent une revue de littérature sur le low pitch et sur les générateurs de la FFR (et la garantie de l'authenticité de la réponse), se basant sur les nombreuses

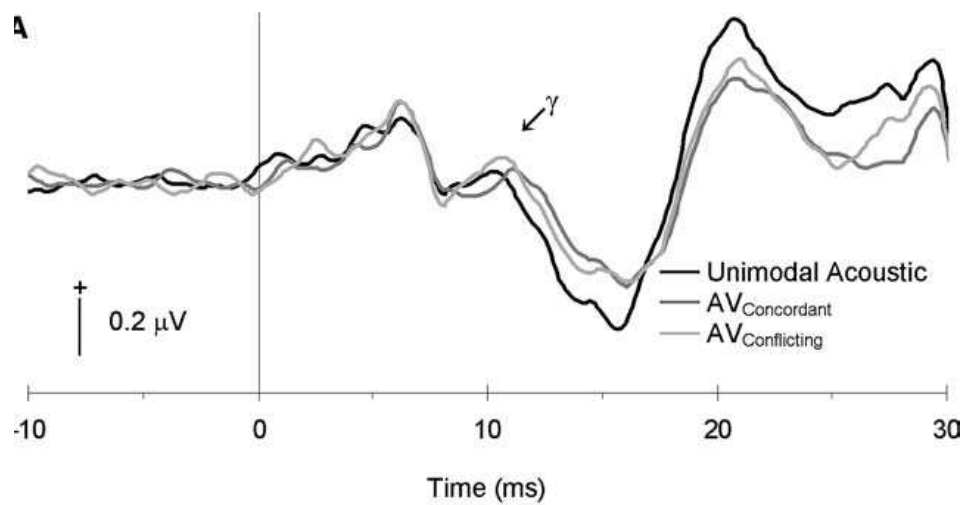


FIGURE 1.17 – Effet Mc Gurk et Speech ABR. Cette différence est minime. [Musacchia 2006]

études animales et humaines de l'équipe, pionnière dans la discipline.

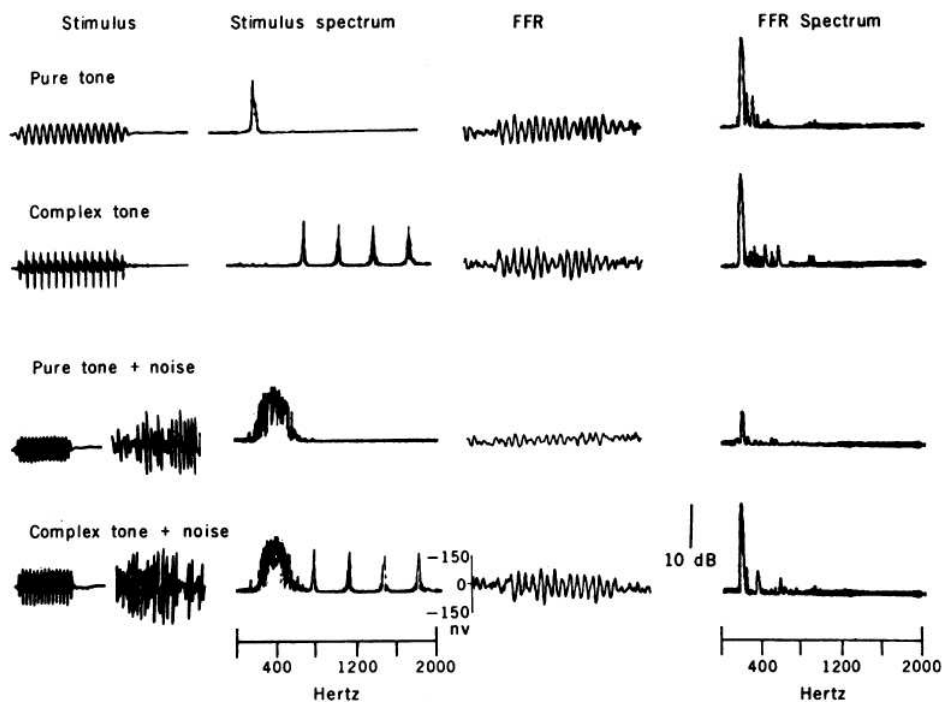


FIGURE 1.18 – Phénomène de missing-fundamental et FFR. Même s'il est physiquement absent, avec du bruit dans sa bande de fréquence, la FFR est bien synchronisée sur le fondamental. [Greenberg 1987]

Ils ont tout d'abord (i) comparé les FFR en réponse à des *sons purs* et des sons complexes harmoniques sans le fondamental *missing fundamental* (masqué par du bruit passe-bas pour la condition 366 Hz), pour des fréquences de 244, 366 et 488 Hz. Les FFR furent similaires (synchronisées sur ces fréquences). Pour la condition 244 Hz, ils n'ont pas trouvé d'effet significatif du déphasage entre les 3 premiers harmoniques sur la synchronisation de la FFR (ii), alors que cette condition déforme beaucoup la forme de l'enveloppe temporelle du stimulus. La latence des FFR fut mesurée aux alentours de 10 ms, valeur habituelle. En revanche (iii), la condition de *déphasage* à 90 degrés atténua la FFR dans le cas des harmoniques haute-fréquence (7, 8 et 9°). Ce qui fait écho aux résultats de [Dau 2003], avec une prépondérance de l'influence des hautes-fréquences dans la formation de la FFR. Pour des sons complexes *inharmoniques* (du type 366, 610 et 854 Hz, harmoniques 3-5-7 de 122 Hz) 3-5-7, 5-7-9 et 7-9-11 les FFR furent ambiguës et enregistrées sur 2 sujets seulement. L'enveloppe des sons est pseudo-modulée (par deux pseudo-fréquences modulantes), et la hauteur tonale est aussi ambiguë ( $p = fc/n$  de [Boer58]). Au passage, ils ont introduit la technique de polarité alternée (par addition – enveloppe FFR, et par soustraction – spectral FFR), en tentant de corrélérer notamment la spectral-FFR au pitch ressenti (dans la plupart des cas, cette technique est sensée isoler les processus sensibles à la phase, apparentés à de l'artéfact), mais là encore, l'expérience ne faisait intervenir que 2 sujets. Enfin (iv), pour mettre en évidence la gamme de *dominance* (la hauteur perçue par missing fundamental est la plus forte lorsque l'énergie est la plus importante sur les harmoniques 2-5) des FFR furent enregistrées à des sons complexes missing fundamental (de  $F_0 = 244, 366$  et 488 Hz) de paires d'harmoniques de plus en plus haut (jusqu'à 2kHz). La FFR est bien restée synchronisée dans tous les cas sur  $F_0$ , avec une amplitude maximale lorsque les harmoniques étaient dans la zone 500 – 1kHz (pour les 3  $F_0$  testées), comportement relié à cette gamme de dominance.

Comme nous l'avons vu ci-dessus dans les effets de la plasticité à long terme, [Krishnan 2004], [Krishnan 2008] ont montré l'effet de l'exposition à une langue maternelle tonale sur le meilleur encodage de la hauteur tonale (mesuré comme une FFR en réponse à des linguistic pitch mieux synchronisée sur  $F_0$ ). Ces résultats sont rejoint par ceux de [Wong 2007] et [Musacchia 2007] avec des musiciens et non-musiciens. De manière analogue, les ASSR ont été appliqués à l'étude de l'encodage du fondamental des *voyelles* avec  $F_0$  dynamique, [Aiken 2006], [Aiken 2008]. Une méthode équivalente au pitch-tracking fut définie pour suivre la synchronisation de l'ASSR sur  $F_0$  [Dajani 2005].

D'autre part, comme nous le mentionnons plus haut, la dynamique de  $F_0$  est importante dans la formation de la FFR, d'après [Xu 2006]. Il a en effet été montré que la FFR engendrée par des analogues (sons purs correspondant

à une portion de la variation de F0) n'était pas différentiable sur les deux populations, montrant comme résultat principal que la synchronisation de la FFR pouvait être améliorée dans le cas de profil curviligne de F0 (cas des phonèmes /ma/), mais pas dans le cas de profils linéaires de F0 (analogues). De plus, la FFR est également moins précises pour les F0 montantes que pour des F0 descendantes que pour des F0 montantes; ceci fut discuté comme le fait qu'en montant, les F0 se déplaçaient vers la région de dominance du pitch (sensation de hauteur tonale plus prononcée lorsque l'énergie des harmoniques 2 – 5 est renforcée), faisant écho aux études précédentes de [Greenberg 1987]. Dans des études antérieures, [Krishnan 2004] avait commencé par comparer les FFR en réponse à ces phonèmes, ainsi qu'à des sons complexes imitant la dynamique de F0, mais précisons dès à présent que ces résultats sont soumis à caution au vu de l'utilisation de casques TDH 49 très susceptibles de donner lieu à une contamination électromagnétique des enregistrements, comme nous le montrons chap.2.2 et [Akhoun 2008d].

[Galbraith 2000] ont enregistré des FFR en réponse à des sons purs entre 200 et 950 Hz, en configuration d'électrodes horizontales et verticales, et des réponses furent obtenues jusqu'à la condition 950 Hz. Auparavant, [Galbraith 1995b] avaient aussi cherché à comparer les FFR en réponse à un son pur et un son complexe *missing fundamental* à 200 Hz enregistrés selon un alignement horizontal vs. vertical des électrodes; il fut montré que la FFR n'encode pas le missing fundamental sur l'orientation horizontale contrairement à la verticale. Ils rapprochèrent les latences observées à des sites générateurs sur le nerf auditif en horizontal et tronc cérébral en vertical. Si on tente de rapprocher ceci du résultat spectaculaire que la forme d'onde du Speech ABR transformée en fichier son peut être intelligible [Galbraith 1995a], [Wever 1930] (Fig. 1.13), on constate alors que ces enregistrements sont sujets à de l'artéfact (voir paragraphe suivant).

- ***ASSR et sonie***

Au laboratoire, Mikaël Ménard a trouvé une corrélation entre la fonction de *sonie* et l'amplitude des ASSR chez le normo-entendant [Ménard 2008], avec des sons de 500 Hz et 2 kHz modulés en amplitude à 80 Hz. Cette corrélation s'est avérée plus forte que celle reliant l'amplitude du son de celui des ASSR.

- ***Encodage de la consonne-stop***

Dans un article récent, [Johnson 2008] a comparé les Speech ABR obtenus en réponse à /ba/, /da/, /ga/, trois phonèmes ayant strictement la même voyelle et ne différant que par la consonne et la transition formantique (F2 et F3). Pour anticiper sur l'effet des formants, l'hypothèse fut faite que la composition spectrale de la FFR ne devrait pas être affectée par les transitions formantiques, qui sont au-delà de la fréquence de coupure du Speech ABR, ce qui va à l'encontre de nos résultats (et ceux de [Dau 2003]) de la section

3.2. Cette hypothèse fut complétée par l'idée que les transitions formantiques seraient répercutées par des différences de latences dans les différents pics de l'onset response et de la FFR qui s'annulent lorsque la voyelle devient steady-state, avec  $/ga/ \leq /da/ \leq /ba/$  car les latences sont les plus grandes pour des formants dans plus dans les basses-fréquences (se basant sur des résultats d'ABR en réponse à des sons purs), ce qu'ils ont effectivement trouvé.

- *Relation avec l'acuité temporelle auditive*

Les envelope-following response (**EFR**), sont une variété d'ASSR, avec des sons modulés en amplitude dont la fréquence modulante augmente / diminue (linéairement) en fréquence au cours du temps (sweeping tones). Du fait de la stimulation en continu, la méthode d'estimation visuelle des latences (en mesurant le décalage d'une onde par rapport à la portion du stimulus qui l'a engendrée) n'est pas envisageable. Dans le cas des EFR, la mesure des *latences apparentes*<sup>2</sup> (apparent latencies) est une alternative intéressante. Les latences apparentes ainsi rapportées sont plus importantes que celles habituellement déduites visuellement en routine clinique, avec les clic-ABR par exemple. [John 2000] a rapporté des latences qui dépendent de la gamme de fréquence où les sweep ont eu lieu, et qui se situent autour de 20.2 ms pour des fréquences modulantes entre 80 et 100 Hz, et 17.7 ms entre 180 et 220 Hz. [Purcell 2004] ont pu corrélérer l'*acuité temporelle* psychoacoustique (gap detection et tMTF, présentées plus haut) aux latences apparentes ainsi déduites.

## Modélisation

Torsten Dau [Dau 2003] a modélisé la formation des ABR sur le scalp, et a comparé cette modélisation à ses résultats précédents faisant intervenir des sons purs, des clics et des chirps (sons dont la fréquence augmente rapidement de 100 à 10.4 kHz, afin de compenser le délai de propagation de l'onde mécanique cochléaire et de stimuler toute la cochlée d'un coup). Si son modèle est imparfait, son approche très rigoureuse éclaire toutefois très significativement les mécanismes neurophysiologiques sous-jacents en jeu dans la formation des ABR ou des FFR.

---

2. La mesure de la phase est possible par analyse de Fourier [Regan 1966], [Picton 2007], Fig.1.19. Pour cela, partant de l'hypothèse de départ des ASSR, on considère que l'EFR est synchronisée sur la fréquence modulante, et donc on concentre la transformée de Fourier sur une seule fréquence, celle de la modulante (au lieu de la transformée de Fourier classique qui multiplie le signal analysé par toutes les fréquences ce qui semble a priori superflu ici); par conséquent l'analyse de Fourier de l'EFR n'a lieu qu'à la fréquence modulante instantanée, et ainsi donne cette phase instantanée (ainsi que l'amplitude, qui n'est pas utilisée ici). La phase instantanée peut être représentée pour chaque fréquence modulante à la fin de l'acquisition. La latence apparente ( $\tau$ ) est le retard de groupe associé au transit dans le système auditif :  $\tau = -\frac{2\pi}{\omega}$ , avec  $\omega = 2\pi f_m$ ; autrement-dit, la latence apparente peut être mesurée comme la pente de la courbe  $\phi = f(f_m)$ .



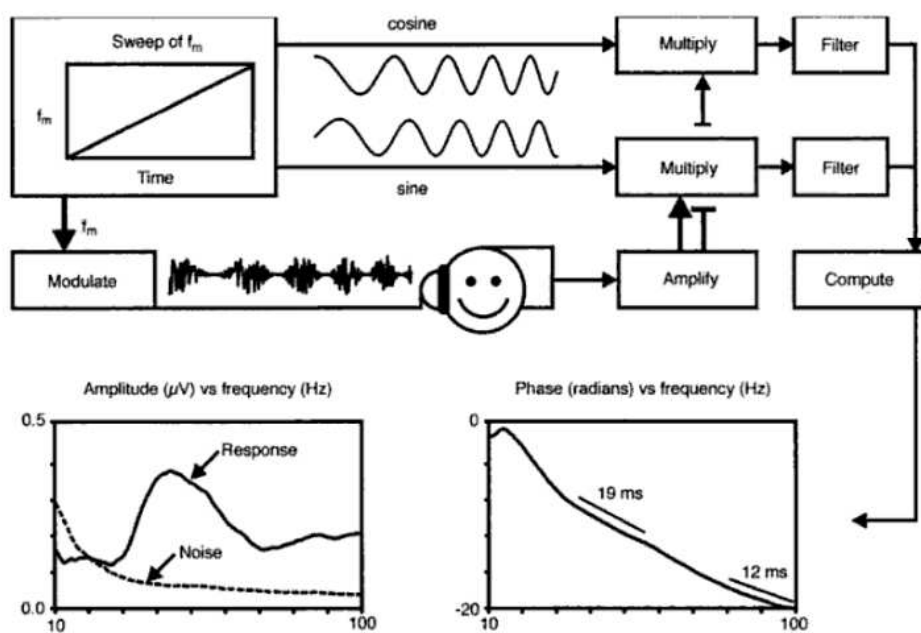


FIGURE 1.19 – Estimation des latences apparentes par retard de phase (**Fourier Analyser**) : Un moyen de mesurer les latences avec une stimulation en continu (modulations d'amplitude). La latence apparente est la pente de la courbe  $\phi = f(f_m)$ , en bas à droite (19 ms pour des fréquences modulantes de 50 Hz, et de 12 pour des fréquences de 80-100 Hz). [Picton 2007]



Le principe de cette modélisation (Fig.1.20) est de partir du modèle de [Melcher 1996]<sup>3</sup> :  $ABR(\bar{x}_1, \bar{x}_2, s) = \sum_{p=1}^P u_p(\bar{x}_1, \bar{x}_2, s) * R_p(t, s)$  (avec  $\bar{x}_1$  et  $\bar{x}_2$  situant la position des électrodes de recueil,  $s$  étant le stimulus,  $u_p$  la réponse unitaire d'un groupe cellulaire  $p$  (parmi les  $P$  considérés) et  $R_p$  le potentiel d'action créé par ces cellules  $p$ ). Les modules de ce modèle reprennent les principales étapes physiologiques : **(a)** transformation dans l'oreille moyenne (filtrage passe-bande 300 – 7kHz), ensuite **(b)** simulation de la décomposition fréquentielle sur la membrane basiliaire et non-linéarité des cellules ciliées internes (redressant et filtrant passe-bas, extrayant l'enveloppe du signal), afin d'en déduire le profil de décharge (time varying discharge rate) du nerf auditif [Heinz 2001] sommé sur tous les canaux de façon à obtenir un potentiel composite (donnant le terme  $R(t,s)$ ). Précisons que l'hypothèse est faite que tous les types cellulaires ont le même profil de décharge que le nerf auditif (primary-like), ce qui est une simplification drastique, comme Dau le précise. Ensuite **(d)**, ce terme  $R(t,s)$  est convolué avec la réponse unitaire  $u(\bar{x}_1, \bar{x}_2, s)$ , qui représente le potentiel créé chaque fois que la cellule décharge avec une cadence de  $R(t,s)$ . Mais auparavant, **(c)** cette réponse unitaire  $u(\bar{x}_1, \bar{x}_2, s)$  est estimée une fois pour toutes au début en déconvoluant l'ABR enregistré pour un clic (à forte intensité) du terme  $R(t, s)$ , qui est la simulation de la décharge du nerf auditif pour ce clic.

Un résultat capital de ce modèle, est la mise en évidence que toutes les parties de la membrane basiliaire n'ont pas le même poids dans le mécanisme de génération de la FFR. Ainsi, ce modèle montre que la partie basale de la cochlée (tonotopiquement responsable du codage des hautes fréquences) est la plus importante pour la FFR. Ce résultat est illustré en comparant la FFR engendrée par un son pur de 300 Hz dans les deux conditions suivantes : d'abord en ne gardant que la décharge du nerf auditif pour les canaux basses-fréquences (100-1.5kHz), puis en ne gardant que les hautes-fréquences (1.5 – 10 kHz). La seconde condition a donné lieu à une FFR bien mieux définie, Fig.1.21. La raison à cela serait la meilleure synchronisation à la base de la cochlée (codant tonotopiquement les hautes fréquences), contrairement à la dispersion des basses-fréquences le long de la transmission de l'onde mécanique jusqu'à l'apex.

Le modèle a bien reproduit la dépendance des ABR et FFR à l'intensité, mais a toujours trouvé les mêmes latences. Ceci doit être dû au fait que les réponses unitaires ne sont estimées qu'à une intensité assez forte, ce qui doit fortement limiter les estimations de non-linéarités du traitement auditif, et donc son effet sur les latences.

## 1.5 ABR : Avancées et Zones d'ombres :

- ***L'origine physiologique de la FFR et des ABR :***

C'est une information capitale en vue d'une utilisation clinique, afin de cor-

---

3. *L'ABR chez le chat est modélisé comme la convolution d'une réponse unitaire par une fonction de décharge instantanée, qui dépend de la population de cellules*

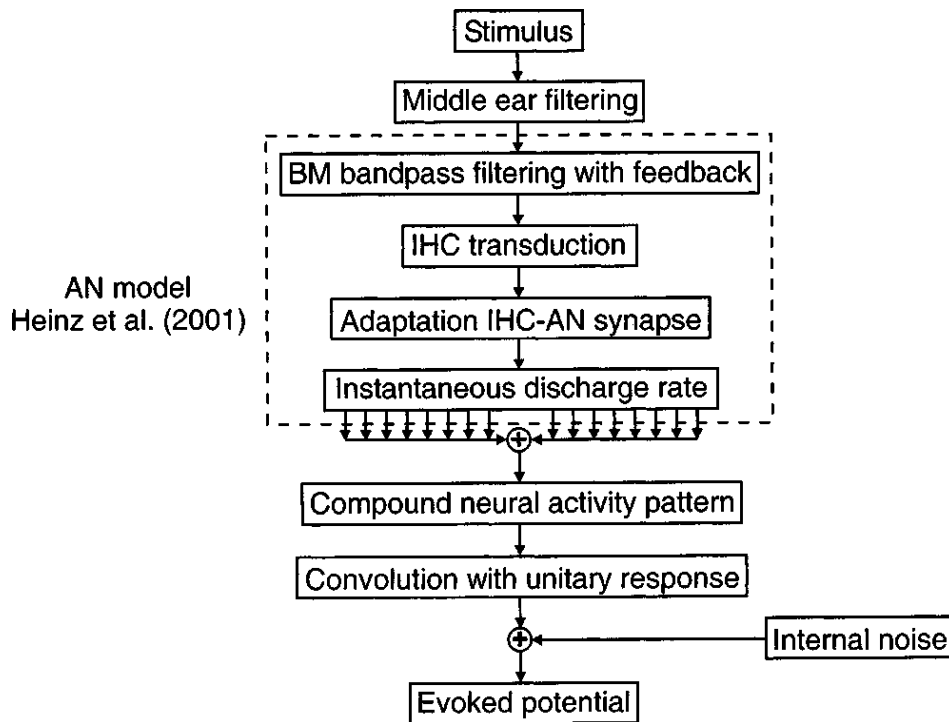


FIGURE 1.20 – Principe du modèle de FFR [Melcher 1996], [Dau 2003]

rectement cibler un éventuel dysfonctionnement. Il est communément admis que la FFR est le résultat d'un phénomène groupé de décharge cellulaire dans le système auditif suite à un son périodique. Peu pensent qu'il s'agit d'une succession d'ondes III et V réagissant à chaque cycle comme à une impulsion acoustique [Eggermont 2004], [Dau 2003], la plupart relie la FFR au phase-locking entre le noyau cochléaire, et/ou dans le colliculus inférieur. Picton et al (et tous les utilisateurs d'ASSR) sont les seuls à proposer une source principalement corticale de la FFR (ASSR) avec des modulations d'amplitude inférieures à 40 Hz, mais ce sont aussi les seuls à utiliser ces stimuli. La modélisation proposée par [Dau 2003], validée par la cohérence de sa simulation avec des enregistrements *in vivo*, indique que l'étape de la transduction cochléaire est fondamentale dans le processus aboutissant à la FFR, et toutes les bandes critiques n'y ont pas le même impact. Toutefois, dans son autocritique très bien documentée, il regrette de n'avoir pas pu modéliser précisément les étages du système auditif au delà du nerf auditif (au vu de la littérature physiologique chap. 1.1, on en imagine la difficulté). Jusqu'à présent (ce travail de thèse y-compris), toute discussion sur les générateurs des Speech ABR n'est que spéculation. Les enregistrements *in-situ* (tels que pratiques entre autre par [Cunningham 2002]) apportent des informations précises anatomiquement, mais qui ne sont transposables à l'Homme qu'en admettant l'extrapolation darwinienne. Or, dans ce cas, les processus de rétrocontrôle

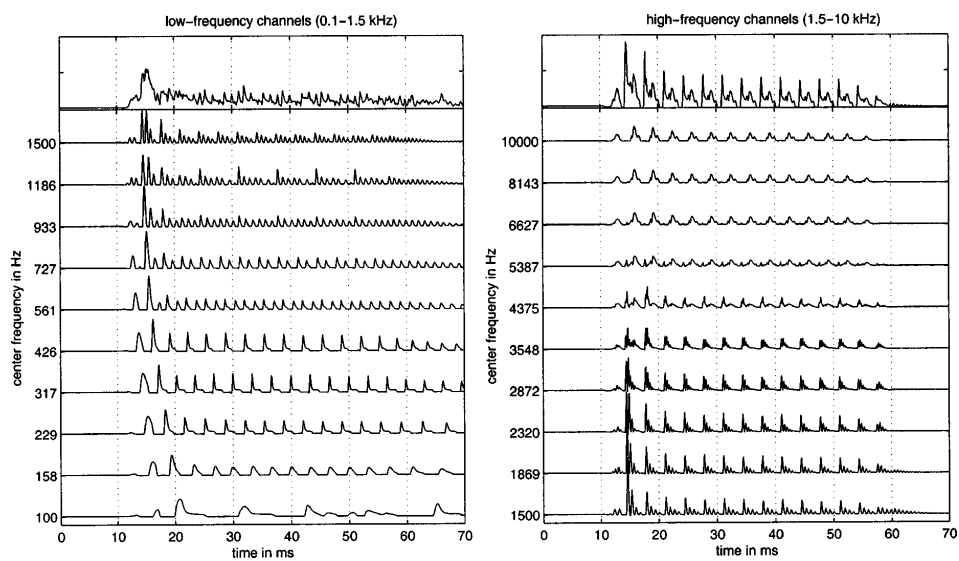


FIGURE 1.21 – Contribution des différentes bandes critiques (filtres auditifs) à la formation de la FFR. La FFR est principalement influencée par les bandes critiques hautes-fréquences, sans doute car elles sont codées en entrée de la cochlée, où les potentiels auditifs créés sont bien synchronisés au contraire des basses fréquences en raison du déphasage dû à la propagation de l'onde vibratoire le long de la membrane basilaire.[Dau 2003]

corticaux sont à coup sur très fortement perturbés par l'anesthésie. L'idée de modéliser, si elle apparaît titanique à implémenter, pourrait éclairer la question. C'est sans doute ce qui fait que les principaux contributeurs dans l'étude de la FFR (tel que Steve Greenberg) se sont longtemps investis dans les études physiologiques et à l'étude de la perception psychophysique de la parole.

- ***Quelles caractéristiques de la réponse ABR ou FFR utiliser ?***

Les avis divergent davantage pour l'utilisation des caractéristiques de la FFR comme fenêtre d'observation du traitement auditif. Kraus et al pensent que l'efficacité du traitement auditif se mesure en quantité de décharges synchronisées (ce qu'ils associent à des latences courtes) aux amplitudes des ondes temporelles ou des pics de la FFT, ou encore à une synchronisation de la FFR la plus proche possible des variations de F0 du stimulus (selon une technique de pitch-tracking issue de l'équipe Krishnan). Mais rien n'a encore montré que cette qualité de synchronisation était liée à un aspect quelconque de la perception (et le relier à l'expérience musicale pourrait être un non-sequitur). L'équipe de Krishnan, quant à elle, ne s'intéresse à aucun moment aux latences. L'équipe Picton est contrainte de se limiter aux informations spectrales, dont ils extraient l'amplitude, la phase (et les latences apparentes qui peuvent en découler dans le cas des sweeps). D'une manière générale toutefois, l'utilisation de l'amplitude pose problème avec la technique de moyennage, en particulier si on décide de normaliser les courbes de chaque sujet en vue d'étudier la grande moyenne.

- ***Dispositif expérimental***

En règle générale, le design expérimental est assez homogène (depuis [Hall 1979]), et ne fait pas l'objet d'une attention particulière. Trois électrodes par voies de recueil et un grand nombre de moyennages par blocs (le même stimulus présenté successivement) sur un sujet somnolant ou regardant un film pour fixer son attention, il est soit sur un fauteuil ou allongé sur un lit. Si, parfois, l'intervalle inter-stimulus est variable, personne n'a encore cherché à reproduire des conditions écologiques de la vie de tous les jours, en tentant de mélanger tous les stimuli d'une session expérimentale, ou en intégrant le stimulus dans des mots signifiants pour ce qui est des phonèmes. Les stimuli acoustiques mis en jeu pour mettre en évidence la FFR sont les sons purs (entre 200 et 975 Hz), les sons complexes harmoniques (avec ou sans F0, entre 122 et 488 Hz), les sons purs modulés en amplitude, les phonèmes consonne-voyelle (voire les mots très courts), les phonèmes de langues tonales (dits linguistic-pitch), leurs homologues IRN, et des notes de violoncelles. Dans tous ces cas, la FFR fut clairement définie, mais avec des amplitudes ou une précision de la synchronisation différentes. De très intéressantes études sur la perception des sons environnementaux captent de plus en plus d'attention, et les perspectives d'utilisation de ce matériel sonore sont déjà envisagées par Kraus.

- *Artéfact ou pas ?*

Sujet assez tabou et généralement éludé dans la plupart des publications (même si elle a interpellé les pionniers Fig. 1.13), l'authenticité de la réponse enregistrée sur le scalp n'est pas toujours assurée. S'agit-il de réponse neuro-physiologique auditive, d'artéfact électromagnétique, d'artéfact physiologique ou d'un mélange de ces éléments ? Dans certains cas, de l'artéfact fut clairement enregistré.

En effet, il est impossible notamment que le Speech ABR soit identifié lorsqu'il est transformé en fichier son [Galbraith 1995a]. En effet, le Speech ABR est passe-bas à 500 Hz (donc, en deçà de la limite inférieure de la gamme des premiers formants), donc non-identifiable si transformé en fichier son. Si on examine les résultats de [Galbraith 1994], l'absence de missing fundamental et les latences propres au nerf auditif sous-entendent que la FFR est apparue en même temps (ou presque – 1ms) que le stimulus, et que sa composition spectrale est la même que le stimulus, ce qui correspondrait à de l'artéfact électromagnétique. En revanche, la présence de FFR sur la disposition verticale des électrodes montrerait que dans ce cas, la FFR fut bien enregistrée, sans artéfact. En somme, de l'artéfact apparaîtrait sur la disposition horizontale, mais pas sur la disposition verticale. Ce risque d'artéfact rend l'exploitation de tout résultat impossible, en effet comme nous le montrons (Chap. 2.2) plus tard, la sensibilité à l'artéfact électromagnétique décroît pour les hautes fréquences, ce qui pourrait artificiellement augmenter l'amplitude des harmoniques supérieurs, par exemple. Une telle situation pourrait notamment biaiser le parallèle effectué par Krishnan [Krishnan 2008] ou Greenberg [Greenberg 1987] entre la prédominance du pic spectral de la FFR pour le second harmonique et le phénomène de dominance de hauteur tonale, où ils mentionnent avoir utilisé des inserts blindés magnétiquement Biologic (ou avec une cage de Mu-métal pour Greenberg, alors que ce matériau est utilisé pour piéger les contaminations magnétiques). Or nous montrons que l'artéfact est plutôt d'origine capacitive (lié aux deux champs électriques et magnétiques), et que la condition inserts sans cage de Faraday reliée à la masse n'est pas garantie sans-artéfact. Même si toutefois, nombre d'études n'ayant pas pris cette précaution drastique ne présentent pas de données artéfactées.

- *Est-ce que les ASSR ont rendu la FFR obsolète ?*

L'utilisation de sons modulés en amplitude, et l'hypothèse que la fréquence modulante (absente du spectre acoustique) va apparaître dans l'ASSR, rend de fait la méthode immune contre une influence directe des artéfacts.

L'hypothèse de départ des ASSR est assez séduisante. En effet, on peut être pratiquement sûrs qu'une haute fréquence va stimuler tonotopiquement un point précis de la cochlée, et que dans les fibres nerveuses partant de cet endroit précis, le signal auditif sera synchronisé par phase-locking sur la fréquence modulante. Toutefois, s'il est vrai que la limite du phase-locking di-

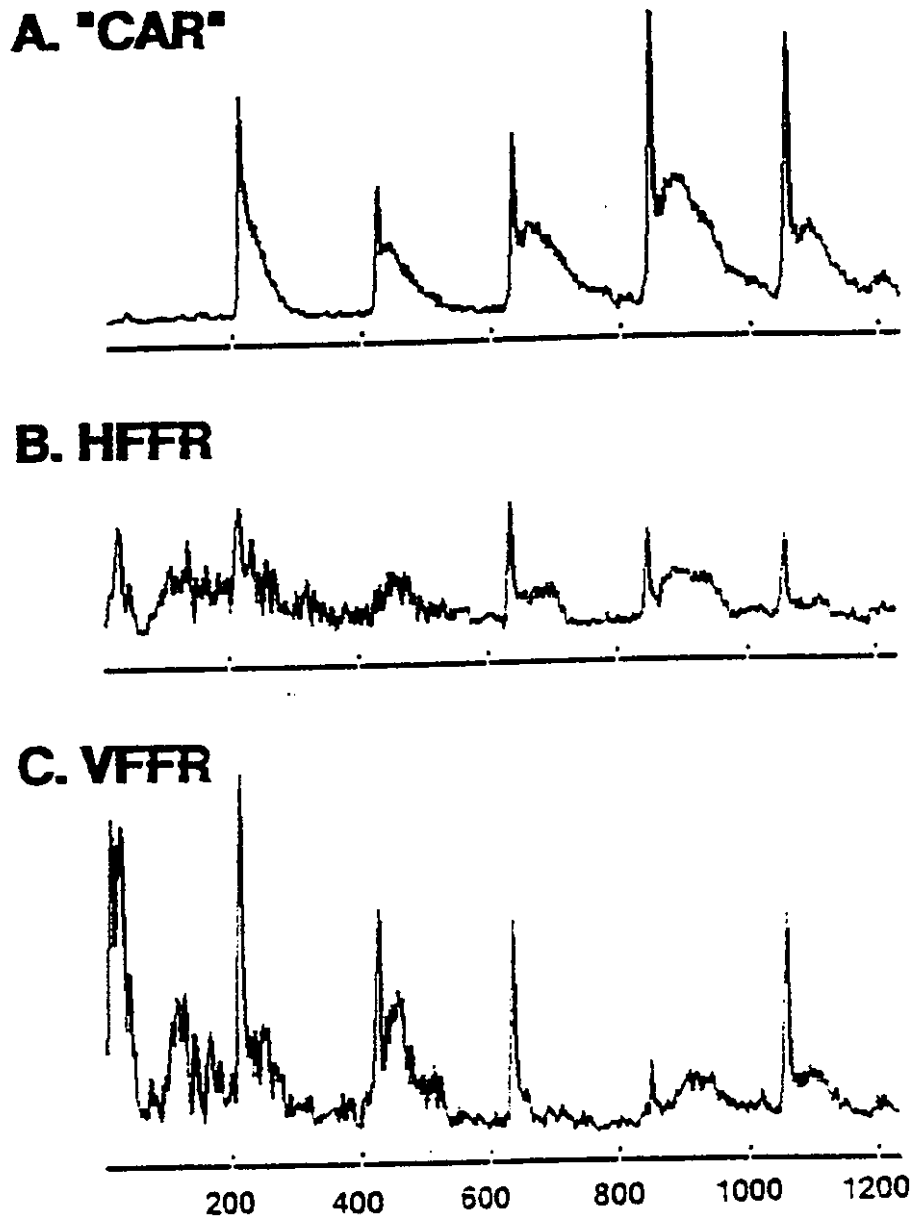


FIGURE 1.22 – Comparaison entre le spectre de la FFR et celui du stimulus /car/, dans l'étude où il est dit que la FFR était intelligible lorsque transformée en piste son. La (trop) remarquable correspondance spectrale est un signe de possible artéfact électromagnétique. [Galbraith 1995a]

minue du nerf auditif vers le cortex, il s'agit d'un phénomène passe-bas, donc une fréquence modulante de 40 Hz pourrait a priori être relevée dans n'importe quelle cellule, voire des le nerf auditif. Mais surtout, la provenance neurophysiologique de l'enregistrement des ASSR semblerait plus à même d'être conditionnée par les propriétés bioélectriques et l'organisation anatomique des cellules auditives [Eggermont 2007]. D'autre part, dans la technique d'EFR (et même pour les ASSR en général), toute détection du signal, ou mesure de latences doit se faire « à l'aveugle » (ce qui requiert une grande confiance dans les techniques de traitement du signal). Toutefois, l'hypothèse que le système auditif est synchronisé sur la fréquence modulante à « chaque instant » n'est valable que pour des sweeps très lents (comme ce fut le cas chez [Purcell 2004] : de 60 à 600 Hz en 30 s, 19Hz/s). Des cas d'adaptation et de fatigue neurale pourraient se faire sentir. L'utilisation des sweeps est assez avantageuse car le grand nombre de fréquences modulantes testées permet de mieux pouvoir gérer les ambiguïtés du à la circularité ( $2\pi$ ) de la phase, en repérant des tendances dans la courbe  $\phi = f(f_m)$ .

La critique principale vient du fait que l'utilisation de l'amplitude des ASSR est mise en cause par le principe de stopper l'enregistrement des ASSR dès que la réponse a été obtenue, il semblerait plus rigoureux de procéder au même nombre de moyennages dans toutes les conditions. Bien que la technique de moyennage par elle-même ne soit guère plus satisfaisante.

- **Correlats psychoacoustique de la FFR** : Hormis la revue de Greenberg sur la mesure de la hauteur tonale par la FFR, très peu de communications ont fait clairement le lien entre les caractéristiques du Speech ABR et la perception. Toutefois, la revue de Greenberg ne présente que des résultats préliminaires avec très peu de sujets, et même si le matériel a peu évolué, cette étude date. L'équipe Picton a montré des corrélations (timides) entre acuité temporelle (gap detection, tMTF) et les caractéristiques des EFR [Purcell 2004]. Une analogie très intéressante est celle entre l'importance des hautes fréquences dans la formation de la FFR [Dau 2003], [Greenberg 1987], et les travaux de [Gilbert 2007] montrant que la structure fine était très importante dans la perception de la parole, du fait de l'extraction de l'enveloppe temporelle (indice acoustique crucial) dans les bandes critiques hautes-fréquences.

- **Influence de la pathologie, de la surdité** :

Le travail de l'équipe Kraus est conséquent dans l'étude des troubles d'apprentissage, et leurs résultats ainsi que leurs limites sont très agréablement décrites dans des revues [Banai 2006], [Banai 2007].

Très peu de résultats font intervenir la perte auditive, hors [Dimitrijevic A 2004].

## 1.6 Objectif du travail

Pris dans leur ensemble, ces études sur les complex-ABR montrent des avancées importantes, et de larges zones d'ombre dans la connaissance des mécanismes physiologiques sous-jacents, ainsi que dans la recherche de corrélats psychoacoustiques ou dans l'étude de la perte auditive. Il semblerait que l'étude des FFR manque d'une approche pluridisciplinaire d'ampleur, comme le montre le peu d'envergure du matériel sonore ou des conditions expérimentales (voire des techniques de traitement du signal) mises en œuvre.

Comme ce premier chapitre le laisse présager, ce travail de thèse s'inscrit dans une thématique d'équipe globale qui est l'étude des mesures objectives de l'audition appliquées au diagnostic de la surdité et au réglage de l'implant cochléaire. A l'heure actuelle, les besoins sont très grands en mesures objectives fiables, en complément de la perception du patient, en particulier dans le cas de l'implant cochléaire, en vue de diagnostiquer les très jeunes enfants par exemple. Parmi elles, les ABR en réponse à des sons complexes sont un outil diagnostique très prometteur qui pourrait permettre d'évaluer la qualité du traitement de la parole au niveau du tronc cérébral.

L'objectif concret de ce travail aura été d'adapter le dispositif expérimental clinique Centor USB, habituellement utilisé en clinique pour le recueil des clic-ABR, au recueil d'ABR en réponse à des sons complexes. Dans une première approche expérimentale, le but aura été d'évaluer la fiabilité instrumentale du dispositif expérimental de mesure des Speech ABR, et de montrer l'authenticité neurophysiologique de la réponse Speech ABR. Puis dans une partie plus exploratoire, nous aurons évalué l'effet de différents paramètres acoustiques sur le Speech ABR, tels que l'intensité, enveloppe / structure fine, structure harmonique et nature des stimuli. Cette partie exploratoire permettant de croiser ces données avec les résultats physiologiques et psychoacoustiques, afin d'obtenir des indices permettant d'en savoir plus sur le processus sous-jacent qui lui donne naissance.





# Etude expérimentale : authentification du Speech ABR.

---

## Sommaire

---

<b>2.1</b>	<b>Dispositif expérimental de recueil des Speech ABR . . . . .</b>	<b>48</b>
2.1.1	Chaîne de stimulation : envoyer une stimulation acoustique . . . . .	48
2.1.2	Chaîne de recueil : enregistrer l'activité électrophysiologique . . . . .	50
2.1.3	Trigger : synchroniser le recueil sur la stimulation . . . . .	51
2.1.4	Principales méthodes d'analyse des signaux . . . . .	51
<b>2.2</b>	<b>Authentification des Speech ABR 1. Etude paramétrique de quantification de l'artéfact . . . . .</b>	<b>52</b>
2.2.1	Introduction . . . . .	52
2.2.2	Matériel et Méthodes . . . . .	54
2.2.3	Résultats . . . . .	55
<b>2.3</b>	<b>Authentification des Speech ABR 2. Etude comparative chez cophotiques unilatéraux . . . . .</b>	<b>63</b>
2.3.1	Introduction . . . . .	63
2.3.2	Matériel et Méthodes . . . . .	63
2.3.3	Résultats . . . . .	63
<b>2.4</b>	<b>Bilan : précautions à prendre et indices d'authenticité du Speech ABR . . . . .</b>	<b>68</b>
<b>2.5</b>	<b>Article Journal of Neuroscience Methods [Akhoun 2008d] . . . . .</b>	<b>69</b>

---

Notre travail portant sur les Speech ABR, nous allons envisager en détail cette technique. La technique de recueil des Speech ABR est assez simple, se rapprochant de près de l'enregistrement de potentiels évoqués auditifs, classique en routine clinique. On peut décomposer le dispositif expérimental sous la forme d'une chaîne de stimulation (qui délivre le stimulus), d'une chaîne de recueil (qui acquiert les enregistrements électrophysiologiques : le Speech ABR). Le trigger permet la synchronisation entre ces deux chaînes Fig. 2.15.

Notre problématique est que les électrodes sont soumises à deux sources électromagnétiques concurrentes : (i) le Speech ABR recherché, et (ii) les artéfacts électromagnétiques issus du couplage avec la chaîne de stimulation. Cette concurrence pose un grand problème ici, en raison de la simultanéité entre la stimulation et le Speech ABR, mais aussi vu la ressemblance entre la FFR et le stimulus. D'autre part, il s'agit d'un sujet relativement tabou, généralement éludé dans la littérature,

hormis la référence à la méthode de polarité alternée pour éliminer l'artéfact (ce qui, comme nous le montrons, n'est pas toujours suffisant), et l'usage de transducteurs dits-magnétiquement blindés (alors que, comme nous le montrons également, le type de couplage est capacitif, donc sensible à la composante électrique de l'onde électromagnétique rayonnée par les transducteurs). Du fait de ces approximations, il n'est actuellement pas possible de savoir si les données présentées dans la littérature sont libres de toutes influences artéfactuelles, comme les auteurs le sous-entendent de façon peu-franche. Les seuls indices allant dans le sens de données non-artéfactées sont les latences des ondes observées par rapport au stimulus, et l'aspect passe-bas des ABR ; et de ce fait, la plupart des données présentées dans la littérature semblent sans artéfact.

Après avoir présenté le dispositif expérimental, une étude paramétrique nous permet de quantifier l'artéfact électromagnétique (en fonction des transducteurs et de la valeur des impédances de contact). Enfin, dans une étude comparative entre les cophotiques unilatéraux et des normo-entendants contrôlés, nous montrons qu'il n'y a de Speech ABR, que si le stimulus a bien été entendu.

## **2.1 Dispositif expérimental de recueil des Speech ABR**

### **2.1.1 Chaîne de stimulation : envoyer une stimulation acoustique**

- **Les transducteurs**

Le son est délivré par un un casque, ou une paire d'insert-earphones Fig.2.1, qui transforment le signal audio en onde de pression. Il s'est avéré que la seule solution possible pour recueillir des Speech ABR était l'utilisation des insert-earphones Bio-Logic, que nous avons de surcroit blindés, afin de limiter les artéfacts électromagnétiques.

Les insert-earphones ont la particularité de maintenir le transducteur électro-acoustique à distance des électrodes de recueil du Speech ABR, et ainsi de limiter les effets perturbateurs d'interférences artéfactuelles (les électrodes enregistrent alors le signal audio véhiculé le long de la chaîne de stimulation). Le transducteur est du même type que ceux utilisés dans les prothèses auditives (Knowles). L'onde acoustique ainsi créée est ensuite véhiculée jusqu'aux oreilles du sujet par un tube acoustique, avec un dispositif en plastique à l'extrémité qui filtre le signal pour imiter le conduit auditif externe. Notons d'ores et déjà que deux désavantages des inserts sont, d'une part que l'éloignement de la source électromagnétique n'élimine pas totalement l'artéfact, et que l'onde acoustique se distord dans le conduit d'air, ce qui peut fortement déformer des signaux complexes (parole ou musique). Le principe de calibration des transducteurs, ainsi que la vérification du signal audio délivré par les insert-earphones sont présentés en Annexe A.

- **L'audiomètre**

L'audiomètre est un pré-amplificateur qui adapte le niveau sonore, et gère la binauralité de la stimulation. Le modèle utilisé fut un Interacoustics AC-30,

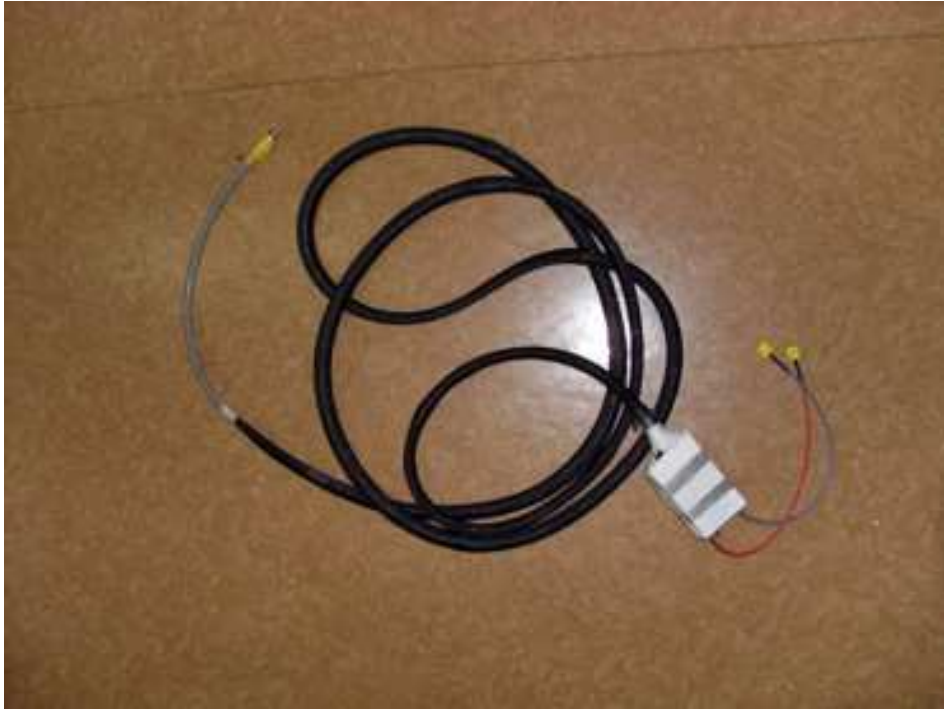


FIGURE 2.1 – Les transducteurs acoustiques insert-earphones avec le système de blindage que nous avons confectionné. Le son est véhiculé depuis l'audiomètre jusqu'aux transducteurs où la vibration de la membrane génère le stimulus à la base des deux tubes d'air. L'onde de pression acoustique se propage alors jusqu'aux oreilles du sujet, en passant à travers un filtre acoustique (en noir avant les earplugs jaunes) imitant les caractéristiques du pavillon de l'oreille externe. L'avantage des inserts est d'éloigner la source de rayonnement électromagnétique des électrodes en électrophysiologie. Toutefois, selon les dispositifs expérimentaux, cette précaution peut s'avérer insuffisante. Nous avons encastré les fils, ainsi que l'habitacle des inserts dans une cage de Faraday (tresse métallique et boîtier en aluminium), ne laissant dépasser que les tubes de caoutchouc.

1 voie.

- **La carte son**

La carte son (une Soundblaster Audigy), a pour but de convertir le stimulus à délivrer en signal analogique (16 bits, 96 kHz), exploitable par l'audiomètre.

- **Le logiciel de son**

Le freeware Audacity permet de synthétiser des signaux de base, ainsi que de manipuler de façon simple des sons pré-existants (synthétisés sous Matlab, par exemple). Les stimuli sont joués sous Audacity. Leur synthèse est décrite au cas par cas dans le détail des études ci-dessous.

### 2.1.2 Chaîne de recueil : enregistrer l'activité électrophysiologique

Cette chaîne de recueil est gérée par le Centor USB. De manière analogue, en remontant la chaîne de recueil, on trouve :

- **Les 3 électrodes**

Les électrodes utilisées sont des électrodes de contact de 10mm en Argent de Nicolet. Les électrodes sont fixées sur le scalp du sujet à l'aide d'une pâte EEG D.O. Weaver, après avoir ôté l'excès de sébum avec une pâte abrasive NuPrep. Ces électrodes furent tressées pour limiter leur sensibilité couplage électromagnétique inductif ambiant.

- **La tête**

Les électrodes sont branchées sur la tête, qui est un pré-amplificateur analogique (gain = 80 dB) suivi d'un convertisseur analogique numérique, qui conditionne les enregistrements sous un format prêt à être stockés, puis exploités informatiquement.

- **La console d'acquisition**

La tête est reliée (USB) à la console d'acquisition elle-même reliée au PC où le logiciel Centor représente l'acquisition de Speech ABR en continu tandis que les Speech ABR sont stockés dans la base de données associée (BDScan). La console d'acquisition reçoit également les impulsions du trigger afin de déclencher chaque enregistrement.

- **Le logiciel de pilotage du Centor**

Les paramètres de recueils sont fixés par l'intermédiaire du logiciel Centor. On peut ainsi choisir : le nombre de moyennages à effectuer, la durée de la fenêtre d'acquisition, la gamme de filtrage passe-bande, le seuil de réjection, et de suivre la valeur des impédances entre les électrodes. Notons au passage que le mode de filtrage est un filtrage en direct qui n'occasionne aucun retard. Parmi ces paramètres, les plus importants sont la sensibilité (qui doit être homogène sur une série de mesures), et la taille de la fenêtre d'acquisition (qui doit être la plus petite possible, et légèrement plus petite que la taille du stimulus).

Notons que le nombre de points de l'acquisition est fixe (4000 points), et que la fréquence d'échantillonnage est directement fonction de la taille de la fenêtre d'acquisition ( $F_{ch} = \frac{4000_{pts}}{W_{acquisition}}$ ). Enfin, la gamme de filtrage [80, 3200] Hz est sensée être centrée sur la gamme de réponse des cellules du tronc cérébral et d'éliminer les composantes basses fréquences, supposées provenir du cortex, et les composantes hautes fréquences, supposées provenir du bruit de mesure ambiant, ou de l'activité périphérique.

### 2.1.3 Trigger : synchroniser le recueil sur la stimulation

La technique de moyennage nécessite une synchronisation parfaite entre la fenêtre d'acquisition et le stimulus. Dans ce dispositif, nous avons choisi d'intégrer un marqueur de synchronisation à l'intérieur du fichier son, qui est ensuite directement relié à l'entrée de synchronisation du Centor. En pratique, on dispose une impulsion (signal carré biphasique) sur la piste de gauche, et le stimulus sur la piste de droite. En sortie (prise jack) de la carte son, une connectique (jack – double RCA) permet de séparer les deux pistes droites (stimulus) et gauche (trigger), l'un étant relié à l'entrée auxiliaire de l'audiomètre et l'autre à l'entrée de synchronisation du Centor, via un amplificateur analogique. De cette façon nous sommes assurés du calage entre la stimulation et l'enregistrement.

### 2.1.4 Principales méthodes d'analyse des signaux

La stimulation a systématiquement été faite en **polarité alternée** : une série moyennée sur une polarité du stimulus, et une autre sur l'autre. Cette technique a l'avantage d'éliminer une partie de l'effet de l'artéfact électromagnétique et microphonique cochléaire, car les deux polarités sont encodées identiquement, au contraire de l'artéfact 2.11.

Notons bien que les réponses neurales qu'on enregistre avec cette technique sont très faibles. Par conséquent, il ne faut pas chercher à trouver des composantes du Speech ABR pertinentes sur des traces individuelles. Il faut au contraire moyenner les traces individuelles (qui sont elle-même un millier de traces instantanées moyennées) afin de faire ressortir les composantes du Speech ABR. D'autre part, il convient de **normaliser** les traces individuelles (par loi centrale normée ( $\frac{(x-\bar{x})}{\sigma(x)}$ )), afin d'en harmoniser le poids. Notons par ailleurs que l'amplitude des Speech ABR n'est pas un indice pertinent en soi, car le moyennage et la normalisation les modifient. En revanche, les latences (abscisses) sont primordiales car elles sont directement reliées au site sous-jacent dans les voies auditives.

Pour le filtrage a posteriori (par exemple pour obtenir l'enveloppe par filtrage passe-bas) nous utilisons les fonctions *butter* et *filtfilt* sous Matlab, que ce soit pour les stimuli ou les Speech ABR, nous avons choisi les filtres de Butterworth, d'ordre 2 et de bande passante à notre guise. La fonction *butter* donne les coefficients du filtre à réponse impulsionnelle finie qui est ensuite généré par la fonction *filtfilt*. Cette fonction *filtfilt* est dite à correction de retard de phase. En effet, le signal, une fois

filtré est automatiquement retourné et filtré à nouveau afin de corriger le retard de phase, d'où un ordre en fait 2 fois plus important que celui demandé.

Une alternative intéressante au type d'analyse que nous avons effectuée est proposée par le système d'analyse des données comportementales et électrophysiologiques très ergonomique, facile de suivi et rigoureux : **OpenElectrophy**. Ce logiciel utilise un environnement de calcul Python et de base de données MySQL, et si ces langages nécessitent des bons pré-requis et un art de la débrouille en cas de bug, on gagne beaucoup de temps et d'efficacité dans le traitement des données par la suite. Libre à l'utilisateur ensuite de développer leurs propres scripts sous Matlab (ou Python) pour eux-mêmes. Pour notre part, nous sommes restés sur nos propres scripts Matlab, faute de temps pour mettre en place la configuration d'Open Electrophy pour notre utilisation.

## 2.2 Authentification des Speech ABR 1. Etude paramétrique de quantification de l'artéfact

### 2.2.1 Introduction

De façon générale, toutes les techniques d'enregistrements électrophysiologiques sont sensibles aux conditions de recueil, et doivent composer avec des perturbations ambiantes (artéfacts) qui menacent d'en fausser l'interprétation. Ces artéfacts peuvent être de différents types : musculaires, liés à l'alimentation électrique, dûs à d'autres phénomènes physiologiques sans intérêt vis-à-vis de l'étude, ... Dans la plupart des cas, des techniques d'instrumentation permettent de limiter leurs effets. Si ce n'est toujours pas suffisant, des techniques de traitement du signal peuvent isoler le signal recherché de l'artéfact, comme l'analyse par composantes principales (ou indépendantes), ou encore les connaissances a priori de l'artéfact et du signal recherché [Delorme 2004].

Dans le cas du Speech ABR, la simultanéité entre la stimulation et l'enregistrement fait que le risque d'artéfact électromagnétique est très grand. Il s'agit de la principale source d'artéfact particulièrement nuisible, du fait de la forte ressemblance entre le Speech ABR et le stimulus. Cet artéfact est lié aux problèmes de compatibilité électromagnétique entre la chaîne de stimulation (en majeure partie les transducteurs acoustiques) et les électrodes de recueil. Notons au passage que cet artéfact n'existe pas avec les clic-ABR, car dans ce cas, l'ABR est enregistré après que le stimulus impulsionnel ait été délivré. Cet artéfact résulte en la superposition du signal neural et du couplage électromagnétique des transducteurs, Fig. 2.2. Par ailleurs la précision, souvent spectaculaire, de la synchronisation entre la FFR et la voyelle pose la question de l'authenticité de cette réponse neurale [Marsh 1970]. Certains l'accusent de couplage électromagnétique entre la chaîne de stimulation et les électrodes de recueil. Est-ce que la méthode de polarité alternée, seule solution technique mentionnée dans la littérature, suffit à annuler efficacement les effets de l'artéfact ? Nous montrons que ce n'est pas le cas. Par conséquent, il est

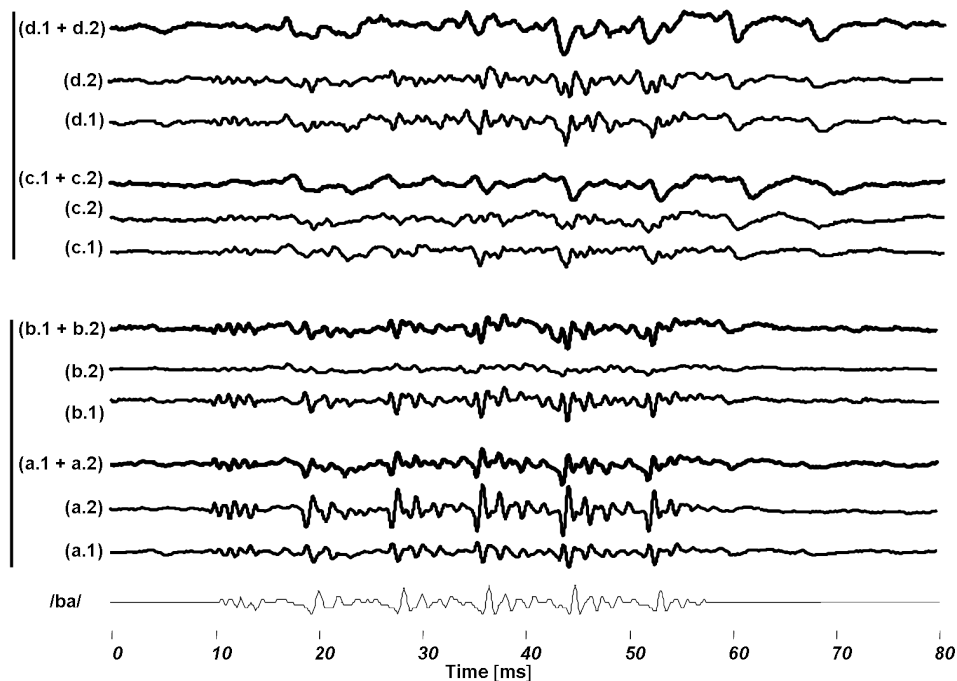


FIGURE 2.2 – Exemple de recueils de Speech ABR artéfactés, montrant la limite de la polarité alternée. Les Speech ABR sont représentés pour chaque polarité (positive, négative et alternée), pour 4 sujets (a, b, c et d), sur le même principe que la Fig.3.3. De la même façon que sur la Fig.2.6, l'artéfact s'est caractérisé par une copie conforme du stimulus, non décalé ni filtré sur les enregistrements, et par une inversion avec la polarité (a1 et a2 sont par exemple clairement en opposition de phase). Deux conditions se dessinent pour ces 4 sujets : les sujets (c et d) ont eu autant d'artéfact sur les 2 polarités et par conséquent, la polarité alternée (c1+c2 et d1+d2) a bien réussi à éliminer l'artéfact ; les sujets (a et b) quant à eux ont eu plus d'artéfact sur une polarité que sur l'autre, ce qui a donné lieu à de l'artéfact résiduel sur leurs polarités alternées (a1+a2 et b1+b2). [Akhoun 2008b]



important de vérifier l'absence d'artéfact avant de procéder à des enregistrements de Speech ABR, sachant que les transducteurs, ainsi que des impédances de contact sont les principaux facteurs influant sur l'apparition de l'artéfact électromagnétique.

### 2.2.2 Matériel et Méthodes

L'effet des transducteurs et des impédances de contact entre les électrodes fut évalués en fonction de la fréquence et de l'intensité de stimulation. Quinze situations furent testées : cinq conditions de transducteurs, et trois conditions d'impédances. Nous avons choisi un modèle très simple pour se placer dans une configuration électromagnétique assez proche du cas réel, mais de telle sorte à ne recueillir que de l'artéfact et de se passer de toute réponse neurale. A ce titre, nous aurions pu opter pour des expérimentations post-mortem, mais notre choix s'est porté sur une pastèque, dont le diamètre se rapproche des distances inter-aurales, et dont l'impédance mesurée avec le Centor USB donne des valeurs de l'ordre du  $k\Omega$ , dans l'ordre de grandeur de ce qui est habituellement observé chez les sujets humains.

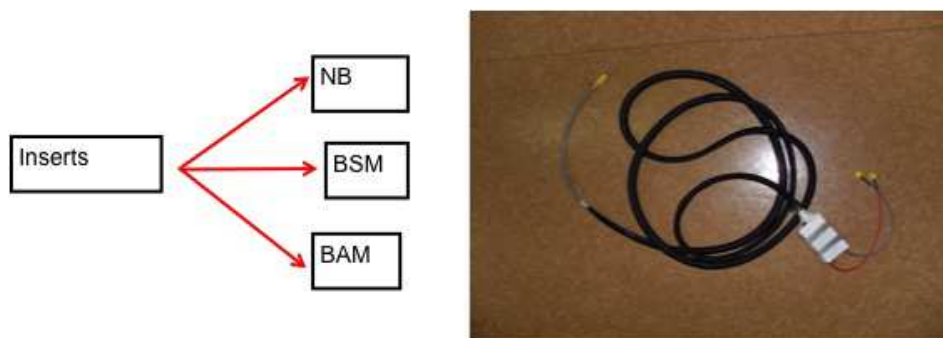


FIGURE 2.3 – **Conditions testées concernant les insert-earphones.** Les inserts furent testés dans 3 conditions de blindage : blindé avec blindage relié à la masse (BAM, dit aussi « Faraday-on »), idem sans relier à la masse (BSM ou « Faraday off »), ou sans aucun blindage (NB, ou « non-shielded »).[Jeanvoine 2008]

**Quinze situations** furent donc testées, avec *cinq cas de transducteurs audio* (1, 2, 3a, 3b et 3c) différents, et *trois conditions d'impédances* (i, ii, iii). Trois transducteurs parmi les plus couramment utilisés en audiologie, furent testés : deux casques Sennheiser Linear 250 (1) et TDH 39 (2), ainsi qu'une paire d'insert-earphones. Ces insert-earphones, Fig.2.3, étaient soit tels-quels libre de tout blindage électromagnétique (3a), soit encastrés dans un blindage qui n'était pas relié à la masse commune du dispositif (3b), soit avec le blindage relié à la masse (3c). Le couplage électromagnétique émanant de ces 3 transducteurs fut mesuré selon 3 impédances de contact limites Fig.2.4 : électrodes plongées en court-circuit dans un bac rempli de liquide physiologique (i, impédances nulles),

électrodes fixées sur un melon avec de la pâte de contact EEG (ii, impédances de l'ordre du  $k\Omega$ ), et enfin électrodes à l'air libre en circuit-ouvert (iii, impédances quasi-infinies). Nous avons mesurés les seuils de niveau sonore d'apparition de

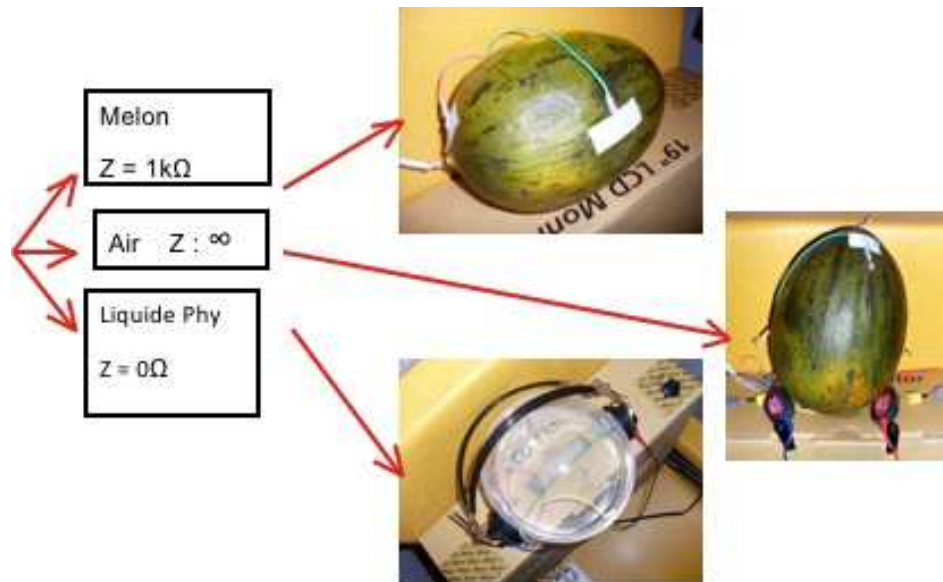


FIGURE 2.4 – Les 3 conditions d'impédances testées. Voir texte. [Jeanvoine 2008]

l'artéfact dans chacune de ces 15 conditions, et en fonction de la fréquence. L'artéfact n'était pas apprécié en fonction de son amplitude, mais de façon « binaire » : présent ou absent Fig.2.5. Pour cela, nous avons vérifié sur le spectrogramme du signal recueilli sur les électrodes que la composante fréquentielle relevée à la fréquence du stimulus était au moins 2 fois plus importante (6dB) par rapport au niveau spectral moyen du signal au voisinage (méthode du F-test [Picton 2003a]).

### 2.2.3 Résultats

Les résultats clés de cette expérimentation sont : d'une part que l'artéfact électromagnétique apparaît d'autant plus facilement que l'impédance de contact des électrodes est élevée, et d'autre part que les transducteurs les moins susceptibles de donner lieu à du couplage électromagnétique sont les insert-earphones 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. Tous les autres transducteurs risquent d'induire de l'artéfact à des fréquences et intensités de stimulation habituellement utilisées, Fig.2.7. Il est bien clair que le fait de blinder les inserts, d'améliorer le contact des électrodes et d'éloigner les sources de courants externes ne peut être que bénéfique... mais il convient de quantifier l'apparition de l'artéfact en fonction du niveau sonore, ainsi que la mise en évidence

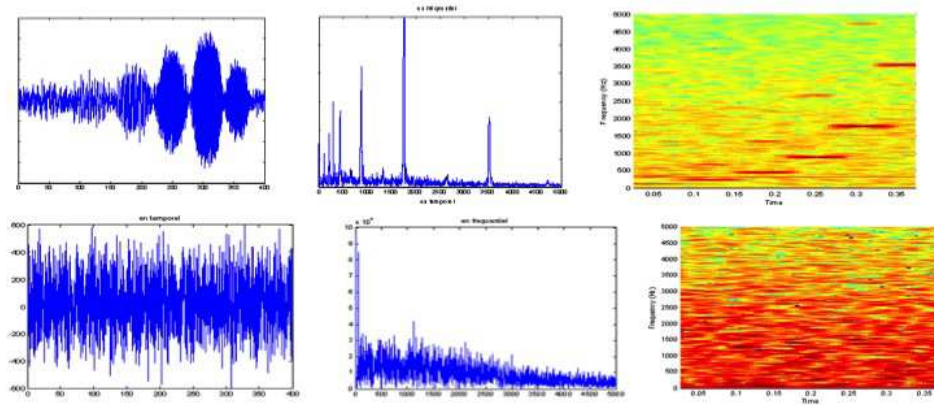


FIGURE 2.5 – **Détection de l'artéfact sur les enregistrements.** L'apparition de l'artéfact fut mesurée en fonction de la fréquence : le son utilisé pour cette manipulation était une succession de 6 sons purs balayant le spectre de la parole : 110, 220, 440, 880, 1760, 3520 Hz. Chaque amplitude de ces sons purs était adaptée en fonction de la courbe de réponse des différents transducteurs évaluée grâce à l'appareil de calibration Larson-Davis, afin d'avoir des niveaux sonores identiques pour chaque fréquence, que nous pouvions dès lors exprimer rigoureusement en dB Sound Pressure Level (SPL = par rapport à la valeur de référence  $2 \cdot 10^{-5}$  Pa). Dans chaque condition (impédance \* transducteur) testée, et pour chaque intensité, l'artéfact est jugé présent dès qu'une des composantes du stimulus (temporel (gauche), spectral (centre), temps-fréquence (droite) en haut) est présente dans l'enregistrement. Lorsqu'aucun pic n'est obtenu au delà de 6 dB par rapport au bruit de mesure, aucun artéfact n'est considéré présent (en bas). [Jeanvoine 2008]

de la présence d'artéfact même pour les inserts lorsqu'ils ne sont pas blindés. Nous proposons ici une alternative intéressante à l'utilisation d'inserts avec de très longs tubes (et donc une très grande distorsion du signal, et une très forte atténuation au-delà de 1kHz) du type Etymotic ER 30.

Les courbes d'apparition de l'artéfact (détection visuelle) furent comme prévu sensibles à la nature des transducteurs, à l'impédance de contact et aussi de façon moins attendue, à la fréquence du stimulus. Pour ce qui est de la nature des transducteurs Fig.2.7, la tendance n'est pas aussi simple que de dire que les insert-earphones sont par nature plus robustes que les casques à l'artéfact. En effet, nous avons obtenu de l'artéfact le plus facilement à de faibles intensités (aussi basses que 45 dB SPL) pour les insert-earphones. En somme, il ressort que la Cage de Faraday est un moyen unique pour ne pas avoir d'artéfact. Notons par ailleurs que le blindage s'il n'est pas relié à la masse (condition Faraday-off) donne lieu à un artéfact important. Les casques habituellement utilisés en audiologie (TDH 39 ou Sennheiser 250) sont donc à proscrire dans l'enregistrement de Speech ABR, et ce malgré leurs caractéristiques de réponse très performantes, en raison du risque d'artéfact qui peut en résulter, et ceci, en particulier à haute intensité sonore, avec des patients malentendants par exemple.

Cette expérimentation confirme le résultat bien connu, selon lequel de hautes impédances de contact des électrodes favorisent l'apparition d'artéfact, Fig.2.8. Mais ce qui est frappant, c'est de constater que le cas réel (impédances proches du  $k\Omega$ ) est plus proche du circuit ouvert (impédances infinies) que du cas court-circuit (impédances nulles). Ceci est cohérent avec le type de couplage capacitif, qui est prépondérant dans ce cas d'artéfact étant donné la gamme de fréquence étudiée (aux alentours du kilohertz). Il sera donc bon de toujours veiller à conserver les impédances les plus basses possibles lors de l'acquisition de Speech ABR.

Le niveau limite d'apparition de l'artéfact est également sensible à la fréquence du stimulus, malgré la correction apportée suite à la prise en compte de la courbe de réponse en fréquence des transducteurs. Les courbes d'apparition de l'artéfact semblent décroître avec la fréquence : avec un son complexe harmonique, on peut avoir de l'artéfact pour les harmoniques élevées (au delà d'1kHz), sans pour autant en avoir pour le fondamental (voire les premiers harmoniques). Notons les conséquences que de tels artéfacts peuvent avoir dans une étude comparant l'amplitude des harmoniques de la FFR (voir chap. ?? ).

Le modèle de la pastèque que nous avons choisi pour simuler la susceptibilité électromagnétique de la tête humaine s'est avéré assez satisfaisant. En effet, en comparant les seuils d'apparition de l'artéfact moyens sur 6 sujets auxquels les insert-earphones étaient disposé normalement sans avoir relié les tubes d'air aux sujets (condition « inserts-out-of-the-ear », Fig.2.9), les courbes d'apparition de l'artéfact sur la pastèque ou chez les sujets (en moyenne) affichèrent deux profils très proches et très semblables en fonction de la fréquence, Fig.2.10. Ces courbes ne furent jamais plus éloignées que 12 dB-SPL à 440 et 3520 Hz. De plus, les erreurs standards entre le niveau sonore d'apparition de l'artéfact sur la pastèque et sur les sujets varia peu pour les inserts non-blindés entre 2 (pour 440 Hz) et 4 dB SPL (pour 220 Hz). Et

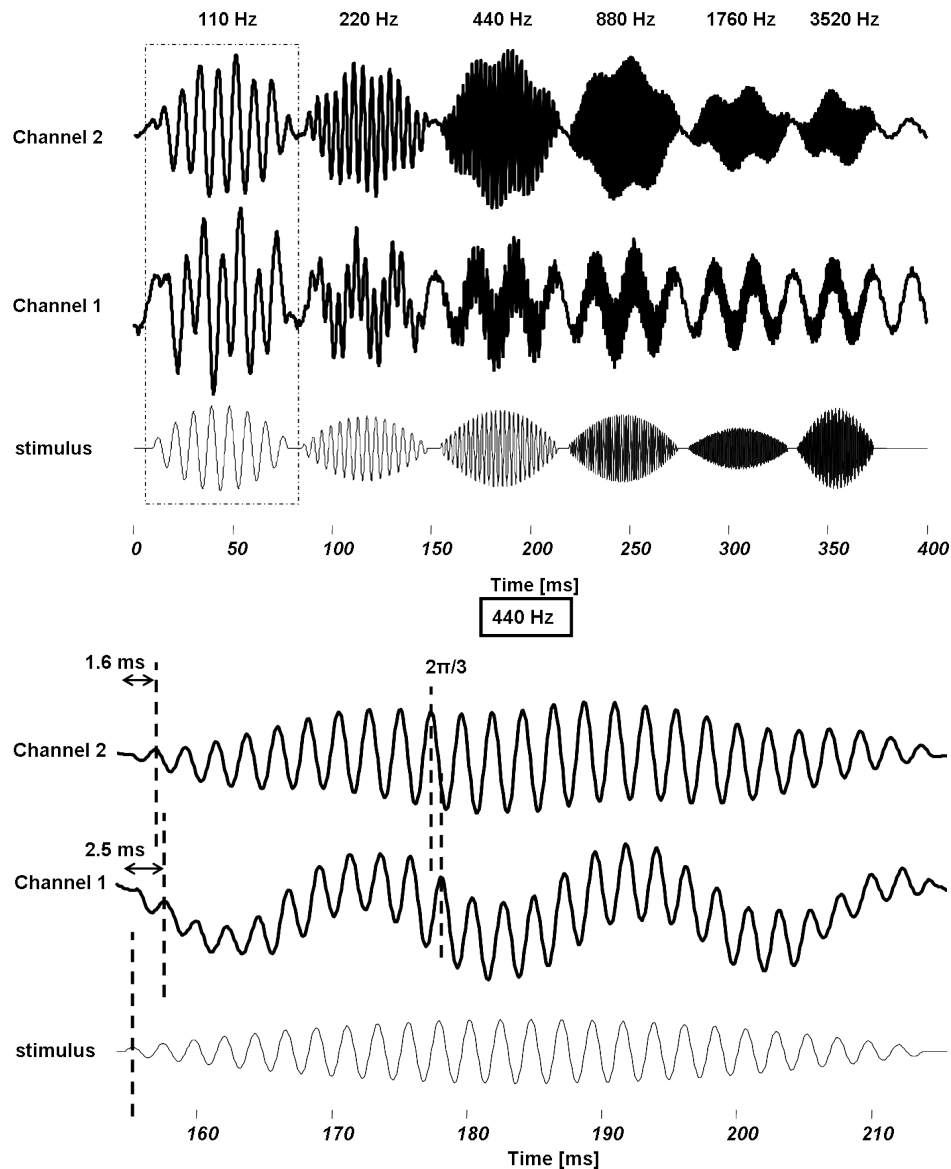


FIGURE 2.6 – Un exemple de recueil d’artéfact sur le modèle de la pastèque, et sur 2 voies. La forme exacte du stimulus se retrouve dans l’enregistrement. On remarque aussi une modulation de l’enregistrement dû à l’influence du rayonnement des alimentations électriques (50 Hz). De plus, on constate un retard de phase (mesuré à 3.5 ms) entre l’enregistrement et le stimulus, particulièrement visible à 110, 220 et 440 Hz (zoomé en bas), qui semble très difficile à expliquer, tout en sachant que ce retard de phase n’existe plus pour les hautes fréquences... On constate que l’artéfact obtenu sur la voie 1 et celui obtenu sur la voie 2 sont différents (et même en fort déphasage de  $2\pi/3$  à 440 Hz à droite, mais pas pour les fréquences plus hautes, non-représentées). [Akhoun 2008b]

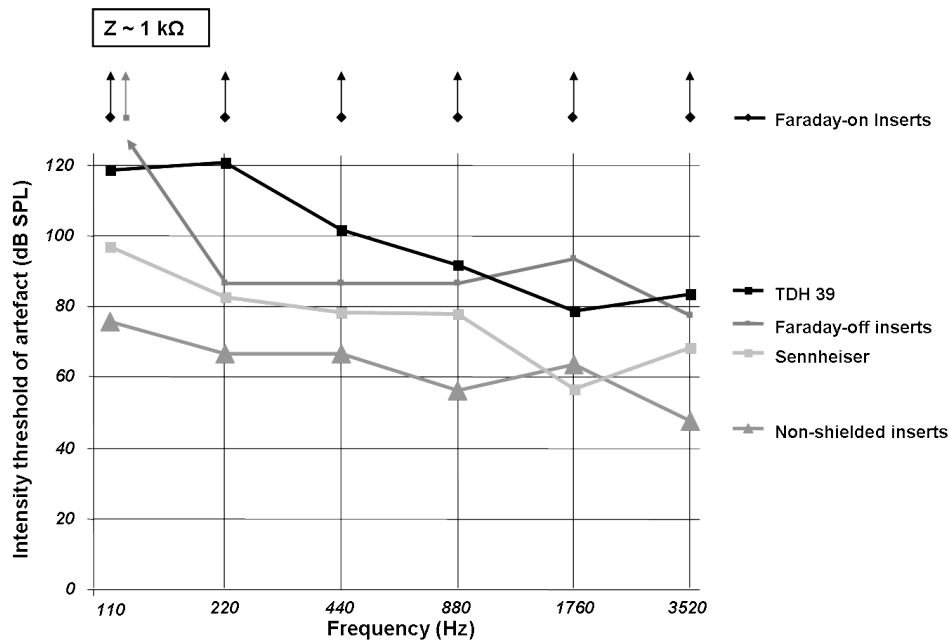


FIGURE 2.7 – Effet de la nature des transducteurs sur l'apparition de l'artéfact. Chaque point représente le seuil d'intensité au delà duquel de l'artéfact fut constaté, en fonction de la fréquence, et pour chacune des 5 conditions de transducteurs considérées. Sur cette figure, seul les seuils obtenus dans la condition « impédances réelles » sont représentés : ils étaient plus bas dans la condition circuit-ouvert, et plus élevés dans la condition électrodes en court-circuit. Notons que seul les inserts blindés n'ont donné aucun artéfact, ce qui fut vérifié dans les 3 conditions d'impédances.[Akhoun 2008b] NB : Voir la vidéo de démonstration en annexe électronique.

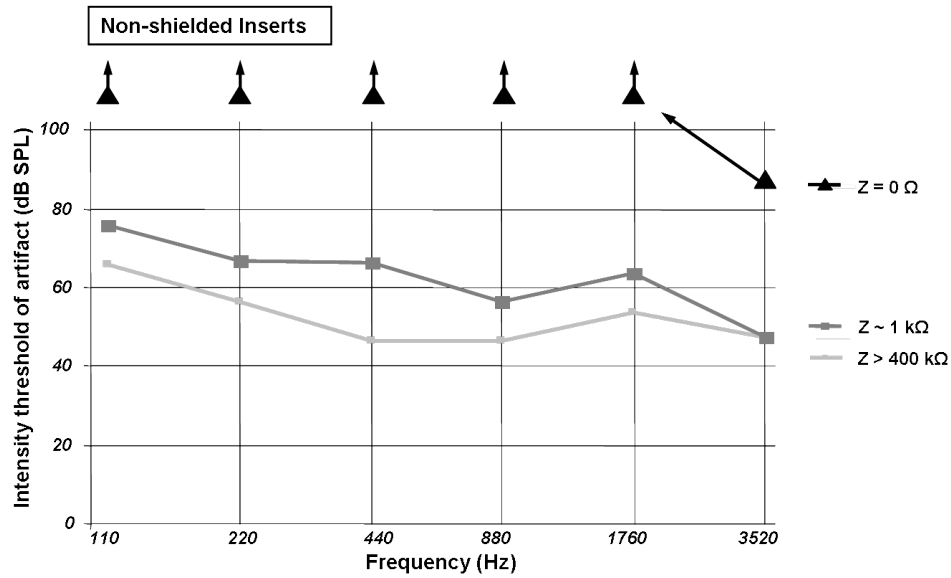


FIGURE 2.8 – **Effet des impédances de contact sur l’artéfact électromagnétique** Chaque point représente le seuil d’intensité au delà duquel de l’artéfact a été enregistré, pour chaque fréquence (échelle logarithmique), et cela pour les 3 conditions d’impédances considérées. Sur cette figure, ne sont représentés que les seuils d’apparition de l’artéfact pour les inserts non-blindés (tendances similaires pour les autres transducteurs). Plus les impédances sont élevées, plus facilement est recueilli l’artéfact. [Akhoun 2008b]

dans la condition inserts blindés mais sans que le blindage ne soit relié à la masse, l'erreur standard varia entre 4 (pour 440 Hz) et 7 dB SPL (pour 880 Hz). S'il est vrai que l'allure globale des courbes d'apparition de l'artéfact en fonction de la fréquence s'apparient bien entre les deux substrats (pastèque et scalp), nous avons plus de mal à expliquer pourquoi inserts non-blindés et inserts blindés mais non-reliés à la masse eurent des courbes quasi-superposées pour les sujets humain, tandis qu'un écart de 20 dB fut trouvé pour la pastèque. Cette incertitude, si elle ne va pas à l'encontre de notre hypothèse, confirme la nécessité d'étudier plus en détail le comportement des champs électromagnétiques en jeu (au travers d'une modélisation par éléments finis, par exemple).



**Insert-out-of-the-ear =  
Insert non positionné  
dans l'oreille**

FIGURE 2.9 – La condition contrôle « insert-out-of-the-ear ». Dans ce cas, les inserts sont positionnés normalement, mais les tubes d'air ne sont pas reliés aux oreilles du sujet, qui n'entend pas la stimulation. Dans cette situation, si aucun Speech ABR n'est obtenu contrairement à la condition où le sujet entend, la méthode de recueil est validée, et on peut dire qu'on enregistre bien une réponse liée à l'audition du stimulus. [Jeanvoine 2008]

Il est important de noter que l'artéfact recueilli varia assez notablement sur les deux voies de recueil Fig.2.6. Si ce n'est guère surprenant pour le Sennheiser, dont seul un côté délivrait la stimulation, une telle disparité s'explique moins facilement pour les autres conditions plus symétriques. Une autre limite de cette approche peut aussi résider dans le fait qu'on considère l'artéfact seulement en termes de présence ou absence, et qu'on ne prend pas en compte l'amplitude de cet artéfact. D'autre part, des techniques de modélisation de problèmes directs (où les sources sont connues) pourraient nous permettre d'évaluer plus précisément ce qui relève



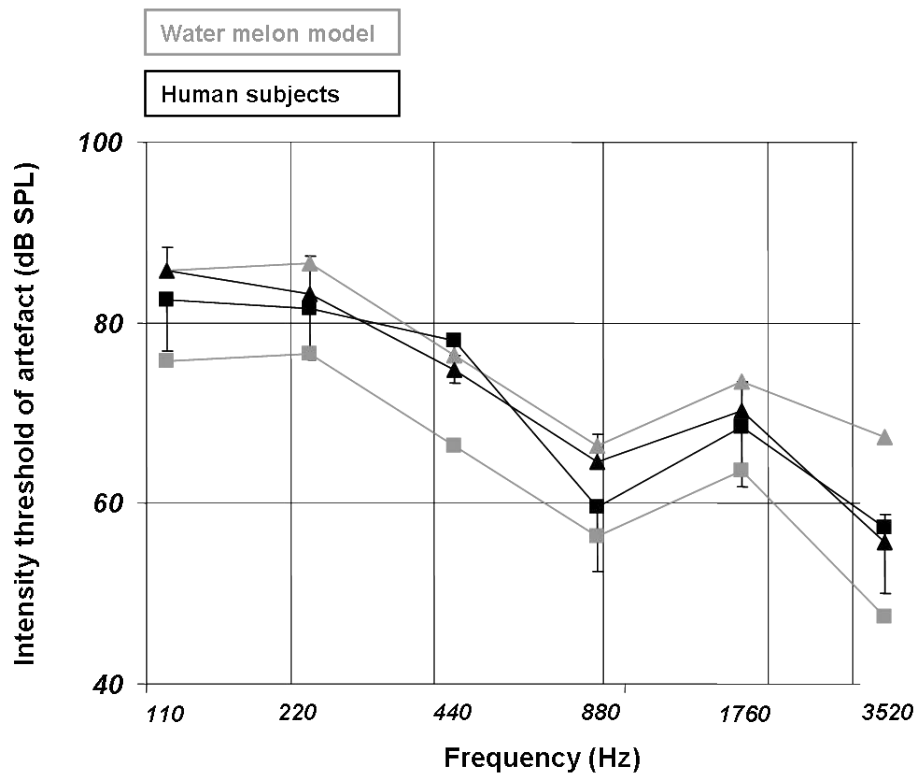


FIGURE 2.10 – **Le modèle de la pastèque est une bonne approximation de la susceptibilité d’une tête humaine.** Comme sur les Fig. Fig.2.7 et 2.8, chaque point représente le seuil où de l’artéfact a été enregistré, en fonction de la fréquence, et ce, pour les 3 conditions d’insert ( $\triangle$  : inserts non-blindés,  $\square$  : cage de Faraday non-reliée à la masse). Les courbes correspondant aux enregistrements sur la pastèque sont en clair, et les moyennes (ainsi que les erreurs standards) de ceux obtenus chez 6 sujets normo-entendants sont en noirs. On remarque une nette proximité, et un comportement identique en fonction de la fréquence, entre les courbes noires et claires, pour chaque condition d’insert. En particulier, aucun artéfact ne fut recueilli dans la condition inserts blindés (non-représentée ici). Notons toutefois, qu’il est difficile d’expliquer l’écart d’environ 20 dB entre les 2 condition d’insert ( $\square$  et  $\triangle$ ) pour la pastèque, alors que l’artéfact apparait quasiment pour les mêmes valeurs pour ces 2 conditions chez les sujets. D’après [Akhoun 2008b]

du couplage inductif et du couplage capacitif. L'artéfact électromagnétique est sans doute un mélange des deux, et en voulant réduire l'un, on pourrait favoriser l'autre. Nous avons trouvé une bonne solution en utilisant la cage de Faraday pour limiter le couplage capacitif, et nous avons tressé les électrodes pour limiter le couplage inductif. Pour conclure, nous avons quantifié la présence d'artéfact électromagnétique, mis en évidence une façon sûre de ne pas avoir d'artéfact, qui permet de s'affranchir de l'utilisation de longs inserts qui distordent très fortement le signal audio et rendent l'utilisation de sons complexes (parole, musique) quasiment impossible.

## **2.3 Authentification des Speech ABR 2. Etude comparative chez cophotiques unilatéraux**

### **2.3.1 Introduction**

Réalisée dans les conditions optimales trouvées précédemment (inserts earphones blindés par une cage de Faraday), cette étude nous a permis d'apporter des preuves supplémentaires que le Speech ABR était bien une réponse neurale. Pour cela, le but était de montrer que le Speech ABR était enregistré dans l'oreille entendante, et que le Speech ABR obtenu lorsque le son n'était pas entendu n'était pas différenciable du bruit ambiant (qui est estimé pendant la période de silence avant le moyennage : pré-moyennage). Les patients sourds unilatéraux ont l'avantage d'être leurs propres témoins (oreille saine vs. oreille sourde).

### **2.3.2 Matériel et Méthodes**

Les Speech ABR furent mesurés dans chaque oreille chez 6 normo-entendants. L'enregistrement de Speech ABR dans chaque oreille d'un cophotique unilatéral (6 sujets) eut l'avantage de confirmer que le Speech ABR ne peut être recueilli que s'il y a bien eu perception sonore. D'autre part, nous avons enregistré les Speech ABR en condition contrôle insert-out-of-the-ear (Fig. 2.9). Le stimulus utilisé était la même syllabe /ba/ de 60 ms que celle utilisée dans les études Chap : 3.1 et 3.5, précédée d'une période de silence de 10 ms (afin d'estimer le bruit de mesure ambiant).

### **2.3.3 Résultats**

Les détails statistiques de cette étude, ainsi que les Speech ABR individuels recueillis, sont présentés dans l'article [Akhoun 2008d]. Les résultats ont mis en évidence que chaque polarité du stimulus (/ba/ ou -/ba/) donnait lieu à des Speech ABR très semblables Fig.2.11, ce qui suppose un encodage cochléaire identique, comme on s'y attend d'après la mécanique cochléaire de transduction du stimulus acoustique, qui fait que l'oreille n'est que très peu sensible à la phase du son. Aucune

différence ne fut visible entre les Speech ABR obtenus dans l'oreille droite ou gauche d'un normo-entendant.

Le rapport signal sur bruit (SNR, en dB) du Speech ABR fut calculé comme le rap-

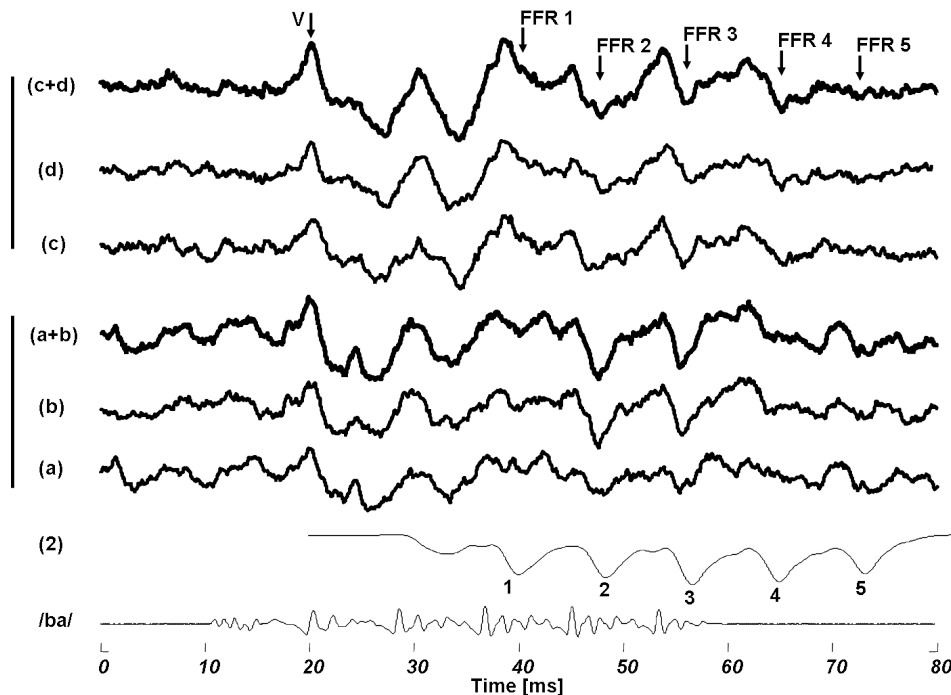


FIGURE 2.11 – **Grande-moyenne (n=6) des Speech ABR obtenus dans chaque oreille d'un normo-entendant.** Cette figure fait écho aux données de l'étude (4.1), Fig.3.3 et 2.11 . Les Speech ABR induits via l'oreille gauche (a, b) et via l'oreille droite (c, d) sont représentés dans les 2 polarités. Les polarités positives (a, c) et négatives (b, d) donnent la polarité alternée (a + b, c + d). Les composantes (onset response et FFR) sont précisées par les flèches. La grande similarité entre toutes ces courbes traduit d'une part que le Speech ABR est insensible à l'inversion de polarité, et d'autre part qu'il est identique sur les deux oreilles. On retrouve l'enveloppe (2) décalée dans la forme de la FFR (mais pas de 14.6 ms, comme c'est le cas Fig.3.3). Comme nous le précisons dans l'étude 3.1, chaque onde de la FFR (FFR 1 à FFR 5) est supposée être engendrée par chaque dépression de l'enveloppe (1 à 5), théorie par ailleurs remise en cause dans la dernière étude 3.2.2. [Akhoun 2008b]

port de la moyenne quadratique de l'amplitude du signal correspondant au Speech ABR sur la moyenne quadratique durant la période de pré-moyennage (Root-Mean-Square, RMS). Ce SNR (Fig.2.12) était significativement plus grand pour un Speech ABR généré dans l'oreille normo-entendante que dans l'oreille sourde d'un cophotique unilatéral, ou encore dans la condition insert-out-of-the-ear. Ce SNR était en revanche proche de celui obtenu chez le normo-entendant (voire même un peu supé-

rieur). Ces observations furent confirmées par le calcul de la grande moyenne : les SNR étaient encore plus grands lorsque le stimulus était perçu, et il était plus petit, voire négatif dans le cas de l'oreille sourde ou avec les inserts déconnectés de l'oreille. Notons toutefois que dans le cas 'none-NH', la grande moyenne a augmentée le SNR de cette condition contrôle, sans doute dû à l'aspect aléatoire du bruit.

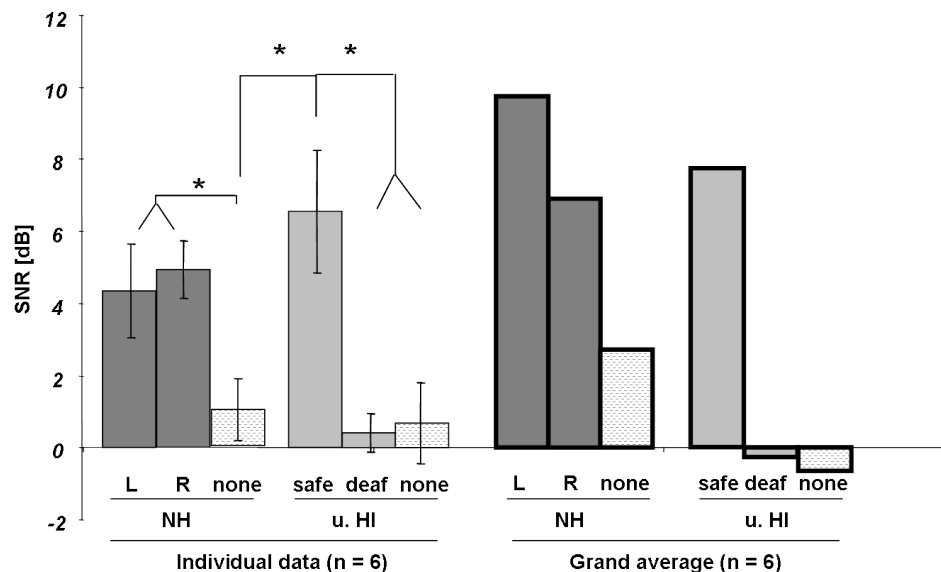


FIGURE 2.12 – Comparaison du rapport signal-sur-bruit (SNR) des Speech ABR obtenus lorsque le stimulus a été entendu ou pas. A gauche sont représentés les données individuelles et à droite les grandes moyennes (traits épais). Le RSB obtenu dans chaque oreille des sujets, ainsi que dans la condition contrôle insert-out-of-the-ear ('none', remplissage en pointillés, Fig. 2.9) furent comparés : chez les normo-entendants ('NH' - sombre) et chez les sourds unilatéraux ('u. HI' - clair). Le RSB fut significativement plus élevé lorsque le stimulus avait bien été entendu que lorsqu'il ne l'avait pas été (oreille sourde, ou condition contrôle chez les malentendants ou chez les normo-entendants). Les grandes moyennes (qui augmentent le RSB du Speech ABR) ont souligné ces résultats : les RSB des grandes moyennes sont plus hauts que la moyenne des individuels lorsque le son a bien été entendu, ils étaient plus faibles (voire même négatifs) dans le cas contraire, sauf pour 'none-NH', voir texte. [Akhoun 2008b]

Le seul recours aux SNR pour décrire les Speech ABR est un peu réducteur : qu'en est-il des paramètres très importants du Speech ABR, tels que la latence et de la qualité de synchronisation entre la FFR et la voyelle ? L'observation des traces temporelles du Speech ABR obtenu chez les cophotiques unilatéraux semble très prometteuse. En y regardant de près, la réponse de l'oreille saine d'un cophotique unilatéral ne semble pas du tout être la même que celle évo-

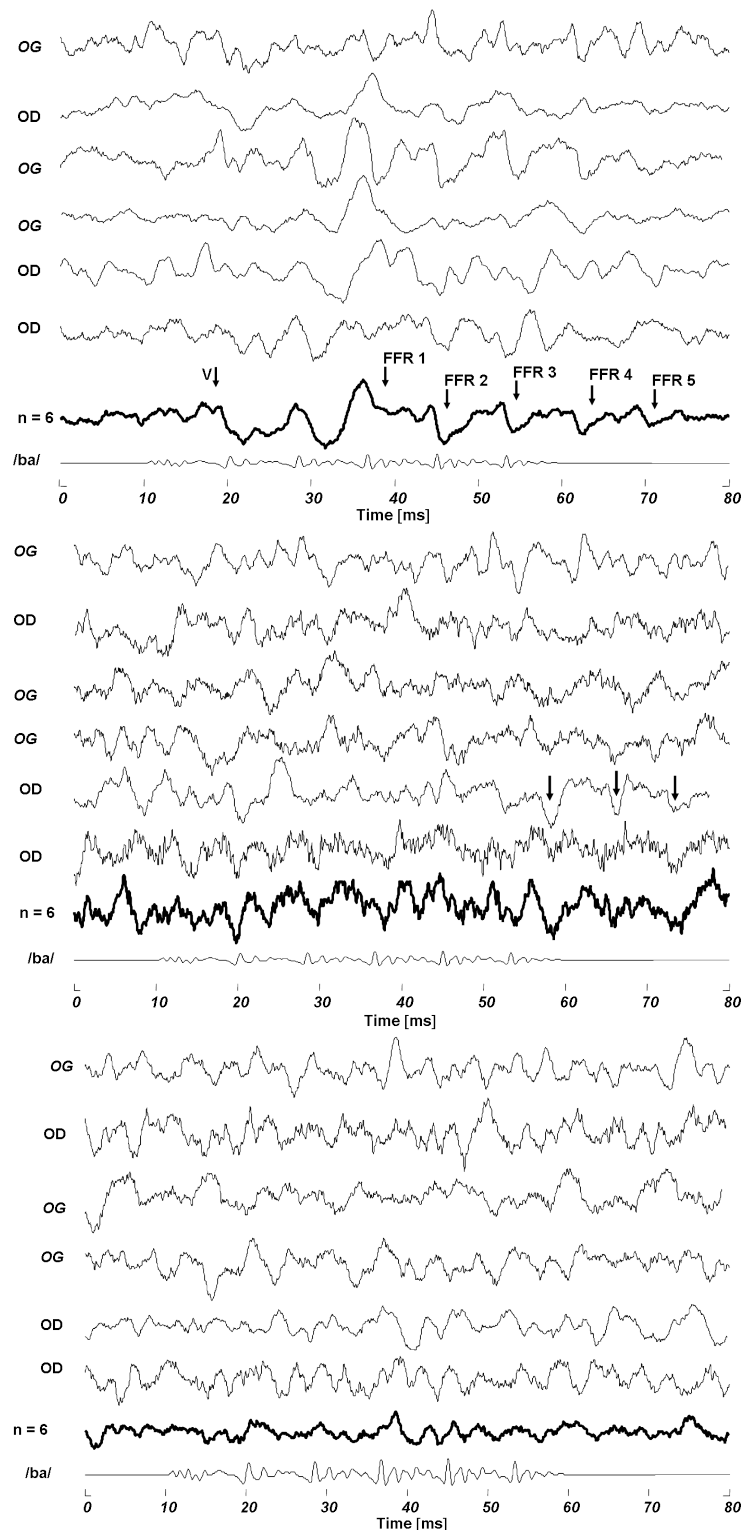


FIGURE 2.13 – Speech ABR en réponse à /ba/ dans l’oreille sourde (milieu) et l’oreille saine (haut), et condition contrôle insert-out-of-the-ear (bas) des 6 patients, sur la voie de droite. L’oreille saine a donné lieu à des Speech ABR assez différents, mais globalement bien définis. En particulier de très larges réponses furent obtenus chez un sujet, avec des latences notablement plus courtes (3<sup>e</sup> en partant du haut). Une réponse peut être obtenue dans l’oreille sourde (par transfert transcrânien?). (OD = oreille droite sourde, OG idem pour la gauche)

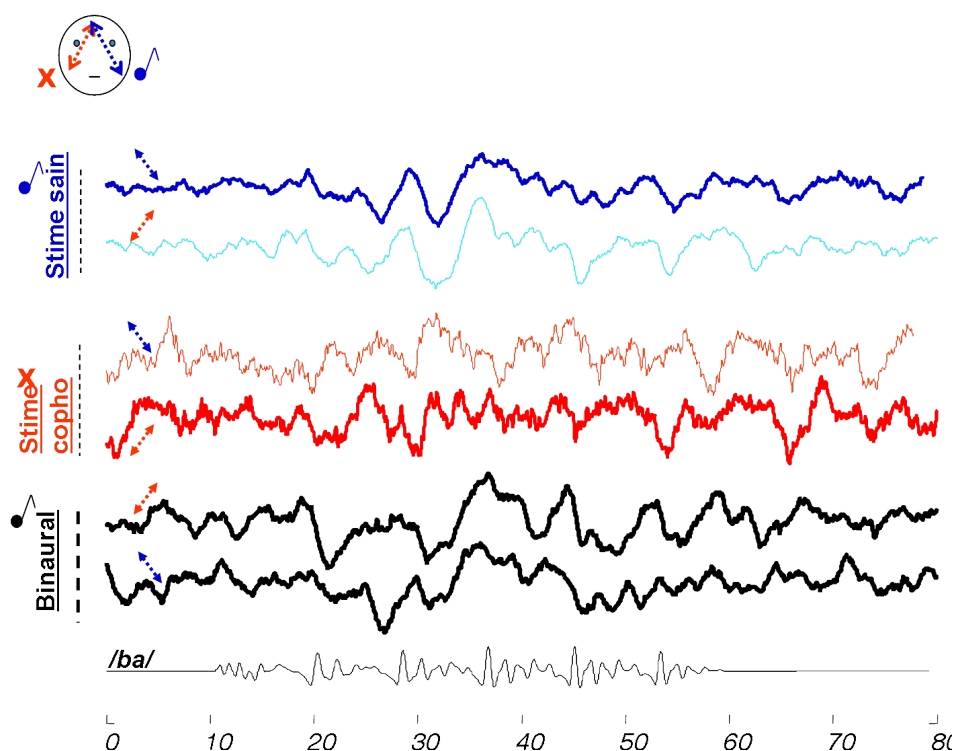


FIGURE 2.14 – Grande moyenne des Speech ABR obtenus chez les patients cophotiques. Les 3 conditions furent testées (i) binaural, (ii) stimulation côté oeil cophotique (X) et (iii) oreille saine (note). Dans chaque cas, les électrodes étaient orientées côté oreille saine (double-flèche en pointillés vers la droite) ou côté oreille cophotique (gauche). Si la stimulation en cophotique ne donne pas de réponse claire et un fort niveau de bruit, il est intéressant de remarquer la différence entre la condition binaurale et l'oreille saine (alors que ces deux conditions donnent la même sensation chez ces sujets sourds unilatéraux), et il est intéressant de constater la différence entre ces courbes et celles obtenues chez les normo-entendants 2.11 .

quée dans l'oreille d'un normo-entendant Fig.2.13. En effet, une grande disparité inter-sujets fut observée (comme le montre la grande-moyenne Fig. 2.14 qui ne fait pas ressortir les caractéristiques du Speech ABR habituel Fig. 2.11), ce qui semblerait être relié à une possible réorganisation des voies auditives suite à la surdité. En particulier, les Speech ABR obtenu chez un sujet (jeune et avec une excellente audition dans l'oreille saine) étaient d'une amplitude remarquablement élevée (3e en partant du haut sur la Fig. 2.13), et avec une latence plus courte que celles obtenues chez les normo-entendants dans les études précédentes. D'autre part, quelques ondes repérées suite à une stimulation dans l'oreille sourde pourraient faire penser à une FFR partielle : serait-elles dues à un transfert transcrânien ? En bref, ce premier aperçu devrait donner suite à une investigation clinique plus poussée de l'effet de la surdité unilatérale dans la réorganisation du traitement de la parole dans les voies auditives centrales des Speech ABR.

### 2.4 Bilan : précautions à prendre et indices d'authenticité du Speech ABR

La seule condition sans artéfact nous a apparu être dans le cas d'inserts blindés dans une cage de Faraday (ne laissant dépasser que les tubes en caoutchouc), bien reliée à la masse commune du dispositif expérimental, Fig. 2.15. La rupture de cette cage de Faraday donne lieu à un artéfact très fort de façon spectaculaire si les conditions sont défavorables (voir fichier video, enregistré dans le cas d'une stimulation à 130 dB SPL pour des impédances infinies). Ce résultat a pour double conséquence de (i) remettre en cause l'assertion largement répandue selon laquelle les inserts earphones éloignant la source électromagnétique des électrodes, ils sont de facto immunisés contre les artéfacts électromagnétiques ; et de (ii) se dispenser de l'utilisation d'inserts à tubes très longs (type Etymotic ER30), qui distordent le signal, et qui ont une courbe de réponse passe-bas à 1000 Hz, ce qui rend l'utilisation de sons de parole ou de musique impossible.

Les indices d'authenticité neurophysiologiques des Speech ABR sont :

- L'indépendance du Speech ABR à la phase du stimulus (i.e. Speech ABR identique dans les deux polarité).
- L'aspect passe-bas du Speech ABR, qui doit être inférieure à la limite de phase-locking dans le site neurophysiologique étudié.
- La latence du Speech ABR doit être de l'ordre d'une dizaine de millisecondes, conformément au délais de traitement du message auditif

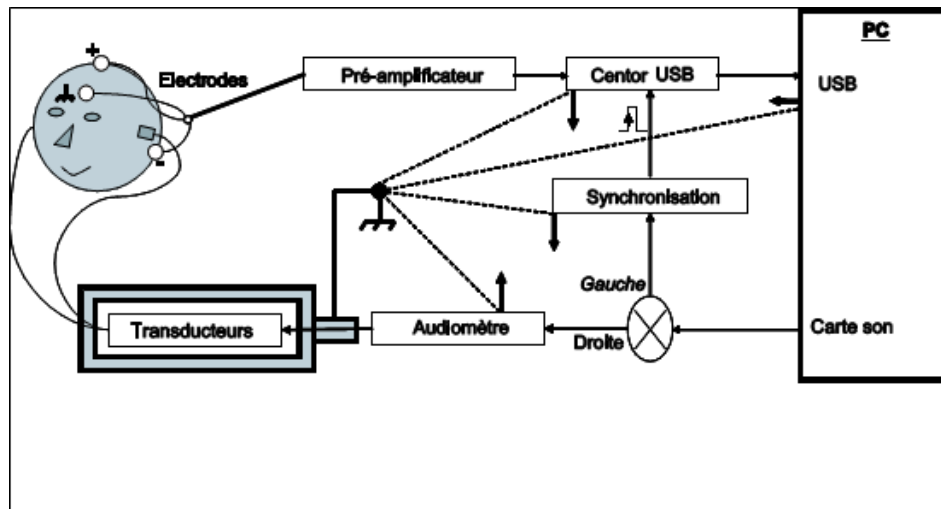


FIGURE 2.15 – Synoptique du dispositif expérimental conseillé d’enregistrement des Speech ABR. Un ordinateur contrôle à la fois l’acquisition des Speech ABR (Centor USB) et la stimulation sonore : ces deux chaînes sont reliées par un module de synchronisation (trigger) [Akhoun 2008d]

2.5 Article Journal of Neuroscience Methods  
[Akhoun 2008d]





## Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status An experimental parametric study

Idrick Akhoun<sup>a,b,d</sup>, Annie Moulin<sup>a,b,c,d</sup>, Arnaud Jeanvoine<sup>a,b,d</sup>, Mikael Ménard<sup>a,b,d</sup>, François Buret<sup>a,e</sup>, Christian Vollaire<sup>a,e</sup>, Riccardo Scorretti<sup>a,e</sup>, Evelyne Veuillet<sup>a,b,d</sup>, Christian Berger-Vachon<sup>a,b,c,d</sup>, Lionel Collet<sup>a,b,c,d</sup>, Hung Thai-Van<sup>a,b,c,d,\*</sup>

<sup>a</sup> Université de Lyon, Lyon F-69003, France

<sup>b</sup> CNRS, UMR 5020, Neurosciences Sensorielles, Comportement, Cognition, Lyon F-69007, France

<sup>c</sup> Hospices Civils de Lyon, service d'audiologie et d'explorations orofaciales, hôpital Edouard Herriot, Lyon F-69003, France

<sup>d</sup> Institut Fédératif des Neurosciences de Lyon, Lyon-Bron F-69677, France

<sup>e</sup> Ecole Centrale Lyon; CNRS, UMR 5005 Laboratoire Ampère, Ecully F-69134, France

### ARTICLE INFO

#### Article history:

Received 22 May 2008

Received in revised form 18 July 2008

Accepted 24 July 2008

#### Keywords:

Speech

Auditory brainstem response

Electromagnetic leakage

Artefact

Unilateral deafness

### ABSTRACT

Speech elicited auditory brainstem responses (Speech ABR) have been shown to be an objective measurement of speech processing in the brainstem. Given the simultaneous stimulation and recording, and the similarities between the recording and the speech stimulus envelope, there is a great risk of artefactual recordings. This study sought to systematically investigate the source of artefactual contamination in Speech ABR response.

In a first part, we measured the sound level thresholds over which artefactual responses were obtained, for different types of transducers and experimental setup parameters. A watermelon model was used to model the human head susceptibility to electromagnetic artefact. It was found that impedances between the electrodes had a great effect on electromagnetic susceptibility and that the most prominent artefact is due to the transducer's electromagnetic leakage. The only artefact-free condition was obtained with insert-earphones shielded in a Faraday cage linked to common ground.

In a second part of the study, using the previously defined artefact-free condition, we recorded speech ABR in unilateral deaf subjects and bilateral normal hearing subjects. In an additional control condition, Speech ABR was recorded with the insert-earphones used to deliver the stimulation, unplugged from the ears, so that the subjects did not perceive the stimulus. No responses were obtained from the deaf ear of unilaterally hearing impaired subjects, nor in the insert-out-of-the-ear condition in all the subjects, showing that Speech ABR reflects the functioning of the auditory pathways.

© 2008 Elsevier B.V. All rights reserved.

### 1. Introduction

Speech auditory brainstem response (Speech ABR) appears to be a very promising audiological technique to investigate the brainstem temporal encoding of speech. In recent studies, Speech ABR has brought further insight in temporal encoding of amplitude modulations (AM) (Purcell et al., 2004) and speech (Wible et al., 2004; Johnson et al., 2005 for a review). Speech ABR was shown to

be related to tonal language processing skills (Krishnan et al., 2004; Wong et al., 2007; Musacchia et al., 2007; Song et al., 2008) and to be correlated to temporal acuity (Purcell et al., 2004; reviewed in Banai and Kraus, 2006), hence its potential usefulness as an objective measurement of subcortical speech and music processing (Kraus and Banai, 2007 and Galbraith, 2008 for reviews).

The Speech ABR waveform elicited by a consonant-vowel (CV) contains a transient onset response related to consonant-vowel transition, immediately followed by a sustained component tuned to the vowel periodicity: the frequency following response (FFR). The FFR has been shown to result from the compound brainstem phase-locking activity, synchronized on the periodicity of the acoustic stimulus (Moushegian et al., 1973; Daly et al., 1976; Faingold and Caspary, 1979; Hall, 1979; Greenberg et al., 1987; Eggermont and Schmidt, 1990; Galbraith et al., 1995; Krishnan,

\* Corresponding author at: CNRS, UMR 5020, Neurosciences Sensorielles, Comportement, Cognition, Lyon F-69007, France. Tel.: +33 472110503; fax: +33 472110504.

E-mail address: [hung.thai-van@chu-lyon.fr](mailto:hung.thai-van@chu-lyon.fr) (H. Thai-Van).

URL: <http://olfac.univ-lyon1.fr> (H. Thai-Van).



2002; Johnson et al., 2005). In addition, the FFR temporal waveform mimics the temporal envelope of the stimulus with a time lag corresponding to physiological processing, and is shown to be sensitive to the stimulus sound level and pitch (Akhoun et al., 2008).

However, Speech ABR timing, and signal to noise ratios have been found to be sensitive to the experimental conditions, and a wide range of artefact may pollute the recordings. These various artefact influences are numerous and must be carefully avoided by the experimenter. Various instrumental techniques have been developed to help immunize the setup against artefact influences, and if instrumental techniques alone fail, signal processing can isolate and cancel part of artefactual effects (Delorme and Makeig, 2004; Felix et al., 2005). Since the beginning of ABR studies, special care has been taken to check electrodes impedances, and to prevent any acoustic noise contamination by placing subjects in sound booth cabins. Moreover, these cabins are often electromagnetically shielded against the electromagnetic effects of power supplies. Indeed, it is common knowledge that power supply can induce a strong frequency component on the recordings (Sadafi et al., 2005), at frequencies corresponding to the alternating frequency of the power supplied (60 Hz for instance in the USA, 50 Hz in most European countries). Another kind of artefact is due to other physiological activities (muscular spikes, other cerebral nervous activities), which are reduced by arithmetic or by sorted averaging techniques (Don and Elberling, 1994; Muehler et al., 1998; Sanchez and Gans, 2006; Luecke et al., 2007). Additionally, timing errors between different computers participating in the recording might result from the mismatch of sample clocks (Jing et al., 2004).

In Speech ABR recording, the worse artefactual response is due to electromagnetic leakage from the transducers: their electrical signal (coding the vowel in the transducer wires and spires) and the FFR may be superimposed. Another possible contamination could be the cochlear microphonic response, a neurophysiological potential induced by the outer-hair cells of the cochlea. Such microphonic response might be recorded by the scalp electrodes, especially if those are close to the stimulated ear. Just as the electromagnetic artifact met in Speech ABR recordings, the microphonic also mimics the stimulus waveform, with no delay and no filtering (Chimento and Schreiner, 1990). In this case, the microphonic is not wanted, since it might overlap with the expected FFR. The similarity of FFR and the vowel periodicity makes the use of signal processing techniques such as principal or independent component analysis (Delorme and Makeig, 2004) impossible. Several studies stressed the importance to cancel this artefact at the source, such as to use insert-earphones (Hall, 1979; Stakenburg and Wit, 1983; Picton and John, 2004; Akhoun et al., 2008), but never systematically investigated potential factors leading to artefactual speech ABR responses. Although being a close family member of speech ABR, auditory steady state responses (ASSR) are, interestingly, immune from electromagnetic artefact, since the stimuli used are AM tones, without modulation frequency peak ( $f_m$ ) in the spectrum (the three frequency components correspond to the 3 carrier frequencies ( $f_c$ ),  $f_c + f_m$ ,  $f_c - f_m$ ). Hence, the modulation is extracted in the auditory system, and may be recorded on the ASSR spectrum, with no risk of misinterpretation due to a stimulation induced electromagnetic artefact (Picton et al., 2003; Small and Stapells, 2004; Picton and John, 2004; Ménard et al., 2008). However, Speech ABR responses are more difficult to distinguish from artefactual responses, since the first frequency peaks of the stimulus correspond to the FFR peaks. Ultimately, the aim of such complex ABRs would be to objectively assess temporal processing along the auditory pathway, and thus to assist hearing aids and cochlear implants fitting. However, another artefactual issue appears, linked to the electrical stimulation of hearing aid or cochlear implant devices (Firzst et al., 2002a,b; Ménard et al., 2004).

Alternate polarity is the usual way to get rid of the artefactual responses: Speech ABR is recorded in response to the stimulus (positive polarity) and the inverted stimulus (negative polarity). The positive and negative polarities induce similar Speech ABR, given the insensitivity of the ear to the phase, and the opposite effects of artefact are cancelled (Greenberg et al., 1987; Chimento and Schreiner, 1990; Johnson et al., 2005). Here, we aimed at checking whether this method is sufficient to totally and reliably eliminate artefactual responses and, in what experimental conditions, artefactual free speech ABR responses can be confidently obtained. In a first part, we systematically investigated artefactual leakage using different transducers, shielding setups and electrodes impedances. In a second part, using the best setup obtained, we checked that the response recorded in human subjects was not contaminated by artefactual leakage, by showing the absence of any response in the deaf ear of unilaterally deafened subjects and in normal hearing subjects tested without the earphones inserted in their ear (i.e. excluding perception of the stimulus).

## 2. Materials and methods

The artefactual and real speech ABR measurements, were processed using a similar method as previously defined (Akhoun et al., 2008). All components were monitored by a single computer (Fig. 1). The stimulation chain started from an Audigy® Soundblaster sound card, whose signal was amplified by an Interacoustics® AC 30 audiometer and delivered to the transducers (either headphones or insert-earphones).

In parallel, the neural responses were recorded with electrodes placed by 4 conventional 10 mm silver-cup EEG electrodes (Nicolet®), preamplified and online processed through a Racia Alvar® Centor USB recording chain. Electrodes wires were twined up to 30 cm before the electrode cup, to enhance their shield against electromagnetic leakage. Speech ABRs were recorded on 2 channels (as described below). The electrodes' contact places on the subject's scalp were first lightly rubbed with Nuprep (D.O. Weaver®) skin abrasive gel. Electrodes were then fixed with Ten 20 conductive EEG paste (D.O. Weaver®), and the impedances between electrodes were maintained below 5 k $\Omega$ .

The stimulation and recording chains were synchronized by a trigger device. All components were stacked, and linked to common ground. In particular, all powered components were connected to a single multi-plug as shown in Fig. 1.

### 2.1. Influence of transducer type on electromagnetic artefact

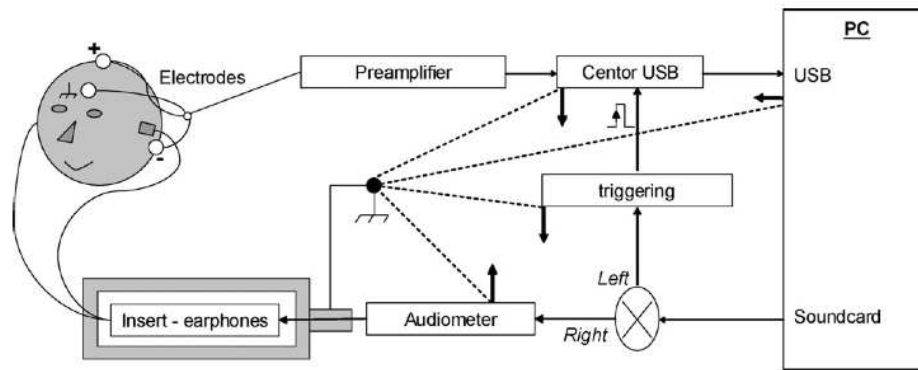
#### 2.1.1. Transducers tested and electrodes substrates used

Eleven artefactual testing conditions were defined, differing by electrodes substrates (scalp models) and transducers used.

Three substrates were used for the electrodes: [1] Electrodes in physiological liquid (short-circuit condition, zero impedances ( $Z = 0 \Omega$ )), [2] electrodes fixed on the surface of a watermelon with the EEG contact paste usually used in clinical routine ABR, to get impedances close to real case ( $Z \sim 1 \text{ k}\Omega$ ) and [3] electrodes not in contact with the watermelon (open-circuit condition,  $Z > 400 \text{ k}\Omega$ ).

Three transducers were tested with the 3 substrates: [A] Teletronics® TDH 39 headphones, [B] Sennheiser® Linear 250 headphones and [C] inserts earphones (Biologic®). Two additional conditions were added for the insert-earphone transducers. Insert-earphones were (i) completely shielded in a custom-built Faraday cage encompassing the insert-earphones transducers and the electric wires (the rubber tubes were the only part remaining out of the cage) and the metallic shield was linked to common ground (first condition termed hereafter "Faraday-on"). For the second condi-





**Fig. 1.** Overview of the experimental setup of Speech ABR recording (in the best artefact-free condition). All components are monitored by a single PC. The stimulation chain starts from the sound card, whose signal is amplified by the audiometer, and delivered to the insert-earphones. In parallel, the Speech ABR is scalp-recorded on the electrodes (placed between vertex (+) and mastoid (–), with forehead as ground). To immunize the electrodes against inductive-leakage induced electromagnetic artefact, they are twined as long as possible from the preamplifier (thicker line). The Speech ABR is then preamplified in the Centor headset, and processed online through the Centor USB recording chain. The stimulation and recording chains are synchronized by a trigger device. All components are stacked and linked to common ground (dotted lines). In particular, all powered components (thicker arrows) are connected to a single multi-plug. To cancel any electromagnetic leakage from the stimulation chain, the insert-earphones (and most of the wires) were embedded in a metallic Faraday box (grey), which was also linked to common ground.

tion (ii) (hereafter “Faraday-off”), the Faraday cage was not linked to common ground anymore. In the initial condition (iii), the insert-earphones were left without any shield. For safety purpose, an extra plastic box was added to prevent the subject from being in direct contact with the Faraday cage.

We made sure to keep the same electrodes configuration as we would have with a human subject (as described for channel 1 in part 2.1.2 of Section 2). The watermelon model was chosen so that its diameter was close to human interaural distance, and placed on a box mimicking a subject’s upper-body. Our rationale was that the electromagnetic susceptibility of Speech ABR recordings with this model would be very close to the susceptibility obtained with a real subject. To check the relevance of the use of a watermelon model, we compared the occurrence of artefact between our model and a real condition in 6 normal hearing subjects, in a control condition where the stimulus was played and not heard: i.e. the insert-earphones were disconnected from the subject’s ears and were lying on the subjects chest, so that the subject could not hear the stimulus (described below as *insert-out-of-the-ear* condition). Therefore, this insert-out-of-the-ear condition cannot induce any auditory evoked potential, and should lead to a recording similar to an ambient recording noise.

### 2.1.2. Stimulation and recording

To assess the effect (on the recorded responses) of the frequency of the acoustic stimulus over the speech range, the sound file was 360 ms long (delivered with a 2.5 stimulus per-second rate) and consisted of a set of 6 ramped-sweeping pure tones, between 110 and 3520 Hz, spread over 6 octaves. Each transducer was calibrated with a Larson Davis® (824) microphone system, including 2cc coupler for the insert-earphones. The sound files were then adapted accordingly, and played for each transducer, between 30 and 130 dB SPL, by 10 dB steps.

Electrodes were placed on the electrode substrate (depending of the impedances wished, as described above) and connected to the ABR recording system Centor USB, as usual. Recording parameters were fixed to record 4000 points for 400 ms (sampling frequency = 10 kHz). Impedances between electrodes were automatically measured by the Centor USB system. The recordings were then filtered online between 2 and 4000 Hz and up to 600 epochs were averaged (less if the artefact was obvious on the traces).

### 2.1.3. Data analysis

The presence of an artefactual response was visually detected in the recorded spectrum (or spectrogram), calculated by Fast Fourier Transform (or Short Time Fourier Transform) with Matlab (The Mathworks®). In a similar technique as the *F*-test (Picton et al., 2003), artefactual contamination was considered as present when, at least, one component, amongst the six components corresponding to the stimulus frequencies, was observed 6 dB above the ambient noise level. The ambient noise level was measured between 2 successive stimulus components.

## 2.2. Speech ABR in normal hearing (NH) and unilateral hearing impaired (uHI)

### 2.2.1. Subjects

Speech ABR were recorded in 2 populations: one group of normal hearing subjects (NH) and one group of unilateral hearing impaired subjects (uHI). Hearing thresholds to sweeping pure tones were measured in sound booth audiometric cabins, for octaves frequencies ranging from 256 to 8000 Hz, by 5 dB step, using an interacoustics AC30 audiometer.

In the NH group, 6 adults (3 men and 3 women, age 19–27 years old (y.o.),  $M = 21 \pm 5$  y.o.) had normal hearing threshold (below 25 dB-HL) for all tested frequencies. In the uHI group, 6 uHI (27–63,  $M = 51 \pm 12$  y.o.) had normal hearing thresholds (below 25 dB-HL) in the normal hearing ear, and were severely hearing impaired in the other ear (thresholds above 90 dB-HL). However, 50 dB-HL notches were noticed at 4 kHz for subjects 3 and 6. Three subjects had left ear impairment and three had right ear impairment.

This study was carried out in accordance with the Declaration of Helsinki, and the French Huriet act pertaining to biomedical research in humans, and was approved by the “South-East IV” ethics committee (CCPPRB # 05-024). Informed written consent was obtained for each participant.

### 2.2.2. Stimulation and recording

Acoustic stimulation was played with “Faraday-on” insert-earphones (the best artefact-free condition, determined in the first study). The sound was played at 45 dB SL on the NH ear and at the same level in the deaf ear of uHI (below hearing threshold in this condition). The stimulus was the consonant-vowel (CV) syllable /ba/ described in our previous study (Akhoun et al., 2008).



The phoneme /ba/ was 60 ms long, with 10 ms of silence before the beginning of the sound, in order to assess the ambient noise level in the first 10 ms of the recordings. The stimulus was delivered with an 11.1 syllable per second rate.

The subjects were stimulated monaurally in the *right ear* and in the *left ear*, and in the control *insert-out-of-the-ear* condition.

Speech ABR were recorded in 2 channels: the electrodes were placed between vertex and right mastoid for channel 1, and vertex and left mastoid for channel 2, with forehead as ground. Inter-electrode impedance was maintained below 5 k $\Omega$ . Fifteen hundred epochs were averaged, with an 80 ms analysis window (sampling frequency = 50 kHz). Speech ABRs were filtered online by an 80–3200 Hz bandpass Finite Impulse Response filter (order 2). Each epoch with amplitude greater than 75  $\mu$ V was discarded. One set of 1500 epochs was averaged for each polarity (positive, negative).

### 2.2.3. Data analysis

Grand-average curves were calculated as the algebraic summation of the individual data on a given polarity (alternate polarity is the algebraic sum of blocs for both polarities, i.e. the average of a total of 3000 epochs). The baselines of individual waveforms were corrected by subtracting the mean value to remove DC offset, and the data were divided by the standard error to balance the influence of each individual waveform (on each polarity) on the grand-average. This technique might however yield a bias in the respective amplitudes of Speech ABR on each polarity, and also on each subject. Therefore, as we are not able to compare directly Speech ABR amplitudes, we chose to compare signal-to-noise ratios (SNR), as described below.

To assess the signal energy, root-means square (RMS) values were calculated either for the whole signal or for the silence period before averaging. Accordingly, the SNR was calculated as the ratio (in dB) between the RMS on the whole Speech ABR and the RMS on the pre-averaging silence.

The function *corrcoef* under Matlab gave intercorrelation coefficients ( $R$ ) as well as the  $p$ -value to test the hypothesis of no-correlation. The intercorrelation coefficient between Speech ABR of each positive and negative polarity are noted  $R$ -*pol*, those between Speech ABR on both ears  $R$ -*ears* and the intercorrelation coefficient between stimulus and Speech ABR was noted  $R$ -*stim*. The temporal shift of Speech ABR with the periodicity of /ba/ was calculated as the maximum absolute  $r$ -value of the  $R$ -stim cross-correlogram. This cross-correlogram was calculated as 1000 shifted-coefficients between the resampled envelope (resampled from 44.1 to 50 kHz to match the Speech ABR sampling frequency) and the FFR (between 10 and 60 ms, i.e. 2500 points). Intercorrelation coefficients were also calculated between individual (and grand-averaged) Speech ABR versus the periodicity of /ba/.

Statistic analysis was performed using SysStat<sup>®</sup>, with a significance level ( $p$ ) < 0.05. Analysis of SNR was performed using a one-way analysis of variance (ANOVA).

## 3. Results

### 3.1. Characterization and quantification of the transducer's type on electromagnetic artefact

First, the artefact presence was tracked down all along the stimulation chain, in order to assess if other elements as the transducers might induce electromagnetic leakage. Indeed, an artefact was recorded even when the transducers were not connected to the audiometer. This leakage was due to a 2-meter long RCA rolled wire, which was used to make the link between the output of the sound

card and the auxiliary input of the audiometer (Fig. 1). However, this artefact disappeared as a shorter 30 cm RCA link was chosen instead.

In line with previous descriptions (Picton and John, 2004; Akhoun et al., 2008), the artefact recorded had an exactly similar waveform as the sound played (Figs. 2 and 6). In particular, this waveform was not delayed and not low-pass filtered, as usually described for physiological responses. Another important result is the strong modulation at 50 Hz, induced by the influence of (French) power supply. Last but not least, the sinusoid components were distorted, and therefore had harmonics that could lead to overestimating the influence of upper frequency components. For example, if an artefact was recorded at 220 Hz, we might have obtained a harmonic peak at 440 Hz (during the period corresponding to the 220 Hz tone), which does not correspond to an artefact at 440 Hz. With our sweeping stimuli, it was possible to avoid being misled by using time–frequency spectrograms that clearly differentiate harmonics to pure tone frequencies.

The sensitivity of the experimental setup in regards to electromagnetic artefact was dependant on the impedances between the electrodes. This is illustrated in Fig. 3 (where only the non-shielded insert-earphones artefact-thresholds were plotted). The artefact was less present in short circuit than when the electrodes were in open circuit. In addition, artefact induced by the transducers had a similar frequency-behavior, in all 3 impedances conditions. Furthermore, the real condition (with the watermelon), was closer to the open circuit, than to the short-circuit condition. Incidentally, it is noticeable (Fig. 2), that the artefact amplitude was not equal for all frequencies despite the ramped-sweeping stimulus calibration, for each transducer.

In addition, the different transducers had different artefactual influences. The headphones also induced artefact, even at moderate sound levels. This is summed up in Fig. 4, showing sound level thresholds, over which artefact was obtained for each transducer, as a function of frequency (in the real-case condition). The only absolute artefact-free condition was obtained with insert-earphones shielded in the Faraday cage, linked to common ground.

In regards to electromagnetic susceptibility, the watermelon was found to model accurately the human head. Indeed the mean values of sound level thresholds of artefactual recordings were found to be similar in the watermelon model and in the 6 human subjects, in the 3 *insert-out-of-the-ear* conditions tested (for Faraday-on, Faraday-off, and non-shielded insert-earphones). In the Faraday-on condition, no artefact was recorded, except once for the first subject at 3520 Hz, for 110 dB SPL (this condition was then not plotted in Fig. 5). This similarity was consistent for all frequencies tested, for each Faraday-off and non-shielded insert condition. Fig. 5 shows that the subjects' threshold behaved similarly to the watermelon model, and that the differences between those thresholds were never greater than 12 dB SPL. In addition, very small inter-subject differences were noticed in the sound level thresholds of artefact detection: individual discrepancies were noticed in non-shielded insert-earphones: standard error ranged from 2 (at 440 Hz) to 4 dB SPL (at 220 Hz). In the Faraday-off condition, standard deviation ranged from 4 (at 440 Hz) to 7 dB SPL (at 880 Hz).

During the experimental measurements described in part 1, a spectacular artefact appeared during the recording when the inserts were not shielded (i.e. by disconnecting the Faraday cage from common ground). Accordingly, an artefact was recorded on four additional NH subjects (a–d) stimulated through non-shielded inserts, with the previously described CV /ba/ stimulus. Here again, the artefact was characterized by a waveform very close to the stimulus, but undelayed and unfiltered (Fig. 6). Two points are worth noticing:

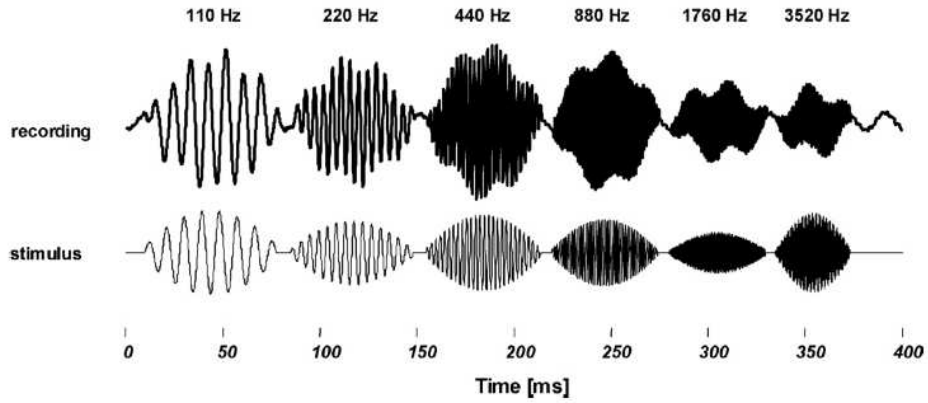


Fig. 2. The artefactual recording temporal waveform matches exactly the successive sinusoid stimuli used. This artefact was recorded in open circuit electrodes configuration, with non-shielded insert-earphones, at very high sound level (around 100 dB SPL). Note that the power supply induces a 50 Hz modulation.

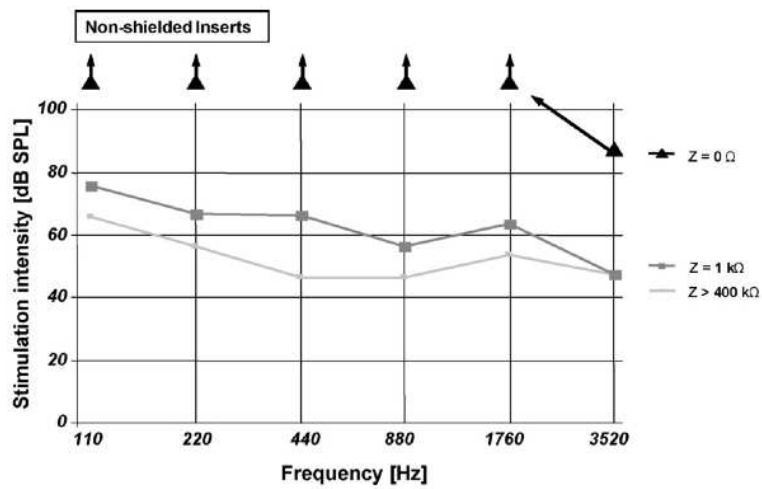


Fig. 3. Effect of different impedances between the electrodes on ABR sensitivity to electromagnetic leakage, with non-shielded insert-earphones, in function of frequency (logarithmic scale). Each point represents the lowest sound level intensity (in dB), at which artefact was seen, for a given frequency (in Hz), and for a given electrode impedance. The open circuit condition ( $Z > 400 \text{ k}\Omega$ ) was the most sensitive to artefact. In contrast, the short-circuit condition ( $Z \sim 0 \Omega$ ) was the least sensitive to artefact: the artefact was only seen at 3520 Hz, at the highest sound level tested. Note that the real case ( $Z = 1 \text{ k}\Omega$ ) is closer to the open circuit than the short-circuit condition.

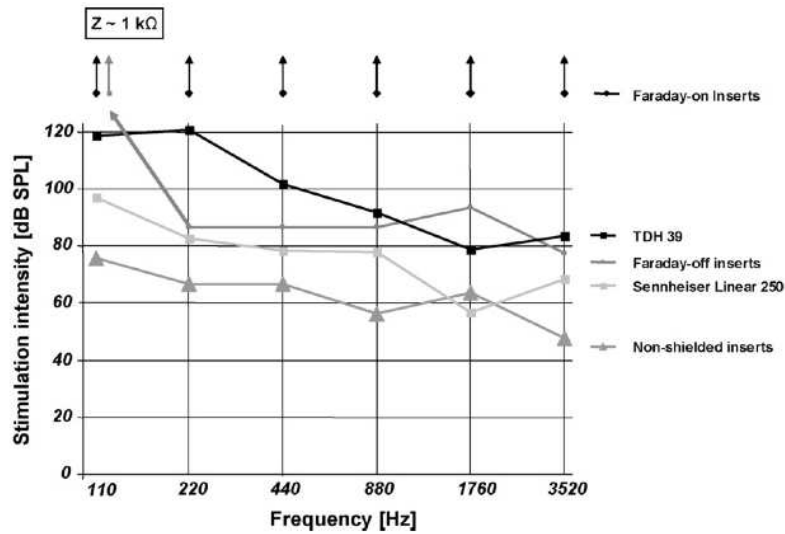
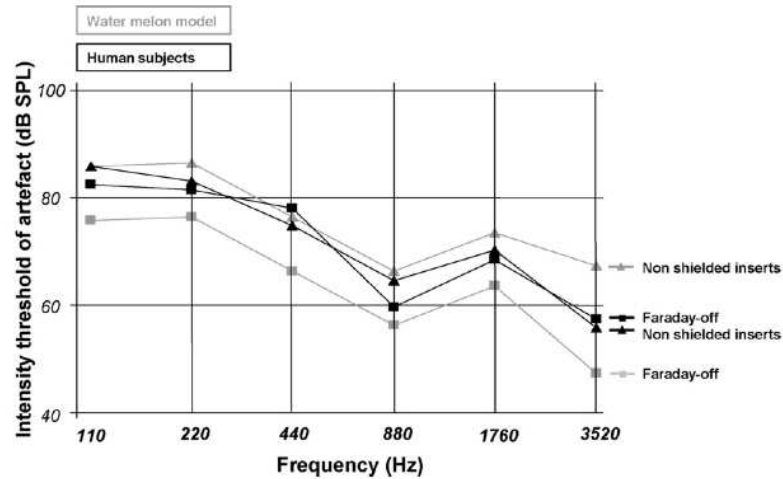


Fig. 4. Effect of different transducers on electromagnetic leakage as a function of frequency (logarithmic scale), in the real-case ( $Z = 1 \text{ k}\Omega$ ) condition. This plot is similar to the previous one (Fig. 3). Here, hardly any artefact was recorded with insert-earphones shielded in a Faraday box linked to common ground.





**Fig. 5.** Comparison of artefact thresholds measured with the watermelon model (grey), and mean values obtained with 6 human subjects (black). Stimulation was made in the 3 insert-earphone conditions: Faraday-on (not plotted), Faraday-off (squares) and inserts non-shielded (triangles), in the *insert-out-of-the-ear* condition. The "Faraday-on" condition was not plotted, since no artefact was recorded in both human subjects and watermelon model (except for subject 1 at 110 dB SPL for 3520 Hz).

First, on two subjects (c) and (d), the effect of artefact was comparable on both polarities (e.g. d.1 versus d.2), hence its suppression with alternate polarity recording (e.g. d.1 + d.2). It appeared that the electromagnetic artefact and Speech ABR had comparable magnitude in this condition. Note that, after cancellation, the alternate polarity Speech ABR waveform shape was similar to usually reported Speech ABR.

Second, for (a) and (b) subjects, the artefact magnitude was greater on one polarity, rendering alternate polarity useless to completely cancel artefactual responses.

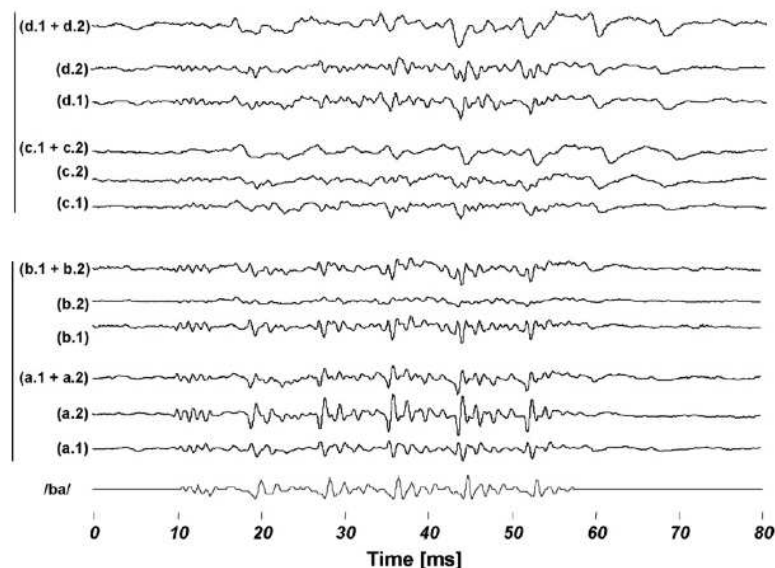
### 3.2. Comparison of Speech ABR in unilateral hearing impaired patients (uHI) and bilateral normal hearing (NH)

In this part, the stimuli were delivered by insert-earphones shielded in the Faraday box, linked to common ground, using

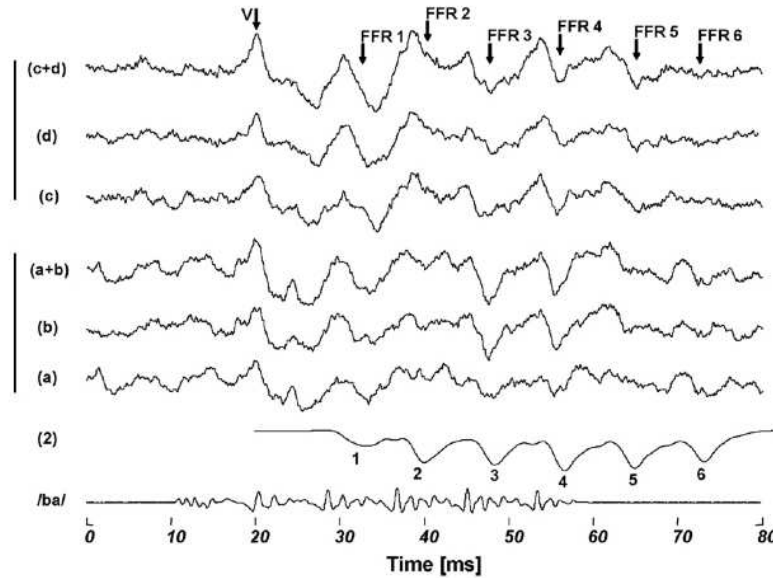
an intensity of 45 dB SL (corresponding to a level of 60–65 dB SPL depending of the subjects individual thresholds). Thus, only genuine neural responses should be recorded, and no artefactual responses were expected here.

#### 3.2.1. Speech ABR in NH

Following Akhoun et al. (2008), the /ba/ stimulus was similarly encoded in positive or negative polarity (Fig. 7). This was illustrated by the greatest correlation coefficient  $R_{\text{pol}}$  found in the absence of any shift between both polarities, for each subject, and whatever the ear stimulated ( $R_{\text{pol}}$  recorded on channel 2, whilst stimulating the left ear: mean ( $M$ ) = 0.72; standard deviation (S.D.) = 0.11; and on channel 1 whilst stimulating the right ear:  $M$  = 0.42; S.D. = 0.21). This was also clear on grand-average data ( $R_{\text{pol}}$  = 0.73 for the left ear;  $R_{\text{pol}}$  = 0.5 for the right ear). The same trends were noticed on each ear, and the greatest correlation coefficient  $R_{\text{ears}}$  between



**Fig. 6.** Example of artefactual Speech ABR recordings on 4 subjects (a, b, c and d), despite the use of alternate polarity. Speech ABR are plotted in each polarity, positive (a.1, b.1, c.1 and d.1), negative (a.2, b.2, c.2, d.2) and alternate (a.1 + a.2, b.1 + b.2, c.1 + c.2, d.1 + d.2). The temporal waveform of the stimulus (/ba/) is present in each Speech ABR, except on (c.1 + c.2) and (d.1 + d.2). In addition, this artefact follows the polarity of the stimulus (most clearly seen on a.1 and b.2), contrary to what is expected from cochlear transduction.



**Fig. 7.** Grand-average ( $n=6$ ) Speech ABR elicited by the /ba/ stimulus in the left (a and b) and right (c and d) ears in normal hearing subjects. The Speech ABR is plotted in positive (a and c), negative (b and d) and alternate (a + b and c + d) polarities. Speech ABR components are specified by arrows. Wave V corresponds to transient onset response, and FFR to the waves of the Frequency Following Response synchronized on the stimulus envelope. The fact that both polarities are encoded similarly and the fact that the stimulus envelope had to be shifted forward by around 10 ms (2) to most accurately match the FFR waveform (Akhoun et al., 2008), are the two main evidence of physiological origin of Speech ABR. It is assumed that each FFR peak (FFR 1 to FFR 6) is related to a peak in the reversed envelope (1–6).

right ear and left ear speech ABR waveforms, was also found in the absence of any shift ( $M=0.53$ ;  $S.D.=0$ ). Here again, grand-average data confirmed this result ( $R\text{-ears}=0.53$ ).

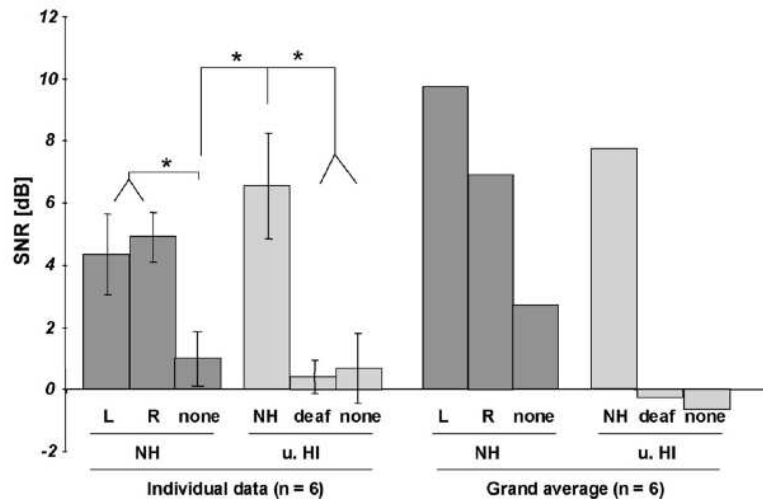
Furthermore, the FFR was shown to match the stimulus waveform, up to a previously described cut-off frequency around 500 Hz. This goodness of fit was observed, with an expected physiological lag of the FFR behind the vowel. For the left ear, this lag varied between 10.3 and 14.8 ms ( $M=12.7$ ;  $S.D.=1.3$  ms), with  $R\text{-stim}$  ranging between 0.21 and 0.48 ( $M=0.27$ ;  $S.D.=0.1$ ). For the right ear, the lag varied between 10.1 and 14.0 ms ( $M=11.48$ ;  $S.D.=1.1$  ms), with  $R\text{-stim}$  ranging between 0.15 and 0.48 ( $M=0.31$ ;  $S.D.=0.1$ ). The good-fit between FFR and the /ba/ envelope was also mea-

sured on grand-average data: for the left ear the lag was 12.7 ms ( $R\text{-stim}=0.27$ ), and for the right ear, it was 11.1 ms ( $R\text{-stim}=0.34$ ).

**3.2.2. Speech ABR in uHI**

Speech ABR elicited in the normal hearing ear was similar for both polarities: the maximum  $R\text{-pol}$  was obtained in the absence of shift ( $M=0.3$ ;  $S.D.=0.2$ ), with  $R\text{-pol}=0.49$  on grand-average data. For the deaf ears,  $R\text{-pol}$  were far lower ( $M=0.03$ ,  $S.D.=0.14$ ), with  $R\text{-pol}=-0.18$  on grand-average.

A poor maximal  $R\text{-ears}$  was measured between alternate polarity Speech ABR in the normal-hearing and the deaf ear ( $M=-0.03$ ,  $S.D.=0.17$ ), for a lag of 0 ms ( $M=0.01$ ,  $S.D.=0$ ). In addition, the



**Fig. 8.** Comparison of SNR in Speech ABR in both ears of NH and uHI subjects, and in the control insert-out-of-the-ear condition (none). The SNR in the uHI normal-hearing ears were significantly greater than the SNR measured in deaf ears or in insert-out-of-the-ear condition. In addition, the NH left or right ear SNR were also significantly greater than both insert-out-of-the-ear conditions, and to the deaf ear condition. Grand-averaging SNR, calculated as the SNR on grand-averaging Speech ABR (Fig. 7, for NH), enhanced the SNR calculated for the uHI's normal-hearing and both left and right NH ears, whereas it reduced the SNR in the deaf ear, or insert-out-of-the-ear condition (at least the one recorded in uHI).



normal-hearing ear FFR lag from the stimulus envelope, was measured between 9.6 and 16.5 ms ( $M = 12.49$ ,  $S.D. = 2.23$  ms), with R-stim ranging from 0.27 to 0.51 ( $M = 0.37$ ,  $S.D. = 0.08$ ). In comparison, the deaf ear FFR lag was more disparate between 5.1 and 14.8 ms ( $M = 11.47$ ,  $S.D. = 4.3$  ms), with R-stim varying between 0.11 and 0.29 ( $M = 0.19$ ,  $S.D. = 0.06$ ).

### 3.2.3. Comparison of Speech ABR SNR between NH and uHI subjects

Fig. 8 shows that the SNR obtained with the insert-out-of-the-ear condition were rather low, either for NH, varying between  $-2.41$  and  $3.76$  dB ( $M = 1.01$ ;  $S.D. = 0.96$  dB), and uHI, between  $-3.5$  and  $4.12$  dB ( $M = 0.69$ ,  $S.D. = 1.12$  dB). Furthermore, the SNR of NH left, between  $1.96$  and  $7.16$  dB ( $M = 4.35$ ;  $S.D. = 1.3$  dB), or right ear condition, between  $1.68$  and  $8.00$  dB ( $M = 4.95$ ,  $S.D. = 0.79$  dB) were in the same range as for the uHI normal-hearing ear, between  $1.96$  and  $12.59$  dB ( $M = 6.57$ ;  $S.D. = 1.70$  dB). In contrast, the SNR of the deaf ear were lower, varying between  $-1.8$  and  $1.9$  dB ( $M = 0.42$ ;  $S.D. = 0.53$  dB).

For uHI subjects, the Speech ABR SNR were significantly greater in the NH ear than in the deaf ear (ANOVA:  $F(1,10) = 11.8$ ;  $p < 0.05$ ), and the uHI's or NH's insert-out-of-the-ear conditions (ANOVA:  $F(1,10) = 8.32$  and  $7.49$ , respectively; both  $p < 0.05$ ).

For NH subjects, SNR were greater than the uHI deaf ear condition (ANOVA:  $F(1,10) = 20.8$ ;  $p = 0.001$ ), and also higher than the NH inserts-out-of-the-ear condition (ANOVA:  $F(1,16) = 7.73$ ;  $p < 0.05$ ).

No significant differences in SNR were obtained between the different normal hearing ear groups (left ear of NH subjects, right ears of NH subjects, NH ears of uHI subjects) ( $F(1,10) = 0.06$ , NS). Likewise, none of the 2 insert-out-of-the-ear (for NH or uHI) conditions lead to SNR significantly different from the deaf ear condition.

The SNR was enhanced on grand-average data (Fig. 8) for normally hearing ears (NH left:  $9.74$  dB; NH right:  $6.93$  dB; NH of uHI:  $7.75$  dB), and decreased for deaf ear ( $-0.26$  dB) and insert-out-of-the-ear ( $-0.63$  dB) conditions for uHI patients.

## 4. Discussion

As for all electrophysiological recordings, special care must be taken to decrease the recording noise. This is especially important for Speech ABR as the recorded response has the particularity to be simultaneous to stimulation. In addition, there is a high similarity between the stimulus and the recorded response, with possible artefactual contamination due to transducers leakage. Hence, this study sought to systematically define, for different stimulus frequencies, the threshold of occurrence of artefactual responses, using different setup parameters and transducer types. We showed possible strong artefactual responses when using any transducer type, with, for instance, a threshold for artefactual response as low as  $80$  dB SPL for TDH 39. Even with insert-earphones, an artefact was present in some cases in normal hearing subjects: it could easily be recognized that the exact temporal waveform of the stimulus was present on the recordings, without any filter, nor delay. Artefact also followed the stimulus polarity. Usually, the alternate polarity method is used to cancel the artefact effects, but this was not sufficient here due to the different artefact amplitudes obtained on each polarity (Chimento and Schreiner, 1990). Such examples underline the need to cancel the electromagnetic artefact at its source, in order to get an artefact-free response.

Amongst all the transducers tested, only the insert-earphones could lead to artefactual free responses, with the added condition of the presence of a full-shielded system. The Faraday cage was found here to be a very efficient way to shield the transducers. The efficiency of the Faraday cage was underlined by the low artefact

threshold (as low as  $85$  dB-SPL at  $220$  Hz) obtained when the Faraday box was inactivated (when the metallic box was not linked to common ground), and a much greater threshold when the Faraday cage was activated.

Test-retest experiments with non-shielded and Faraday-off insert-earphones showed important discrepancies in artefact-thresholds. However, in each case, no artefact was ever recorded in the Faraday-on condition. Despite the fact that stimuli amplitude were adapted in regards to transducers calibration in order to result in similar SPL, the electromagnetic artefact was frequency-sensitive. For each transducer, the artefact threshold most often decreased with increasing stimulus frequency, probably due to the frequency-sensitivity of the recording setup's electromagnetic susceptibility. For instance, artefact threshold was  $95$  dB-SPL at  $110$  Hz and  $55$  dB-SPL for  $1760$  Hz with Sennheiser Linear headphones.

The impedances between the electrodes were shown to influence recording setup sensitivity to electromagnetic leakage. In the Speech ABR frequency range, centered on the brainstem phase-locking boundaries (between  $80$  and  $3200$  Hz), the electromagnetic leakage is rather capacitive than inductive. Thus, electromagnetic leakage is most important for high impedances (open circuit), and most limited for zero impedances (short-circuited electrodes). For the non-shielded insert-earphones, this is illustrated by the great differences in artefact thresholds between open and short-circuited electrodes. Indeed, no artefact was obtained with short-circuited electrodes in physiological liquid (except over  $85$  dB-SPL at  $3520$  Hz), whereas very low artefact thresholds were measured in open-circuit (from  $45$  dB-SPL at  $440$ ,  $880$  and  $3520$  Hz to  $65$  dB-SPL at  $110$  Hz). Interestingly, a close one-to-one relationship was found on the watermelon condition (modeling real-case condition) and on the open-circuit condition. Consequently, the effects of the transducers type on electromagnetic artefact, in physiological liquid, were under-estimated in our previous study (Akhoun et al., 2008).

The relevance of the watermelon model was checked by comparing its thresholds to the same thresholds obtained with the insert-out-of-the-ear condition on NH subjects. In each condition, a close similarity was observed between artefact-thresholds obtained with human subjects and with the watermelon. Both were decreasing with frequency, and differences in artefact thresholds remained smaller than  $15$  dB SPL (at  $440$  Hz) for Faraday-off inserts, and also smaller than  $15$  dB SPL (at  $3520$  Hz) for non-shielded insert-earphones. Here again, no artefact was recorded for the Faraday-on condition (except over  $95$  dB-SPL at  $3520$  Hz for one out of the six tested subjects). The watermelon model was especially useful for testing artefact thresholds with headphones. Indeed, it was possible with insert-earphones to deliver the stimulus without the subject able to hear it (insert-out-of-the-ear), and with a similar "electromagnetic configuration". However, with headphones, it would have been impossible to ensure the non-perception of the stimulus by the subjects, even with earplugs.

Our results point out several ways to discard a possible artefact contamination. It is necessary to devise a common ground system on the experimental setup to have an efficient Faraday cage. It is also important to avoid any loop with the electrical wires, and to control contact impedances during the recording sessions. Moreover, the shielded insert-earphones are an interesting ways to avoid the use of insert-earphones with very long tubes, whose response characteristics induce strong distortions to the stimuli after  $1$  kHz, compromising the use of speech (or music) stimuli. To go further, one final aim of Speech ABR would be to assess objectively speech, or music, temporal encoding in hearing aid and cochlear implant recipients. However, in this condition, another kind of stimulation artefact is raised by the electrode array electrical stimulation. Simple instrumental in situ solutions seem difficult to devise, and much



hope resides in signal processing techniques, such as optimal use of weighted alternate polarity (Alvarez et al., 2007).

In the best artefact-free condition, Speech ABR could be expected only if the auditory pathway was functional (reviewed in Greenberg et al., 1987). Accordingly, three main results of the second part point towards a neural origin (from the auditory pathway) of Speech ABR.

First, the similarity of Speech ABR in both polarities relates to the insensitivity of cochlear transduction to stimulus polarity (Patuzzi, 1987; Chimento and Schreiner, 1990). This phenomenon is the reason why the alternate polarity method is supposed to cancel the artefactual recording (provided that artefact magnitude on each polarity are similar), and to leave the Speech ABR unchanged (Small and Stapells, 2004; Russo et al., 2004). Unlike Speech ABR, the recorded artefact always followed the stimulus polarity. In this study, we recorded different artefacts on each polarity (Fig. 6), when using non-shielded inserts, probably because the effect of electromagnetic leakage was different, due to variations in the electromagnetic features on each polarity block. Alternate polarity should be used, provided that the electromagnetic influences, as well as cochlear microphonics, are similar on both polarities. Chimento and Schreiner (1990) warned against the technique of alternate polarities in FFR responses to pure tones, and advised to use the subtraction of tones-elicited FFR (FFR and cochlear microphonics) to forward-masking tones elicited FFR (which result in cochlear microphonics only).

Second, the FFR lag behind the stimulus envelope (around 10 ms) is related to the neurophysiological transmission time of the message from the ear to the upper brainstem nuclei, supposed to elicit Speech ABR. In addition, the cut-off frequency of Speech ABR (around 500 Hz) is within the frequency range of phase locking for the brainstem nuclei. The link between this lag, the cut-off frequency and the supposed FFR generators has been previously shown (Faingold and Caspary, 1979; Hall, 1979; Greenberg et al., 1987; Eggermont and Schmidt, 1990; Krishnan, 2002; Russo et al., 2004; Akhoun et al., 2008). Here again, the recorded artefact always occurred in the absence of any shift, with no low-pass filtering.

Third, Marsh et al. (1970) proved that FFR was present, only if the brainstem nuclei underlying generators were functional. In line with this study, we confirmed here that normal Speech ABR pattern could not be obtained in the deaf ear of unilaterally hearing impaired subjects, in an artefact-free experimental setup. As expected, the uHI's normal-hearing ear SNR was close to those obtained in any ear of NH, whereas the deaf ear SNR was similar to the insert-out-of-the-ear condition. In addition, those results were underlined on grand-average data: the SNRs were enhanced for NH and uHI's normal-hearing ear, whereas it was reduced (and even became negative) for the deaf ear, and for the insert-out-of-the-ear conditions. The absence of any reliable response in the insert-out-of-the-ear conditions confirms the absence of contribution from any non-physiological source.

## 5. Conclusion

This study provides the thresholds of occurrence of artefactual recordings as a function of frequency, for several transducers commonly used in auditory experiments, and different setup parameters. Insert-earphones shielded in a Faraday cage (linked to common ground) were found to be an efficient way to discard artefact. In this artefact-free condition, three points were provided to show that Speech ABR is of neurophysiologic origin, as opposed to artefactual origin: (i) Speech ABR's insensitivity to stimulus polarity, (ii) Speech ABR's lag and low-pass filtering and (iii) the absence of Speech ABR with non-functional auditory pathway (in uHI).

## Acknowledgements

The authors would like to thank all the subjects who kindly participated in this study. The authors acknowledge: Guy Clerc and Noel Burais (CNRS, UMR 5005); Samuel Garcia, Michel Vigouroux and Bernard Bertrand (CNRS, UMR 5020); and also A&K Nya-may'Antu, Stéphane Gallégo, Arnaud Norena, Céline Richard and Clémence Pavillon. We are grateful to Professor Paul Avan (Service de Biophysique, Université d'Auvergne, Clermont-Ferrand I), for providing helpful advice on that work, and to the two anonymous reviewers for their precious input. The first author was sponsored by a CIFRE Ph.D.-grant from Cochlear France and the University of Lyon.

## References

- Akhoun I, Gallégo S, Moulin A, Veuillet E, Berger-Vachon C, Collet L, et al. The temporal relationship between Speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/. *Clin Neurophysiol* 2008;119:922–33.
- Alvarez I, de la Torre A, Sainz M, Roldan C, Schoesser H, Spitzer P. Generalized alternating stimulation: a novel method to reduce stimulus artifact in electrically evoked compound action potentials. *J Neurosci Methods* 2007;165(1):95–103.
- Banai K, Kraus N. The neurobiology of central auditory processing disorder (CAPD), language impairment and learning disability. In: Chermak GD, Musiek FE, editors. *Handbook of central auditory processing disorder: from science to practice*. Plural Publishing Inc.; 2006.
- Chimento TC, Schreiner CE. Selectively eliminating cochlear microphonic contamination from the frequency-following response. *Electroenceph Clin Neurophysiol* 1990;75:88–96.
- Daly D, Roeser R, Moushegian G. The frequency-following response in subjects with profound unilateral hearing loss. *Electroenceph Clin Neurophysiol* 1976;40(2):132–42.
- Delorme A, Makeig SJ. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 2004;134(1):9–21.
- Don M, Elberling C. Evaluating residual background noise in human auditory brainstem responses. *J Acoust Soc Am* 1994;96(5 Pt 1):2746–57.
- Eggermont JJ, Schmidt PH. The auditory brainstem response. In: Colon EJ, Visser SL, editors. *Evoked potential manual*. Kluwer Academic; 1990. p. 41–77.
- Faingold CL, Caspary DM. Frequency-following responses in primary auditory and reticular formation structures. *Electroenceph Clin Neurophysiol* 1979;47(1):12–20.
- Felix LB, Moraes JE, Miranda de Sá AM, Yehia HC, Moraes MF. Avoiding spectral leakage in objective detection of auditory steady-state evoked responses in the inferior colliculus of rat using coherence. *J Neurosci Methods* 2005;144(2):249–55.
- Firzst JB, Chambers RD, Kraus N, Reeder RM. Neurophysiology of cochlear implant users I: effects of stimulus current level and electrode site on the electrical ABR, MLR and N1-P2 response. *Ear Hear* 2002a;23(6):502–15.
- Firzst JB, Chambers RD, Kraus N. Neurophysiology of cochlear implant users II: comparison among speech perception, dynamic range, and physiological measures. *Ear Hear* 2002b;23(6):516–31.
- Galbraith G, Arbagey P, Branski R, Comerci N, Rector P. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 1995;6(17):2363–7.
- Galbraith GC. Deficient brainstem encoding in autism. *Clin Neurophysiol* 2008;119(8):1697–700.
- Greenberg S, Marsh J, Brown W, Smith J. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear Res* 1987;25(2–3):91–114.
- Hall J. Auditory brainstem frequency following responses to waveform envelope periodicity. *Science* 1979;205(4412):1297–9.
- Jing H, Chojnowska C, Heim S, Thomas J, Benasich AA. Timing errors in auditory event-related potentials. *J Neurosci Methods* 2004;138(1–2):1–6.
- Johnson KL, Nicol TG, Kraus N. Brainstem response to speech: a biological marker of auditory processing. *Ear Hear* 2005;26(5):424–34.
- Kraus N, Banai K. Auditory-processing malleability. Focus on language and music. *Curr Directions Psychol Sci* 2007;16(2):105–10.
- Krishnan A. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 2002;166(1–2):192–201.
- Krishnan A, Xu Y, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004;189(1–2):1–12.
- Luecke D, Stockmanns G, Gallinat M, Kochs EF, Schneider G. Auditory evoked potentials for the assessment of depth of anaesthesia: different configurations of artefact detection algorithms. *Biomed Tech (Berl)* 2007;52(1):90–5.
- Marsh JT, Worden FG, Smith JC. Auditory frequency following response: neural or artifact? *Science* 1970;169(3951):1222–3.
- Ménard M, Gallégo S, Berger-Vachon C, Durrant J, Collet L. Auditory steady-state response evaluation of audiometric thresholds in cochlear implants. *Int J Audiol* 2004;43:s39–43.

- Ménard M, Gallégo 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(1–2):105–13.
- Moushegian G, Rupert AL, Stillman RD. Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroenceph Clin Neurophysiol* 1973;35(6):665–7.
- Muehler R, von Specht H, Pethe J. Comparison of methods for reducing residual noise in suprathreshold early auditory evoked potential registration. *Laryngorhinootologie* 1998;77(7):382–7.
- Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 2007;104(40):15894–8.
- Patuzzi RB. A model of the generation of the cochlear microphonic with nonlinear hair cell transduction and nonlinear basilar membrane mechanics. *Hear Res* 1987;30:73–82.
- Picton TW, John MS, Dimitrijevic A, Purcell D. Human auditory steady-state responses. *Int J Audiol* 2003;42(4):177–219.
- Picton TW, John MS. Avoiding electromagnetic artifacts when recording auditory steady-state responses. *J Am Acad Audiol* 2004;15(8):541–54.
- Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am* 2004;116(6):3581–93.
- Russo NM, Nicol TG, Musacchia G, Kraus N. Brainstem responses to speech syllables. *Clin Neurophysiol* 2004;115(9):2021–30.
- Sadafi HA, Cadusch P, Wood AW. Real-time recording of neuropsychophysiological parameters during 50 Hz magnetic field exposure. *Australas Phys Eng Sci Med* 2005;28(1):43–50.
- Sanchez JT, Gans D. Effects of artifact rejection and bayesian weighting on the auditory brainstem response during quiet and active behavioral conditions. *Am J Audiol* 2006;15(2):154–63.
- Small SA, Stapells DR. Artifactual responses when recording auditory steady-state responses. *Ear Hear* 2004;25(6):611–23.
- Song JH, Skoe E, Wong PCM, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 2008;20(10):1892–902.
- Stakenburg M, Wit HP. Piezoelectric earphones for artefact-free recording of auditory brainstem responses (ABR). *Scand Audiol* 1983;12(1):79–80.
- 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.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 2007;10(4):420–2.



# Influences acoustiques sur les ABR.

---

## Sommaire

<b>3.1 Etude pilote : Speech ABR à /ba/</b> . . . . .	<b>82</b>
3.1.1 Matériel et méthodes . . . . .	82
3.1.2 Résultats : . . . . .	84
<b>3.2 Effets comparés de l’enveloppe et de la structure fine sur le Speech- (et Analogue-) ABR</b> . . . . .	<b>87</b>
3.2.1 Parole vs Analogues-en-enveloppe-temporelle, inharmoniques et spectralement appauvris 1. . . . .	87
3.2.2 Parole vs Analogues-en-enveloppe-temporelle, inharmoniques et spectralement appauvris 2. Variation sur les analogues. . .	104
3.2.3 Parole vs Analogues-en-structure-harmonique, et spectralement riches. . . . .	117
<b>3.3 Effet de la structure harmonique : sons purs vs sons complexes harmoniques.</b> . . . . .	<b>120</b>
3.3.1 Matériel et Méthodes . . . . .	120
3.3.2 Résultats . . . . .	120
3.3.3 Discussion . . . . .	122
<b>3.4 Effet de la nature du stimulus et de la hauteur tonale sur le Speech-, Music-, Analogue- ABR</b> . . . . .	<b>123</b>
3.4.1 Objectifs : . . . . .	123
3.4.2 Matériel et méthodes : . . . . .	123
3.4.3 Résultats : . . . . .	127
<b>3.5 Effet du niveau sonore sur le Speech ABR</b> . . . . .	<b>136</b>
3.5.1 Matériel et méthodes : . . . . .	136
3.5.2 Résultats : . . . . .	136
<b>3.6 Bilan : indices acoustiques influençant les ABR</b> . . . . .	<b>139</b>
<b>3.7 Article Clinial Neurophysiology [Akhoun 2008c]</b> . . . . .	<b>140</b>
<b>3.8 Article Proceedings of Meetings on Acoustics [Akhoun 2008a]</b> . . . . .	<b>141</b>

---

Ayant établi les conditions expérimentales optimales pour réaliser des enregistrements de Complexes-ABR, nous avons mené cette série d’investigations exploratoires (en manipulant des sons de parole), avec pour but de retrouver les paramètres



acoustiques qui influencent la génération de ces Speech-ABR. Nous avons procédé à nos premiers enregistrements (i) en réponse à une syllabe /ba/, puis nous avons (ii) étudié les effets respectifs de l'enveloppe (en dessous de 500 Hz) et de la structure fine (au delà de 500 Hz), appauvrie ou pas, puis (iii) les effets de la structure harmonique sur des sons complexes harmoniques et sons purs. Puis, nous avons voulu connaître (iv) l'effet de la nature du stimulus. Nous avons également recherché les effets de paramètres acoustiques élémentaires comme (v) la hauteur tonale ou (vi) le niveau sonore.

### 3.1 Etude pilote : Speech ABR à /ba/

Pour débiter l'enregistrement des ABR en réponse à des sons complexes avec le système Centor USB, encore jamais enregistrés au laboratoire, nous avons voulu, dans cette étude pilote, retrouver les Speech ABR tels qu'ils furent décrits par l'équipe Kraus, Fig.1.12.

Nous montrons ici l'aspect passe-bas ( $F_c = 500$  Hz) du Speech ABR, et la spectaculaire corrélation entre l'enveloppe temporelle du stimulus /ba/ et la FFR.

#### 3.1.1 Matériel et méthodes

Les Speech ABR furent enregistrés d'après le protocole expérimental décrit plus haut, étude 2 et dans [Akhoun 2008c].

**Sujets :** Les sujets inclus furent 23 adultes normo-entendants.

#### **Stimulation :**

Le stimulus utilisé fut une syllabe /ba/, d'une durée totale de 50 ms. Ce son fut synthétisé à partir d'un synthétiseur vocal de formants, et il est constitué d'une consonne très brève suivie d'une voyelle. Les paramètres spectraux de cette voyelle /a/ étaient stationnaires dans le temps, avec un fondamental ( $f_0$ ) à 120 Hz, et deux formants autour de 720 et 1200 Hz Fig.3.3. Ce stimulus fut synthétisé via le synthétiseur de Klatt Fig.3.1 à Calgary par l'équipe Eggermont [Tomita 2004]. Nous l'avons tronqué afin de le rendre reconnaissable, mais très court pour satisfaire aux contraintes liées à la fréquence d'échantillonnage du Centor USB ( $\frac{4000pts}{W_{ms}}$ ) et à la durée totale des expérimentations.

Le stimulus était délivré en monaural dans l'oreille droite à 60 dB Sensation Level (SL - au dessus du seuil auditif du sujet).

**Acquisition :** Mille cinq cent moyennages furent effectués pour chaque polarité (plus, /ba/, et moins, -/ba/). Auparavant, pour normaliser l'effet des Speech ABR des différents sujets sur la grande moyenne, la loi centrale normée ( $\frac{(x-\bar{x})}{\sigma(x)}$ ) fut

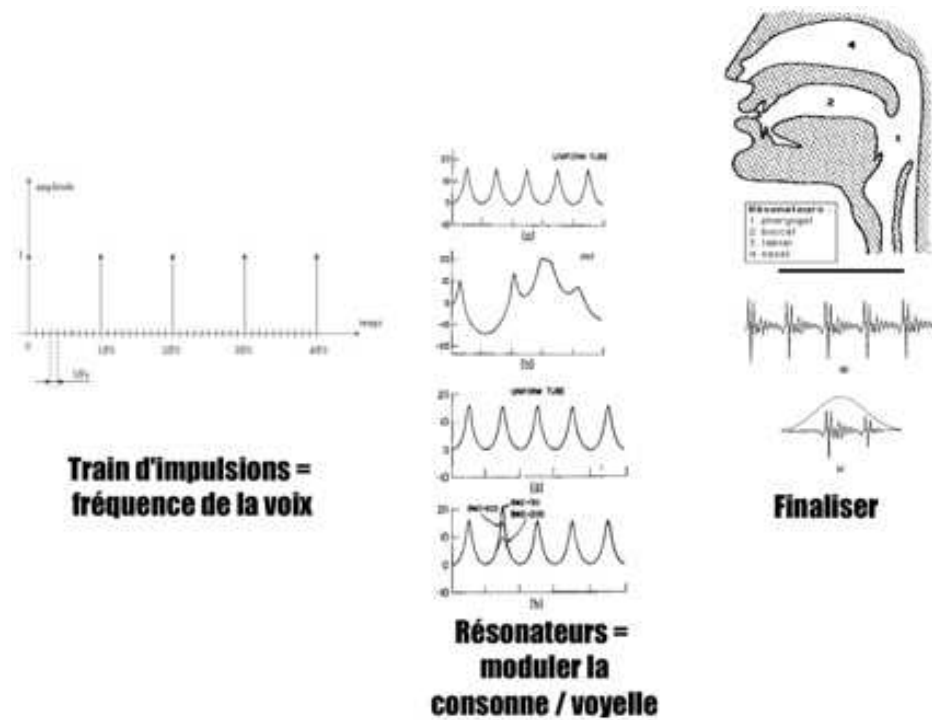


FIGURE 3.1 – **Principe du synthétiseur de voyelles de Klatt.** Il s'agit de créer un train d'impulsions (spectrales) espacées de la fréquence fondamentale (qui détermine le voisement de la voyelle). Ensuite, des filtres spécifiques à la voyelle souhaitée est appliqué à ces trains d'impulsions, afin de former les formants (qui reproduisent les phénomènes de résonance appliqués dans le conduit vocal). Certains harmoniques sont donc amplifiés, d'autres sont atténués. Ensuite, une Transformé de Fourier Inverse est appliquée à ce signal, et il est modulé en amplitude pour reproduire l'enveloppe temporelle souhaitée.

appliquée aux Speech ABR individuels. La fenêtre d'acquisition commençait 10 ms avant le début du stimulus (pré-moyennage) et se terminait 25 ms après, afin de pouvoir apprécier le rapport signal sur bruit. Le Speech ABR était enregistré avec une fréquence d'échantillonnage de 50 kHz, et il était filtré en temps réel par le Centor entre 80 et 3200 Hz, afin de se focaliser sur la gamme préférentielle de réponse temporelle (phase-locking) du tronc cérébral, voir étude 1.1. Tous les Speech ABR supérieurs à 75 micro-Volts étaient considérés comme des artéfacts musculaires et étaient rejetés.

Pour évaluer les latences du Speech ABR par rapport à l'enveloppe Fig.3.3 du stimulus (calculée comme la valeur absolue du stimulus filtrée à 500 Hz, fréquence de coupure du Speech ABR 3.2), nous avons utilisé la fonction `corrcoef`. Le filtre utilisé était un filtre à réponse impulsionnelle finie à correcteur de phase (*filtfilt*), et pour s'adapter au speech ABR, l'enveloppe fut ré-échantillonnée à 50 kHz, et elle fut inversée. Mille coefficients furent calculés entre le Speech ABR et l'enveloppe, décalée point par point. La latence du Speech ABR a été calculée comme l'abscisse du maximum de la courbe ainsi obtenue.

### 3.1.2 Résultats :

Le Speech ABR obtenu en réponse au son /ba/ fut conforme aux descriptions précédentes [Russo 2004],[Johnson 2005]. Le Speech ABR recueilli fut constitué des trois parties successives précédemment citées : le pré-moyennage, l'onset response et ensuite la FFR. Les résultats clés de cette étude sont, d'une part, la forte ressemblance du Speech ABR, avec l'enveloppe du stimulus décalée de 14.6 ms Fig.3.3, et d'autre part l'aspect passe-bas du Speech ABR (avec une fréquence de coupure de 500 Hz).

Il est tout d'abord intéressant de noter que la forme du Speech ABR fut très semblable pour les deux polarités de stimulation (i.e. en réponse à /ba/ ou son opposé -/ba/, Fig. 3.3), comme on s'y attendrait d'après le mécanisme d'encodage neural lors de la transduction cochléaire qui est insensible à la phase, voir section 1.1.3, [Hewitt 1991]. Par conséquent, la polarité alternée (somme des Speech ABR obtenus pour chacune des polarités) est inchangée par rapport au Speech ABR obtenu dans les 2 polarités, mais elle est dépourvue d'influence artéfactuelle dû au potentiel microphonique ou au couplage électromagnétique résiduel (s'il y en avait eu), car cet artéfact suit la polarité de stimulation, contrairement au Speech ABR qui est identique pour les 2 polarités. Ceci s'est vérifié Fig. 3.2 et Fig. 3.3, où les composantes artéfactuelles autour de 900 Hz s'annulent sur la polarité alternée.

Ensuite, on remarque que l'OR est située environ 7-8 ms après le début de la stimulation, ce qui correspond aux latences de l'onde V habituellement observées à ces niveaux sonores (60 dB SL). La FFR, quant à elle, se caractérise par une série de pics espacés périodiquement à une fréquence correspondante au fondamental laryngé de la voyelle de /ba/. De plus, les spectres fréquentiels indiquent que le

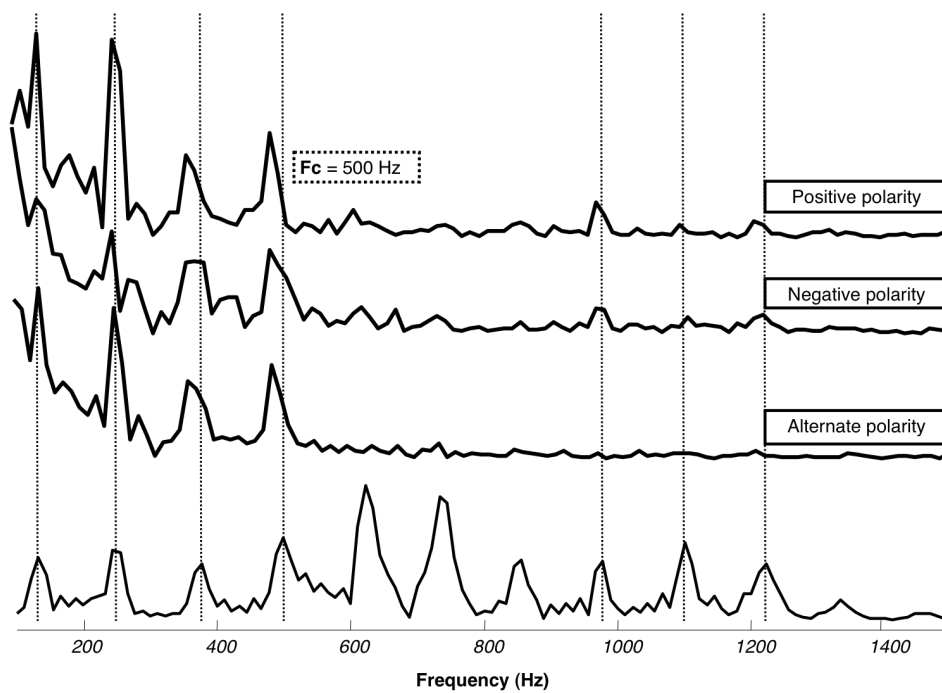


FIGURE 3.2 – **Spectre du Speech ABR pour chaque polarité (échelle linéaire)**. Les Speech ABR et le stimulus acoustique (en bas) ont des pics spectraux en commun (spécifiés par les lignes en pointillés) dans les basses fréquences jusqu'à 500 Hz ( $f_c$  = fréquence de coupure du Speech ABR). Notons que quelques harmoniques peuvent être observés sur les polarités positives et négatives, mais que ces pics avaient disparus sur la polarité alternée : ainsi, ces pics seraient des reliquats d'artéfact électromagnétique ou de potentiel microphonique.



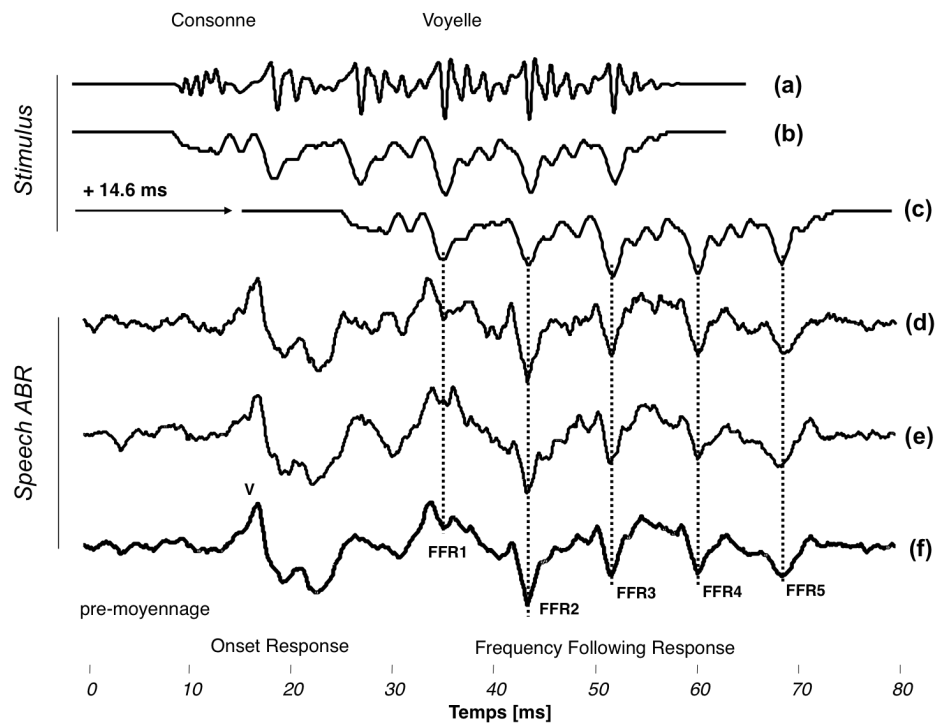


FIGURE 3.3 – L’enveloppe temporelle du stimulus est contenue telle quelle dans le Speech ABR. Les deux traces du dessus sont celles du stimulus acoustique (a) et de son enveloppe (b). L’enveloppe est obtenue par filtrage passe-bas (à 500 Hz, qui est la fréquence de coupure du Speech ABR) de la valeur absolue du stimulus. Cette enveloppe, décalée de 14.6 ms (c), se retrouve de façon très distincte dans le Speech ABR, comme en témoignent les lignes en pointillés.

Speech ABR correspond très bien à celui du stimulus jusqu'à 500 Hz (Fig. 3.2). Mais plus encore, nous avons mis en évidence une bonne correspondance de phase entre l'enveloppe temporelle inversée du stimulus (calculée comme étant la valeur absolue filtrée passe-bas à 500 Hz), lorsque celle-ci est décalée de 14.6 ms et la FFR (Fig. 3.3). Contrairement aux latences de l'OR, cette latence de FFR est supérieure à ce que nous aurions attendu d'une réponse du tronc cérébral. Mais des latences apparentes de cet ordre ont été rapportées avec des modulations d'amplitudes (qui peuvent être considérées comme un modèle de voyelle), et ([John 2000], [Purcell 2004]) les ont reliées à des interactions intra-tronc cérébral ou à des mécanismes efférents (top-down) de rétroaction corticales.

## 3.2 Effets comparés de l'enveloppe et de la structure fine sur le Speech- (et Analogue-) ABR

### 3.2.1 Parole vs Analogues-en-enveloppe-temporelle, inharmo- niques et spectralement appauvris 1.

Intrigués par la ressemblance entre la FFR et l'enveloppe du stimulus dans l'étude précédente, nous avons voulu savoir si réciproquement, l'enveloppe était le trait acoustique nécessaire et suffisant pour engendrer une FFR. Nous montrons ici que ce n'est pas le cas.

#### 3.2.1.1 Matériel et Méthodes

- *Sujets* :

Les ABR ont été enregistrés sur 15 sujets normo-entendants.

- *Stimuli acoustiques* : Dans cette expérimentation, les stimuli étaient (i) deux syllabes enregistrées /ba/ et /pa/ situées aux deux extrémités d'un continuum de Voice Onset Time (VOT) [Veuillet 2007], et (ii) leurs analogues. Le VOT est l'écart de temps séparant l'ouverture du conduit vocal du début du voisement. Ces deux syllabes avaient la même voyelle, et ne différaient l'une de l'autre par la durée de la consonne et la zone de coarticulation (VOT de -55 et 10 ms respectivement). Ces phonèmes faisaient 230 ms de long, avec la voyelle qui commence au même moment à 140 ms. Le pré-moyennage était de 10 ms. Le fondamental (F0) était stable proche de 200 Hz, et les 4 formants (F1 à F4) variaient quant à eux autour de 750, 1660, 2800 et 4400 Hz.

Les **analogues** furent déduits de ces caractéristiques acoustiques pour avoir la même enveloppe temporelle, mais une structure fine appauvrie et très simplifiée (où ils sont comparés en temporel Fig. 3.4, et en spectral Figs.3.8 et 3.10 sur les deux fenêtres du haut). Ces analogues furent synthétisés sous Matlab (the Mathworks ®) comme la somme de 5 sinusôides (de phase 0) à la fréquence de F0 et des formants de la voyelle, et pondérées

par l'amplitude des formants telles qu'elles étaient lues sur la FFT globale de la syllabe correspondante. Enfin, cette somme de sinusoides était multipliée par l'enveloppe temporelle de la syllabe correspondante (valeur absolue, filtrée passe-bas à 500 Hz par un filtre à correction de phase de Butterworth d'ordre 4). La formule de ces analogues est donc de la forme :  $analogue(t) = Env(t) * \frac{1}{751} (751 \sin(2\pi F_0 t) + 193 \sin(2\pi F_1 t) + 33 \sin(2\pi F_2 t) + 4.2 \sin(2\pi F_3 t) + 4.2 \sin(2\pi F_4 t))$ .

Notons que l'analogue ainsi synthétisé ne ressemble perceptivement pas à la syllabe de départ (ce qui apparaît peu surprenant au vu de l'appauvrissement spectral au delà de 500 Hz); gardons à l'esprit que ces analogues furent synthétisés uniquement dans le but d'avoir la même structure temporelle (i.e. la même enveloppe) et d'en déduire son influence sur l'ABR engendré.

- **Mesure des ABR :**

Les ABR ont été enregistrés avec l'appareillage Centor USB (Racia-Alvar <sup>®</sup>), par ailleurs utilisé dans le service audiolgique pour l'acquisition des clic-ABR en routine clinique. Les ABR furent enregistrés par 4 électrodes Ag/AgCl sur le vertex (sommet du crâne), les électrodes de références étant sur la mastoïde droite (voie 1), et sur la mastoïde gauche (voie 2). L'électrode de masse fut placée sur le front. Lors des enregistrements, la valeur des impédances fut maintenue en dessous de 5 k $\Omega$ . Le dispositif expérimental utilisé est optimisé pour éliminer toute influence des artéfacts électromagnétiques (Fig.2.15). C'est celui que nous avons précédemment décrit [Akhoun 2008d].

- *Chaîne de stimulation* : Les stimuli utilisés étaient joués par le logiciel Audacity <sup>®</sup> par blocs de 800 traces, à une cadence de 3 stimuli par seconde. La stimulation était binaurale à un niveau de 45 dB Sensation Level (au dessus du seuil du sujet), ce qui correspond à un niveau entre confortable-plus et fort, en fonction des sujets.

Le niveau sonore était maîtrisé par un audiomètre AC-30 (Interacoustics <sup>®</sup>), qui reliait la carte-son Soundblaster Audigy <sup>®</sup> aux transducteurs acoustiques : des inserts-earphones Biologic <sup>®</sup>. Les inserts étaient blindés par une cage de Faraday reliée à la masse électrique du dispositif expérimental, qui diffusent le son au bout d'une paire de tubes de caoutchouc. Notons que le transfert de la stimulation acoustique dans les tubes d'air induit 1 ms de retard, qui a été ensuite retranchée dans les enregistrements, Fig. Annexe A.3.

- *Chaîne d'acquisition* : Pour l'expérience 1, chaque ABR fut enregistré dans une fenêtre d'analyse de 300 ms, composée de 4000 points; ce qui fait une fréquence d'échantillonnage de 13.3 kilohertz. Les ABR

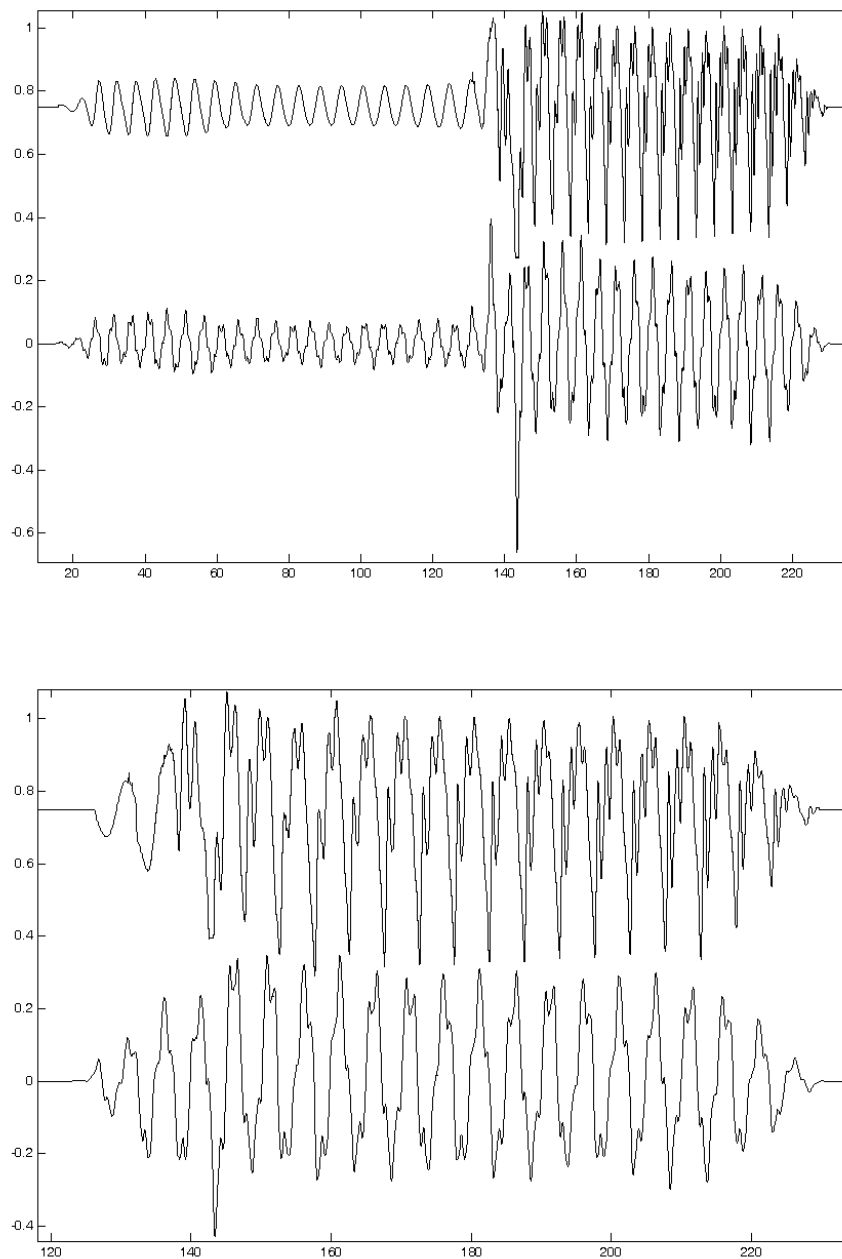


FIGURE 3.4 – /ba/, /pa/ et leurs analogues. La syllabe /ba/ est issue d'un enregistrement (voix de femme). Notons que /pa/ et /ba/ ont strictement la même (portion steady-state de) voyelle. L'analogue du /ba/, synthétisé comme une somme de sinus (aux fréquences des formants, mais non-harmoniques) modulés par l'enveloppe du /ba/, présente une structure temporelle (en dessous de 500 Hz) identique à celle du /ba/, mais très fortement appauvrie spectralement au delà de 500 Hz (c'est pourquoi il n'est pas significatif). L'abscisse est en millisecondes, l'ordonnée en unités arbitraires.

ainsi enregistrés furent ensuite filtrés en temps réel entre 80 et 3200 Hz. Chaque trace trop ample (au-delà de  $75 \mu V$ ) était assimilée à un artéfact musculaire et rejetée.

Dans le but d'éliminer tout artéfact lié au potentiel microphonique cochléaire [Verpy 2008], nous avons enregistrés les ABR engendrés par chaque polarité positive et négative (le stimulus et son opposé). La transduction sensorielle dans l'organe de Corti est très peu sensible à la polarité, tandis que les artéfacts reproduisent le stimulus tel-quels. Par conséquent, la somme des ABR dans chaque polarité (dite polarité alternée) réduit les artéfacts sans affecter la réponse neurophysiologique auditive. Un ABR en polarité alternée est en fait le moyennage de 1600 traces (800 traces de chaque polarité).

- **Analyse a posteriori des ABR :**

Les ABR furent analysés visuellement d'après leurs traces temporelles, puis leur contenu spectrotemporel fut analysé en temps-fréquence.

Les grandes moyennes temporelles combinent les ABR de tous les sujets. Elles furent calculées comme la somme algébrique des données individuelles dans une polarité donnée (positive, négative, alternée). Auparavant, les ABR furent normalisés (composante continue retranchée et correction des erreurs standard par loi normale centrée,  $\frac{(x-\bar{x})}{\sigma(x)}$ ).

Les spectrogrammes (d'énergie) furent obtenus sous Matlab (par transformée de Fourier glissante de 1024 points à court-terme, se recouvrant à 90%), à partir des traces temporelles normalisées. Pour les ABR, la précision fréquentielle était de 13 Hz, et de 11 Hz pour les stimuli acoustiques. Les grandes-moyennes des spectrogrammes furent obtenus en faisant la moyenne des spectrogrammes individuels. D'après la méthode du F-test décrite par [Picton 2003a], les composantes spectro-temporelles furent considérées présentes si l'énergie dans cette région est supérieure de 6 dB à celles environnantes. Des coupes horizontales de ces spectrogrammes, la durée et l'amplitude de la FFR peuvent être déduite des bandes de fréquences correspondantes ( $F_0$  et ses harmoniques). Notons que les informations de phase ont été ignorées, ce qui représente un avantage, car nous avons constaté que certains décalages (dits jitter) peuvent exister entre les FFR des différents sujets, et ainsi artificiellement annuler les grandes moyennes temporelles des FFR [Akhoun 2008a]. Enfin, l'activité globale dans les bandes de fréquences du spectrogramme a été estimée en sommant algébriquement les énergies contenues dans ces coupes horizontales (ce qui donne un nombre sans dimension pertinente [dB \* ms]). Nous avons de même déduits la représentation spectrale des coupes verticales des spectrogrammes à des moments pertinents (consonne / transition consonne-voyelle / voyelle steady-state). De la même façon, les coupes verticales des spectrogrammes donnent des représentations des spectres large-bande centrées sur ces moments pertinents.

- **Description du modèle auditif :**

Nous avons utilisé le modèle Auditory Image Model (AIM) [Patterson 1995] afin de simuler la réponse électrophysiologique du système auditif périphérique (entre la cochlée et le noyau cochléaire). Le modèle AIM (version 2006) se comporte comme une série de modules en cascade qui simulent les différentes étapes du traitement de l'information auditive. Notons que cette simulation, basée sur des études animale in-situ, ne peut se borner qu'à donner une idée des mécanismes en jeu. Cette série de modules est organisée de la façon suivante [Bleack 2004].

(i) Le stimulus acoustique qui est importée sous forme de fichier wave subit tout d'abord un traitement pré-cochléaire (**Pre-cochlear processing** - PCP) qui reproduit les étapes de filtrage passe bande entre 500 et 15000 Hz du conduit auditif externe (pavillon de l'oreille externe et chaîne d'osselets dans l'oreille moyenne), [Moore 2003].

(ii) Dans le module suivant (**Basilar Membrane Motion** - BMM), l'étrier entraîne la vibration de la membrane basiliaire, où le stimulus est décomposé spectralement (des hautes fréquences à la base de la cochlée aux basses fréquences à l'apex). Cette décomposition spectrale opère sous la forme d'un banc de 50 filtres gammatones se recouvrant, et répartis entre 100 et 6000 Hz [Patterson 1995], présentés au 1.1. Notons que les filtres hautes-fréquence sont stimulés avant ceux basses-fréquence, en raison du retard de la propagation de l'onde mécanique le long de la membrane basiliaire.

(iii) Lors de cette décomposition fréquentielle, les cellules ciliées interne traduisent l'amplitude de la vibration mécanique cochléaire en un train de potentiels d'action se propageant sur le nerf auditif dans chaque canal (filtre) auditif. C'est le troisième module (**Neural Activity Pattern** - NAP). Une particularité de cette transduction est que les cellules ciliées redressent le signal filtré dans le module PCP (les décharges de potentiels d'action n'ont lieu que dans un sens du mouvement de la membrane basiliaire) et lui appliquent un filtrage passe-bas, vu leur temps de réponse limité pour les vibrations trop rapides en raison de la nécessité de renouveler les transmetteurs synaptiques [Gaumond 1982], [Hewitt 1991]. Le module NAP tient également compte des phénomènes d'adaptation et de réfraction des nerfs, qui ont tendance à moins décharger quand ils sont trop sollicités. De plus, la disparition du phase-locking dans les hautes-fréquences (au-delà de 5kHz) est reprise dans ce module. Notons enfin le parallèle entre ce mécanisme et la façon dont nous avons calculé l'enveloppe de nos stimuli dans l'expérience 2, ainsi que le rationnel de l'utilisation de la polarité alternée.

(iv) Le modèle propose ensuite de modéliser l'action du tronc cérébral supérieur et du cortex auditif, à travers les images auditives stabilisées (**Stabilised**

**Auditory Images** - SAI). Ces centres nerveux intégrés sont sensés assimiler l'information véhiculée dans tous les filtres auditifs, afin d'en extraire les informations les plus pertinentes pour comprendre le son. Par exemple une auto-corrélation de l'activité neurale conglomérée de toutes les sorties des filtres auditifs du module NAP donne des maxima dont l'écart est relié à la période de hauteur tonale perçue par le sujet [Hewitt 1991], [Lyon 1996]. Toutefois, nous ne nous y intéressons pas, car c'est précisément cette activité que nous pensons enregistrer à travers les Speech ABR. Autrement dit, nous considérons la sortie du module NAP comme une simulation de l'entrée de notre *système* (générateurs neurophysiologiques des Speech ABR).

### 3.2.1.2 Résultats

Comme les études précédentes l'avaient suggérées, nous n'avons observé aucun effet de la polarité et de la latéralité.<sup>1</sup> Par conséquent, toutes les données sont présentées en polarité alternée, sur la voie 1. Le résultat phare est que les ABR obtenus en réponse à /ba/, comme à /pa/ ont montré des FFR bien mieux définie que pour leurs analogues (Fig.3.5).

En y regardant de plus près, les analogues ont donné lieu à des Onset Response, situées peu après l'apparition de l'équivalent de la voyelle, entre 130 et 150 ms. Toutefois, malgré la présence visible de cette onde, le calcul de l'énergie locale (par moyenne quadratique - RMS) n'a pas donné de pic locaux dans cette portion de l'ABR. A l'inverse, peu d'Onset Response fut obtenue en réponse au /ba/ (le pic situé vers 50 ms ressemble à une Onset Response, mais s'apparente davantage à une middle-latency response comme dans le cas des sons purs), alors qu'un pic existe à 170 ms pour le /pa/. Cette Onset Response est également visible sur les spectrogrammes, et se présente comme un rectangle clair (donc énergétique) dans les basses fréquences entre 130 et 150 ms, et entre 50 et 200 Hz, Fig.3.6. Dans la lignée des études précédentes, la FFR (lorsqu'elle fut enregistrée pour /ba/ ou /pa/, mais pas pour leurs analogues) fut concentrée dans les basses fréquences, avec une forte présence à F0, et à ses harmoniques jusqu'à environ 600 Hz. Ces mesures sont représentées sur les spectrogrammes par les barres horizontales blanches dans les bandes de fréquences correspondantes, pendant la durée de la voyelle. Toutefois, dans la condition /pa/, on a constaté une synchronisation jusqu'à 1000 Hz, allant ainsi au-delà des fréquences de coupure habituellement reportées. Notons enfin

1. En effet, les coefficients d'intercorrélation entre les grandes moyennes pour les deux polarités fut maximal pour des décalages de 0 ms pour tous les stimuli ( $R = 0.27, 0.13$  et  $0.37$  for /ba/, analogue /ba/ and analogue /pa/). Notons que pour la condition /pa/, le coefficient de corrélation maximal ( $R = 0.12$ ) fut mesuré pour un décalage de 4.8ms, mais ce pic était proche d'un autre centré en 0 ( $R = 0.11$ ). De plus, les ABR ainsi obtenus furent similaires sur les deux canaux d'enregistrements ("droite", voie 1 et "gauche" voie 2 en fonction de la position des électrodes). Ceci reste visible sur les traces temporelles (Fig.3.5), et cette conjecture est confirmée par intercorrélation entre les ABR de chaque voie, qui fut maximal pour un décalage de 0 ms ( $R = 0.47, 0.43, 0.36, 0.63$  pour /ba/, /pa/, analogue /ba/, et analogue /pa/). Ceci suggère une absence de latéralisation visible, tout du moins avec la technique des ABR.

que les composantes fortement énergétiques à 900 ou 1200 Hz sont le résultat de l’influence de l’artéfact électromagnétique lié à l’alimentation électrique (50 Hz), distordue.

Nous n’avons pas observé de tendances différentes pour les FFR engendrés en réponse à /ba/ et à /pa/ (Fig.3.5). En revanche, la différence entre la FFR obtenue pour les vrais sons de parole par rapport à leurs analogues fut très marquée (Fig. 3.12). Nous avons tenté d’analyser la répartition fréquentielle de ce clivage, dans le but de mieux comprendre les paramètres acoustiques auxquels la FFR est sensible. Pour cela, nous avons représentés la distribution d’énergie dans les bandes de fréquences les plus intéressantes du spectrogramme ( $F_0 = 200$  Hz, et ses deux harmoniques suivants), sur la Fig.3.12. Il est bien visible sur la partie gauche de la Fig.3.12 que les syllabes (en gras) /ba/, comme /pa/, et leurs analogues (en traits fins) ont la même quantité d’énergie à  $F_0$  pendant la portion voisée du son, mais que les syllabes sont plus énergétiques pour les fréquences supérieures. A l’inverse, sur la partie droite de la Fig.3.12, on remarque que la FFR était bien plus énergétique dans la bande de fréquence de  $F_0$  (200 Hz) lorsqu’elle est engendrée par les syllabes /ba/, comme /pa/ (traits gras), par rapport aux analogues (traits fins) dans la zone de la voyelle (rectangles en pointillés). Pour les fréquences supérieures, les syllabes comme les analogues présentent des répartitions énergétiques comparables. Ces deux tendances furent confirmées sur les histogrammes de la Fig.3.13 (et idem dans la Fig.3.14). Dans ces deux figures, les différences d’énergie dans 8 bandes de fréquences données (100 (bruit de recueil), 200 ( $F_0$ ), 400, 600, 800, 1000, 1200 et 1400 Hz) sont mesurées entre (i) les grandes moyennes des ABR obtenues en réponse au /ba/ (resp. /pa/) et à son analogue, (ii) les ABR individuels, (iii) les stimuli et (iv) entre l’activité neurale modélisée par AIM. Conformément à ce qui est décrit pour la Fig.3.12, on remarque un contraste à 200 Hz ( $F_0$ ) entre la FFR, la stimulation acoustique et AIM.

D’un côté, les ABR sont significativement plus énergétiques pour /ba/ par rapport à son analogue (ANOVA :  $F(1,118) = 4.811$ ,  $p=0.03$ ), et même de façon encore plus franche pour le /pa/ (ANOVA :  $F(1,118) = 16.6$ ,  $p<0.001$ ) ; la grande moyenne accentue ce résultat. De l’autre côté, acoustiquement, les /ba/, comme /pa/ furent similaires pour  $F_0$  (ils furent synthétisés de fait pour avoir la même enveloppe, ce qui se voit le mieux sur les 3 premiers pics communs dans la zone encadrée en pointillé sur les Fig. 3.8 et 3.10). Il est intéressant de constater que le modèle neural AIM (module NAP) simule en effet une activité neurale différente pour les syllabes et leurs analogues, Annexe (B, B.2). Ce modèle prédit une activité quasi-semblable pour le /ba/, comme pour le /pa/, et leurs analogues aux alentours de  $F_0$ , tandis que plus de disparités est attendue pour les fréquences plus élevées. Notons toutefois qu’il ne s’agit pas de code neural temporel mais de code spatial, à l’inverse de la FFR : les fréquences mentionnées ne représentent pas (nécessairement) une activité oscillatoire des neurones impliqués, mais elles représentent la fréquence centrale des filtres auditifs (issus de la tonotopie cochléaire).



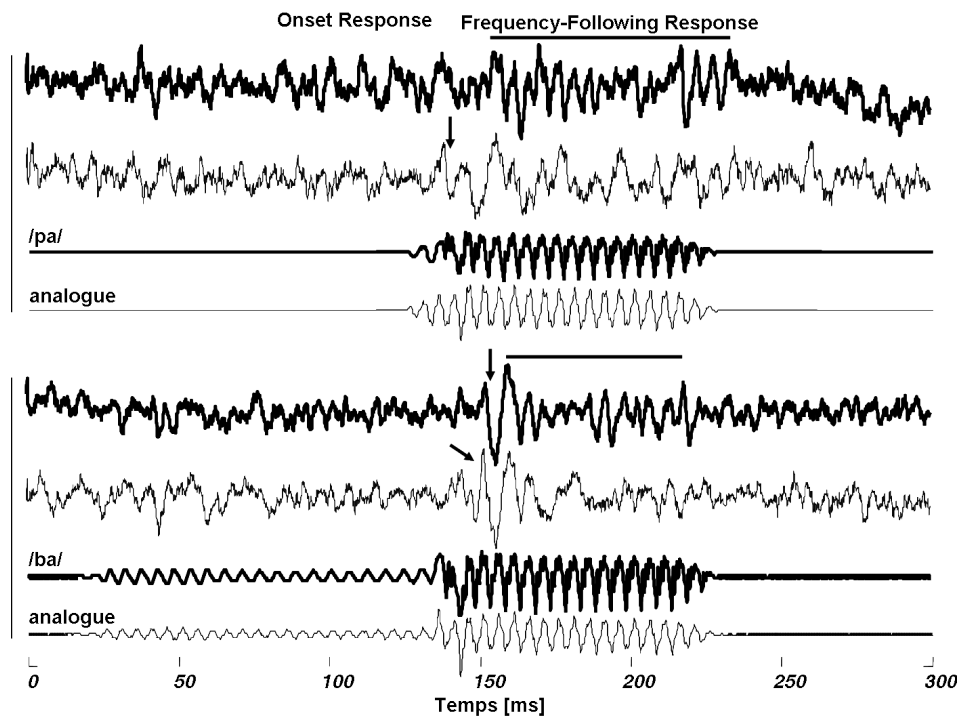


FIGURE 3.5 – **Comparaison (temporelle) des speech- et analogues-ABR.** La syllabe /pa/ est représentée en gras, ainsi que la grande moyenne ( $n=15$ ) des Speech ABR engendrés. L'analogue du /pa/ (structure temporelle identique, mais très fortement appauvrie spectralement) est représenté de la même façon en trait fin. Tout ce qui correspond au /ba/ est en-dessous.

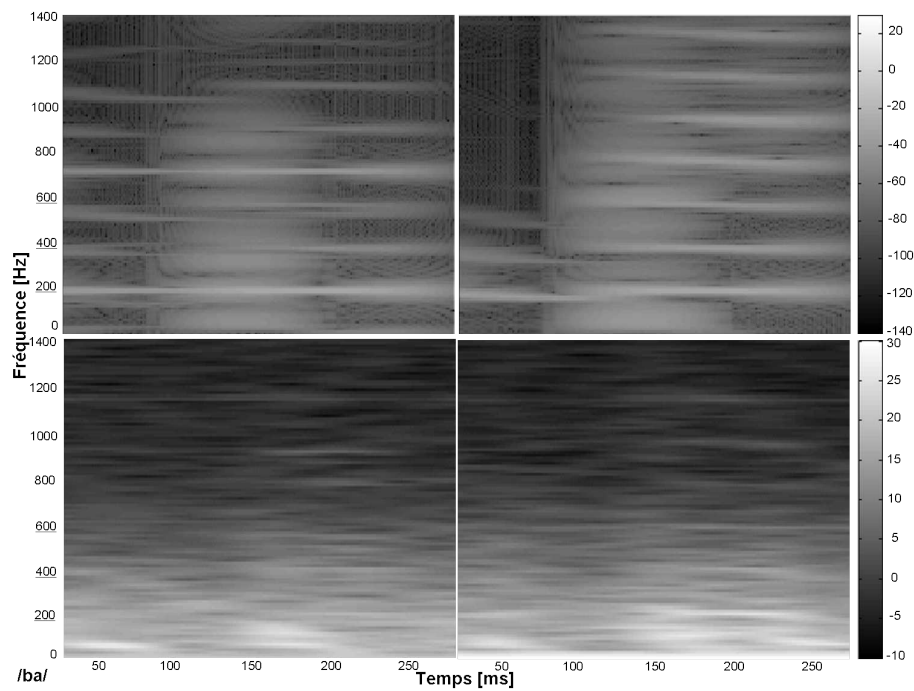


FIGURE 3.6 – **Comparaison spectro-temporelles des speech- et analogues-ABR, pour le /ba/.** Représentation temps-fréquence du vrai /ba/ (haut-droite) et de son analogue (haut-gauche), ainsi que des grandes moyennes des spectrogrammes des ABR correspondants (en dessous). Le vrai /ba/ et son analogue ont la même enveloppe temporelle, ce qui se voit ici dans la similarité entre leurs composantes à la fréquence du fondamental, tandis que la structure fine (harmoniques, formants) est beaucoup plus dynamique et progressive pour le vrai /ba/. Le /ba/ a donné lieu à une FFR bien définie (bande fréquentielle blanche à 200 Hz entre 150 et 250 ms).

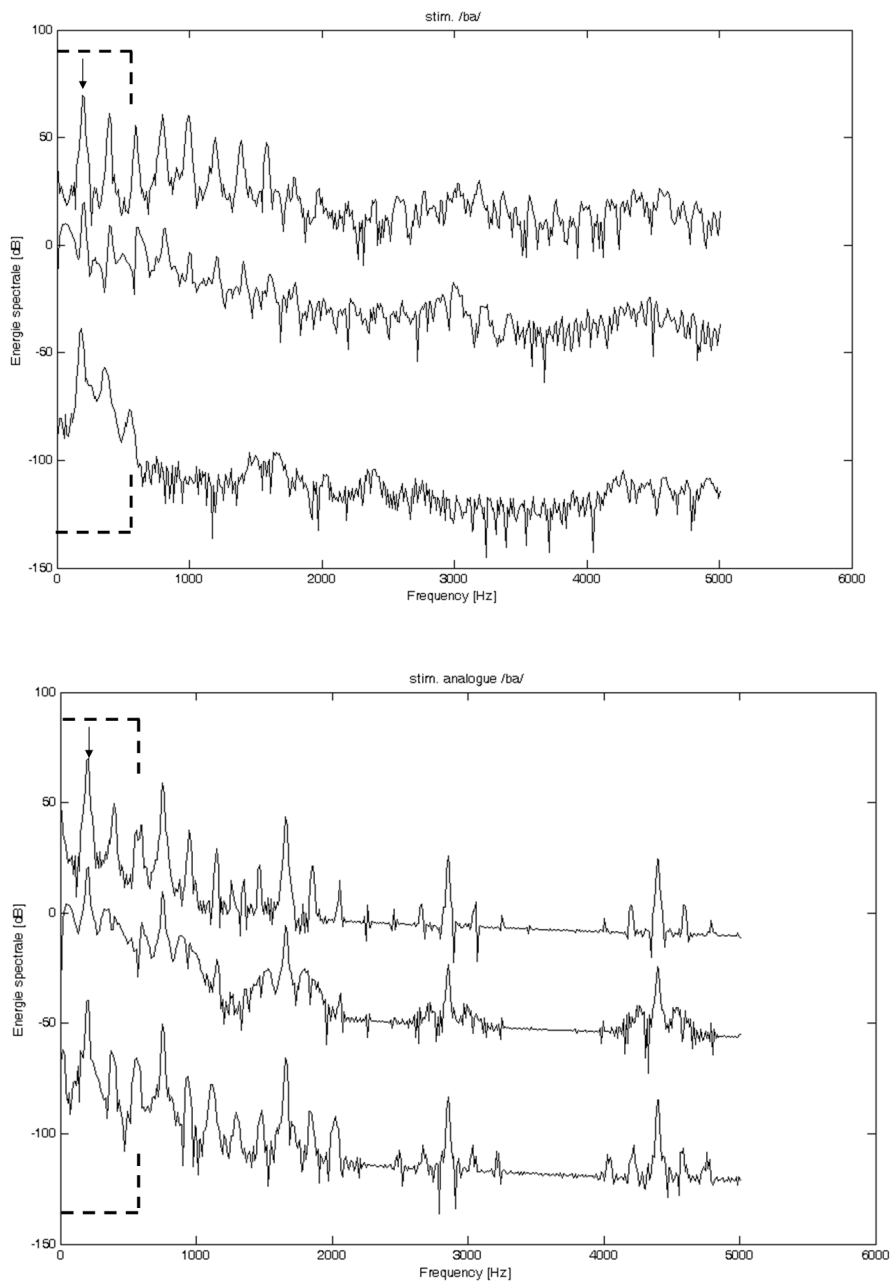


FIGURE 3.7 – /ba/. Coupe spectrales (verticales) des spectrogrammes ci-dessus (Fig.3.6). L'intérêt ici est de pouvoir repérer les harmoniques présents ou pas dans 3 zones temporelles d'intérêt : avant la voyelle (en bas), pendant la transition formantique consonne-voyelle (milieu) et pendant la portion steady-state de la voyelle (en haut). La gamme fréquentielle où /ba/ est identique à son analogue sont mis en évidence par la barre horizontale en pointillés (met en jeu  $F_0$  et ses 2 premiers harmoniques). Attention à la différence d'échelle en abscisse entre les stimuli et les FFR.

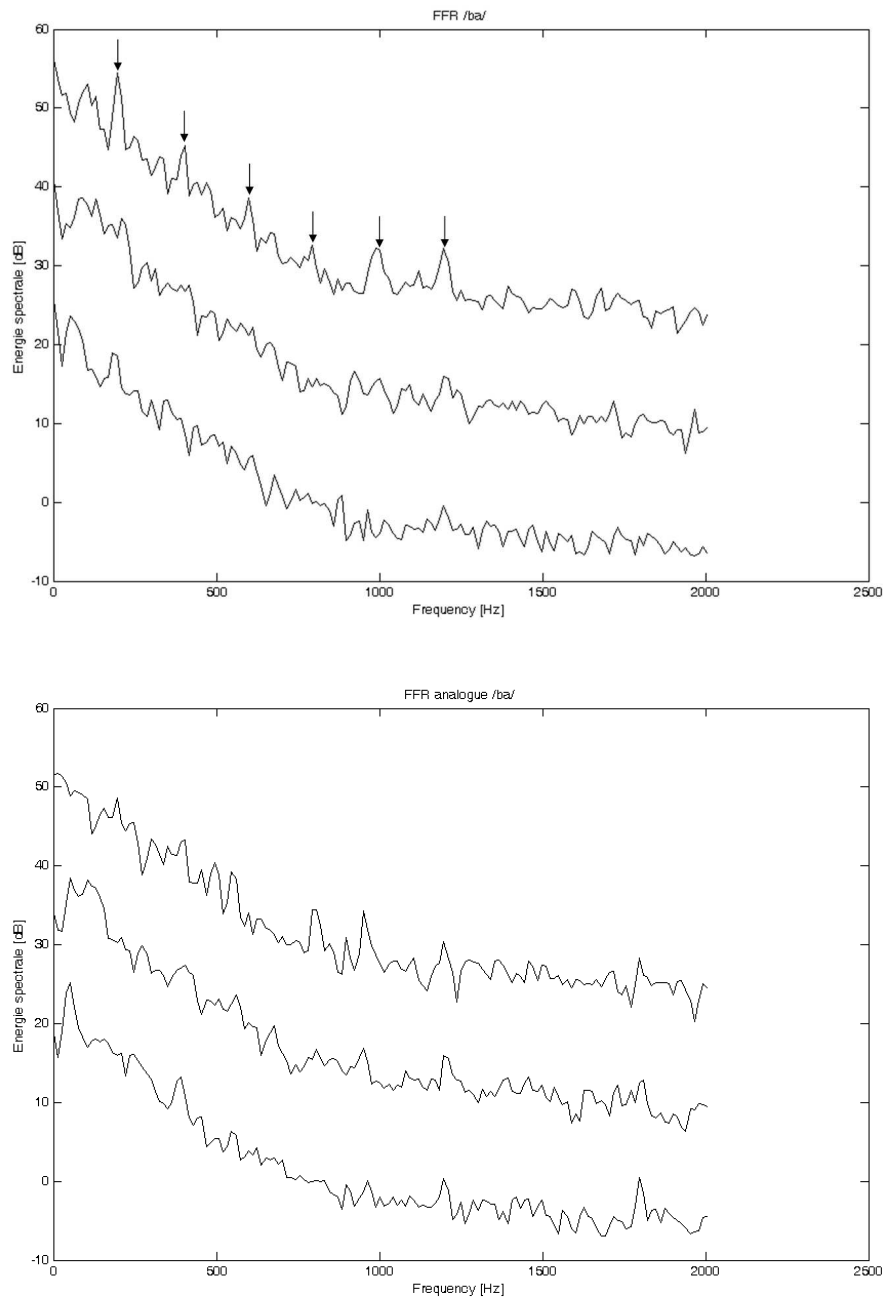


FIGURE 3.8 – /ba/. Coupe spectrales (verticales) des spectrogrammes des FFR ci-dessus (Fig.3.6). Les deux fenêtre du bas représentent les FFR, et les pics correspondant à  $F_0$  et à ses harmoniques sont signalés par des flèches lorsqu'ils sont présents. Attention à la différence d'échelle en abscisse entre les stimuli et les FFR.

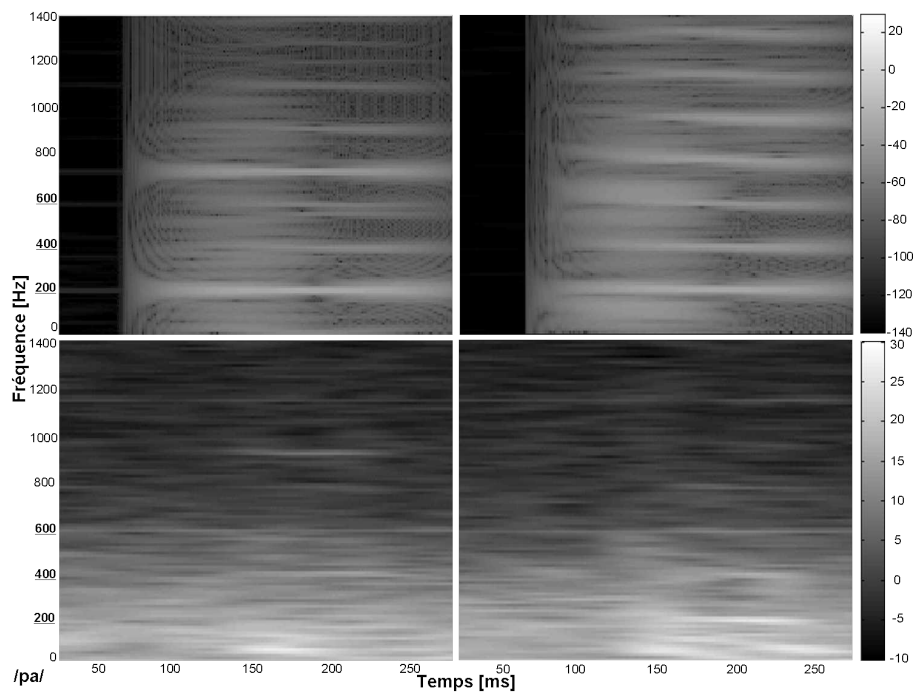


FIGURE 3.9 – (idem Fig.3.6) Comparaison spectro-temporelles des speech et analogues-ABR, pour le /pa/. La similarité des voyelles entre le /ba/ et le /pa/ est bien visible en comparant les fenêtres en haut à droite sur cette figure et la 3.6. On pourrait ajouter qu'on ne voit pas particulièrement d'effet de la consonne en comparant les Fig.3.6 et 3.9.

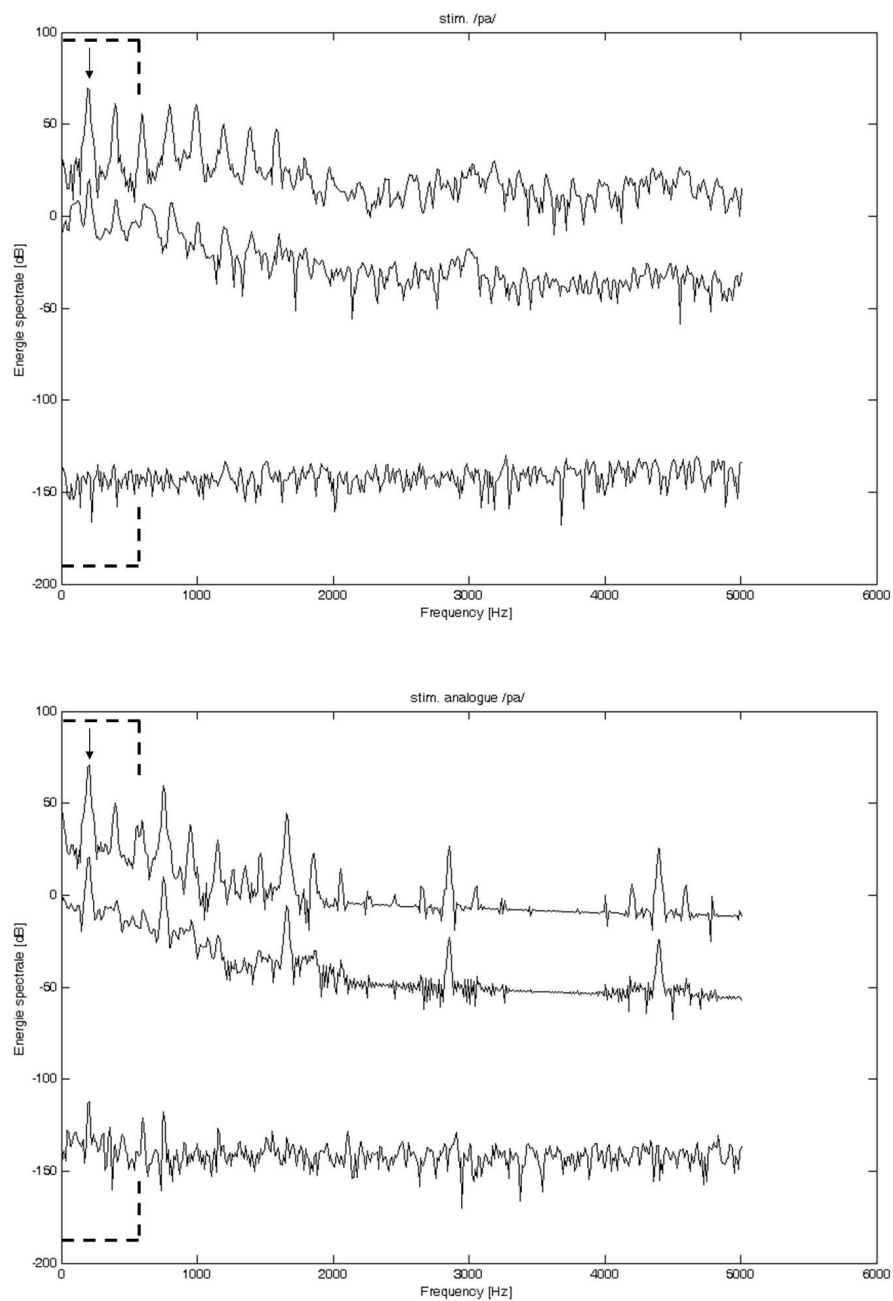


FIGURE 3.10 – /pa/. Coupe spectrales (verticales) des spectrogrammes des stimuli ci-dessus (Fig.3.9). Idem Fig.3.8

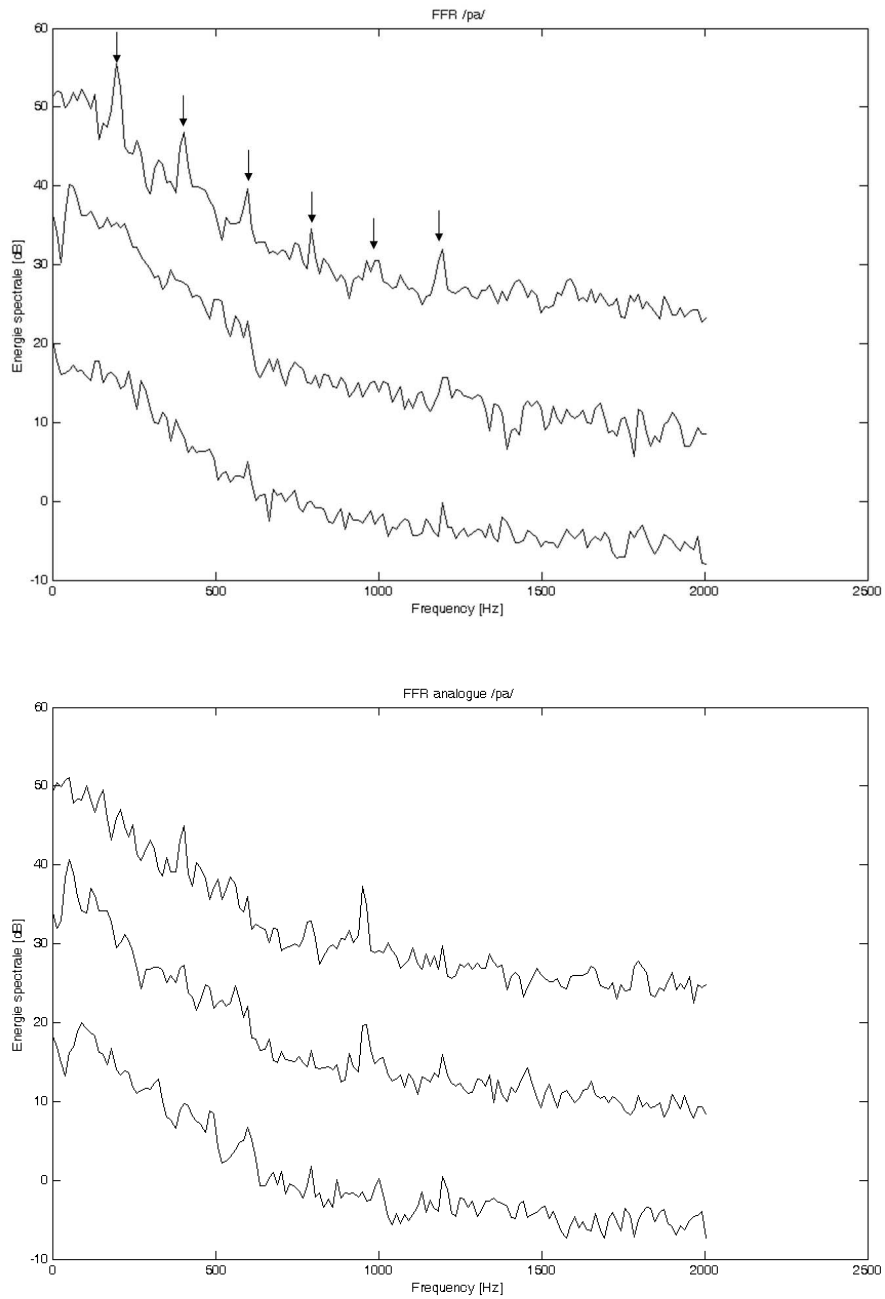


FIGURE 3.11 – /pa/. Coupe spectrales (verticales) des spectrogrammes des FFR ci-dessus (Fig.3.9). Idem Fig.3.8

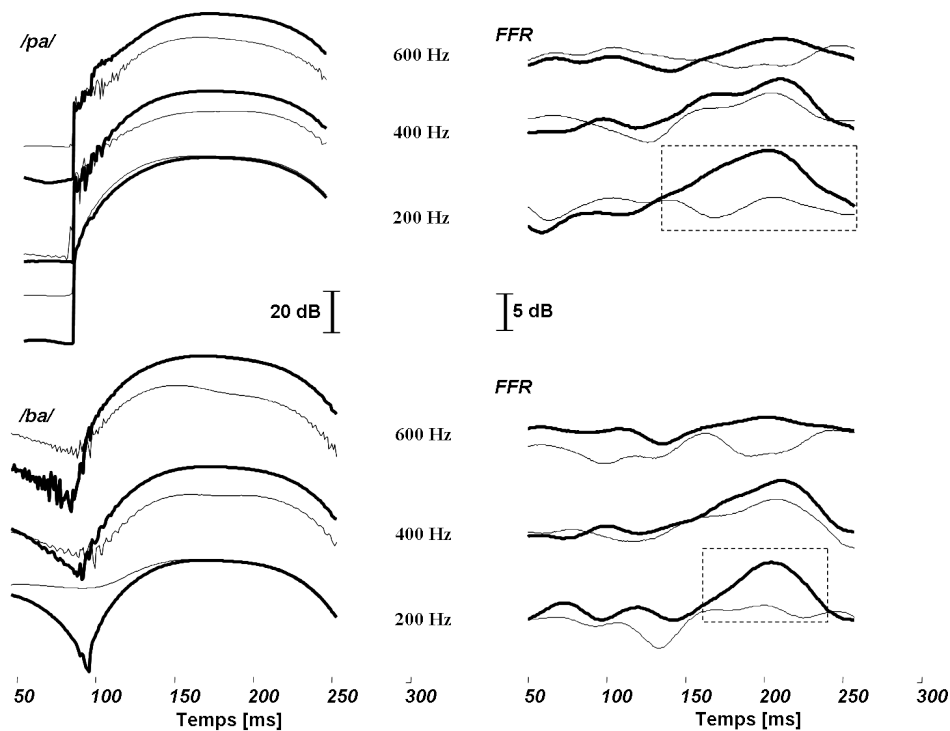


FIGURE 3.12 – **Comparaison des coupes des spectrogrammes ci-dessus (Fig.3.6 et 3.9).** L'intérêt ici est de pouvoir suivre la dynamique spectrale (évolution de l'énergie des bandes de fréquences pertinentes). Comparaison de l'énergie présente dans les bandes de fréquences correspondant au fondamental (200 Hz), et les deux premiers harmoniques (400 et 600 Hz), (i) pour la vraie parole (gras) et l'analogue (fin), et pour les stimuli (à gauche) ainsi que (ii) pour la FFR (droite), et enfin (iii) pour le /ba/ (en haut) et le /pa/ (en bas). La similarité des enveloppes des analogues et des syllabes se retrouve dans la superposition entre les courbes d'énergie à 200 Hz (que ce soit pour le /ba/ en bas ou pour le /pa/ en haut) ; en revanche, la distribution d'énergie est différente pour les harmoniques supérieures. A l'inverse, l'énergie contenue dans le Speech ABR à 200 Hz est beaucoup plus important en réponse au vrai son de parole (comme souligné par les rectangles en pointillés) ; tandis que l'énergie du Speech ABR était semblable dans les deux conditions pour les harmoniques 400 et 600 Hz.



### 3.2.1.3 Discussion

Nous avons montré ici que les caractéristiques de ces ABR (Onset Response et FFR) étaient sensibles aux caractéristiques acoustiques du stimulus, ce qui nous permet de faire le parallèle avec l'aspect fonctionnel des générateurs neurophysiologiques sous-jacents. En particulier, le résultat-clé de cette étude est que nous avons obtenus des FFR très franches en réponse à des syllabes /ba/ et /pa/, mais pas de FFR en réponse à leurs analogues (malgré le fait que leurs enveloppes soient identiques). Ce distinguo entre FFR et caractéristiques acoustiques fait pencher la balance vers une hypothèse de codage global de toutes les composantes harmoniques sous la forme d'une synchronisation sur F0. Nous nous sommes servis de ces indices de sensibilité des caractéristiques des ABR pour cibler les groupes cellulaires dans le tronc cérébral qui seraient les plus à-même d'être à l'origine des ABR ainsi recueillis. Une meilleure connaissance des origines anatomo-fonctionnelles des ABR est d'une grande utilité en vue d'implémenter cet outil dans le diagnostic des pathologies auditives, mais c'est également un des rares moyens techniques à disposition pour enregistrer l'activité neurale auditive chez l'humain éveillé.

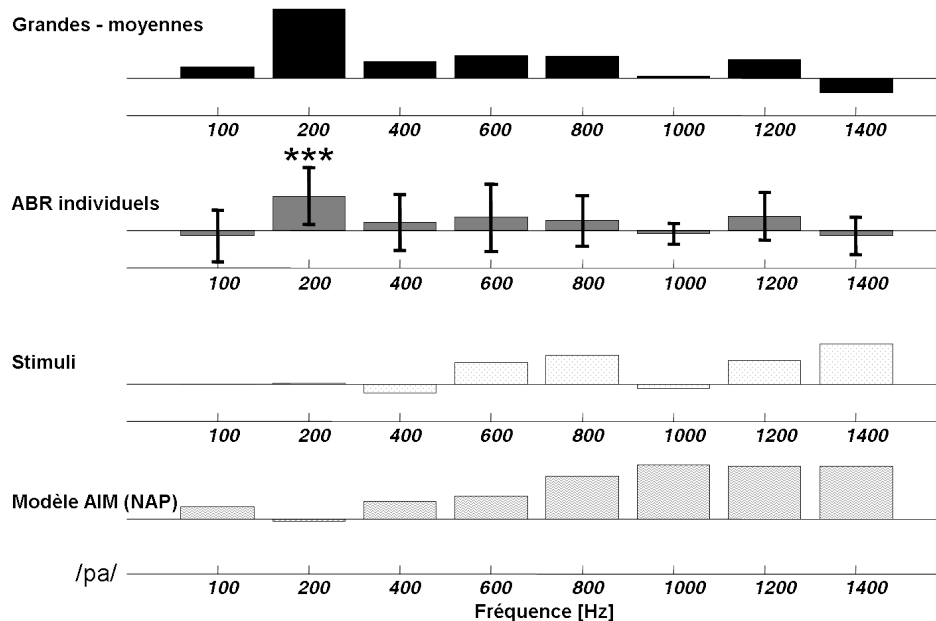


FIGURE 3.13 – Comparaison globale des composantes fréquentielles entre la /pa/ et son analogue pour 4 situations : (i) grande moyenne des ABR, (ii) des ABR individuels, (iii) des stimuli, (iv) du modèle AIM (module NAP). Ces valeurs furent mesurées aux 8 fréquences 100 (en dessous de  $F_0$ ), 200 ( $F_0$ ), 400, 600, 800, 1000, 1200 et 1400 Hz. Chaque barre dans ces histogrammes sont les différences entre les coupes des spectrogrammes pour (i) les grande moyennes des ABR du /pa/ et de son analogue, (ii) des ABR individuels, (iii) des stimuli acoustiques (Fig.3.12), et (iv) de l'activité neurale simulée par NAP en Annexe B, Fig.B.2. On remarque que la composante 200 Hz ( $F_0$ ) est très marquée pour les ABR en grande-moyenne ou en individuel (ce qui signifie que la FFR était plus énergétique à 200 Hz en réponse au vrai /pa/ par rapport à son analogue) contrairement aux stimuli acoustiques et au modèle AIM. La modélisation NAP (qui représente l'entrée supposée du système neurophysiologique de synthèse des ABR) montre davantage d'activité du code neural spatial dans les hautes fréquences, alors que l'activité en réponse aux analogues ou au /pa/ est similaire dans les basses-fréquences.

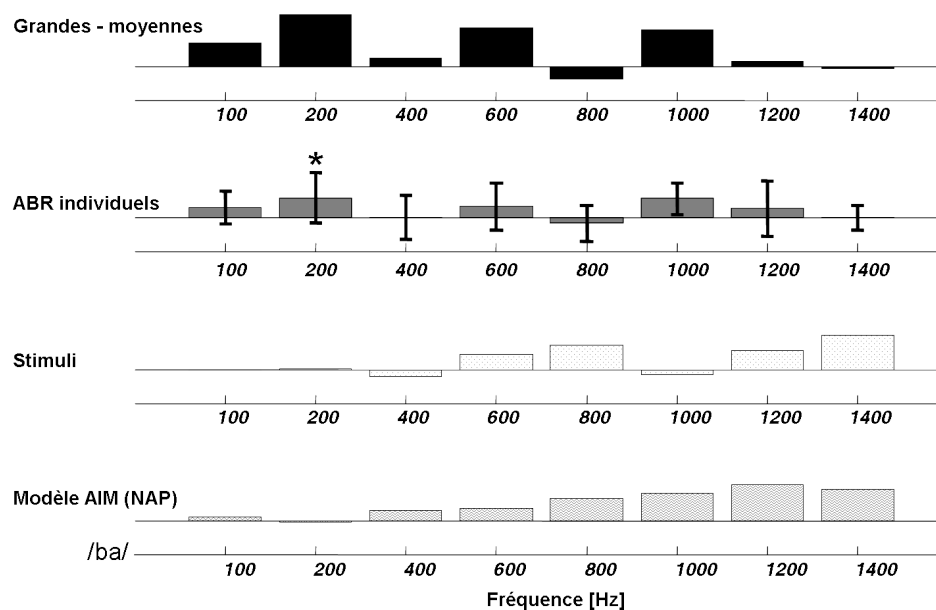


FIGURE 3.14 – idem Fig.?? pour /ba/

### 3.2.2 Parole vs Analogues-en-enveloppe-temporelle, inharmo- niques et spectralement appauvris 2. Variation sur les analogues.

Dans l'étude précédente 3.2, nous avons volontairement construit des analogues pour qu'ils aient bien la même forme temporelle que le phonème /ba/, avec une structure fine (la somme de sinus) stationnaire qui ne reproduit pas la dynamique formantique nécessaire à la bonne intelligibilité du son de parole. En pratique, les analogues n'étaient pas du tout perçus comme le /ba/ de référence utilisé dans l'étude 3.2, mais c'est précisément ce que nous recherchions.

Après avoir constaté que le vrai son de parole engendrait une FFR bien synchronisée, contrairement aux analogues, nous avons cherché à produire des analogues intermédiaires qui ressembleraient plus ou moins au /ba/ de départ, afin d'observer le lien éventuel entre la qualité de perception de ces analogues (plus ou moins ressemblants au /ba/ de départ). Cette ressemblance a été estimée par un jury d'une quinzaine de personnes, dans un processus de notation (0-7) inspiré de celui utilisé par [Kuhl 1991] lorsqu'elle voulait évaluer la ressemblance d'une voyelle dont la position de formants était décalée par rapport à une voyelle de référence.

Cette pré-étude exploratoire nous aura permis de conjecturer sur la question de savoir s'il y a un lien entre la qualité de la perception de l'analogue, les caractéristiques acoustiques retenues pour la synthèse de ces analogues, et de l'autre côté, la qualité de synchronisation de la FFR ?

### 3.2.2.1 Matériel et méthodes :

**sujets :** Neuf sujets adultes normo-entendants furent inclus dans cette étude.

**stimulation :** Les analogues furent synthétisés selon la même procédure que celle décrite précédemment dans le chap.3.2, selon l'équation  $analogue(t) = Env(t) * \frac{1}{a_0}(a_0 \sin(2\pi F_0 t) + a_1 \sin(2\pi F_1 t) + a_2 \sin(2\pi F_2 t) + a_3 \sin(2\pi F_3 t) + a_4 \sin(2\pi F_4 t))$ , où la position et le poids des formants étaient estimées sur le spectre du signal de parole de référence. Ici, nous avons voulu synthétiser, de façon empirique, des stimuli analogues intermédiaires avec plus ou moins de ressemblance avec le /ba/ de départ. Cependant, nous avons tenu à conserver le même mode de synthèse des analogues (avec les sommes de sinusoides, y-compris le fondamental) et à chaque fois à ne manipuler que certains paramètres acoustiques : précisons que dans cette étude initiale, notre but n'était pas de synthétiser un son le plus ressemblant possible au /ba/ de référence. Mais des stimuli intermédiaires d'un point de vue perceptif entre le vrai phonème et l'analogue utilisé dans l'étude précédente, tout en conservant la même méthode de synthèse.

Nous avons donc réfléchi à la meilleure manière de calculer les amplitudes et les fréquences à appliquer aux sinusoides. Nous avons testé plusieurs fréquences de coupure, et nous avons évalué l'intérêt de déphaser ces sinusoides les unes aux autres vis-à-vis de la ressemblance des analogues. En revanche, contrairement à la méthode des sine-wave speech (qui reproduit les formants par des sinusoides de fréquences variables en conséquence au cours du temps [Remez 1981]), dans notre cas, les sinusoides étaient constantes sur tout le signal, seul l'enveloppe pouvait éventuellement procurer une quelconque dynamique à l'analogue.

Une interface graphique sous Matlab nous a permis de diffuser les 43 analogues synthétisés dans un ordre aléatoire, et de les soumettre à *notation subjective* sur un jury d'étude de 15 personnes normo-entendantes, entre 1 (ressemble très peu à un vrai /ba/) et 7 (ressemble très exactement au /ba/). Enfin, à chaque analogue correspond une note moyenne, et nous avons retenus les 5 analogues jalonnant cette échelle. Ceci nous a donné les **5 stimuli** suivants, représentés temporellement Fig.3.15, et spectralement (spectre issu du spectrogramme large fenêtre centré sur consonne, transition formantique et voyelle ou leurs équivalents chez les analogues) Fig. 3.16 :

- stimulus 1 *très mauvais* : (note moyenne :  $\frac{1.5}{7}$ )  
 enveloppe filtrée à : 100 Hz  
 position des formants : 200, 800, 1200 , 3200, 4400 Hz  
 amplitude des sinusoides : 1, 0.8, 0.6, 0.46, 0.4  
 phase = 0

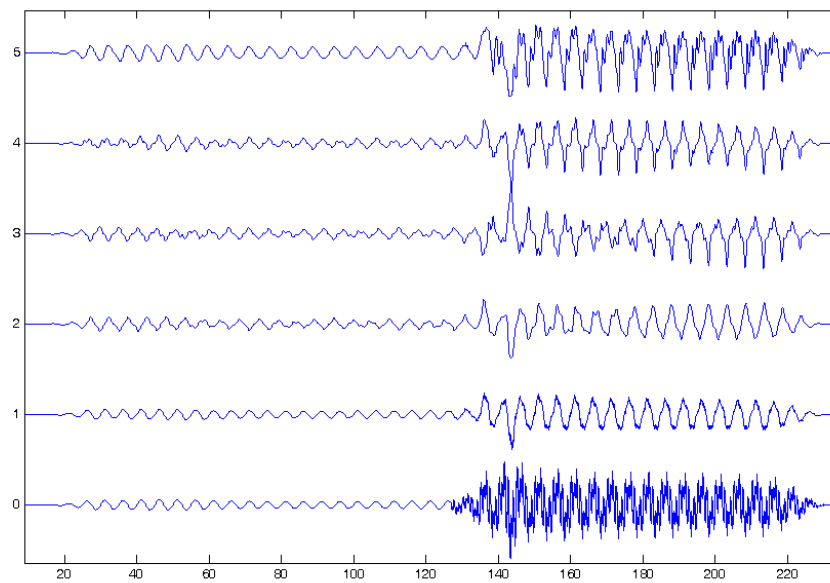


FIGURE 3.15 – **Stimuli retenus parmi la batterie d’analogues.** Ces analogues ont été respectivement perçus du plus mauvais (en bas) au meilleur ressemblant au /ba/ (en haut, en fait le /ba/ de référence, celui utilisé dans l’étude 3.2, Fig. 3.4). Les abscisses sont en millisecondes, les ordonnées en unités arbitraires

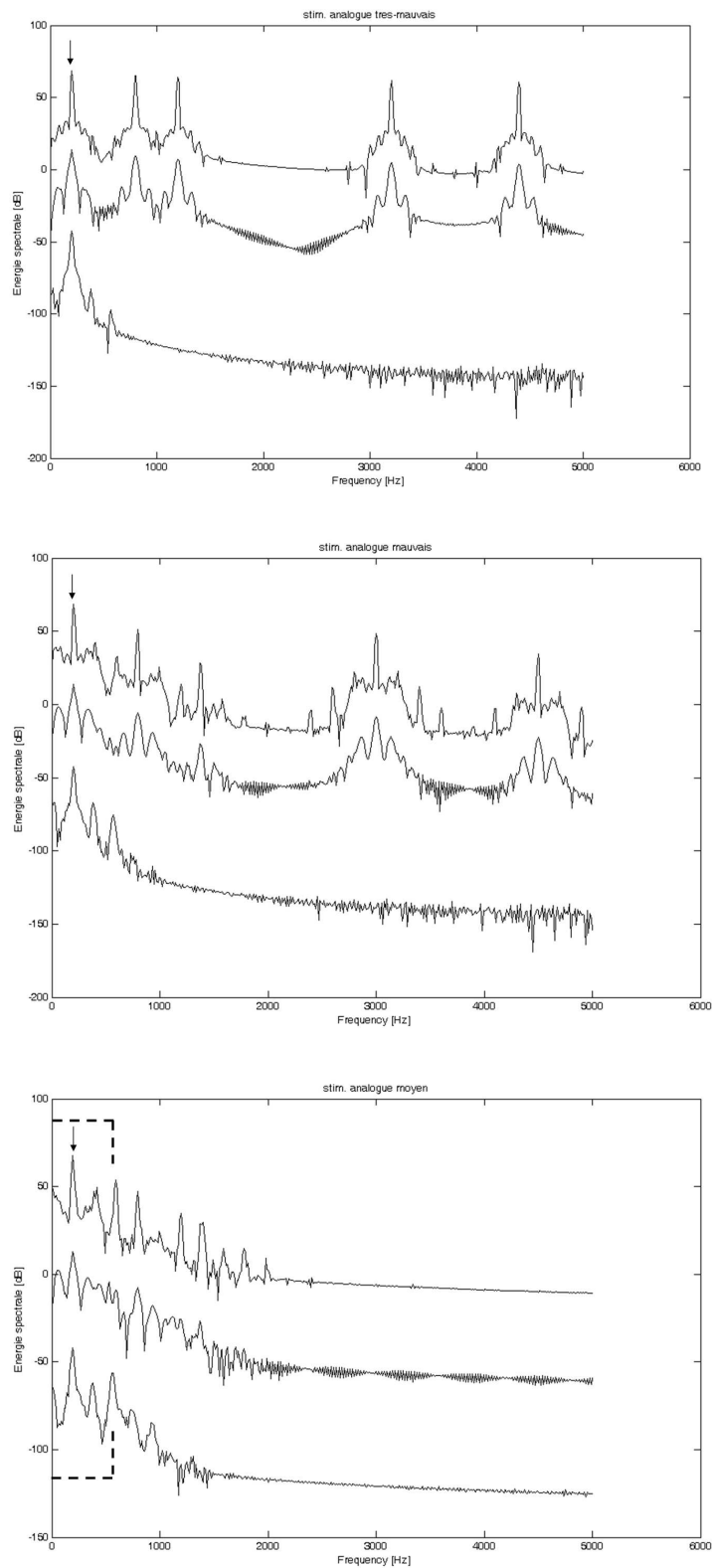


FIGURE 3.16 – Coupe spectrale (verticale) des spectrogrammes (non-représentés) des stimuli Fig.3.15. Sur chaque figure, le profil spectral centré sur les portions *consonne*, *transition formantique* et *voyelle steady-state*, comme dans l'étude 3.2.

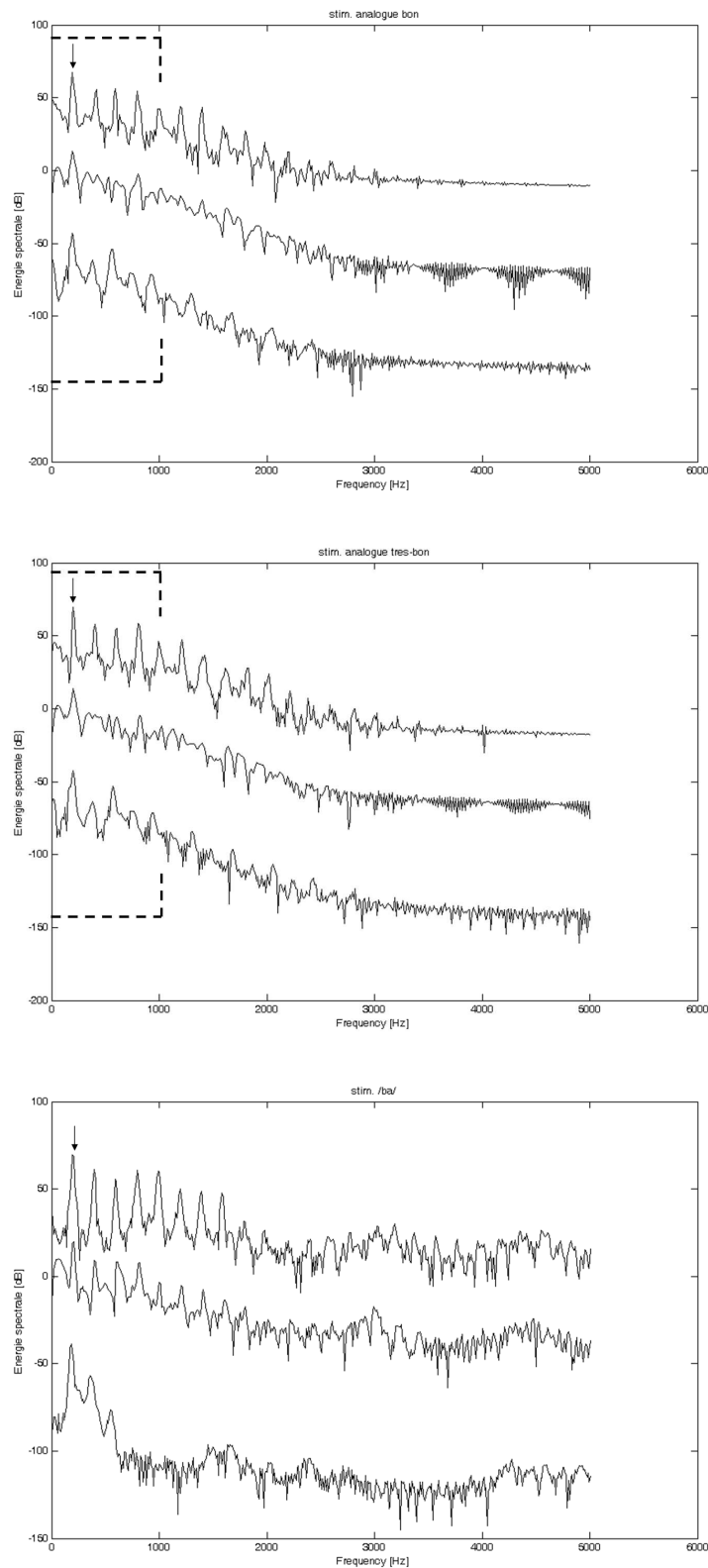


FIGURE 3.17 – Coupe spectrale (verticale) des spectrogrammes (non-représentés) des stimuli Fig.3.15. idem Fig.3.16, avec les 3 meilleurs stimuli (bon - très bon et le /ba/ de référence).

- stimulus 2 *mauvais* (note moyenne :  $\frac{2.6}{7}$ ),  
enveloppe filtrée à : 200 Hz  
position des formants : 200, 796 ; 1378, 3000, 4500, 5254 Hz  
amplitude des sinusoïdes : 0.13, 0.01, 0.1, 2.10-2, 1.5.10-2  
phase : 0
- stimulus 3 *moyen* (note moyenne :  $\frac{3.6}{7}$ ),  
enveloppe filtrée à : 500 Hz  
position des formants : 200, 796, 1378, 3000, 4500, 5254 Hz  
amplitude des sinusoïdes : 0.14, 0.01, 2.8.10-5, 2e-5, 1.5e-5  
phase : 0
- stimulus 4 *moyen-bon* (note moyenne :  $\frac{3.9}{7}$ ),  
enveloppe filtrée à : 1000 Hz  
position des formants : 200, 796, 1378, 3000, 4500, 5254 Hz  
amplitude des sinusoïdes : 0.14, 0.01, 2.9e-5, 2e-5, 1.5e-5  
phase : 0.47 1.027 0.86 -1.277 -1.16 -1.17
- stimulus 5 *très bon* (note moyenne :  $\frac{4.4}{7}$ ),  
enveloppe filtrée à : 1000 Hz  
position des formants : 200, 818, 1378, 3000, 4500, 5254 Hz  
amplitude des sinusoïdes : 0.14, 0.01, 2.9e-5, 2e-5, 1.5e-5  
phase : 0 0 0 -274.14 390.60 -90
- stimulus 6 *vrai /ba/* (note de référence :  $\frac{7}{7}$ ).  
Signal /ba/ d'origine.

Ces stimuli furent joués en binaural à 45 dB-SL, via les inserts-earphones blindés.

**Acquisition :** L'enregistrement des Speech ABR a été fait exactement dans les mêmes conditions que celles décrites dans le chap.3.2.

### 3.2.2.2 Résultats :

Le premier résultat de cette étude est que nous avons confirmé la bien meilleure synchronisation de la FFR pour le vrai /ba/ comparé à tous les 5 autres analogues, ce qui correspond bien aux observations de l'étude 3.2. Mais ici, il fut intéressant de constater que dans certaines conditions, un analogue peut engendrer une FFR bien synchronisée. Plus précisément, nous avons montré une tendance globale



(significative, tableau 3.23) à ce que la qualité de la synchronisation soit reliée à la qualité de la perception de ces analogues.

En effet, l'observation temporelle des FFR Fig.3.18, tout comme celle spectrotemporelle Fig. 3.19, nous suggère que la qualité de synchronisation de la FFR est graduelle (meilleure pour les analogues ressemblant le mieux au /ba/ de départ, celui-ci ayant la FFR la plus ample). Nous avons utilisé l'énergie située dans la bande de fréquence correspondant à  $F_0$  pour quantifier cette synchronisation et la comparer (tableau 3.23) à la répartition fréquentielle acoustique des stimuli, Fig.3.22. Notons dès à présent que le /ba/ de référence avait une énergie supérieure à tous les analogues dans cette bande de fréquence (ce qui représente un biais pouvant favoriser la formation de la FFR, même si il semblerait plutôt que celle-ci soit sensible aux hautes-fréquences [Dau 2003]). La synchronisation graduelle de la FFR s'est retrouvée dans les quantités d'énergie significativement plus grandes pour le /ba/ que pour n'importe quel analogue, et du meilleur analogue par rapport à la plupart des autres analogues.

Pour relier la qualité de la perception aux caractéristiques acoustiques des analogues, il semblerait que la fréquence de coupure retenue pour le calcul de l'enveloppe ait un poids conséquent sur le rendu de l'analogue : plus cette fréquence est élevée, meilleure est la perception de l'analogue. Dans le cas des meilleurs analogues (avec une fréquence de coupure à 1kHz), les premier formant (750 Hz) est totalement représenté, et le second formant (1300 Hz, juste après la fréquence de coupure) n'est que peu atténué. Or, la resituation des 2 premiers formants est habituellement suffisante pour une bonne intelligibilité. Le poids relatif des sinusoïdes est aussi importante : lorsque les sinusoïdes ont à peu près le même poids (cas de l'analogue très-mauvais), la perception est la plus mauvaise.

D'autre part, il ressort, comme dans la totalité de nos études, que la synchronisation de la FFR, si elle existe, a lieu sur le fondamental (sauf pour l'analogue très-mauvais, voir ci-dessous), et que le Speech ABR est passe-bas en dessous de 1000 Hz. Il est intéressant de constater que cette synchronisation est conditionnée par la structure fine (comme [Greenberg 1987] et [Dau 2003], et comme dans notre précédente étude 3.2), sans quoi tous ces analogues auraient donné la même FFR. Mais cette étude semble nous indiquer notamment que la synchronisation de la FFR est renforcée lorsque des informations acoustiques spectrales sont pourvues entre 500 et 1000 Hz, au vu de la meilleure synchronisation des analogues ayant une enveloppe à 1000 Hz, par ailleurs assez proche (même si significativement inférieures) de celle observée pour le vrai /ba/.

On remarque en effet que les analogues les mieux perçus sont ceux obtenus avec une modulation par une enveloppe filtrée à une fréquence de 1kHz. Du point de vue physiologique, il est donc vraisemblable que la dynamique spectrale conservée dans ces meilleurs analogues (et supprimée dans le cas des analogues obtenus avec une modulation par une enveloppe filtrée à une fréquence plus basse) favorise la

formation de la FFR, comme suggéré par [Krishnan 2008], ou [Xu 2006], même si cette dynamique dont ils parlent concernait  $F_0$ , et même si [Dau 2003] attribue plus d'importance aux bandes critiques au delà de 4-5 kHz, voir chap.???. 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 (où le pitch du missing fundamental est le plus prononcé), et il ne serait sans doute pas incohérent de penser que cette gamme de fréquence a de ce fait une importance capitale dans la formation des objets auditifs et en particulier la perception de la parole. Il serait donc probablement très intéressant de corréler la FFR (en amplitude, pitch-tracking ou pitch-strength) au pitch perçu en missing fundamental, en particulier chez des sujets ne la percevant pas (malentendants, amusiques,...).

Enfin, anecdotiquement, le cas obtenu avec l'analogue très-mauvais (aucune information d'enveloppe et un poids spectral très fort sur les formants) est frappant, car la synchronisation a lieu non pas sur le fondamental mais sur le premier harmonique à 400 Hz (voir flèche sur la Fig.3.20, en haut à gauche). Or, il s'avère que c'est le seul des 5 stimuli à avoir une structure pseudo-harmonique; et il est composé des fréquences suivantes (avec chacune une amplitude comparable) : 200 - 800 - 1200 - 3200 - 4400 Hz Fig.3.16, c'est-à-dire les harmoniques suivant de 400 Hz :  $1/2$  - 2 - 3 - 8 - 11. Il s'avère que cette anecdote est lourde de sens, car dans ce cas, la FFR n'a pas été synchronisée sur le fondamental (200 Hz, comme l'indique le spectre 3.16), mais que la FFR a été semble-t-il le plus influencée par la structure harmonique dont la périodicité de 400 Hz a été extraite quelque part à travers les harmoniques 2-3-8 et 11 (par des phénomènes de battements dans les bandes critiques?). Cette donnée isolée ne nous permet pas d'aller plus loin que ces quelques conjectures. Néanmoins, il serait très intéressant de procéder à une étude paramétrique faisant varier systématiquement le nombre d'harmoniques pour évaluer leur impact dans la formation de la FFR.

Soulignons, pour conclure, que cette pré-étude exploratoire avait principalement pour but de satisfaire notre curiosité vis-à-vis du rapport entre les caractéristiques acoustiques de ces analogues, la façon dont ils sont perçus, et la qualité de la FFR qui en découle. Rien, toutefois ne peut nous permettre d'affirmer que c'est la qualité de perception qui a influencé la qualité de la FFR (même si c'est ce qui est observé ici). Il se pourrait en effet que les aspects physiologiques (FFR) et psychoacoustiques (qualité de perception de la parole) soient influencés par les mêmes aspects acoustiques, mais de façon totalement indépendantes et décorrelées (ou peut-être pas?).

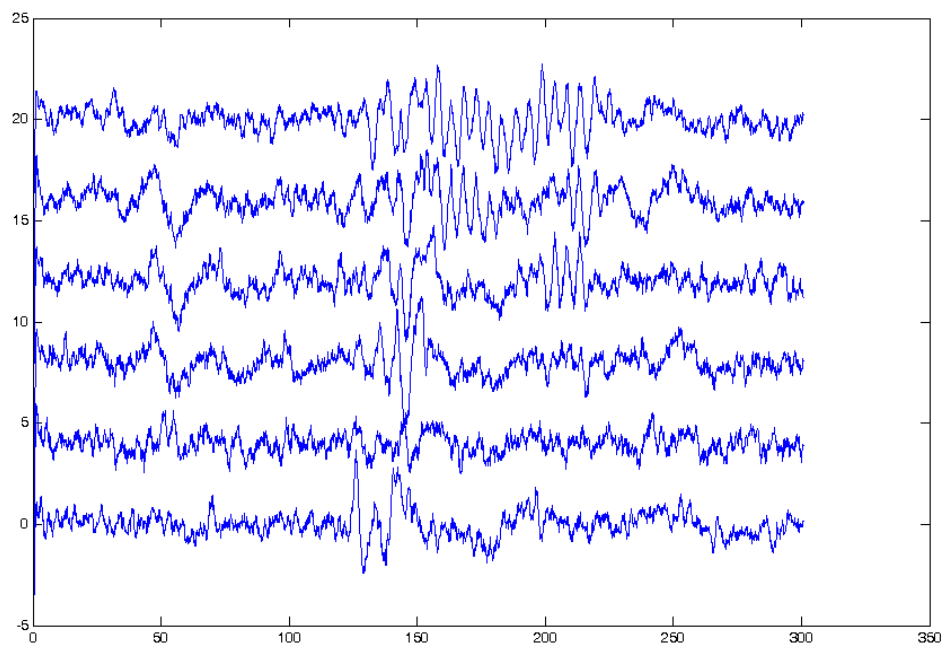


FIGURE 3.18 – Grande moyenne ( $n=10$ ) des FFR obtenues pour les stimuli **Fig.3.15**. Les FFR obtenues varient, et on remarque la tendance à graduellement obtenir une meilleure FFR pour les analogues les mieux notés

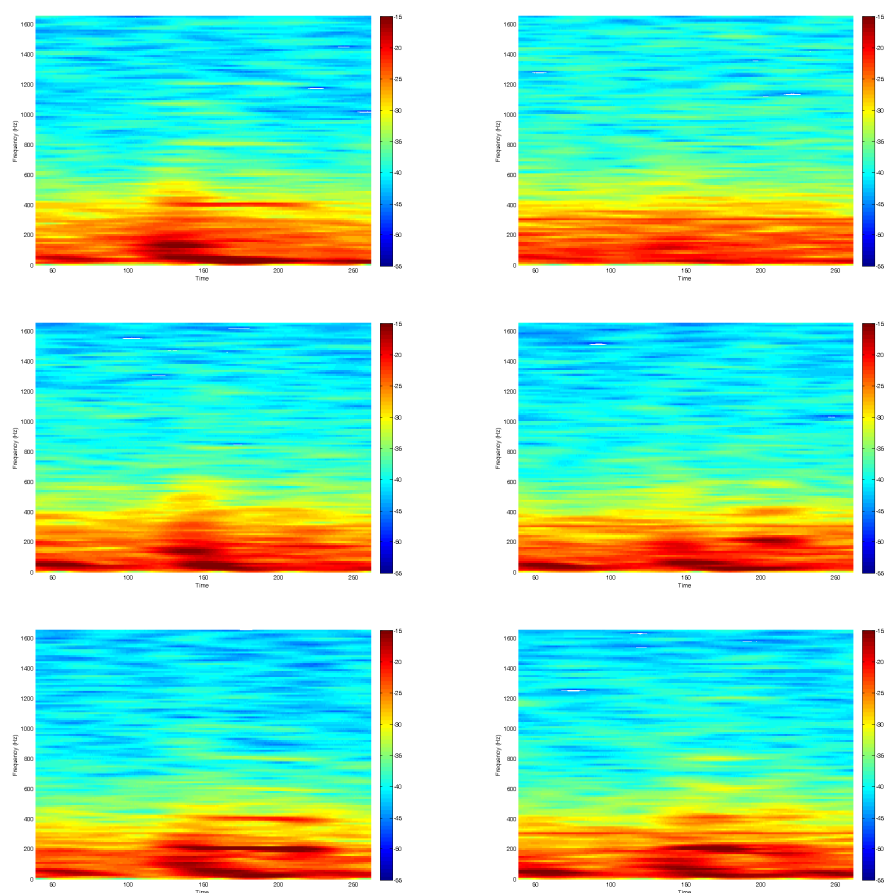


FIGURE 3.19 – Grande moyenne (n=10) des spectrogrammes des FFR précédentes Fig.3.19. La synchronisation graduelle est là encore visible.

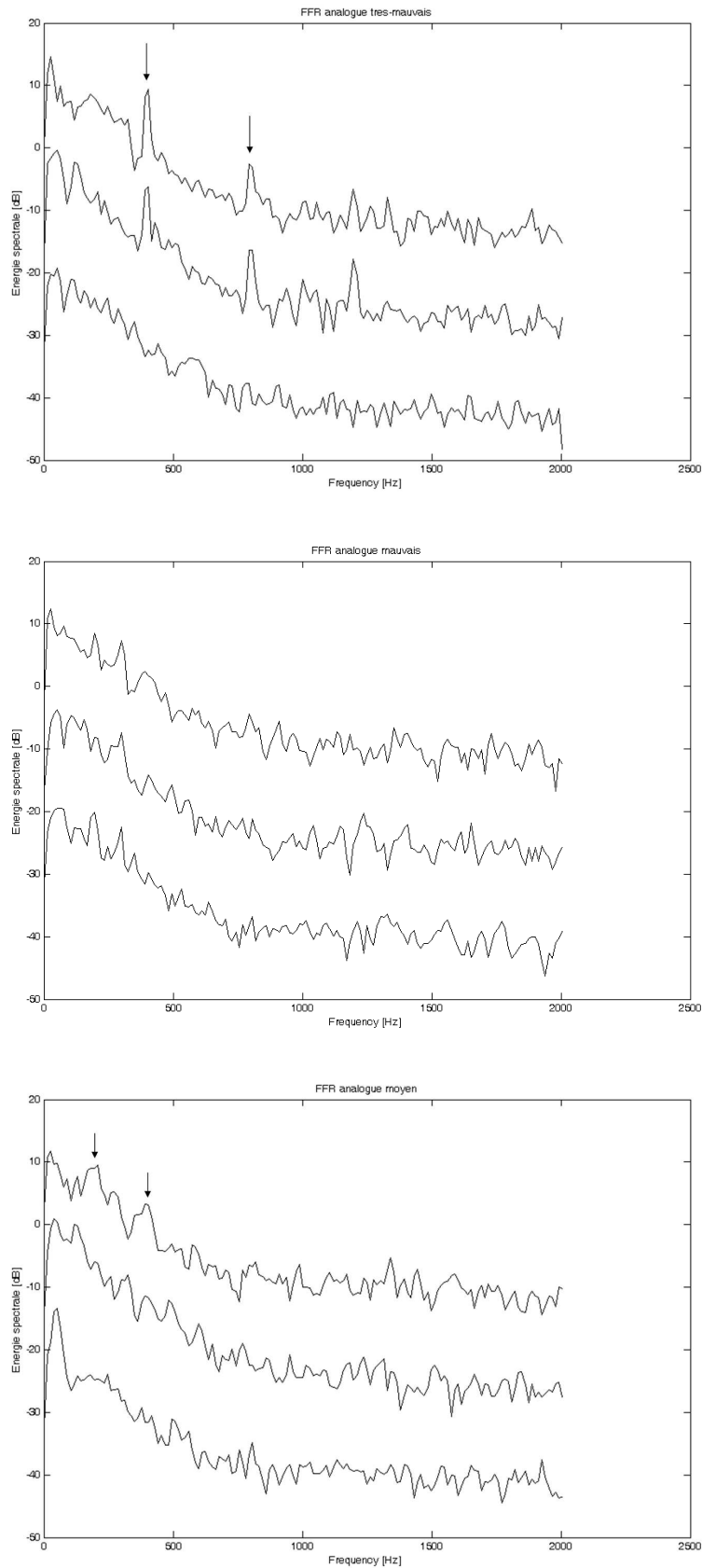


FIGURE 3.20 – Coupes spectrales (verticales) des spectrogrammes des FFR de la Fig.3.19. Notons la surprenante synchronisation sur 400 Hz de l’analogique jugé très-mauvais (en haut, voir texte).

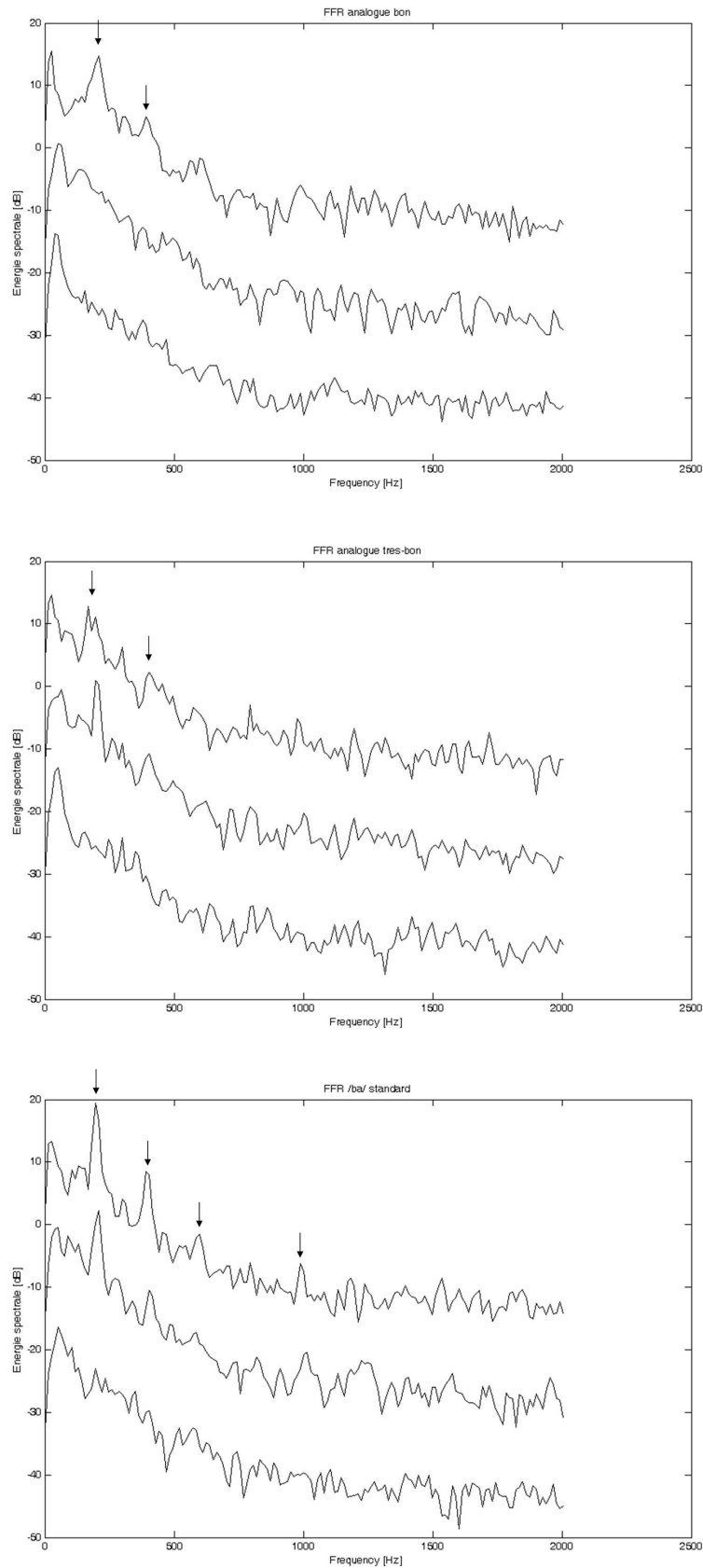


FIGURE 3.21 – Coupes spectrales (verticales) des spectrogrammes des FFR de la Fig.3.19. Les 2 meilleurs analogues et le /ba/ de référence.



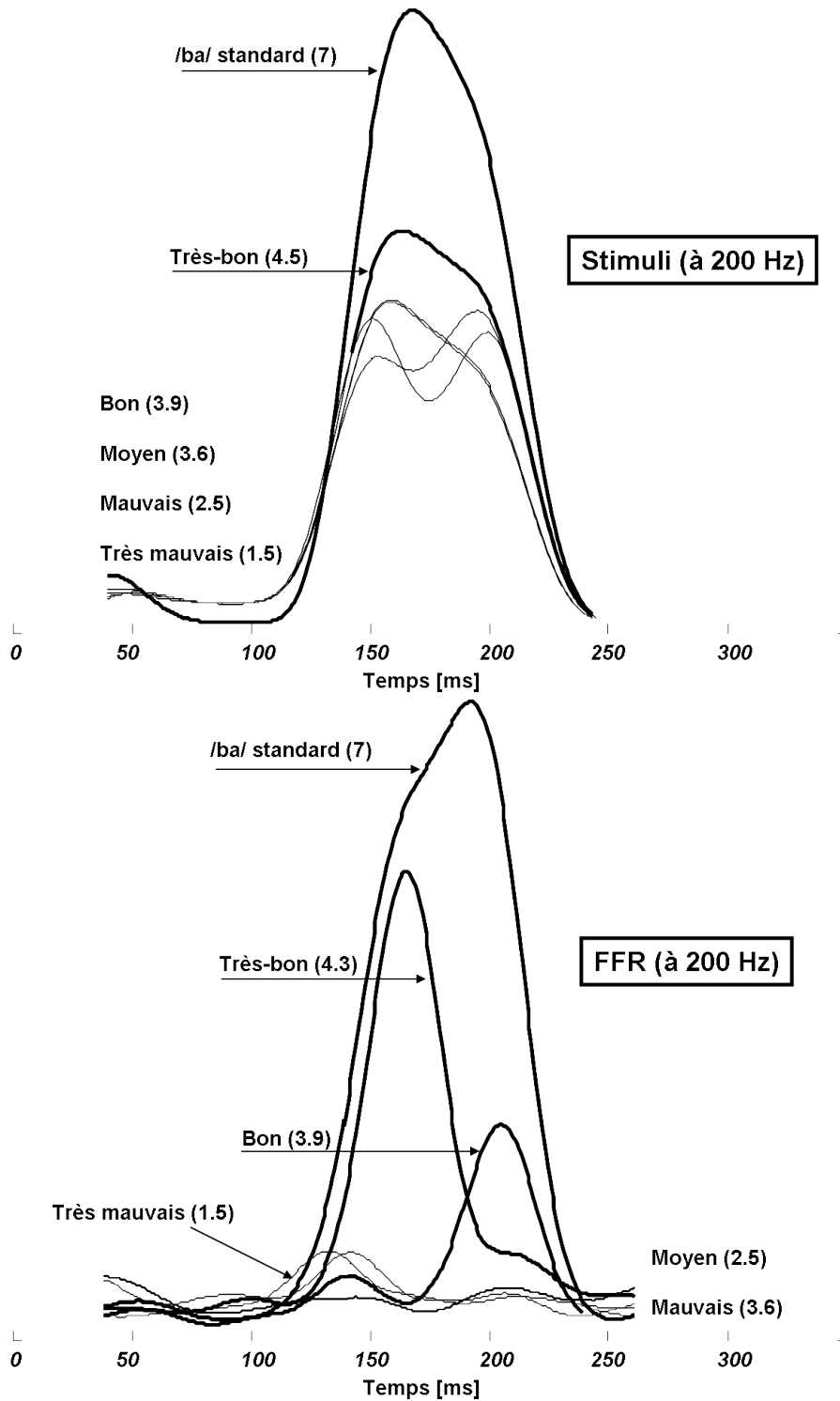


FIGURE 3.22 – Coupe horizontales des spectrogrammes des FFR précédentes Fig.3.19 dans la bande de fréquence correspondant à  $F_0 = 200$  Hz. La synchronisation graduelle entre les analogues de mieux en mieux perçus (les notes moyennes sont entre parenthèses) est confirmée ici, on peut se servir des aires sous ces énergies dans la bande de fréquence du spectrogramme centré sur  $F_0$ , pour effectivement quantifier la synchronisation de la FFR sur  $F_0$ .

	<i>Très mauvais</i> <b>p = 0.086</b>	<i>Mauvais</i>	<i>Moyen</i>	<i>Bon</i>	<i>Très bon</i>	<i>/ba/</i>
<i>Bon (/7)</i>	<b>F(1,16) = 3.4</b> <b>p=0.016</b>	NS	NS	x		
<i>Très bon (/7)</i>	<b>F(1,16) = 7.3</b> <b>p&lt;0.001</b>	<b>F(1,16) = 6.4</b> <b>p = 0.02</b>	<b>F(1,16) = 6.2</b> <b>p = 0.02</b>	F(1,16) = 2.9 <b>p = 0.1</b>	x	
<i>/ba/ (7 / 7)</i>	<b>F(1,16) = 19.4</b>	<b>F(1,16) = 18.4</b>	<b>F(1,16) = 18.2</b>	<b>F(1,16) = 13.5</b>	<b>F(1,16) = 4.4</b>	x

FIGURE 3.23 – Comparaison des synchronisations des FFR d’après le critère des énergies présentes dans la bande de fréquence des spectrogrammes Fig.3.19 centrées sur 200 Hz Fig.3.22 dans la bande de fréquence correspondant à  $F_0 = 200$  Hz. Ces valeurs ont été obtenues par ANOVA.

### 3.2.3 Parole vs Analogues-en-structure-harmonique, et spectrale-ment riches.

Dans les deux études précédentes, on constate que les analogues sont moins bien encodés dans la FFR que les sons de parole originaux. Serait-ce dû à leur structure fine très appauvrie? Serait-ce une conséquence de l’inharmonicité des composantes spectrales? Pour apporter un peu de lumière sur cette question, nous avons utilisé des phonèmes /ka/, synthétisés par Klatt. Dans ce cas, les analogues furent obtenus en n’appliquant pas la partie amplification / atténuation des harmoniques qui confèrent la structure formantique dans le synthétiseur de Klatt (Fig.3.1 fenêtre du bas).

Par conséquent, /ka/ et son analogue ont une structure harmonique très proche, mais la seule différence est que /ka/ possède des formants (maxima locaux dans le spectre), qui en font un signal de parole, tandis que l’analogue n’est pas un signal de parole en étant dépourvu Fig.3.25 (Fig.3.24).

Dans ce cas, une très bonne synchronisation de la FFR sur  $F_0$  fut observée, aussi bien pour l’analogue que pour /ka/. Ce qui confirme que la différence de synchronisation observée précédemment est plus à relier à la structure harmonique (ou à la richesse de la structure fine, ce que cette étude ne permet pas de cibler), qu’à la nature du stimulus. Nous allons voir ce point par la suite.

Ces résultats sont par ailleurs présentés en détail au sein de l’étude Chap.3.4, où est investigué l’effet de la nature du stimulus, et celui de la hauteur tonale.

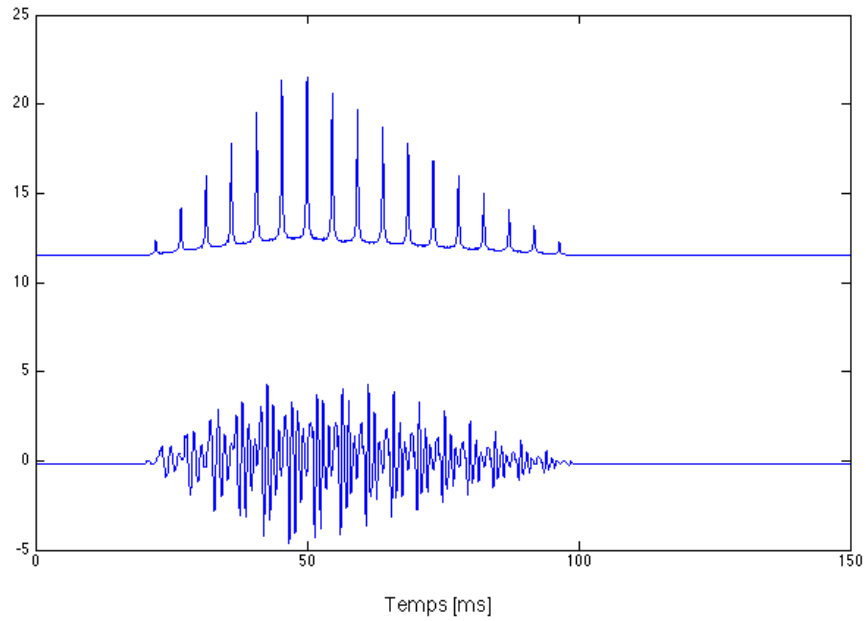


FIGURE 3.24 – /Ka/ et son analogue. En temporel.

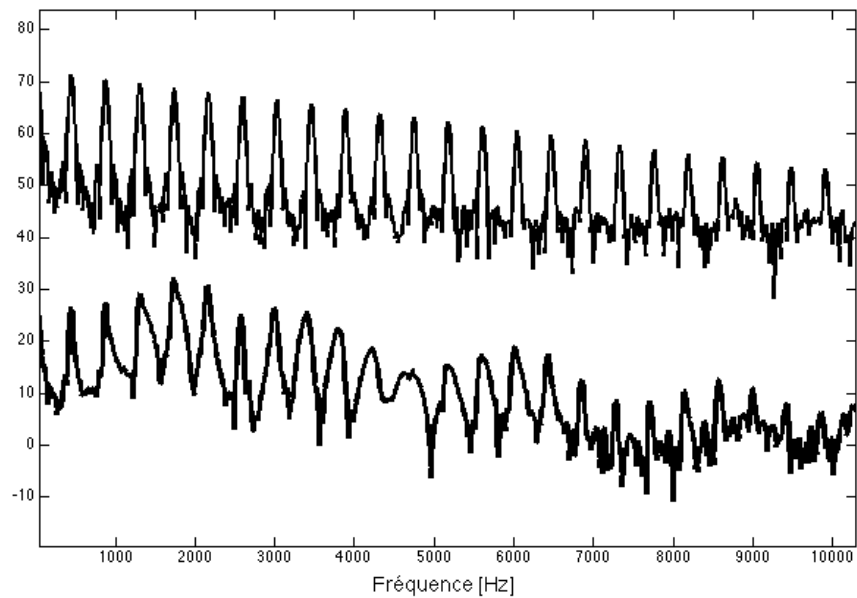


FIGURE 3.25 – /Ka/ et son analogue. En fréquence.

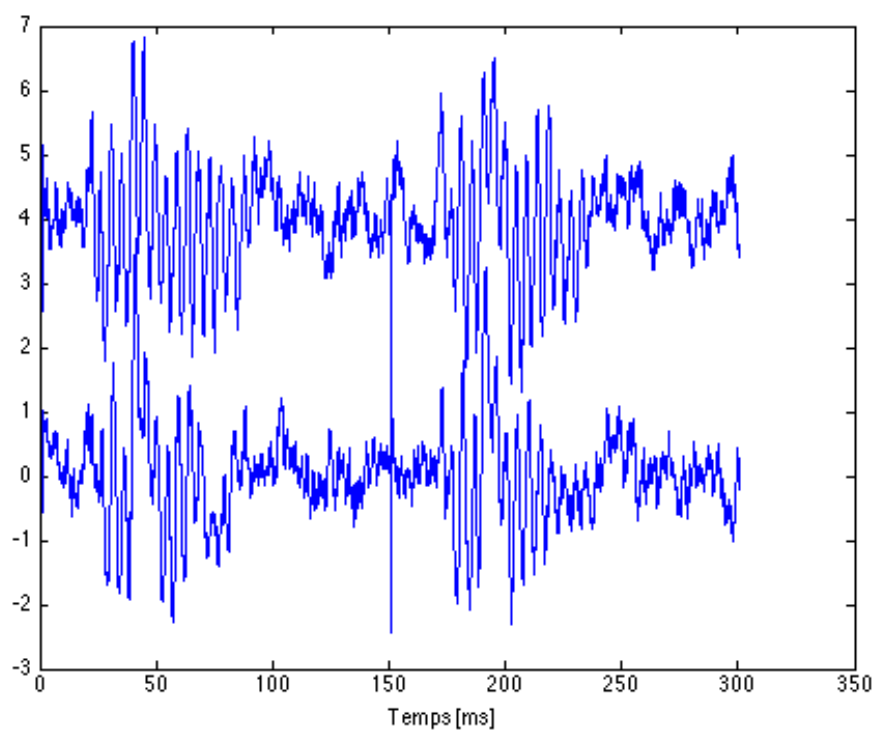


FIGURE 3.26 – FFR du /Ka/ et de son analogue. En temporel.

### 3.3 Effet de la structure harmonique : sons purs vs sons complexes harmoniques.

Intrigués à la fois par la signification de la fréquence de coupure du Speech ABR (500 Hz dans l'étude pilote, jusqu'à 1200 Hz pour le /pa/), nous avons voulu savoir s'il existait une sorte de fonction de transfert de la FFR, et s'il était possible de l'estimer en réponse à des sons purs balayant la zone fréquentielle la plus importante pour la parole (100 - 4000 Hz). D'autre part, nous avons voulu évaluer la linéarité de la FFR, et savoir si la FFR en réponse à un son complexe harmonique (ici,  $F_0 = 130$  et 220 Hz) était la somme des FFR obtenues pour chacune de ses composantes. Nous avons obtenu un résultat très tranché, avec aucune réponse quel que soit le son pur présenté, tandis que pour les deux sons complexes harmoniques, une très nette FFR fut obtenue, synchronisée sur  $F_0$ .

#### 3.3.1 Matériel et Méthodes

- *Sujets :*

Les ABR ont été enregistrés sur 10 sujets normo-entendants.

- *Stimuli acoustiques :*

Les stimuli utilisés furent (i) 6 **sons purs** sinusoïdaux et (ii) deux **sons complexes** de fréquence fondamentale 110 et 220 Hz. Chaque son durait 100 ms et était fondu en ouverture et en fermeture sur 10 ms.

La fréquence des sons purs (choisies pour balayer la gamme de fréquence la plus importante pour la perception de la parole) était : 130, 220, 440, 880, 1760 et 3520 Hz. Les sons complexes étaient composés de 10 harmoniques sommés sans déphasage (sinus).

#### 3.3.2 Résultats

Quels que soient les sons purs utilisés, aucune réponse caractéristique du Speech ABR ne fut obtenue, tandis que les deux sons complexes harmoniques ont donné lieu à une FFR très nette. Il est frappant qu'absolument aucune FFR ne fut enregistrée pour les sons purs, comme l'atteste la platitude de la courbe Fig.3.27, et ceci, quelle que soit la fréquence du stimulus. Ceci est contraire à la plupart des données de la littérature [Greenberg 1987] et (Galbraith2000), mais est-ce que toutes les conditions pour se prémunir de l'artéfact avaient été bien prises? Toutefois signalons que les latences trouvées (10 s pour [Greenberg 1987]) plaident en la faveur d'enregistrements sans artefact (le recul de leur équipe de même force le respect). En revanche, quelques pics suggèrent la présence d'une Onset Response, entre 12 ms (pour 110 Hz) et 22 ms (à 3520 Hz). Sachant que le stimulus démarrait après 10 ms de pré-moyennage, les Onset Response (attendues une dizaine de millisecondes après la transition acoustique du début du stimulus) des conditions 110 et 220 Hz

sont peu vraisemblables. Hormis ces Onset Response, l'onde la plus remarquable dure une dizaine de ms et se situe entre 30 ms (à 110 Hz) et 45 ms (à 3520 Hz). Elle correspond à une réponse d'origine thalamo-corticale dite de latence moyenne (middle-latency response) [Xu 1997]. Il est intéressant de noter que la latence de cette middle-latency response semble augmenter graduellement avec la fréquence des stimuli, ce qui rejoint l'hypothèse de [Johnson 2008].

Dans le cas des sons complexes harmoniques, l'ABR présenta une forme classique, avec une Onset Response suivie d'une FFR. La FFR fut périodique, bien synchronisée sur F0, comme c'est très bien visible Fig.3.28. L'analyse spectrotemporelle (non-reportée ici), montra que les harmoniques de F0 furent présents jusqu'à une fréquence de coupure entre 400 et 600 Hz.

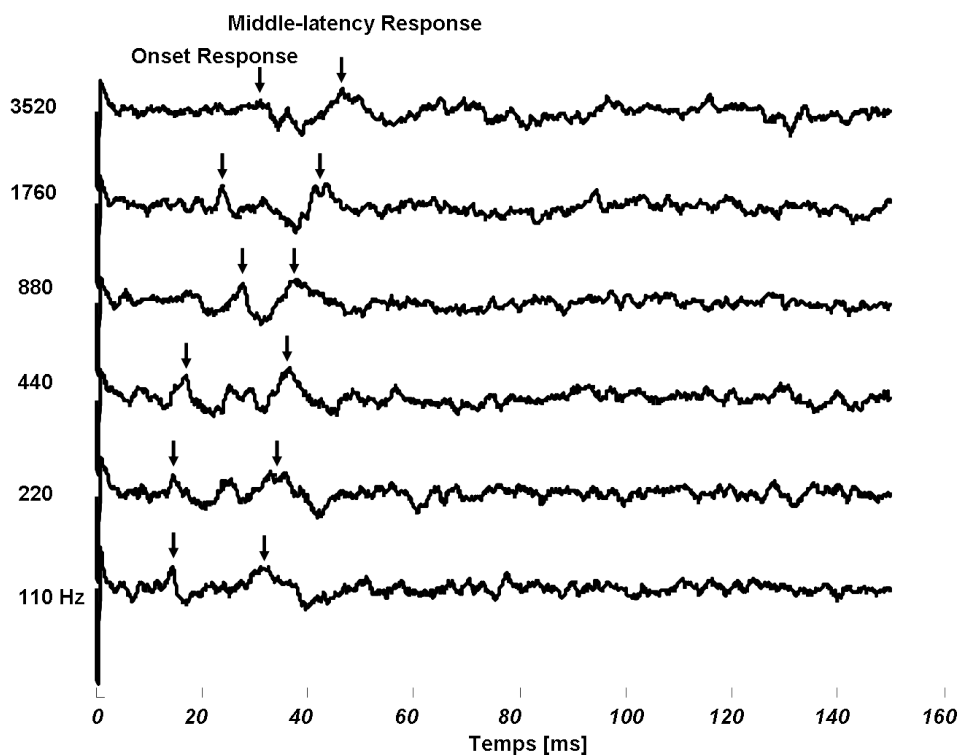


FIGURE 3.27 – Grande moyenne temporelle (n=10) des ABR en réponse à des sons purs. La fréquence des sons purs ayant engendré ces ABR est indiquée à gauche. La déflexion au début est due à un artefact du système d'acquisition. La seule réponse identifiable est l'onde positive de latence moyenne (dite Middle-latency response) située entre 35 (à 110 Hz) et 55 ms (à 3520 Hz). Elle ne correspond ni à une Onset Response, ni à une FFR. Cette dernière est clairement absente. En revanche, les pics situés entre 15 (à 110 Hz) et 25 ms (à 3520 Hz) s'apparentent à une Onset Response.

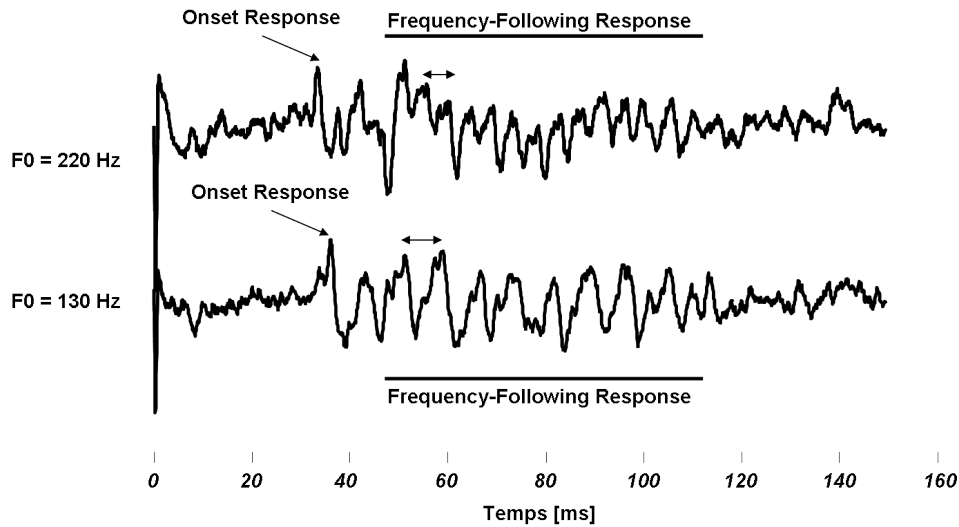


FIGURE 3.28 – Grande moyenne temporelle ( $n=10$ ) des ABR en réponse à des sons complexes harmoniques. A l'inverse de la figure précédente, avec les sons purs, la FFR est très clairement présente ici. Elle est synchronisée sur la fréquence fondamentale des sons complexes harmoniques (précisée par les double-flèches). L'Onset Response est présente au début du stimulus. On ne remarque pas de Middle-latency response ici.

### 3.3.3 Discussion

Le fait que nous n'ayons trouvé aucune réponse quel que soit le son pur présenté, tandis que pour les deux sons complexes harmoniques, une très nette FFR fut obtenue, synchronisée sur  $F_0$  appuie l'idée que la FFR est issue d'un phénomène d'interaction entre les harmoniques, qu'on présume dans les hautes-fréquences. Il s'agit d'un résultat qui diffère de ce qui est généralement présenté dans la littérature [Greenberg 1987]. On pourrait penser à des influences artéfactuelles, mais les latences obtenues dans ces études plaident en la faveur d'enregistrements authentiques...



## 3.4 Effet de la nature du stimulus et de la hauteur tonale sur le Speech-, Music-, Analogue- ABR

### 3.4.1 Objectifs :

Pour faire écho au résultat chap. 3.2, qui montrait un encodage différent du vrai son de parole par rapport à l'analogue, la question se posait de savoir si le Speech ABR était sensible à la nature du stimulus (parole, musique, analogue sans-signification); d'autant plus qu'une large partie de la littérature issue de la neuroimagerie situe la spécialisation du traitement de la parole dans les aires corticales plus intégrées [Zatorre 2008], sans décrire la contribution sous-corticale, principalement faute de données à disposition.

D'autre part, nous avons cherché à évaluer l'effet de la hauteur tonale de sons de parole sur les Speech ABR, en prenant le même stimulus de référence et en décalant son fondamental. Cette démarche se distingue de ce qui existe dans la littérature, où ce sont principalement des sons purs et sons complexes harmoniques qui sont utilisés, ou encore des linguistic pitch dont la  $F_0$  varie au sein même du stimulus. En particulier, nous étions intéressé par l'encodage d'un son de parole lorsque le fondamental est hors de la gamme de  $F_0$  habituelle (80-300 Hz). Après les premières observations de la pré-étude ?? et [Akhoun 2008a], avec des /ba/, sur les effets du jitter pour les hautes fréquences (575 Hz), nous avons voulu reproduire l'étude de l'effet de la hauteur tonale sur le Speech ABR, à la fois avec des phonèmes et avec des notes de piano. D'autre part, nos résultats préliminaires [Akhoun 2008a] ont été recueillis en stimulant avec des inserts non-blindés, ce qui porte cette pré-étude à caution étant donné le risque d'artéfact que nous avons mis en évidence dans l'étude 2.2, et surtout le fait que l'artéfact soit plus facilement enregistré pour les plus hautes fréquences. Il était donc souhaitable de reprendre ce protocole dans la condition déclarée sans artéfact (inserts blindés, et reliés à la masse).

Pour répondre à ces questions, nous avons considéré 3 *phonèmes (/ka/)*, leurs 3 *analogues* synthétisés selon le modèle de Klatt avec la totalité des harmoniques présents mais modulation du spectre autour des formants ni transition formantique ([Bélanger 2007], [Tillmann 2008]), comme l'atteste la forme temporelle Fig.3.29 très différente des autres analogues utilisés dans l'étude 3.2 (série de pics espacés temporellement de  $1/f_0$ ), ainsi que 3 *notes de pianos*. Ces 3 catégories de stimuli furent pris à 3 *fréquences fondamentales* à  $F_0$  (**220 Hz**),  $2F_0$ ,  $4F_0$ . Le logiciel Straight (Kawahara2007) permet de procéder à ce genre de transformation.

### 3.4.2 Matériel et méthodes :

**Sujets :** Neuf sujets normo-entendants ont participé à cette expérimentation.

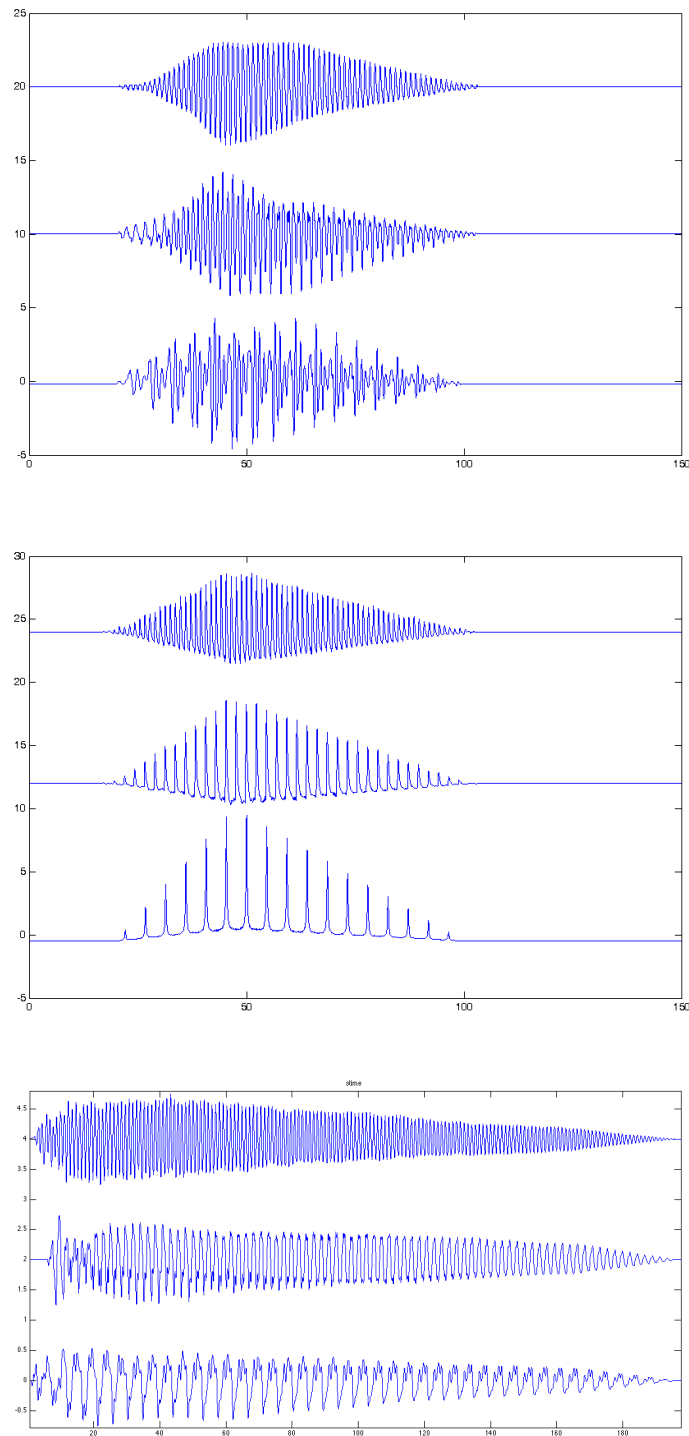


FIGURE 3.29 – /ka/, analogues /ka/ (Buzz) et notes de piano. Les /ka/ ont été synthétisés par Klatt, et les /ka/ à  $2F_0$  et  $4F_0$  ont été obtenus par transposition dans Straight. Les sons de piano ont été synthétisés en format MIDI. Pour ces 3 stimuli,  $F_0 = 215 - 220$  Hz. Les abscisses sont en millisecondes et les ordonnées en unités arbitraires.

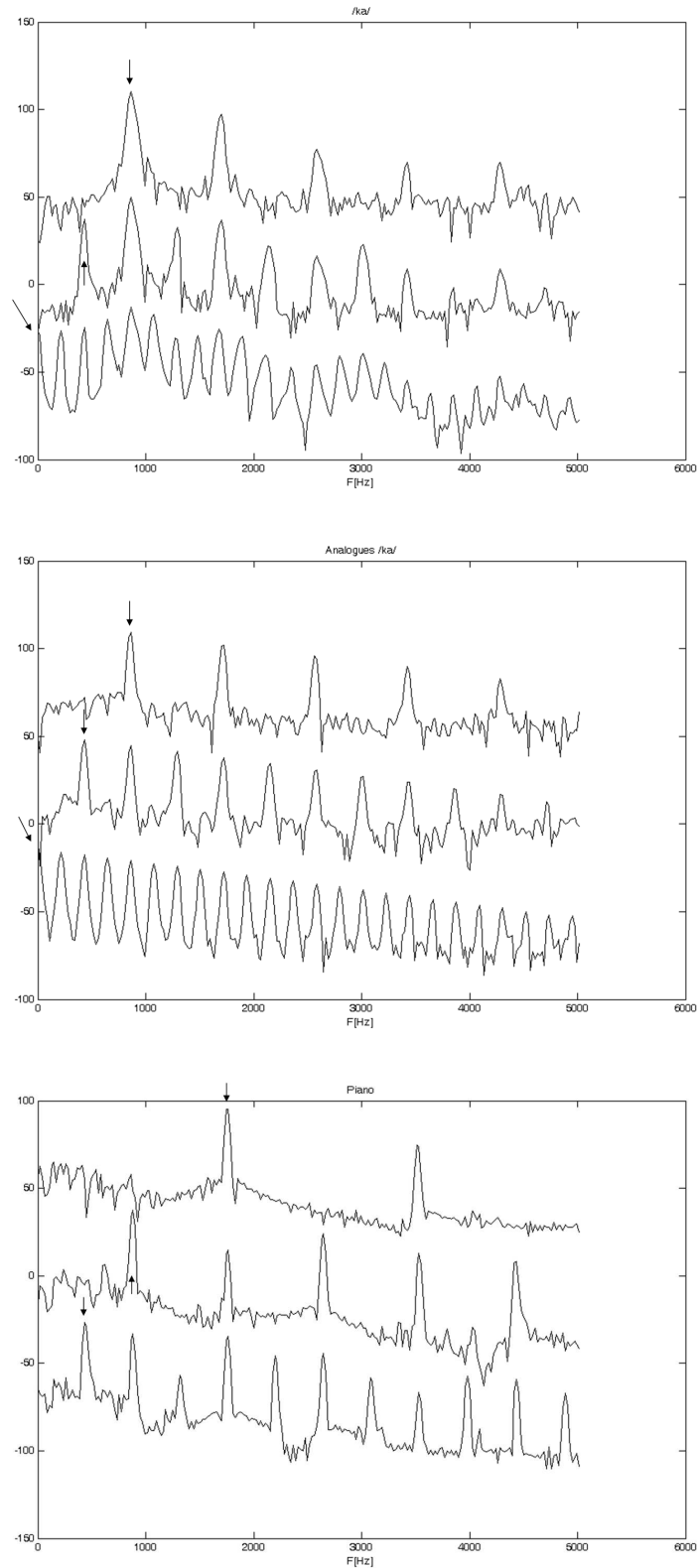


FIGURE 3.30 – Spectres large bande de la partie steady-state des /ka/, analogues /ka/ (Buzz) et notes de piano,  $F_0 = 220$  Hz (trace du bas), 440 (milieu) et 880 Hz (trace du haut). On remarque que les analogues utilisés dans cette étude sont différents de ceux des études 3.2 ou 3.2.2. NB : Dans le chap. 3.2, les 3 traces étaient des spectrogrammes du même stimulus centrés en 3 endroits différents du stimulus.

**Stimulation :** Nous avons considéré comme signal de référence pour la parole une syllabe synthétique /ka/, qui s'apparente davantage au /ba/ que nous avons utilisé dans les premières études (3.1 et 3.5) qu'au /ba/ écologique que nous avons considéré dans l'étude 3.2. Cette syllabe durait 75 ms, sa fréquence fondamentale était de **215 Hz**, avec les formants classiques du /a/.

Les syllabes /ka/, synthétisées selon Klatt Fig.3.1, sont composées de 20 ms de bouffées de bruit (noise-burst) suivies d'un train de pulses qui reproduisent le fondamental laryngé lorsque la syllabe /ka/ est produite. Ensuite, cette succession est passée à travers une série de 4 filtres passe-bandes reproduisant les formants (effets des résonances propres à la voyelle /a/ dans le conduit vocal larynx-lèvres) avant et après la modulation de l'*articulatory gesture* qui inclut les transitions formantiques. Ces formants étaient positionnés autour des fréquences suivantes : 950, 1570, 3150 et 4370 Hz. Ces fréquences furent fixes le long de la voyelle, mais la transition formantique entre la consonne et la voyelle fut respectée.

Les *Buzz - analogues des /ka/* sont déduits d'après le même mode de synthèse que les syllabes, sauf que le dernier module de filtrage du train de pulses reproduisant les formants n'est pas pris en compte. Par conséquent, les analogues et les syllabes /ka/ ont la même enveloppe, la même fréquence fondamentale, mais leur enveloppe spectrale ne présente pas de maxima locaux pour les formants Fig.3.30. C'est sur ce point qu'ils diffèrent des analogues utilisés dans la partie 3.2, qui ont la même enveloppe que les syllabes /ba/ et /pa/, qui ont de l'énergie présente aux formants, mais qui n'ont plus de structure harmonique.

Pour le son de musique, nous avons pris un son MIDI de piano (note La) ayant pour fréquence fondamentale **220 Hz**. Ce son était plus long et durait 200 ms. Les stimuli étaient délivrés à 45 dB-SL (ce qui correspond à un niveau confortable à assez fort selon les sujets) en binaural à un rythme de 2 par secondes pour le piano, et 7 par secondes pour le /ka/. La longueur du stimulus, qui conditionne la cadence de répétition, ainsi que la forme de l'enveloppe du stimulus musical comparé aux syllabes /ka/ et leurs analogues peuvent constituer un biais dans la génération de la FFR. Mais il fallait que le stimulus ne soit pas plus court, sans quoi il aurait été difficile de l'identifier comme un son musical, quant aux sons de parole (et par suite les analogues qui doivent reproduire leurs formes temporelles), nous avons vu qu'ils étaient très bien définis avec des stimuli très courts (les FFR de l'étude 3.1 sont mieux définies que celles de l'étude 3.2).

**Acquisition :** Les conditions d'enregistrement du Speech ABR furent identiques pour les /ka/ et analogue /ka/ par rapport au son de piano, avec une fenêtre d'acquisition de 300 ms afin d'observer 2 epochs en 1. La bande passante du filtre d'acquisition, la sensibilité restait identique à 75 microvolts, entre 80 et 3200 Hz,

et la fréquence d'échantillonnage était toujours de 13.3 kHz. Notons que dans cette condition, deux Speech ABR étaient enregistrés par « epochs » pour le /ka/, et pour leurs analogues. Comme on pouvait s'y attendre, les Speech ABR engendrés par les 2 stimuli étaient très semblables; et dans les résultats ci-dessous, seuls le premier des deux est représenté. La position des électrodes était identique à celle décrite ci-dessus ainsi que dans l'étude 3.2.

### 3.4.3 Résultats :

Ces résultats montrent que pour les 3 types de stimuli (parole, analogues ou musique), des Speech ABR très nets apparaissent, avec des FFR pas seulement synchronisées pour les stimuli basses-fréquences, mais aussi à  $F_0 = 440$  et à 880 Hz (Figs 3.31 à 3.37). Les enregistrements furent très semblables sur les deux voies d'acquisition (même si ce n'est pas représenté ici). En référence aux deux points techniques exposés dans (2.2 et ??), nous sommes assurés que les réponses ne sont pas dues à de l'artéfact, et le moyennage des spectrogrammes nous montre très clairement les bandes énergétiques caractérisant la FFR aux hautes fréquences, alors que la FFR est difficile à apercevoir sur les traces temporelles simples où les FFR individuelles ont probablement été artificiellement atténuées par effet jitter.

Les notes de piano ont engendré une réponse *onset* très nette Fig.3.31, et ont donné lieu à une synchronisation de la *FFR* très précise aussi bien à pour La 220 que La 440 Hz. En revanche, peu de FFR ressort du spectrogramme du La 880 Hz, ce qui correspond bien à ce qu'on attendait des caractéristiques passe-bas de la FFR, Figs. 3.32 et 3.36. Cette précision dans la synchronisation est particulièrement visible sur les spectrogrammes (de manière analogue par exemple à dans l'étude 3.2). De plus, nous n'avons pas trouvé de différences notables entre les FFR enregistrés d'un côté ou de l'autre (mastoïde gauche ou droite comme référence), Fig.3.31<sup>2</sup>.

Les /ka/, comme les analogues /ka/ ont engendré des FFR très nettes que ce soit pour 220, 440 ou 880 Hz, Fig. 3.35 et 3.37 ce qui tranche avec les Speech ABR obtenus pour les analogues dans l'étude (3.2), ce qui n'est pas très surprenant vu la différence de constitution acoustique entre ces deux stimuli. En revanche, très peu d'onset response fut recueilli dans ces deux cas, certainement vu la forme très régulière du fondu en ouverture et en fermeture de l'enveloppe de ces stimuli, qui est tout le contraire de l'impulsion d'un clic (supposé donner l'onset response la mieux définie).

Pour conclure vis-à-vis de cette étude, nous avons obtenu des Speech ABR bien définis dans tous les cas (moins pour le piano à haute fréquence). Le Speech ABR

---

2. La technique de recueil des FFR ne permet donc a priori pas d'observer de phénomène de latéralisation particulier. Toutefois, il s'agit vraisemblablement plus d'une limite technique que d'un phénomène physiologique absent.

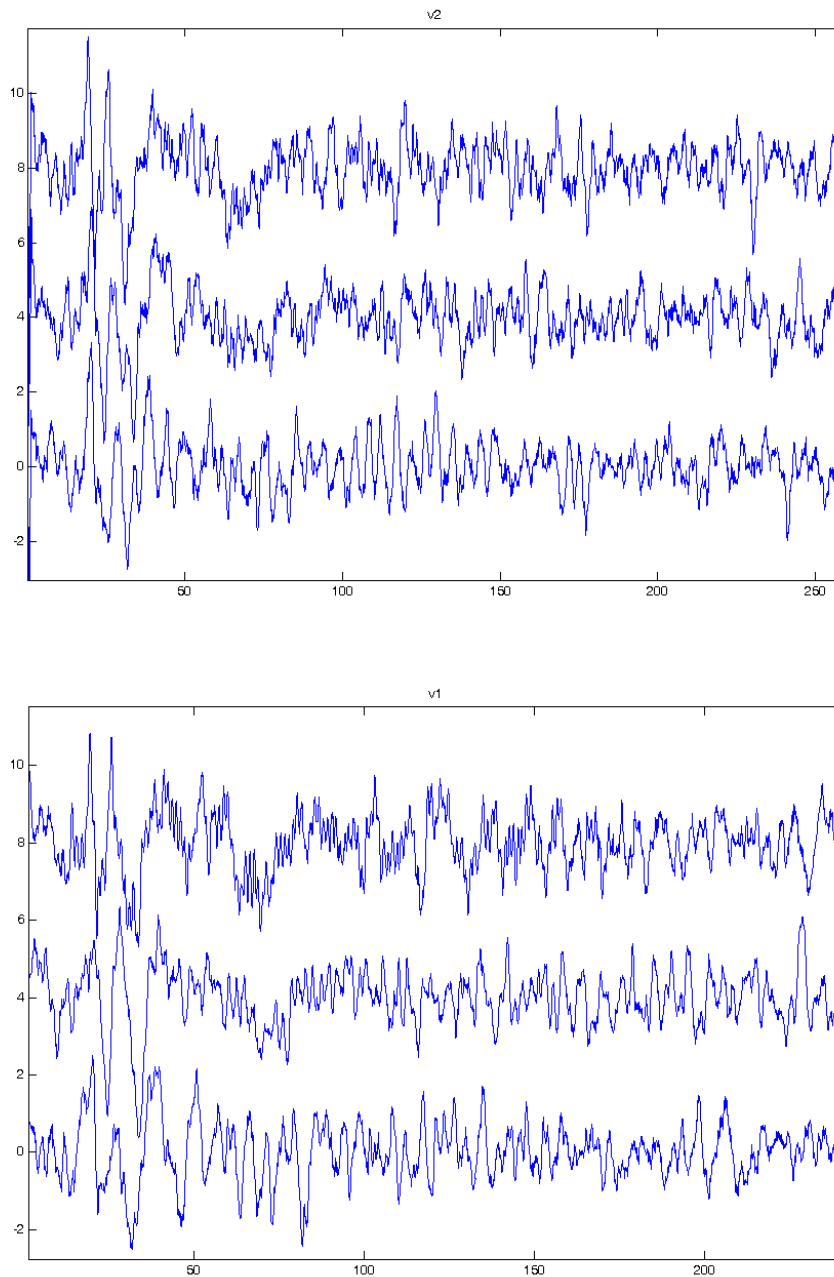


FIGURE 3.31 – **FFR en réponse aux notes de piano (enregistrement gauche et droite)**. L'onset response, large onde suivant le début du stimulus, est bien présente, et est consistante pour les 3 conditions. Les grandes moyennes font clairement apparaître une FFR bien définie ( $n=5$ ) aux trois notes de piano ( $F_0 = 220, 440$  et  $880$  Hz, de bas en haut). Ces deux figures représentent les FFR enregistrées soit entre le vertex et la mastoïde gauche (en haut), soit entre le vertex et la mastoïde droite (en bas).

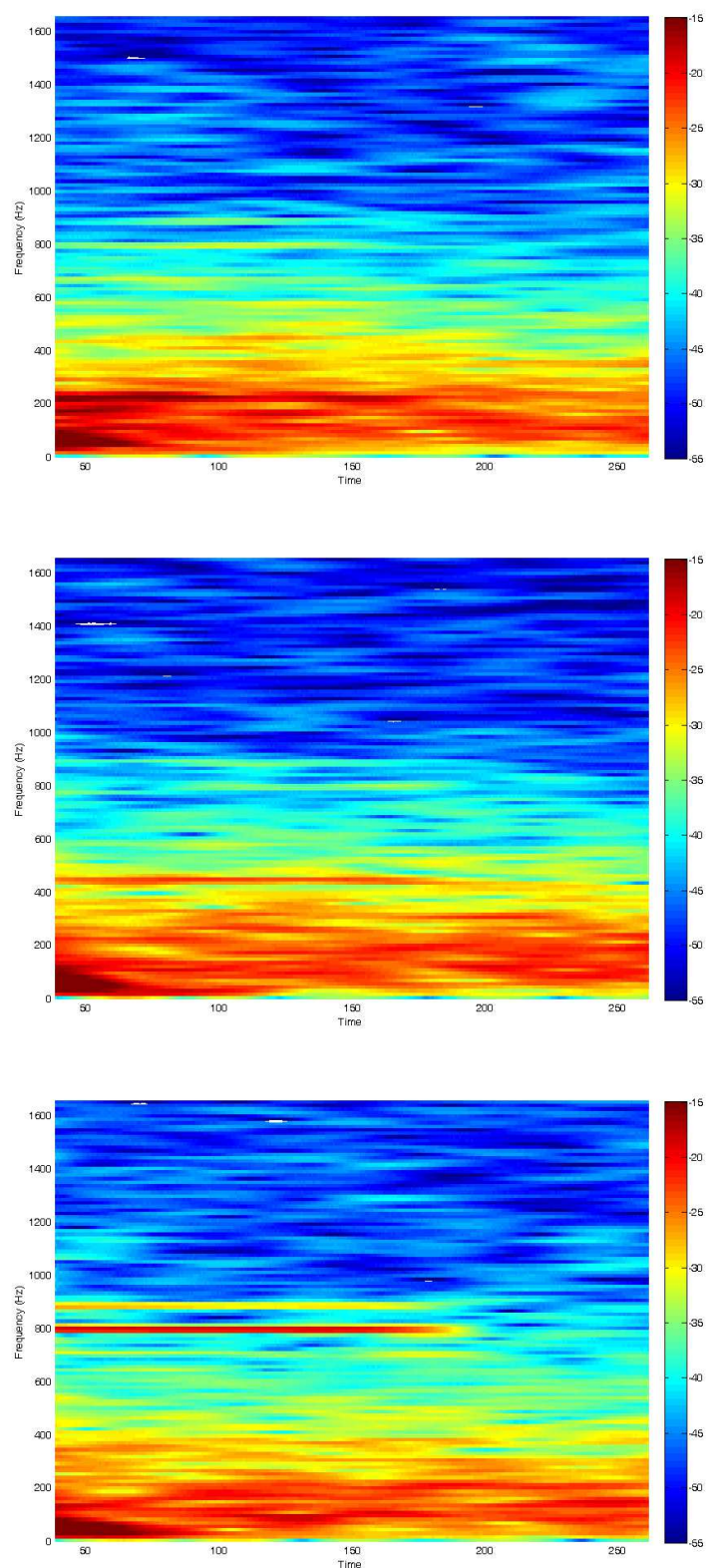


FIGURE 3.32 – FFR en réponse aux notes de piano. Grandes moyennes des spectrogrammes ( $n=5$ ) des FFR en réponse aux sons de piano. La synchronisation de la FFR correspond aux barres d'énergie rouge aux fréquences correspondantes aux F0 (220 en haut, 440 au milieu et 880 en bas).



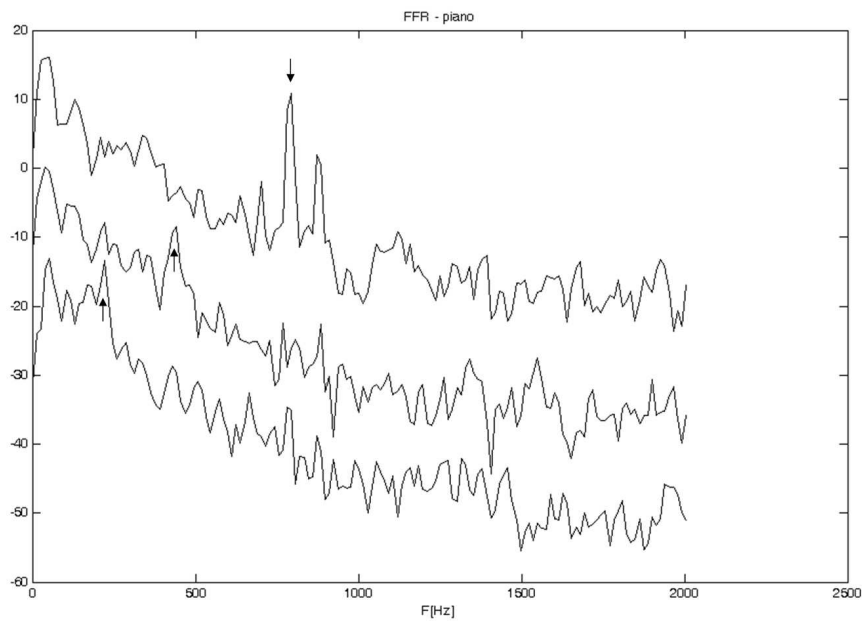


FIGURE 3.33 – Coupes verticales des spectrogrammes de la FFR en réponse aux notes de piano Fig.3.32. Il s'agit du profil spectral du spectrogramme pris à une abscisse de 50 ms, ce qui représente le coeur de la FFR. La bonne synchronisation de la FFR pour les 3  $F_0$  (220 en bas, 440 au milieu et 880 Hz en haut) est visible par le pic à ces fréquences fondamentales, désignés par des flèches.

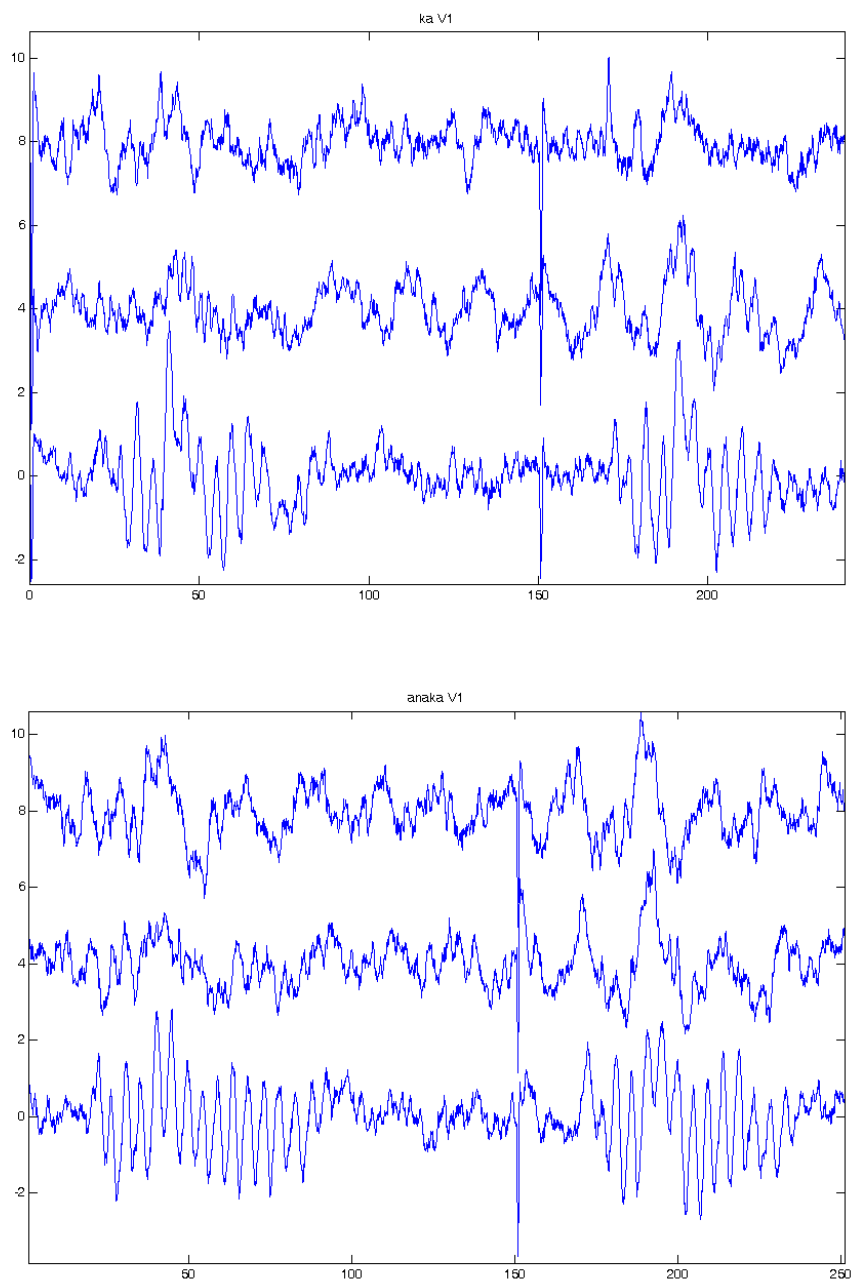


FIGURE 3.34 – FFR en réponse aux /ka/ et aux analogues.

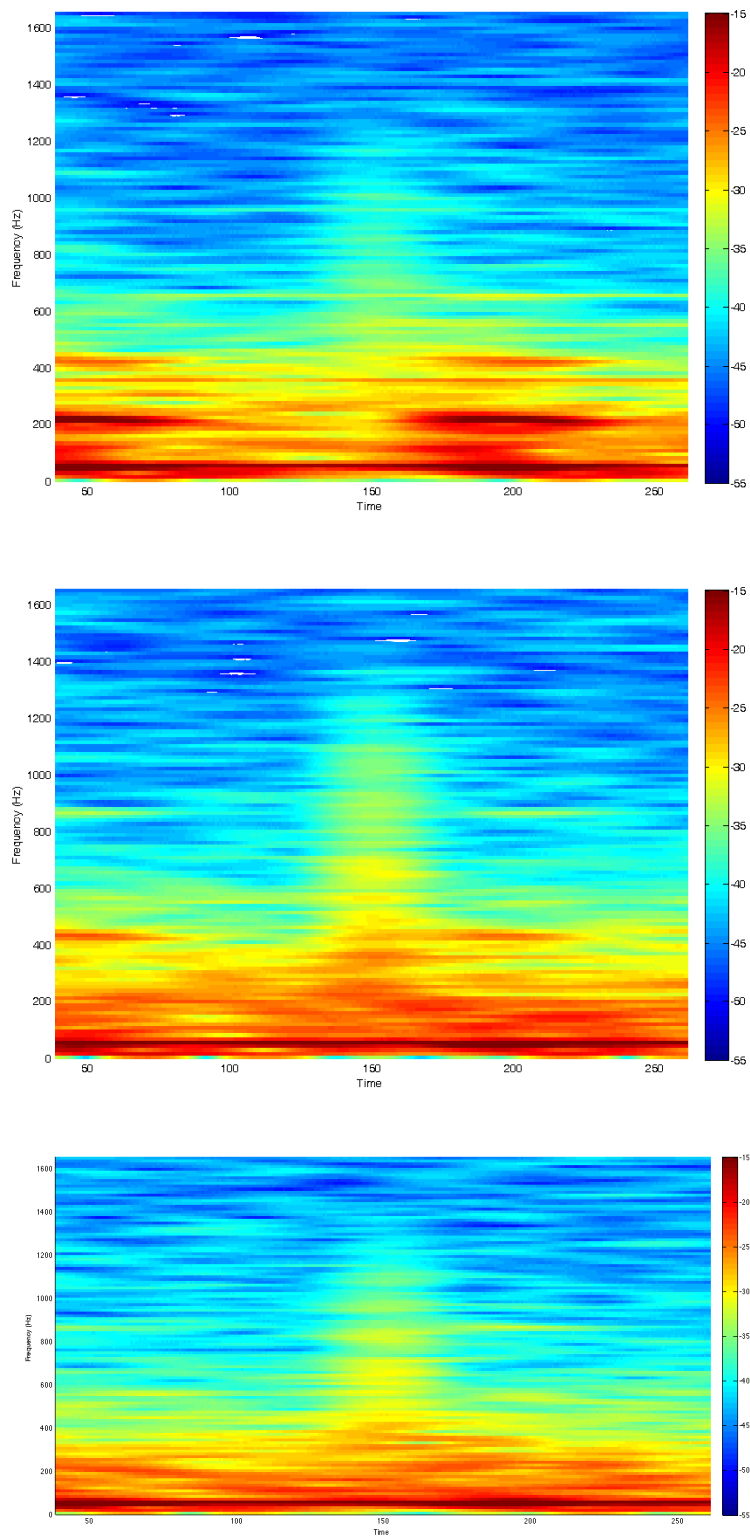


FIGURE 3.35 – Spectrogramme en réponse aux /ka/ à  $F_0 = 215$  (en bas), 440 et 880 Hz.

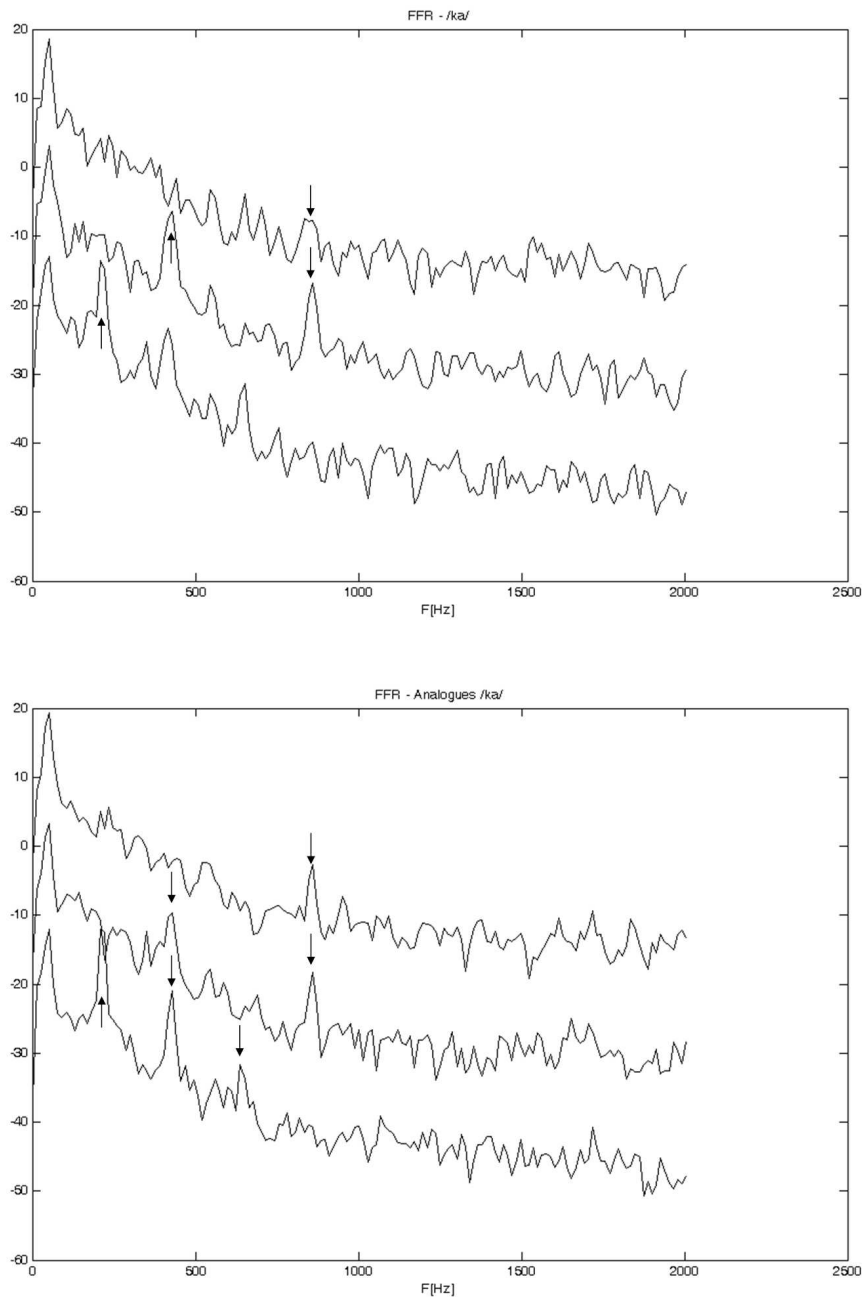


FIGURE 3.36 – Coupes verticales des spectrogrammes de la FFR en réponse aux /ka/ (en haut) et analogues /ka/ (en bas). Fig.3.35. Idem Fig. 3.36. Il s'agit du profil spectral du spectrogramme pris à une abscisse de 200 ms, ce qui représente le coeur de la FFR. La bonne synchronisation de la FFR pour les 3  $F_0$  est visible par le pic à ces fréquences fondamentales, désignés par des flèches.

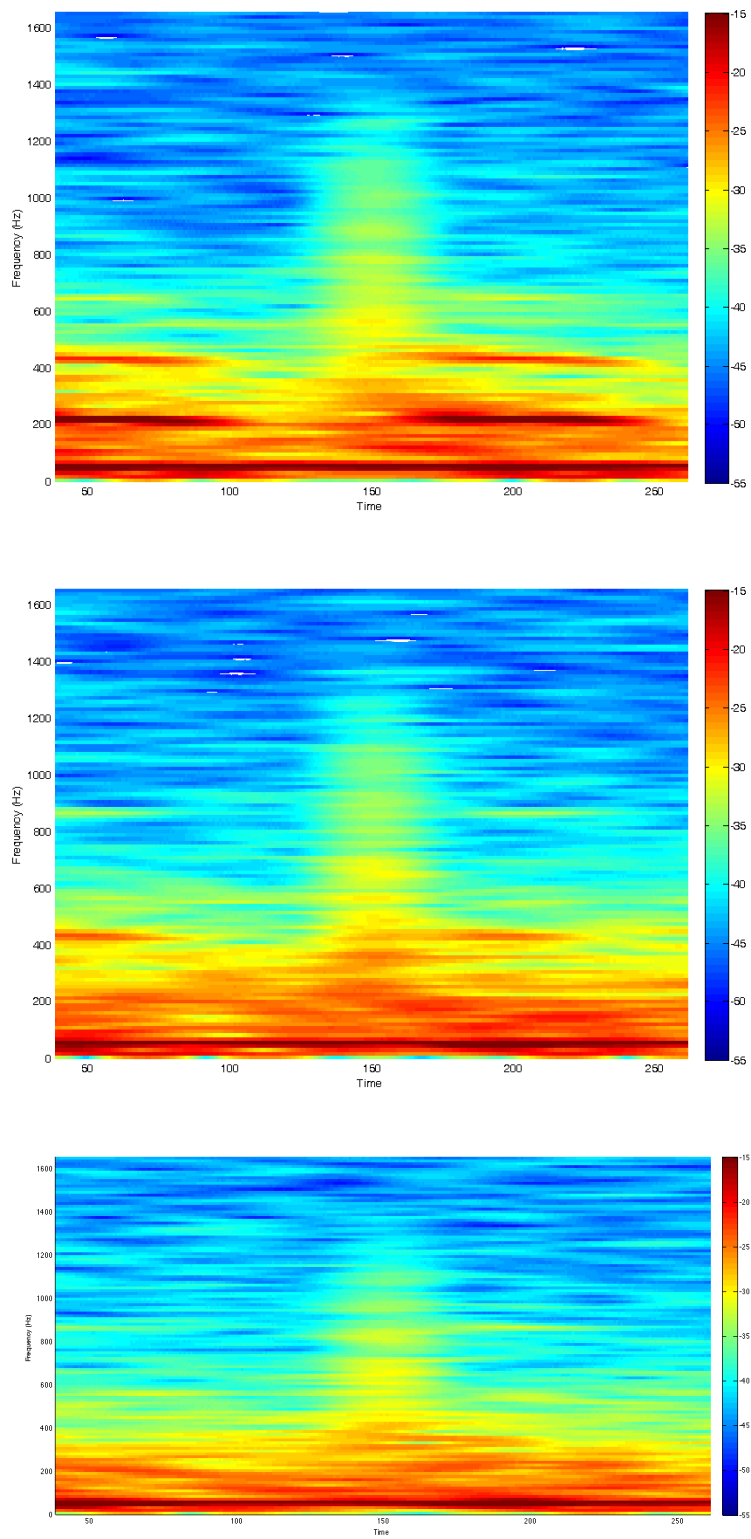


FIGURE 3.37 – Spectrogramme en réponse aux analogues.

semble montrer une sensibilité à la périodicité du stimulus, quel que soit sa nature. Cette expérimentation semble appuyer la thèse selon laquelle le tronc cérébral est davantage sensible à la structure acoustique du stimulus qu'à sa nature (parole, musique ou autre ...). Cela rejoint la très récente étude, où [Krishnan 2008] ont obtenus la même bonne synchronisation sur des IRN comparés aux linguistic pitch (voir ??); même si ces stimuli n'ont pas de structure spectrale et n'introduisent que des sensation de pitch-temporel.

## 3.5 Effet du niveau sonore sur le Speech ABR

Dans l'étude pilote, après avoir obtenu des Speech ABR semblables à ceux décrits, entre autres, par [Russo 2004], nous avons évalué le comportement du Speech ABR en fonction du niveau sonore. Le but était de comparer le comportement des ondes de l'OR et de la FFR en fonction du niveau sonore, avec les résultats bien connus des clic-ABR, dont la latence diminue lorsque l'intensité de stimulation augmente.

### 3.5.1 Matériel et méthodes :

**Sujets :** Les 11 sujets retenus dans cette étude font partie des 23 de l'étude précédente.

**Acquisition :** Les paramètres de recueil sont strictement identiques à ceux de l'étude précédente.

**Stimulation :** La syllabe /ba/ fut présentée de la même manière que dans l'étude précédente, et le niveau sonore fut choisi entre le seuil auditif du sujet (0 dB-SL) et un niveau confortable de 60 dB-SL (situation décrite dans l'étude précédente).

### 3.5.2 Résultats :

Le niveau sonore a une influence sur les latences des composants du Speech ABR. Conformément aux résultats bien connus sur les clics-ABR, ces latences de l'OR et de la FFR diminuent lorsque le niveau sonore augmente (Fig. 3.38). Toutefois, la FFR décroît plus rapidement que l'OR, avec -0.6 ms pour 10 dB supplémentaires, pour l'OR, contre -1.4 pour la FFR. Ceci appuierait les observations de latences de l'étude précédente et orienteraient la discussion vers deux générateurs cellulaires distincts dans le tronc cérébral pour l'OR et pour la FFR (Fig. 3.39).

Il est intéressant de noter Fig.3.38 que la forme des Speech ABR varie très peu pour les intensités variant entre 10 et 50 dB SL, avec une forme assez peu définie. Au contraire, le Speech ABR est quasi-plat au niveau du seuil auditif, et il est très bien défini à 60 dB-SL, avec une OR bien marquée et une FFR très accentuée. De même, la fréquence de coupure du Speech ABR semble fixe aux alentours de 300 Hz (à comparer aux 500 Hz trouvés précédemment).



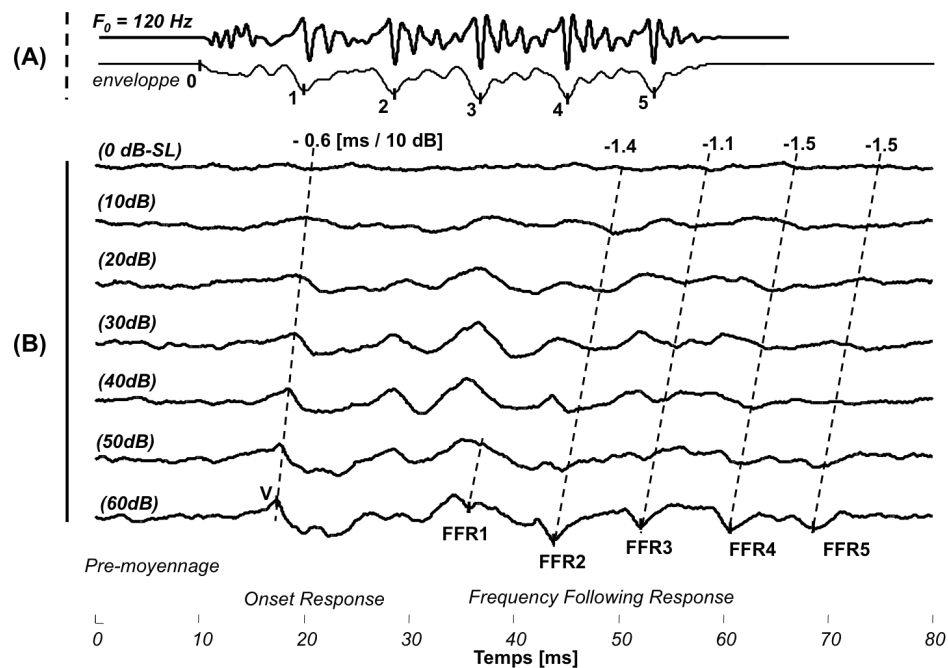


FIGURE 3.38 – **Speech ABR en fonction du niveau sonore (grande moyenne sur 11 sujets)**. Chaque courbe est la grande moyenne en polarité alternée à chaque intensité. L'onde V et la FFR ont des latences décroissantes lorsque l'intensité augmente, comme le soulignent les lignes en pointillés. L'onde V est supposée être engendrée par le début du stimulus, et chaque onde de la FFR par chaque pic de l'enveloppe (1 à 5). Au niveau du seuil auditif (0 dB SL), le Speech ABR est plat, tandis qu'à 60 dB-SL (condition de la Fig.3.3), l'onde V et la FFR sont bien définis. Entre ces deux extrêmes, le Speech ABR conserve une forme globale assez constante, ce qui peut être relié à la stabilité perceptive de perception de la parole en fonction de l'intensité.

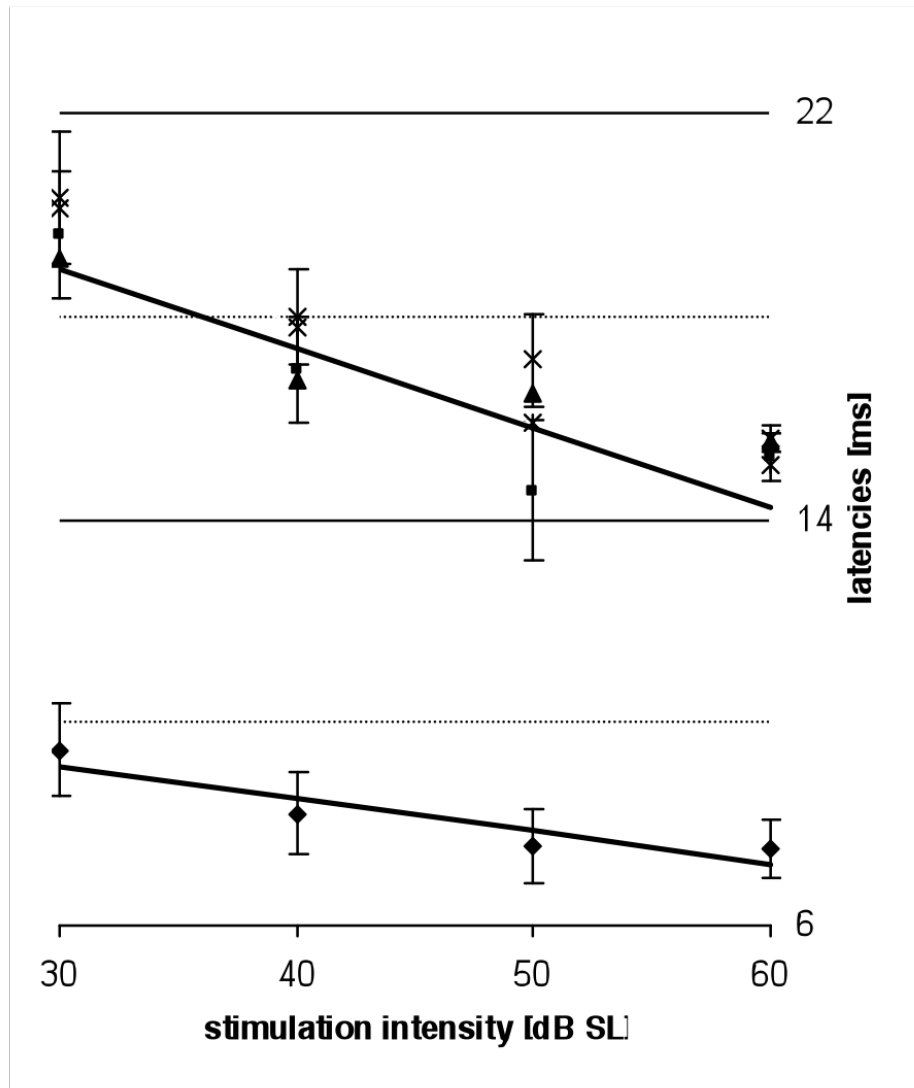


FIGURE 3.39 – **Evolution des latences en fonction de l'intensité.** Chaque point est la latence moyenne (et l'erreur moyenne associée) de chaque ondes, d'après la figure précédente.

### 3.6 Bilan : indices acoustiques influençant les ABR

Dans cette partie d'explorations acoustiques, nous avons mis en évidence une *non-réciprocité* du Speech ABR. Le Speech ABR (la FFR) est le reflet de l'enveloppe temporelle, mais en revanche, l'enveloppe temporelle n'est pas un trait acoustique suffisant pour engendrer une FFR. Au contraire, il semblerait que la FFR soit le reflet d'interactions entre le codage des harmoniques (dans les hautes fréquences?).

Nous avons enfin montré la dépendance du Speech ABR à l'intensité de stimulation, avec deux comportements distincts en latence de l'onset response et de la FFR qui suggèrent deux réseaux différents pour générer ces deux composantes. Nous avons aussi montré que la synchronisation de la FFR suivait la fréquence du stimulus.

Nos données sembleraient montrer que les ABR ne sont pas sensibles à la nature du stimulus, que ce soit parole, musique ou analogues, mais que cette dépendance serait plus liée à la structure acoustique (harmonique en l'occurrence).

### 3.7 Article Clinial Neurophysiology [**Akhoun 2008c**]

3.8 Article Proceedings of Meetings on Acoustics  
[[Akhoun 2008a](#)]



# The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults

I. Akhoun \*, S. Gallégo, A. Moulin, M. Ménard, E. Veuillet,  
C. Berger-Vachon, L. Collet, H. Thai-Van

Université Claude Bernard Lyon I, Université de Lyon, Laboratoire Neurosciences Sensorielles, Comportement, Cognition,  
CNRS UMR 5020, Hôpital Edouard Herriot, Pavillon U – 5, pl. d'Arsonval, F-69003 Lyon, France

Accepted 17 December 2007

## Abstract

**Objective:** To investigate the temporal relationship between speech auditory brainstem responses and acoustic pattern of the phoneme /ba/.

**Methods:** Speech elicited auditory brainstem responses (Speech ABR) to /ba/ were recorded in 23 normal-hearing subjects. Effect of stimulus intensity was assessed on Speech ABR components latencies in 11 subjects. The effect of different transducers on electromagnetic leakage was also measured.

**Results:** Speech ABR showed a reproducible onset response (OR) 6 ms after stimulus onset. The frequency following response (FFR) waveform mimicked the 500 Hz low pass filtered temporal waveform of phoneme /ba/ with a latency shift of 14.6 ms. In addition, the OR and FFR latencies decreased with increasing stimulus intensity, with a greater rate for FFR (−1.4 ms/10 dB) than for OR (−0.6 ms/10 dB).

**Conclusions:** A close relationship was found between the pattern of the acoustic stimulus and the FFR temporal structure. Furthermore, differences in latency behaviour suggest different generation mechanisms for FFR and OR.

**Significance:** The results provided further insight into the temporal encoding of basic speech stimulus at the brainstem level in humans.  
© 2007 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Temporal coding; Speech processing; Consonant–vowels; Periodicity; Auditory brainstem response; Frequency following response

## 1. Introduction

How well are speech-cues encoded in the central nervous system during the processing of a simple syllable?

The relative contribution of spectral and temporal information for the intelligibility of speech has drawn the attention of many researchers in psychophysics (Clark, 2003; Zeng et al., 2005). Degrading information from the spectral distribution of a speech stimulus does not affect speech recognition, provided envelope temporal cues (up to about 50 Hz) are preserved (Shannon et al., 1995; Shannon,

2002). Speech periodicity temporal cues (high amplitude modulation (AM) frequencies between 64 and 500 Hz) are responsible for prosody and voicing, and are thought to be processed in the brainstem (Giraud et al., 2000). Listeners are capable of detecting temporal fluctuations up to 300–500 Hz even in the absence of spectral information (Viemeister, 1979). Furthermore, cochlear implant recipients can perceive pitch based only on the temporal cue of electrical stimulation, with a rate up to 300 Hz (Moore and Carlyon, 2005). Envelope cues can be recovered from the information conveyed through temporal fine structure (Gilbert and Lorenzi, 2006). Assuming that temporal periodicity contains important information needed for speech perception, and given that temporal auditory acuity can be disturbed for language impaired subjects, it is of main interest to have avail-

\* Corresponding author. Tel.: +33 472110526; fax: +33 4372110504.  
E-mail address: idrick.akhoun@chu-lyon.fr (I. Akhoun).  
URL: <http://olfac.univ-lyon1.fr>.



able objective measures of temporal neural coding such as auditory evoked potentials, for assessing how temporal information is conveyed along the auditory pathways (Russo et al., 2004; Purcell et al., 2004). Galbraith et al. (1995) had previously reported that the information contained in word stimuli is so robustly encoded by the brainstem, that the acoustic pattern of these words remains quasi-intact in the neural response. Those studies give evidence that the acoustic signal is already pre-processed in the brainstem before the message is integrated in the cortex. Indeed, acoustic cues are temporally coded at the cell scale in brainstem nuclei thanks to phase-locking phenomenon: neural discharge of a group of cells is synchronized to the periodic part of an acoustic stimulation. This phenomenon is maintained along the auditory pathway, irrespective of the cell characteristic frequencies. The upper boundary of phase-locking has been measured around 3000–5000 Hz for auditory nerve fibers (AN) (Rose et al., 1967; Johnson, 1980; Clark et al., 1995), around 1000 Hz for the cochlear nucleus (CN) (Rhode and Greenberg, 1994), whereas in the cat inferior colliculus (IC), the rate modulation transfer function follows AM up to 300 Hz (Krishna and Semple, 2000; Dicke et al., 2007).

The FFR is the compound neural activity from phase-locking in brainstem nuclei in lower brainstem (CN), IC and medial geniculate body (Galbraith et al., 2001). It was first recorded from the human scalp in response to tones distributed in the speech frequency range between 250 and 2000 Hz (Moushegian et al., 1973), then in response to complex tones (Greenberg et al., 1987). Despite its strong analogies with the stimulus itself, it was shown that the FFR is a true neural mechanism distinct from cochlear microphonics (CM), given its latency, the absence of response far from the neural source, and its behaviour under terminal anoxia (Marsh et al., 1970; Hall, 1979). However, special care must be taken to decrease the effect of electromagnetic artefact which risks contaminating the recordings, since both recordings and stimulations are performed within the same time frame, contrary to click-elicited ABR. To ensure that the recording system is as immune as possible against this kind of perturbation, we assessed, in a first part, the influence of different types of transducers on the recording system.

Speech elicited auditory brainstem response (Speech ABR) to a short synthetic consonant–vowel (CV) syllable is made of two successive components: the onset response (OR), which is a transient component which corresponds to the onset burst of the stimulus, and the FFR which arises after the OR and mimics the periodicity of the signal, reflecting underlying phase-locking (Faingold and Caspary, 1979; Russo et al., 2004; Johnson et al., 2005 for a review). Speech ABR has been described to be a potential useful diagnostic tool. Several studies have linked stimuli patterns to Speech ABR component characteristics and investigated the magnitude and one-to-one correspondence of spectral peaks (Wible et al., 2004; Russo et al., 2004; Krishnan, 2002). The correspondence between high-amplitude formant lines on both spectrograms of FFR and the

vowel was reported by Krishnan et al. (2004) and Wong et al. (2007). In addition, several studies have pointed out the potential usefulness of Speech ABR in the diagnosis of speech processing impairment (Johnson et al., 2005 for a review): a population of learning impaired children was shown to have delayed grand-average OR latencies compared to a population of control subjects (Russo et al., 2004; Johnson et al., 2005). Temporal modulation transfer functions (tMTFs), which are known to be linked to certain speech processing abilities, were correlated to the amplitude of the spectrum in scalp-recorded envelope following responses (EFR) (Purcell et al., 2004). In addition, language experience was also shown to have a significant effect on FFR synchronization (Krishnan et al., 2004, 2005; Chandrasekaran et al., 2007). Finally, Speech ABR to Chinese tones (for English speaking subjects) followed the fluctuations of the periodicity of the stimulus far more accurately in musicians than in non-musicians (Wong et al., 2007). This was suggested to reflect the greater ability of musicians to detect prosody in a foreign tonal language than non-musicians. Furthermore, musicians had enhanced FFR synchronization to consonant–vowels (/da/) and also to music stimuli (G note on a cello) than non-musicians, especially in audiovisual modality (Musacchia et al., 2007).

The question of the generators of Speech ABR remains controversial. In practice, comparing latencies of the Speech ABR cues with those in electrophysiological multi-unitary recordings gives indications regarding the locus where it has been generated (Cunningham et al., 2002). Latencies can be either measured directly or calculated by intercorrelation between the FFR and the vowel pattern (Palaskas et al., 1989; Russo et al., 2004). According to some authors, it is thought to reflect the temporal firing activity of IC auditory neurons (Smith et al., 1975; Sohmer et al., 1977; Greenberg et al., 1987; Cunningham et al., 2002). Many authors including John and Picton (2000) and Eggermont and Schmidt (1990) suggested that the FFR elicited by tone bursts are a succession of waves III (generated in CN) and V (generated in IC) concomitant to the periodic part of the stimulus, suggesting that the FFR may have distinct generators, all located in the brainstem (Bledsoe et al., 1982).

Here we aimed at assessing the quality of encoding of a CV /ba/ syllable in the brainstem in normal-hearing adults. In addition, we studied the possible generators by measuring the cut-off frequency of the Speech ABR (around 500 Hz) and the evolution of latencies as a function of stimulation intensity.

## 2. Methods

### 2.1. Quantitative assessment of electromagnetic artefact

#### 2.1.1. Transducers tested

The artefact test was performed in a simple head model (physiological liquid), around which 5 transducers (3 headphones and 2 insert-earphones) were successively tested.



### 2.1.2. Stimulation and recording

**2.1.2.1. Stimulation.** The sound file consisted of a set of 6 pure tones between 110 and 4120 Hz over 6 octaves. Each transducer was calibrated with the Larson Davis<sup>®</sup> (824) microphone system, including 2 cc coupler for the insert-earphones. The sound files were then played for each transducer, around the liquid, between 30 dB and 130 dB-SPL, by 10 dB steps (see below in Section 2.2.2.1).

**2.1.2.2. Recordings.** EEG silver electrodes were placed in the physiological liquid and connected to the ABR recording system Centor USB (Racia-Alvar<sup>®</sup>). Recording parameters were fixed to record 4000 points for 400 ms (sampling frequency = 10 kHz). The recordings were then filtered between 2 and 4000 Hz and 600 epochs were averaged.

### 2.1.3. Data analysis

The presence of artefact was visually detected in the recorded spectrum, calculated by Fast Fourier Transform with Matlab (The Mathworks<sup>®</sup>). Artefactual contamination was considered as present when, at least, 1 peak, among the 6 corresponding to the stimulus frequencies, was observed above the ambient noise level.

## 2.2. Temporal relationship between Speech ABR and envelope of the /ba/

### 2.2.1. Subjects

Evoked potentials were recorded from 23 adults male and female (age 19–48 years,  $M = 24 \pm 7$ ). Among them, a subgroup of 11 subjects was tested with different stimulus intensities. All had normal hearing threshold below 20 dB HL for octave frequencies ranging from 250 to 8000 Hz. 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.

### 2.2.2. Speech ABR measurements

Speech ABR measurements were performed using a Centor USB (Racia-Alvar<sup>®</sup>) ABR recording system, using three twined silver electrodes placed between vertex (+) and right mastoid (–), with forehead as ground (Galbraith et al., 2001). Electrode impedances were maintained below 5 k $\Omega$ .

**2.2.2.1. Stimulation.** The stimulus was a consonant–vowel syllable /ba/ (Tomita et al., 2004) composed of a release burst and a vowel. The sound file was sampled at 44,100 Hz, and the vowel was shortened from 250 to 60 ms, to obtain a higher recording sampling frequency. The fundamental frequency was constant at 120 Hz over the whole vowel duration, whereas the first and second formants were around 700 and 1200 Hz, respectively (Fig. 3). The sound was played at 60 dB SL. In a subgroup of 11

subjects, stimulus intensities varied from 0 to 60 dB SL, by 10 dB steps.

Subjects were stimulated monaurally on the right ear. Ten ms of silence was recorded before the beginning of the sound in order to assess the ambient noise level in the first 10 ms of the recordings. Stimuli were presented at a rate of 11.1 syllables per second.

The sound file was obtained from a wave file (44.1 kHz, 8 bits) played by a sound card SoundBlaster<sup>®</sup> Audigy (96 kHz, 16 bits) through an AC 40 (Interacoustics<sup>®</sup>) audiometer and BioLogic<sup>®</sup> insert-earphones. We checked that the signal characteristics, at the insert-earphones output, were a (temporal and frequency) quasi-perfect fit to the wave-file waveform, with the expected delay of 1 ms due to air conduction through the 30 cm-rubber tubes. The 1 ms delay has been subtracted from the Speech ABR recordings.

**2.2.2.2. Recording procedure.** A thousand and five hundred epochs were averaged, with an 80 ms analysis window (sampling frequency = 50 kHz). Speech ABRs were filtered online by an 80–3200 Hz bandpass Finite Impulse Response filter (order 2). Each epoch with amplitude greater than 75  $\mu$ V was discarded. One set of 1500 epochs was averaged for each polarity (positive, negative).

### 2.2.3. Data analysis

Grand-average curves were the algebraic summation of the individual data on a given polarity (alternate polarity is the algebraic sum of blocks for both polarities, i.e. the average of a total of 3000 epochs). The baselines of individual waveforms were corrected by subtracting the mean value to remove DC offset, and the data were divided by the standard error to normalize each individual waveform.

To assess the ambient noise, root-means square (RMS) values were calculated either for the whole signal or for the silence period before averaging.

Spectra were calculated as the module of the Fast Fourier Transform (FFT). The frequency analysis precision was 12.5 Hz (half the sampling frequency 50 kHz divided by 4000 points). Speech ABR cut-off frequency was measured as the frequency over which harmonics were at least 2 times (6 dB) lower than the previous one.

The envelope of the /ba/ was obtained by filtering its rectification using Matlab. The envelope was inverted and resampled at 50 kHz (to match the sampling frequency of the Speech ABR). The filter used was an order 3 zero-phasing Butterworth FIR with a 500 Hz cut-off frequency (which was the measured cut-off frequency for the Speech ABR).

Latencies for all components of Speech ABR (wave V and FFR1 to FFR5) were measured directly on individual Speech ABR, and using an intercorrelation method: The function *corrcoef* under Matlab gave intercorrelation coefficients ( $R$ ) as well as the  $p$ -value to test the hypothesis of no-correlation. The temporal shift of Speech ABR with the periodicity of /ba/ was calculated as the maximum



absolute value of the 1000 intercorrelation coefficients, calculated between the resampled envelope and the FFR (between 10 and 60 ms, i.e. 2500 points). Intercorrelation coefficients were also calculated for individual Speech ABR *versus* grand-averaged Speech ABR between both polarities, and between individual (and grand-averaged) Speech ABR versus the periodicity of /ba/. However, for stimulation intensities below 30 dB, individual Speech ABR components were too small compared to ambient recording noise and could not be measured (Fig. 5).

Previous research has evidenced a perfect synchronization between the FFR and the periodicity of the phoneme /da/ (Cunningham et al., 2002; Russo et al., 2004; Johnson et al., 2005). Accordingly, corrected latencies of the Speech ABR components were calculated as the difference between the measured latency and the time of the corresponding peak in the envelope of the /ba/ stimulus. By analogy with the measurement of wave V latency in click-ABR (Picton et al., 1974), the wave V latency was calculated as the difference between the measured OR latency and the stimulus onset.

Statistical analysis was performed using SysStat<sup>®</sup>, with a significance level of 0.05. Analysis of latencies as a function of stimulus intensities was performed using a repeated measure analysis of variance (RMANOVA).

### 3. Results

#### 3.1. Effect of the choice of transducer on the presence of artefact

We reasoned that any spectral peak corresponding to a frequency of the stimulation sound was induced by electromagnetic leakage from the transducers, since the electrodes were placed in the liquid. Fig. 1 shows the intensity threshold at which artefact occurred for each tested frequency.

For the three headphones tested, an artefact was present for intensities commonly used for stimulation in usual ABR recordings. Using BioLogic insert-earphones, less leakage was seen. Although an artefact appeared only at high intensities and for high frequencies, it was still observed in the range of possible stimulation levels. The only artefact-free transducers were insert-earphones (artefact was only recorded for 4 kHz, and at 130 dB SPL).

#### 3.2. Temporal relationship between Speech ABR and the envelope of the stimulus /ba/

The 1500 averaging blocks of Speech ABR were very similar for both polarities (positive or negative): for every individual Speech ABR, intercorrelation coefficients between both polarities were maximal around 0: between 0.02 and 0.68 ms (mean ( $M$ ) =  $0.22 \pm$  standard error of mean (SEM) = 0.10 ms). The coefficients varied between 0.13 and 0.58 ( $M = 0.53 \pm$  SEM = 0.02). Alternate polarity strongly reduces the effects of residual artefact (Fig. 3). All the results are presented in alternate polarity.

Speech ABR was found to be reproducible, and its components were very well defined on individual Speech ABR, and even better on grand-averaged Speech ABR. The Speech ABR waveforms consisted of the following consecutive parts: the pre-averaging line, the OR & the FFR (the real expected auditory responses) and the post-averaging recording (Fig. 2).

The pre-averaging line, obtained during a 10 ms silence preceding the acoustic stimulus, was used to assess the ambient noise level of the recording. The ambient noise was strongly reduced by the averaging method. The RMS measurements, during pre-averaging time interval, were smaller than the RMS during the rest of the waveform: for pre-averaging, RMS varied between 0.35 and 1.43 ( $M = 0.94 \pm$  SEM = 0.06), whereas for the whole Speech ABR, RMS varied between 1.56 and 1.88 ( $M = 1.70 \pm$  SEM = 0.02). For grand average Speech ABR, the RMS data were smaller than for any individual Speech ABR, as expected from the denoising action of averaging (0.52 for the whole signal and 0.11 for pre-averaging).

The OR and FFR peaks were very well defined and reproducible at 60 dB SL. The wave V of the OR was seen in every subject (except subject 5). However, the first and the last peaks of the FFR (FFR1 and FFR5) were less reproducible. The intermediate part of the recording, between the end of the OR and FFR1 (23–32 ms), was not reproducible, and almost random among subjects. Eventually, in the last part of the recording, the FFR continued and ended in the post-averaging time frame (after the end of the stimulus).

Spectral features were expected within the bandpass of the online recording filter, between 80 and 3200 Hz. Some Speech ABR presented clearly visible higher frequency (first-harmonic) component. The Speech ABR was synchronized on the /ba/ stimulus up to a cut-off-frequency ( $F_c$ ). On the grand average module of the spectrum, the  $F_c$  was visually observed at 500 Hz (Fig. 3). At higher frequencies (around 900 Hz), smaller peaks were also visible on both positive and negative polarities, which were cancelled in alternate polarity: these peaks were probably due to residual artefacts. Spectral analysis gave consistent results across subjects, and similar  $F_c$ , except 600 Hz for subject 5 and 400 Hz for subject 10. Note that the slight lack of synchronization between the stimulus peaks and the Speech ABR spectra was probably due to an artefact of the technique, since the frequency precision of our FFT was 12.5 Hz.

We used the  $F_c$  found, to calculate the (inverted) envelope of the phoneme /ba/: when shifted in time by 14.6 ms, it was found to have a very similar shape to the recorded FFR (Fig. 4). This observation was confirmed by the maximal intercorrelation coefficient between envelope and grand-average FFR:  $R = 0.47$  for a shift in time of 14.6 ms. This observation was consistent among subjects: intercorrelation coefficients were maximal for delays varying between 10.54 and 15.76 ms ( $M = 14.7 \pm$



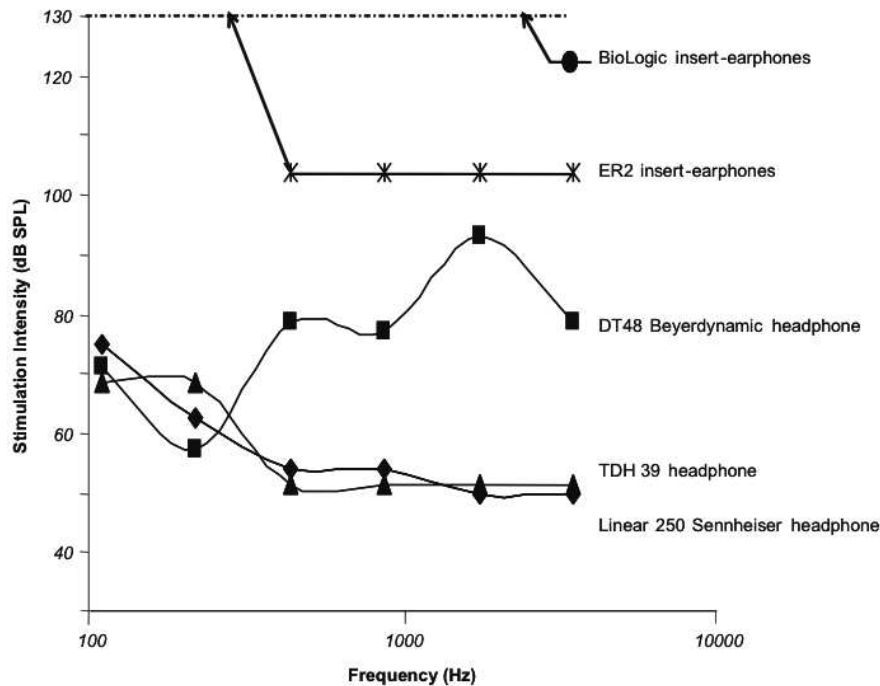


Fig. 1. Electromagnetic artefact for different transducers. Three headphones and two insert-earphones were tested. Each plot represents the intensity threshold at which artefact was observed for a given frequency. The stimuli were a set of 6 pure tones and electrodes were placed in physiologic liquid with the transducer around.

SEM = 0.21 ms), with  $R$  ranging between 0.07 and 0.79 ( $M = 0.35 \pm \text{SEM} = 0.04$ ).

### 3.3. Effect of stimulation intensity on Speech ABR latencies

On grand average data, for the 11 subject subgroup, Speech ABR was best defined for the highest intensity tested (60 dB SL) and was flat at hearing threshold. For every stimulation intensity,  $F_c$  was measured around 300 Hz (including at 60 dB SL), whereas 500 Hz was obtained on the 23-subject grand average. The lower  $F_c$  found for 11-subject grand-average (300 Hz) compared to 500 Hz for 23-subject grand average might be a consequence of poorer high-frequency noise reduction in the former group due to the smaller number of averages. Speech ABR components remained quite robust with the decrease of intensity, and latencies were dependent on stimulation intensity (RMANOVA  $F(3,88) = 2.9$ ,  $p < 0.05$ ): latencies were increasing with decreasing stimulation intensity, as expected (Fig. 5). The FFR 1 peak was not present for stimulation intensities below 50 dB SL, and even not reproducible among subjects at 60 dB SL. Therefore, it was not taken into account for further analysis.

Corrected latencies show two distinct groups, with significant different latencies (Fig. 6) (RMANOVA  $F(1,54) = 614$ ;  $p < 0.001$ ): the OR lagged between 6.4 and 9.2 ms from the stimulus onset, whereas the FFR peak latencies were grouped between 14.6 and 20.6 ms after the stimulus envelope (Fig. 5).

The Speech ABR latencies varied linearly as a function of stimulus intensity for most subjects. The linear regression coefficients found for each subject, for wave V and four FFR peaks were above 0.60 (except for 3 subjects for wave V data, 2 out of 11 for FFR2 and 3 out of 11 for FFR3). The rate of latency decrease with increasing stimulus intensities was significantly different (RMANOVA  $p < 0.001$ ,  $F(4,40) = 5.9$ ), between wave V latency ( $M = -0.64 \pm \text{SEM} = 0.08$  ms/10 dB) and the FFR latencies, ( $M = -1.41 \pm \text{SEM} = 0.16$  for FFR 2,  $M = -1.09 \pm \text{SEM} = 0.15$  for FFR 3,  $M = -1.48 \pm \text{SEM} = 0.12$  for FFR 4,  $M = -1.53 \pm \text{SEM} = 0.18$  ms/10 dB for FFR 5), as shown on Fig. 6.

## 4. Discussion

In the present study, Speech ABR, elicited by a consonant–vowel syllable /ba/, was found to be an objective measurement of temporal encoding in humans. This is in line with previous ABR studies using the phoneme /da/ (Russo et al., 2004), or AM (Purcell et al., 2004). The key result was that the FFR temporal pattern mimicked the periodicity of the phoneme /ba/ for any polarity. In addition, we observed differences in OR and FFR latency behaviour as a function of stimulus intensity.

The Speech ABR  $F_c$  ranged between 300 and 600 Hz, far below the upper limit of the recording bandpass filter (3200 Hz). The  $F_c$  approximated the phase-locking boundary described in animal brainstem nuclei studies (Krishna

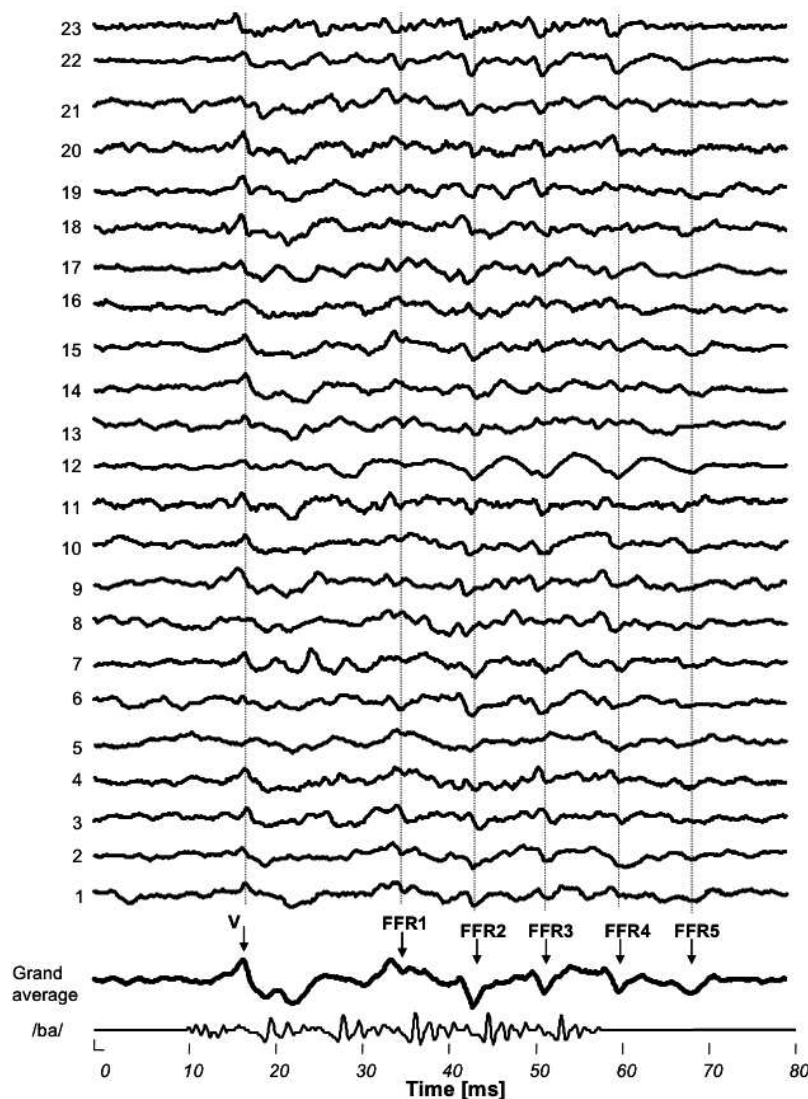


Fig. 2. Individual temporal waveforms of Speech ABR (alternate polarity). Bottom trace represents the /ba/ stimulus played at 60 dB SL, bold trace the grand-averaged Speech ABR. Specific peaks (wave V and FFR peaks) are shown by arrows.

and Semple, 2000). Interestingly, human behavioural studies pointed out 500 Hz as the limit of temporal acuity for modulation amplitude detection in tMTF test (Viemeister, 1979). Russo et al. (2004) and Wible et al. (2004) also reported a 500 Hz  $F_c$  for grand-averaged Speech ABR to a short /da/ in children.

The FFR is thought to have spectral peaks close to the modulating frequency (Purcell et al., 2004), or to the stimulus fundamental frequency and its first harmonics (Krishnan et al., 2004). Investigating the brainstem encoding of steady-state vowels, Krishnan (2002) pointed out that higher frequency spectral peaks might fade out because of the decrease of phase-locking in upper brainstem. In addition, he found that FFR  $F_c$  increases with stimulation intensity and reported much higher  $F_c$  values than here (around 1000 Hz). A possible explanation is that this author did not use insert-earphones, which may have

affected high frequency FFR components. Indeed, these components may be due to another mechanism than just neural response, such as electromagnetic leakage from the transducers. That was checked in our study and we showed that electromagnetic leakage was strongly reduced by the use of insert-earphones.

The OR is robustly present among subjects at 7–10 ms, which agrees with the range of latencies obtained for wave V at equivalent intensities in click evoked ABR (Picton et al., 1974), as well as the OR values reported by Russo et al. (2004). In this study, the consonant of the /ba/ was used as a burst to elicit an obvious OR, like a click would do in standard ABR (Eggermont and Schmidt, 1990; Jewett et al., 1970; Picton et al., 1974). The OR clearly shows a wave V around 7 ms, at 60 dB SL. This is consistent with the study of Song et al. (2006): in children aged 8–12 years, the wave V has been found, in quiet, to arise about 6.7 ms



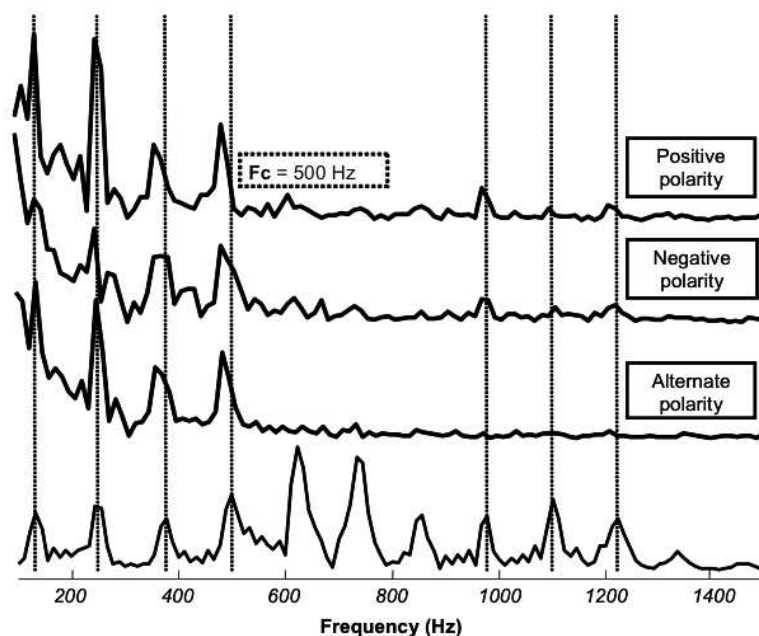


Fig. 3. Spectra of the Speech-ABR for both polarities and alternate polarity (linear scale). Both electrophysiological response and acoustic stimulation (bottom waveform) show common peaks (linked by dotted vertical lines) up to a cut-off frequency ( $F_c$ ) of 500 Hz. Some higher harmonics can also be seen at 900 and 1200 Hz for positive and negative polarities, but not for alternate polarity.

after the stimulus onset in response to the /da/ syllable played at 80 dB SPL, versus 5.9 ms in response to click. The OR complex (positive wave V, and negative wave

A), as described by Russo et al. (2004), may be the result of a massive synchronization in many types of brainstem cells (in CN and IC). Furthermore, we have shown that

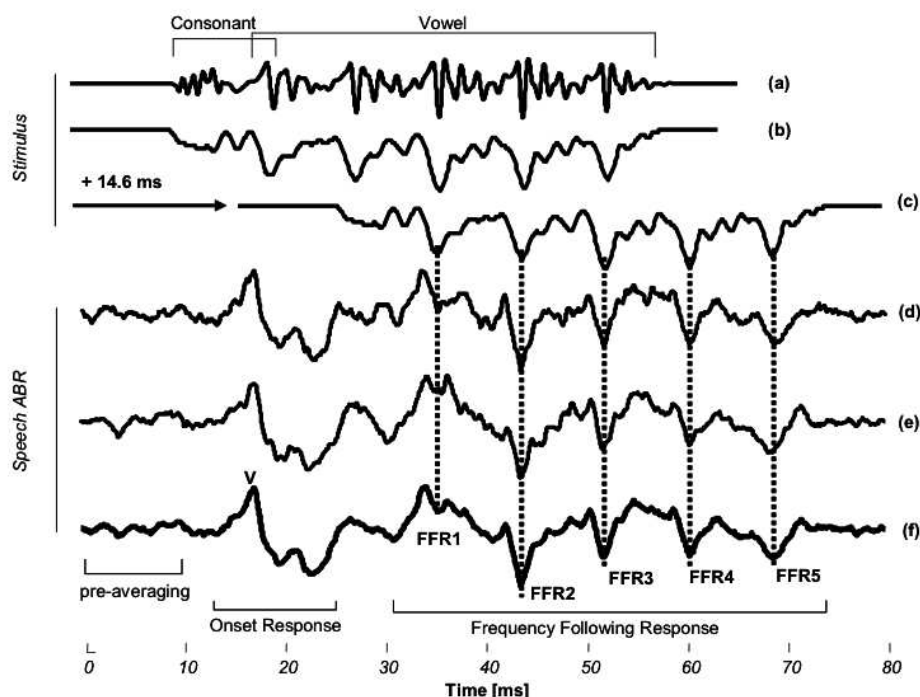


Fig. 4. Good-fit between /ba/ envelope and Speech ABR. The two top waveforms show the temporal waveform of the /ba/ stimulus (a) and its envelope (b). The envelope is extracted by full wave rectification and filtering at  $F_c$  (500 Hz). Note that, when inverted and shifted by a 14.6 ms (c), the envelope is accurately represented in FFR responses obtained for positive (d), negative (e) and alternate polarity (f). Ordinates are in arbitrary units (the Speech-ABR is a voltage ( $\mu\text{V}$ ) and the sound an 8 bits-wave file).

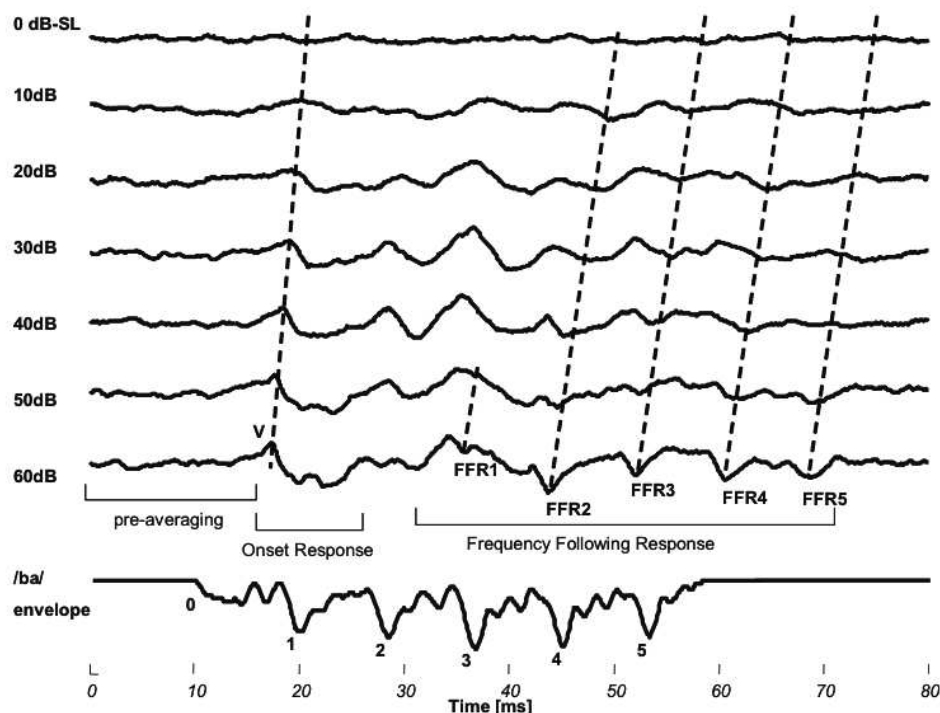


Fig. 5. Speech ABR latency/intensity function. The wave V is supposed to be elicited by the beginning of the stimulus and each FFR wave is supposed to be elicited by the corresponding peak of the envelope (FFR1 generated by '1', and *idem* for FFR2 to FFR5). The bottom waveform is the envelope of the /ba/ (*idem* Fig. 4). Note that the latency of each component (wave V, FFR 1, FFR 2, FFR 3, FFR 4) clearly increases as intensity decreases. The grand-averaged Speech ABR shown here was obtained for alternate polarity in 11 subjects.

OR is sensitive in regard to stimulus intensity (Junius and Dau, 2005), and we found OR latencies that agree with a generator at the CN or IC level (Picton et al., 1974; Wible et al., 2004). Interestingly, octopus cells firing has been shown to be intensity dependent, and to act like a coincidence detector (at least in animal studies) from a large number of AN fibers input within a great span of characteristic frequencies (Rhode and Greenberg, 1994; Eggermont, 2001, chapter 1, p. 3; Carr et al., 2005). Such coincidence detector would fire greatly at the stimulus onset. Our big OR response might be, at least partly, the consequence of firing of octopus cells, located mainly in lateral lemniscus, CN and IC.

Our results about the OR characteristics, or the non-reproducibility of FFR1 and  $F_c$ , are in agreement with Russo et al.'s (2004) and Johnson et al. (2005) studies. However, we observed an FFR synchronized to the waveform of the envelope of the phoneme /ba/, with a time shift of 14.6 ms, which is greater than the latencies reported by Russo et al. (2004) and Johnson et al. (2005), using the same intercorrelation technique (despite similarities between our speech material and the one they used). First, this could be due to the fact that the characteristics of the stimulus were steadier in this study, with no formant transitions during the vowel. Second, children were tested in Russo's study whereas it was adults here. No differences in OR latencies based on anatomical delays could be expected, since it has been shown that the brainstem is

already mature at 3 years of age (Moore et al., 1996). However, speech perception as a whole keeps on changing during childhood. Thus, due to functional maturation, it may be possible to find differences between children and adults speech processing at the brainstem level, such as differences in FFR latencies. Third, the last peak of the FFR (FFR5) could correspond to the *offset response* introduced by Johnson et al. (2005). The absence of this peak in the FFR can have consequences on the latency measured by the intercorrelation technique. However, the steady fundamental frequency did not allow us to consider the last wave as an offset response. Therefore, we chose to take the whole Speech ABR wave as part of the FFR for the autocorrelation, and not to differentiate the last wave as an offset. Further investigation would be useful to detail the behaviour and the role of this offset response for speech processing. Other physiological studies, such as multi-unit neural recordings in guinea pigs IC, have found that the FFR lags behind the synthetic stimulus /ada/ by an offset ranging between 5 and 9 ms (Cunningham et al., 2002). This is consistent with previous FFR offset reported to be of 5.8 ms for human scalp recordings and 5.4 ms for multi-unit neural recordings in cat IC. However, these values concerned FFR to 500–1000 Hz-tones (Smith et al., 1975). More recently, EFR latencies reported in previous research (John and Picton, 2000; Purcell et al., 2004; Aiken and Picton, 2006) are higher than those measured by Russo et al. (2004), and closer to the FFR latencies we measured.



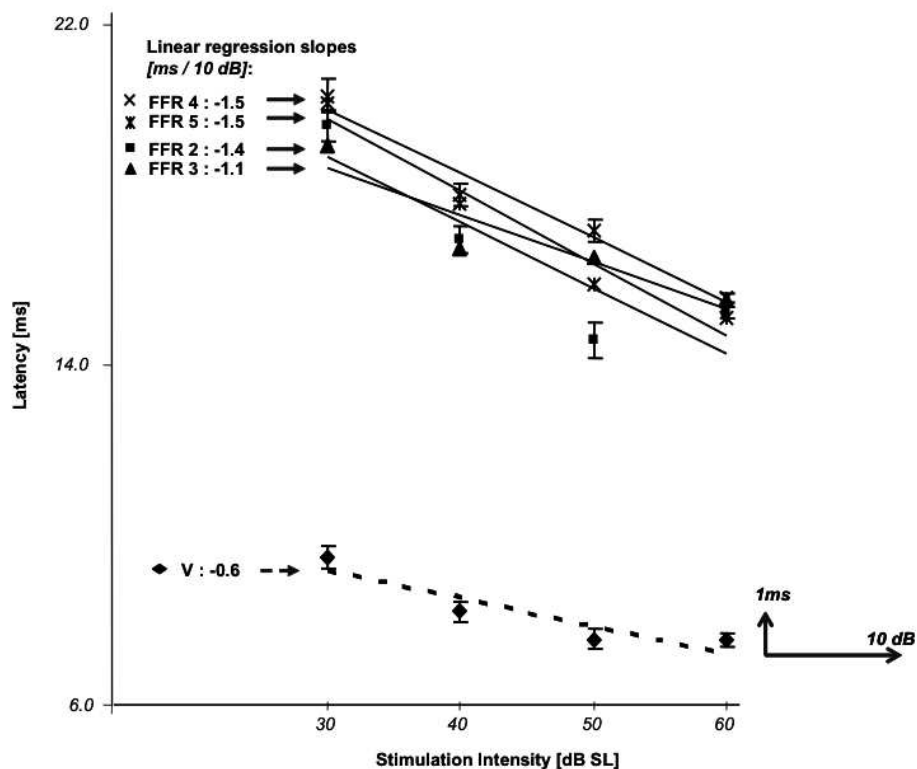


Fig. 6. Differences in latency/intensity function for Speech ABR onset response (OR) and frequency following response (FFR) peaks. The OR corrected-latency was the time lag between 'V' and '0', FFR1 corrected latency was between 'FFR1' and '1' and *idem* for FFR2 until FFR5 (Fig. 5). Mean latency values ( $\pm$  SEM) are plotted as a function of stimulation intensity for each individual Speech ABR component, except FFR 1 (non-reproducible for stimulation intensity below 50 dB, as shown on Fig. 5). Note that the latency/intensity function is well fitted by a linear regression for these components. Steeper slopes are obtained for FFR components than for OR.

Apparent-latencies, defined as the sweeping-EFR group-delay (slope of the phase of Fourier-analyzed EFR *vs.* modulating frequency), depend on stimulation intensity, on the carrier frequencies and modulation rate. Given the similarity of John et al.'s AM stimuli and the vowel of our /ba/, the outcome of their EFR-latency study can be compared to latencies measured in our study. The vowel /a/ can be modelled by the following AM stimuli: a 700 Hz carrier frequency (first formant) added to a 1200 Hz carrier frequency (near the second formant), both modulated by the 120 Hz fundamental. The fundamental of our vowel is in between the two modulating ranges chosen by John et al., i.e. 80–100 and 150–190 Hz. For these two sweeping AM stimuli, they reported EFR latencies varying, respectively, between 20.2 and 17.6 ms for a modulation rate of 80–100 Hz, and between 11.1 and 8.9 ms for a modulation rate of 150–190 Hz (for stimulation intensity increasing from 35 to 75 dB SPL), latencies that agree with the FFR latencies measured in our study (14.6 ms).

As expected, stimulation intensity has a significant effect on the Speech ABR components: OR latencies decrease as intensity rises in a similar way as click ABR do, and so do FFR peaks. At hearing threshold, the Speech ABR waveform is non-surprisingly flat, whereas at 60 dB SL it is best defined. Between 10 and 50 dB SL, the pattern remains the

same, even if latencies are shifted later as intensity decreases, as expected. First, this regularity in the Speech ABR pattern for several intensities can be related to the invariant perception of speech (in quiet) when intensity varies. Second, this can be related to the robustness of the discharge of some auditory brainstem cells, such as the onset chopper group of cells. These cells, also described as "primary-like with notch" cells, are very sensitive to AM and have been shown to result in the most prominent phase-locking, up to 1 kHz for the cat: they are thought to play a major role in FFR generation, by accurate AM temporal coding (Rhode and Greenberg, 1994; Oertel et al., 2000; Kalluri and Delgutte, 2003; Winter, 2007). In addition, IC cells integrate afferences from CN up to 300 Hz (Lorenzi et al., 1995; Krishna and Semple, 2000; Dicke et al., 2007).

Both FFR and OR occur with different latencies, for each stimulation intensity. Furthermore, the two different groups of corrected-latencies of OR and FFR vary differently, respectively, decreasing by 0.6 ms/10 dB between 9 and 6 ms, and by 1.4 ms between 21 and 14 ms, with increasing stimulation intensity between 30 and 60 dB SL. Since waveform latencies are directly related to underlying generators, the difference of decreasing rate of OR and FFR latencies implicates that OR and FFR coding would



separate in 2 distinct neural streams within the same pathway, addressing again the question of the origin of the Speech ABR (Krishnan, 2002; Cunningham et al., 2002; Purcell et al., 2004; Russo et al., 2004). In addition to Speech ABR studies, EFR studies provide further information about physiological generators. To explain the apparent latencies measured, John and Picton (2000) and Purcell et al. (2004) suggested that the physiological response may originate from either above the brainstem in the auditory pathway, or from multisynaptic processes intrinsic to the brainstem. In addition, such a delay could also be due to refractoriness. There is little probability that any component of the Speech ABR could be generated at a higher level than the brainstem, as the  $F_c$  would have been much lower (around 80 Hz for the cortex, Eggermont (2001)). Furthermore, the OR has been described as a direct result from the response of IC cells in analogy with click-ABR (Picton et al., 1974): this supports the idea that the FFR is processed within the same brainstem nuclei as OR, but through a distinct network probably involving different synaptic interactions (Wible et al., 2004; Purcell et al., 2004; Greenberg et al., 1987; Sohmer et al., 1977; Smith et al., 1975).

Kraus and Nicol (2005) developed this hypothesis, i.e., that brainstem encoding is already a precursor to the auditory aspect of the divergent sensory streams identified in the cortex as *what & where* pathways: *what* stands for the speaker's identity (*source*) and *where* for the spoken message (*filter*). To show that, the authors distinguish these two different streams of information in the Speech ABR profile. The *source* is linked to fine timing parameters, harmonic pitch, speaker identification (vocal tract vibration and fundamental frequency): it is encoded in the FFR. The *filter* conveys the global spectral patterns message, the formant information: it is encoded in the onset and offset waves. Those 2 cues are differentiated by the fact that they are not equally robust in adverse conditions (e.g. in noise) (Kraus and Nicol, 2005; Johnson et al., 2005). The OR and FFR thus result from different mechanisms of temporal processing of speech stimuli. The behaviour of these components, depending on stimulation condition, can offer valuable insights in understanding speech integration at the brainstem level (Wong et al., 2007; Russo et al., 2004; Purcell et al., 2004).

However, in our study, FFR synchronizes on frequencies up to 500 Hz, which is smaller than the first formant of our vowel /a/. In Speech ABR, FFR would then be more an objective assessment of the global spectral pattern (i.e. the *source* pathway), than the *filter* pathway, given the lack of synchronization to assess the formant component at high frequencies. However, such high frequency information spatially coded in cochlea and AN may be present in an encoded form in Speech ABR. Furthermore, OR timing would be correlated to voice onset time (VOT). Further investigation should be done to assess the relationship between OR timing and VOT.

We confirmed that the temporal relationship between Speech ABR and the syllable /ba/ is very accurate between FFR and the vowel /a/, whereas the consonant /b/ is poorly represented. Interestingly, the dual speech perception theory of Liberman et al. suggests that consonants are 'encoded' in the speech signal (and the listener needs the help of the context to understand them), unlike vowels which are 'non-encoded' (Liberman et al., 1967; Moore, 2004). The accurate synchronization of FFR to the vowel may show that the acoustic information relative to the vowel is already integrated in the brainstem, and thus the vowel can be understood, whereas the context is important to understand the consonant, which is more likely to be represented in OR.

## 5. Conclusion

In this study, we found a close relationship between the FFR and the envelope of the vowel, whereas the consonant had more influence on the OR. We provided evidence for separate streams to code both components. This raises interest into correlating Speech ABR with behavioural studies of speech perception. The cut-off frequency of the Speech ABR could be linked to the limit of perception for patients; however the averaging technique may cancel high frequency components because of jitter across traces and subjects. Latencies of both OR and FFR could also be interesting cues to assess temporal acuity in speech perception. The use of different stimuli and tests on pathological populations such as hearing impaired subjects (Daly et al., 1976) or learning impaired children (Wible et al., 2004; Russo et al., 2004; Johnson et al., 2005) would be the next step to detail factors influencing Speech ABR.

## Acknowledgements

The authors thank the participants of this study. Warm regards are also expressed to our colleagues A. Jeanvoine, C. Pavillon, C. Richard, M. Vigouroux, F. Buret and C. Vollaire for their work and precious advice on the artefact issue, A. & K. Nyamay'Antu, F. Seldran, E. Truy, for their implication in the study, S. Greenberg, N. Kraus and J. Eggermont for very interesting discussions. We also thank the two anonymous reviewers for their very constructive criticisms. Work supported by CNRS, Univ. Lyon, Univ. Lyon 1, Hospices Civils de Lyon, SFA. I.A. is supported by a CIFRE grant by Cochlear France and UCBL.

## References

- Aiken SJ, Picton TW. Envelope following responses to natural vowels. *Audiol Neurootol* 2006;11(4):213–32.
- Bledsoe Jr SC, Rupert AL, Moushegian G. Response characteristics of cochlear nucleus neurons to 500-Hz tones and noise: findings relating to frequency-following potentials. *J Neurophysiol* 1982;47(1):113–27.
- Carr CE, Iyer S, Soares D, Kalluri S, Simon JZ. Are neurons adapted for specific computations? Examples from temporal coding in the auditory



- system. In: Hemmen LV, Sejnowski T, editors. 23 Problems in systems neuroscience. Oxford; 2005. p. 245–65.
- Clark G. Cochlear implants. Fundamentals & applications, modern acoustic & signal processing coll. Springer; 2003. p. 212.
- Chandrasekaran B, Krishnan A, Gandour JT. Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport* 2007;18(18):1963–7.
- Clark G, Carter TD, Maffi CL, Shepherd RK. Temporal coding of frequency: neuron firing probabilities for acoustic and electric stimulation of the auditory nerve. *Ann Otol Rhinol Laryngol* 1995;113:825–8.
- 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(1–2):97–111.
- Daly D, Roeser R, Moushegian G. The frequency-following response in subjects with profound unilateral hearing loss. *Electroencephalogr Clin Neurophysiol* 1976;40(2):132–42.
- Dicke U, Ewert SD, Dau T, Kollmeier B. A neural circuit transforming temporal periodicity information into a rate-based representation in the mammalian auditory system. *J Acoust Soc Am* 2007;121(1):310–26.
- Eggermont JJ. Between sound and perception: reviewing the search for a neural code. *Hear Res* 2001;157(1–2):1–42.
- Eggermont JJ, Schmidt PH. The auditory brainstem response. In: Colon EJ, Visser SL, editors. Evoked potential manual. Kluwer Academic; 1990. p. 41–77.
- Faingold CL, Caspary DM. Frequency-following responses in primary auditory and reticular formation structures. *Electroencephalogr Clin Neurophysiol* 1979;47(1):12–20.
- Galbraith G, Arbagey P, Branski R, Comerci N, Rector P. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 1995;6(17):2363–7.
- Galbraith G, Bagasan B, Sulahian J. Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations. *Percept Mot Skills* 2001;92(1):99–106.
- Gilbert G, Lorenzi C. The ability of listeners to use recovered envelope cues from speech fine structure. *J Acoust Soc Am* 2006;119(4):2438–44.
- Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, et al. Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 2000;84(3):1588–98.
- Greenberg S, Marsh J, Brown W, Smith J. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear Res* 1987;25(2–3):91–114.
- Hall J. Auditory brainstem frequency following responses to waveform envelope periodicity. *Science* 1979;205(4412):1297–9.
- Jewett D, Romano M, Williston J. Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 1970;167(924):1517–8.
- John MS, Picton TW. Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements. *Hear Res* 2000;141(1–2):57–79.
- Johnson KL, Nicol TG, Kraus N. Brain stem response to speech: a biological marker of auditory processing. *Ear Hear* 2005;26(5):424–34.
- Junius D, Dau T. Influence of cochlear traveling wave and neural adaptation on auditory brainstem responses. *Hear Res* 2005;205(1–2):53–67.
- Johnson KO. Sensory discrimination: neural processes preceding discrimination decisions. *J Neurophysiol* 1980;43:1793–815.
- Kalluri S, Delgutte B. Mathematical models of cochlear nucleus onset neurons. II. Model with dynamic spike-blocking state. *J Comput Neurosci* 2003;14(1):91–110.
- Kraus N, Nicol T. Brainstem origins for cortical what and where pathways in the auditory system. *Trends Neurosci* 2005;28(4):176–81.
- Krishna BS, Semple MN. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J Neurophysiol* 2000;84(1):255–73.
- Krishnan A. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear Res* 2002;166(1–2):192–201.
- Krishnan A, Xu Y, Gandour JT, Cariani PA. Human frequency-following response: representation of pitch contours in Chinese tones. *Hear Res* 2004;189(1–2):1–12.
- 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(1):161–8.
- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychol Rev* 1967;74:431–61.
- Lorenzi C, Micheyl C, Berthommier F. Neuronal correlates of perceptual amplitude-modulation detection. *Hear Res* 1995;90(1–2):219–27.
- Marsh JT, Worden FG, Smith JC. Auditory frequency following response: neural or artifact?. *Science* 1970;169(3951):1222–3.
- Moore BC. An introduction to the psychology of hearing. 5th ed. Elsevier Academic Press; 2004. p. 310.
- Moore BC, Carlyon RP. Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Plack CJ, Oxenham AJ, editors. Springer handbook of auditory research: pitch perception. New York: Springer Verlag; 2005.
- Moore JK, Ponton CW, Eggermont JJ, Wu BJC, Huang JQ. Perinatal maturation of the auditory brainstem: changes in path length and conduction velocity. *Ear Hear* 1996;17(5):411–8.
- 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(6):665–7.
- Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 2007;104(40):15894–8. [Epub].
- Oertel D, Bal R, Gardner S, Smith P, Joris P. Detection of synchrony in the activity of auditory nerve fibers by octopus cells of the mammalian cochlear nucleus. *Proc Natl Acad Sci USA* 2000;97:11773–9.
- Palaskas CW, Wilson MJ, Dobie RA. Electrophysiologic assessment of low-frequency hearing: sedation effects. *Otolaryngol Head Neck Surg* 1989;101(4):434–41.
- Picton T, Hillyard S, Krausz H, Galambos R. Human auditory evoked potentials. I. Evaluation of components. *Electroencephalogr Clin Neurophysiol* 1974;36(2):179–90.
- Purcell DW, John SM, Schneider BA, Picton TW. Human temporal auditory acuity as assessed by envelope following responses. *J Acoust Soc Am* 2004;116(6):3581–93.
- Rhode WS, Greenberg SR. Encoding of amplitude modulation in the cochlear nucleus of the cat. *J Neurophysiol* 1994;71:1797–825.
- Rose JE, Brugge JF, Anderson DJ, Hind JE. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J Neurophysiol* 1967;30:769–93.
- Russo NM, Nicol TG, Musacchia G, Kraus N. Brainstem responses to speech syllables. *Clin Neurophysiol* 2004;115(9):2021–30.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270(5234):303–4.
- Shannon RV. The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. *Am J Audiol* 2002;11(2):124–7.
- Smith J, Marsh J, Brown W. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr Clin Neurophysiol* 1975;39(5):465–72.
- Sohmer H, Pratt H, Kinart R. Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 1977;42(5):656–64.
- Song JH, Banai K, Russo NM, Kraus N. On the relationship between speech- and nonspeech-evoked auditory brainstem responses. *Audiol Neurootol* 2006;11(4):233–41.
- Tomita M, Norena AJ, Eggermont JJ. Effects of pure tone exposure on voice onset time representation in cat auditory cortex. *Hear Res* 2004;193:39–50.
- Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* 1979;66(5):1364–80.

- 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.
- Winter I. The neural representation of pitch in the auditory brainstem (poster 2aPP3). *J Acoust Soc Am* 2007;121(5):3068.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 2007;10(4):420–2.
- Zeng FG, Nie K, Stickney GS, Kong YY, Vongphoe M, Bhargava A, et al. Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci USA* 2005;102(7):2293–8.

# Proceedings of Meetings on Acoustics

---

Volume 1, 2007

<http://asa.aip.org>

---

**153rd Meeting**  
**Acoustical Society of America**  
Salt Lake City, Utah  
4 - 8 June 2007

**Session 2pPP: Psychological and Physiological Acoustics**

---

## **2pPP4. Temporal, Intensity and Pitch Relationship on Speech Auditory Brainstem Response to Phonemes /ba/ in Normal Hearing Adults.**

**Idrick Akhoun, Barbara Tillmann, Annie Moulin, Stéphane Gallégo, Evelyne Veuillet, Christian Berger-Vachon, Lionel Collet and Hung Thai-Van**

The temporal acoustic pattern of a short speech signal is encoded very accurately in the brainstem. This compound remote response can be recorded on the scalp under the form of speech elicited Auditory Brainstem Response (Speech ABR). Speech ABR is described as the succession of a transient Onset Response (OR) and a sustained Frequency Following Response (FFR). It has been shown that the FFR is the compound phase-locking activity of brainstem cells, and that FFR mimics the periodicity of phonemes (Russo et al., 2004). In Experiment 1 reported here, we confirmed this observation (23 subjects). Experiment 2 showed that Speech ABR to /ba/ components, had different latency-intensity functions (around 15ms for FFR and 7 ms for OR), assuming that these two responses were the result of separate processing streams (11 subjects). In Experiment 3, three stimuli with steady fundamental frequencies (F0) at 133, 375 and 575 Hz vowel were used (12 subjects). Speech ABR showed a reproducible onset response (OR) around 6 ms after stimulus onset, whereas FFR seemed to vanish for high F0. This lack in synchronization gives further evidence that upper brainstem may be the main generator for Speech ABR.

---

Published by the Acoustical Society of America through the American Institute of Physics

**Idrick Akhoun, Barbara Tillmann, Annie Moulin, Stéphane Gallégo, Evelyne Veuillet,  
Christian Berger-Vachon, Lionel Collet and Hung Thai-Van**

*Université Claude Bernard Lyon I, Université de Lyon, Laboratoire Neurosciences sensorielles, comportement, cognition.  
CNRS UMR 5020  
Hôpital Edouard Herriot - Pavillon U - 5, pl. d'Arsonval. F-69003 Lyon (France)  
akhoun@laposte.net*

## 1. Introduction

Numerous researchers have stressed the potentially active role of the brainstem in pitch processing. Given the similar frequency range of fundamental frequency and temporal synchronization of brainstem units, one may expect a significant input of the brainstem in speech processing. This issue was investigated here in humans.

Pitch sensation has been shown to be mainly related to the fundamental frequency (F0) in complex sounds such as speech or animal vocalisations (Moore, 2004 for a review). Both formant and fundamental (F0) acoustic information are necessary for speech intelligibility in general, but F0 more specifically conveys information relevant to prosody and gender (Giraud et al., 2000). In addition, pitch is temporally coded at the cell-scale in brainstem nuclei thanks to phase-locking phenomena: neural discharge of a group of cells is synchronized to the periodic part of an acoustic stimulation. The upper boundary of phase-locking has been measured around 3000-5000 Hz for auditory nerve fibers (AN), around 1000 Hz for the cochlear nucleus (CN), whereas in the cat inferior colliculus (IC), the rate modulation transfer function follows amplitude modulation (AM) up to 300 Hz (Eggermont, 2001 for a review).

Scalp recorded auditory brainstem response to speech (Speech ABR) has been described as an objective measurement of temporal encoding of speech (Faingold & Caspary, 1978; Russo et al., 2004; Johnson et al., 2005). Speech ABR to a short synthetic consonant-vowel (CV) syllable is made of two successive components: the Onset Response (OR), which is a transient component consecutive to the stimulus onset burst, and the FFR which arises after the OR and mimics the periodicity of the signal, reflecting underlying phase-locking.

In the present paper, we summarize three experiments investigating several aspects of the speech signal influencing speech ABR. In Experiment 1, we have shown that Speech ABR mimics accurately the envelope of the stimulus, up to 500 Hz (Akhoun et al., 2008). In Experiment 2, we point out that the component had different latency-intensity functions (around 15 ms for FFR and 7 ms for OR) with different decreasing rate, suggesting that they are the result of separate processes. In Experiment 3, we analyzed speech ABR for the consonant vowel syllable /ba/ presented at three pitch heights 133, 375 and 575 Hz.

## 2. Material & Methods

Evoked potentials were recorded from male and female adults (23 for Experiment 1 including the 11 subjects of Experiment 2, and 12 for Experiment 3) in accordance to 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). All had normal hearing thresholds below 20dB HL for octaves from 250 to 8000 Hz.

Speech ABR measurements were performed using a Centor USB (Racia-Alvar<sup>®</sup>) ABR recording system, using 3 twined silver-electrodes placed between vertex (+) and right mastoid (-), with forehead as ground (Galbraith *et al.*, 2001). Electrode impedances were maintained below 5 k $\Omega$ . A thousand and five hundred sweeps were averaged, with an 80 ms analysis window (sampling frequency = 50 kHz). Speech ABRs were filtered online by a 80-3200 Hz bandpass Finite Impulse Response filter (order 2). Each epoch with amplitude greater than 75  $\mu$ V was discarded. For each polarity (positive, negative), 1500 epochs were averaged. By default, all data are presented in alternate polarity, which is the algebraic sum of blocs for both polarities, i.e. the average of a total of 3000 epochs. Residual artifact and cochlear microphonic are supposed to be cancelled out by this technique, as the auditory system encodes both polarities alike.

Three experiments were conducted, involving different stimulations. In Experiment 1, the stimulus was a consonant vowel syllable /ba/ (Tomita *et al.*, 2004), shortened to duration of 50 ms, and composed of a release burst and a vowel. The fundamental frequency was steady at 120 Hz over the whole vowel duration, whereas the first and second formants were around 700 and 1200 Hz. The /ba/ stimulus was played at 60 dB SL. In Experiment 2, the /ba/ syllable of experiment 1

was presented at 7 different intensities, varying from hearing threshold (0 dB SL) to 60 dB SL by 10 dB steps. Lastly, in Experiment 3, three different /ba/ stimuli were used: the /ba/ syllable of Experiment 1 at 120 Hz and two different versions of it, changed in pitch to 375 and 575 Hz. These two higher-pitch syllables were obtained by shifting the first /ba/ frequency components using Straight® software (Kawahara et al., 1999).

In all three experiments, subjects were stimulated monaurally in their right ear. The phonemes were 60 ms long, with 10 ms of silence before the beginning of the sound in order to assess the ambient noise level in the first 10 ms of the recordings. The inter-stimulus interval (ISI) was 90 ms (11 syllables per second). The sound file was played by BioLogic® insert-earphones: the stimulation through the 30 cm-rubber tubes induced a 1 millisecond delay (air transmission) which was subtracted from the Speech ABR recordings.

**3. Results and discussion**

Our present paper brings together three experiments on FFR, investigating basic FFR pattern, the influence of intensity and pitch. Experiments 1 and 2 pointed out that the FFR temporal pattern mimicked the periodicity of the phoneme /ba/, for any polarity, and robustly against stimulation intensity. We further observed differences in OR and FFR latencies, as a function of stimulus intensity (Experiment 2). The key result (Experiment 3) was that pitch had a significant effect on OR latencies as well as a vanishing effect on FFR waves. However, the averaging computation method may prevent from analysing high frequency information by artificially cancelling it out over 500 Hz (notably due to temporal jitter in the responses).

*3.1 Experiment 1: The temporal relationship between Speech ABR and the acoustic pattern of a syllable /ba/ (F<sub>0</sub> = 120 Hz) at 60 dB SL.*

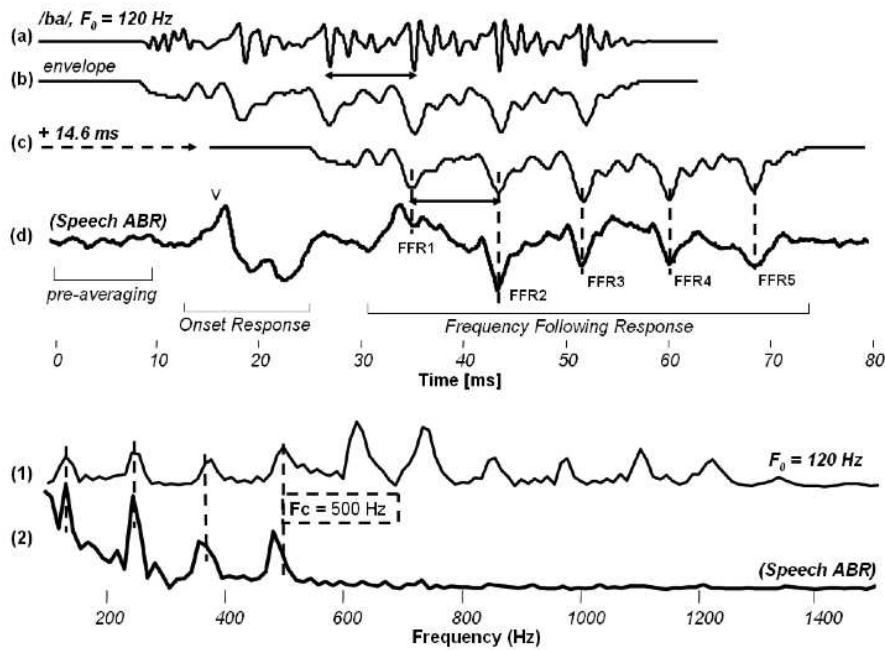


Fig. 1. Temporal (a - d) and spectral (1, 2) comparison of Speech ABR and the acoustic /ba/ stimulus (Experiment 1). The temporal /ba/ syllable waveform used in Experiment 1 is plotted in (a). The /ba/ syllable envelope (b) was extracted by full wave rectification and filtering at F<sub>c</sub> (500 Hz). Interestingly, when inverted and shifted by a 14.6ms (c), the envelope is accurately represented in FFR (d). The good-fit of fundamental periodicity (7.5 ms) between both acoustic stimulus and Speech ABR is underlined by the double arrows. Both Speech ABR (2) and /ba/ spectra (1) show common minima and maxima linked by dotted vertical lines up to a cut-off frequency (F<sub>c</sub>) of 500 Hz. Note that the poor frequency precision of Speech ABR FFT (12.5 Hz) artificially shifted the 2 last maxima of Speech ABR spectrum to the left.



The Speech ABR was synchronized to the /ba/ stimulus up to a cut-off-frequency ( $F_c$ ) which was visually measured at 500 Hz on the grand average Speech ABR spectrum (Fig.1). We used this  $F_c$  to calculate the inverted envelope of the phoneme /ba/. When shifted in time by 14.6 ms, the envelope of the stimulus was found to have a very similar shape to the recorded FFR (Fig.1).

Latencies and FFR  $F_c$  indicate, accordingly to physiological investigations in animals, that the FFR may be the compound phase-locking activity of brainstem cells between CN and IC, such as *octopus* or *stellate cells* (also respectively named pure onset or onset chopper units) in ventral CN (Rhode & Greenberg, 1994, Winter, 2007), in lateral lemniscus, in IC (*sustained cells*) or even in medial geniculate body (*followers units*), (Eggermont, 2001, Palmer & Shamma, 2004, Joris et al., 2005 for reviews). Similarly, OR seems to result from a massive synchronized firing in upper brainstem (lateral lemniscus and IC) cells, or from an enhancement of dynamic discharge pattern in multipolar postero-ventral CN cells, which are sensitive to formant transitions (Miller & Sachs, 1983). However, the task of identifying the possible Speech ABR generators is much compromised, given the lack of access to unitary neurophysiological recordings in human, except during brainstem implantation.

3.2 Experiment 2: The effect of stimulation intensity of a /ba/ ( $F_0 = 120\text{Hz}$ ) on Speech ABR.

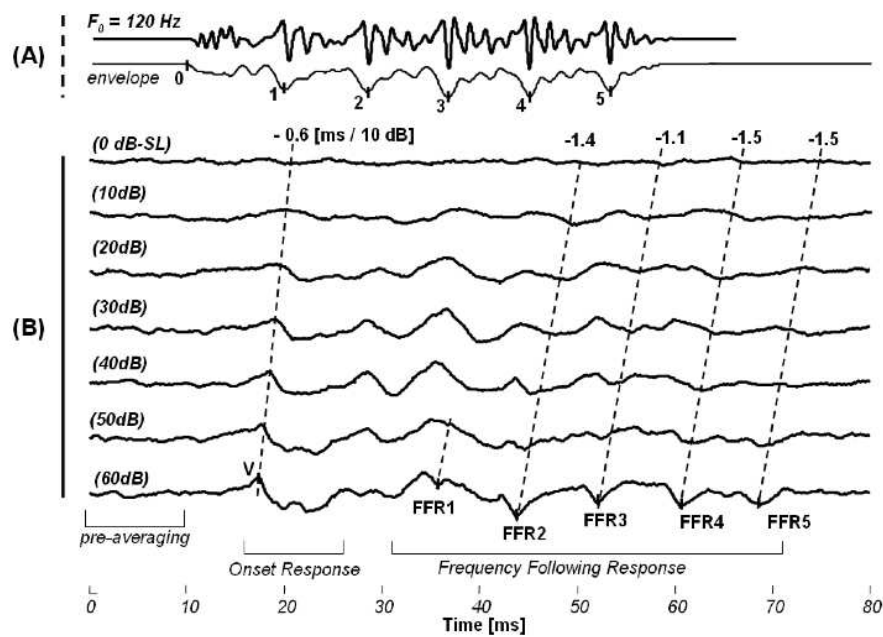


Fig. 2. Grand-averaged Speech ABR ( $n = 11$ ) for stimulation intensity varying between 0 and 60 dB SL (Experiment 2). Both /ba/ and /ba/ envelope waveforms are plotted in (A) and they are similar to (a) and (b) in Fig.1. In (B) are the 7 grand averaged Speech ABR, for each stimulation-intensity tested. Note that the latency of each component (wave V, FFR 1, FFR 2, FFR 3, FFR 4) clearly increases as intensity decreases, as suggested by dotted lines. The latency/intensity function for each Speech ABR peak is well-fitted by a linear regression for these components. Steeper slopes are obtained for FFR components than for OR (figures above, in ms/10 dB).

Corrected latencies of the speech ABR components were calculated as the difference between the measured latency and the time of the corresponding peak in the envelope of the /ba/ stimulus (e.g. FFR 2 is supposed to be elicited by peak 2 in the envelope of the acoustic stimulus - Fig.2). The measured latencies decreased as sound level increased, as expected considering classic click ABR results. But interestingly, these corrected latencies showed 2 distinct groups in the latency-intensity function plot, with significantly different latencies as a function of sound level (Repeated Measured Analysis of Variance - RM-ANOVA  $F(1,54) = 614; p < 0.001$ ): the OR lagged between 6.4 and 9.2 ms from the stimulus onset, whereas the FFR peak latencies were grouped between 14.6 ms and 20.6 ms after the cue in the stimulus envelope, which elicited the FFR peak,

respectively at 60 and at 30dB SL. In addition, the rate of latency decrease with increasing stimulus intensities was significantly different (RM-ANOVA  $p < 0.001$ ,  $F(4,40) = 5.9$ ) for the latency of wave V (-0.64 ms for 10 dB-steps of intensity increases) and for the FFR latencies (-1.4 ms / 10 dB).

The 2 different groups in the latency-intensity functions plot, coupled with a faster decrease of latencies for FFR than for OR, suggest the existence of 2 separate streams within the same pathway. Indeed, extensive substrates have been described in neurophysiological animal studies to process intra or reverse loops in brainstem nuclei when stimulated with complex stimuli such as amplitude modulations, starting from CN (Rhode & Greenberg, 1994). As a consequence, the FFR may be related to the vowel (language specific acoustic stimulus), while the OR, also recorded in click ABR (Picton et al., 1974), seems to be more related to a basic detection of stimulus onset, or to formant transition (Miller & Sachs, 1983). Hence, this outcome suggests the possibility for these 2 neural streams to be related to the origin of cortical what and where separations (Kraus & Nicol, 2005; Shannon, 2002).

3.3 Experiment 3: The effect of the pitch of /ba/ stimuli (at 60 dB SL) on Speech ABR.

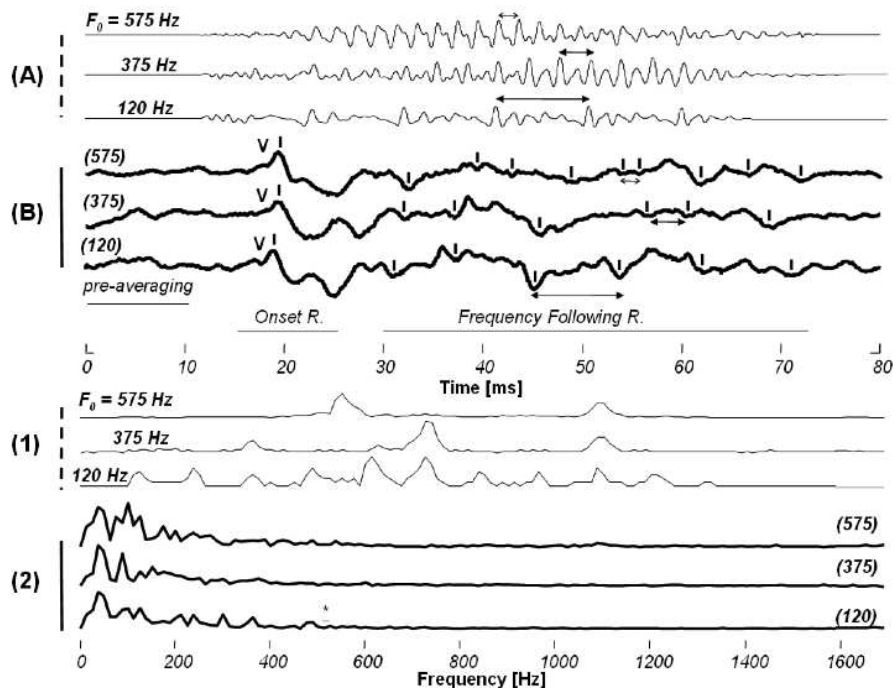


Fig. 3. The effect of syllable pitch on grand-averaged (n=12) Speech ABR (Experiment 3). The 3 stimuli used in this experiment are plotted as a function of time (A), and the 3 F0 are shown. The fundamental period (T0) is indicated by the double arrows. Temporal grand average Speech ABR (B) is plotted with the corresponding stimulus F0 notified. The salient peaks of OR and FFR is notified by markers (I), as well as the double arrows to underline the synchronization of FFR on the corresponding stimulus. Similarly, the spectra of the 3 stimuli are plotted (1), with the spectra of the elicited Speech ABR (2). The corresponding F0 are also notified. Note that the recordings are reproducible: the Speech ABR components obtained here for condition /ba/ (F0 = 120Hz) are very similar to those of Experiment 1, where the same material was used.

Grand-average data suggest that the OR and, to a lesser extent, the last wave of the FFR (offset response), remained visible for each of the stimulus pitch (for /ba/ at 120, 375 or 575 Hz). Nevertheless, pitch had a significant effect on ORs latencies (RM-ANOVA,  $F(2,33) = 5.11$ ,  $p < 0.05$ ). The OR latencies were shorter at 120 Hz (the OR lagged behind the stimulus around a mean latency (m) of 7.4 ms, standard error (sd) = 0.15) compared to 375 (m = 8.0 ms, sd = 0.16) and 575 Hz (m = 8.0, sd = 0.15). For these higher frequency elicited OR, latencies were quite similar (RM-ANOVA,  $F(1,22) = 0.005$ ,  $p = 0.946$ ). Contrary to the well-defined OR, grand averaged FFR waves obviously vanished for higher frequencies (Fig. 3). However, individual Speech ABR clearly showed that FFR does exist at higher frequencies for some subjects, even if more discrepancies and jitter occur. For example, for subject 2 (Fig. 4), the FFR does exist with an Fc as high as 1100 Hz, even if the good fit with stimulus pattern does not last during the entire

vowel duration, as Experiment 1 (or the current 120 Hz condition) would have suggested (Fig. 1 and 3).

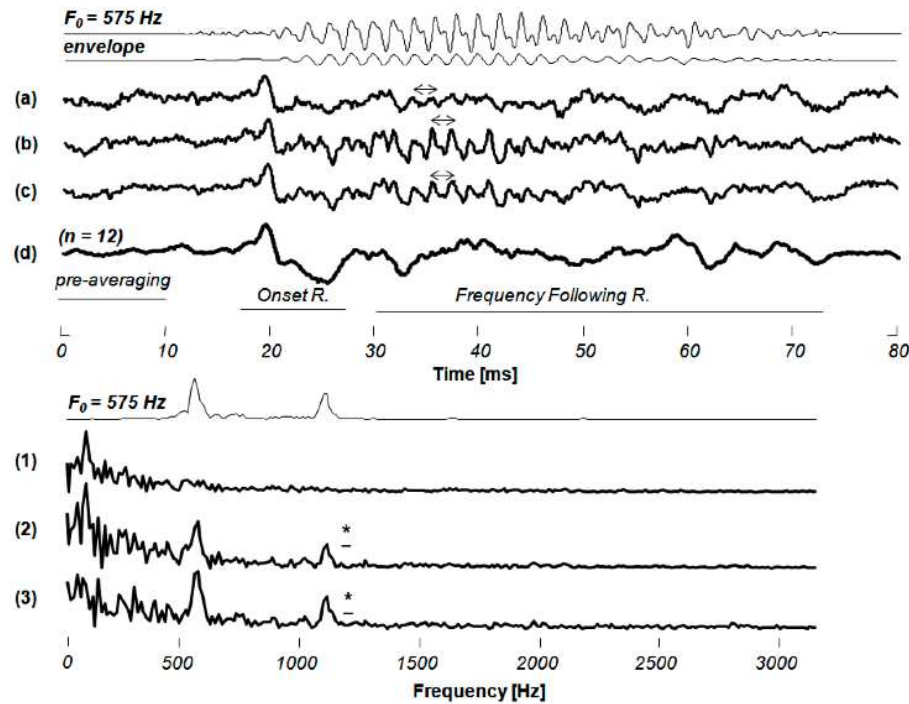


Fig. 4. Individual Speech ABR at 575 Hz. Some subjects showed consistent FFR at 575 Hz: Example of individual Speech ABR for subject 2 (60 dB SL). The 3 bold waveforms illustrate individual temporal (a-c) and spectral (1-3) Speech ABR in positive (a, 1), negative (b, 2) and alternate ( $c = a+b$ , 3) polarity. Contrary to grand-average, individual FFR is well defined (at least on (b) and (c)), even if it does not last during the entire vowel duration. The Speech ABR  $F_c$  are specified by stars (\*). Grand-averaged Speech ABR ( $n=12$ ), already presented on Fig.3 is represented (d) for comparison purpose. The synchronization of FFR on the periodicity of the stimulus appears far more accurate on subject 2 Speech ABR, than it is suggested on grand-averaged data.

These findings suggest that pitch variation has diverse effects on human scalp recorded Speech ABR. Indeed, pitch has a subject-dependant vanishing effect on the sustained FFR, and higher pitches result in longer transient OR latencies. However, recording artefact may also play a role in this pitch effect. Accordingly, the FFR at high frequencies might be, attenuated by the averaging method, which cancels out the high frequencies. This could explain, at least to a certain extent, why the FFR is only partly synchronized to the vowel. Furthermore, both /ba/ stimuli at 375 and 575 Hz are above the limit observed in real speech, and thus rather non-ecological speech stimuli, even though this pitch range is highly used in music.

The observed pitch influence over FFR is in agreement with the work of Wong *et al.* (2007) and Musacchia *et al.* (2007): Speech ABR to Chinese tones (for English speaking subjects) followed the fluctuations of the periodicity of the stimulus far more accurately in musicians than in non-musicians. This was suggested to reflect the greater ability of musicians to detect prosody in a foreign tonal language than non-musicians. In addition, phase-locking physiological animal recordings suggest that, for high  $F_0$  stimuli such as 575 Hz (and even at 375 Hz), synchronisation is not possible anymore in IC, but is still possible below in CN (Eggermont, 2001, Palmer & Shamma, 2004, Joris *et al.*, 2005 for reviews). This may bring further information on the physiological underlying contributors of speech ABR, and more generally on the important temporal processing steps of pitch along the auditory pathway, at the level of the brainstem.

#### Acknowledgement

[This work was presented at the 153 rd. meeting of the Acoustical Society of America (Salt Lake City, Utah): Akhoun *et al.*, poster 2 p PP-4].

The authors would like to thank the subjects who participated in this study, and the involved colleagues: Arnaud Jeanvoine, Mike Ménard, Céline Richard, Clémence Pavillon, Fabien Seldran, Christian Vollaire, François Buret, Etienne Gaudrain, Samuel Garcia, A & K. Nyamay'antu. Best regards to N.Kraus and her team, A.Norena, J.Eggermont, S.Greenberg and R.V. Shannon for fruitful discussions. Frederick Gallun provided helpful comments on this work.

The first author would like to express his gratitude to the SFA, Cochlear France, the ASA and his lab in Lyon for their financial support to the journey in Utah. The first author is sponsored by a CIFRE grant from *Cochlear France & Univ.Lyon 1*.

## References and links

- Akhoun, I., Gallégo, S., Moulin, A., Veuillet, E., Berger-Vachon, C., Collet, L., and Thai-Van, H. (2008) "The temporal relationship between Speech auditory brainstem responses and the acoustic pattern of the phoneme /ba'". *Clin Neurophysiol* [in press]
- Eggermont, J.J. (2001) "Between sound and perception: reviewing the search for a neural code." *Hear Res.* 157(1-2), 1-42.
- Faingold, C.L., and Caspary, D.M. (1979). "Frequency-following responses in primary auditory and reticular formation structures." *Electroenceph clin Neurophysiol.* 47(1):12-20.
- Galbraith, G., Bagasan, B., and Sulahian, J. (2001). "Brainstem frequency-following response recorded from one vertical and three horizontal electrode derivations." *Percept Mot Skills.* 92(1), 99-106.
- Giraud A.L., Lorenzi C., Ashburner J., Wable J., Johnsrude I., Frackowiak R. , and Kleinschmidt A. (2000). "Representation of the temporal envelope of sounds in the human brain." *J Neurophysiol.* 84(3), 1588-98.
- Johnson K.L., Nicol T.G., and Kraus N. (2005) "Brainstem response to speech: a biological marker of auditory processing." *Ear Hear.* 26(5): 424-34.
- Joris P.X., Schreiner C.E., and Rees A. (2004) "Neural processing of amplitude modulated sounds", *Physiol Rev.* 84(2):541-77.
- Kawahara, H., Katayose, H., de Cheveigne, A., and Patterson, R.D. (1999). "Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity", *Proc. EUROSPEECH'99* 6, 2781-2784.  
[http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index\\_e.html](http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_e.html)
- Kraus, N., and Nicol, T. (2005). "Brainstem origins for cortical what and where pathways in the auditory system." *Trends Neurosci.* 28(4), 176-81.
- Krishna, B.S., and Semple, M.N. (2000). "Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus." *J Neurophysiol.* 84(1), 255-73.
- Krishnan, A., Xu, Y., Gandour, J.T., and Cariani, P.A. (2004). "Human frequency-following response: representation of pitch contours in Chinese tones." *Hear Res.* 189(1-2), 1-12.
- Miller M.I., and Sachs M.B. (1983). "Representation of stop consonants in the discharge pattern of auditory nerve fibers". *J Acoust Soc Am.* 74(2): 502:17
- Moore, B.C. (2004). "An Introduction to the Psychology of Hearing", 5th ed. (Elsevier Academic Press), pp 206.
- Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007) "Musicians have enhanced subcortical auditory and audiovisual processing of speech and music." *Proc Natl Acad Sci U S A.* 104(40):15894-8. Epub
- Palmer A.R., and Shamma S.A. (2004). "Physiological representations of speech", in *Springer Handbook of Auditory Research, (vol.18) Speech processing in the auditory system.* edited by Greenberg, S., Ainsworth, W.A., Popper, A.N. and Fay, R.R. (Springer New-York).
- Picton, T., Hillyard, S., Krausz, H., and Galambos, R. (1974). "Human auditory evoked potentials. I. Evaluation of components." *Electroenceph clin Neurophysiol.* 36(2), 179-90.
- Purcell, D.W., John, S.M., Schneider, and B.A. Picton, T.W. (2004). "Human temporal auditory acuity as assessed by envelope following responses." *J Acoust Soc Am.* 116(6), 3581-93.
- Rhode, W.S., and Greenberg, S.R. (1994) "Encoding of amplitude modulation in the cochlear nucleus of the cat." *J. Neurophysiol.* 71, 1797-1825.
- Russo, N.M., Nicol, T.G., Musacchia, G., and Kraus, N. (2004). "Brainstem responses to speech syllables", *Clin Neurophysiol.* 115(9), 2021-30.
- Shannon, R.V. (2002). "The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design." *Am J Audiol.* 11(2),124-7.
- Tomita, M., Norena, A.J., and Eggermont, J.J. (2004). "Effects of pure tone exposure on voice onset time representation in cat auditory cortex." *Hear Res.* 193, 39-50.
- Winter, I. (2007). "The neural representation of pitch in the auditory brainstem" (2aPP3). *J Acoust Soc Am.* 121(5), 3068
- Wong, P.C., Skoe E., Russo, N.M., Dees, T., and Kraus, N. (2007). "Musical experience shapes human brainstem encoding of linguistic pitch patterns." *Nat Neurosci.* 10(4), 420-2.

# Discussion générale

---

Les Speech ABR sont une fenêtre d'observation électrophysiologique sur le traitement auditif de la parole, chez l'humain éveillé. Il se présente comme une succession de réponses transitoires et pseudo-périodiques fortement reliée aux caractéristiques acoustiques du stimulus. Ce travail de thèse avait pour but d'évaluer les opportunités de cette technique en vue d'obtenir un outil d'appoint à l'audiométrie vocale, outil fondamental dans le diagnostic de la surdité et l'évaluation de la réhabilitation auditive. Notre démarche a été conditionnée par l'aspect exploratoire de l'utilisation des Speech ABR, dont la description dans la littérature reste éparse, et qui n'avait jamais été implémentée au laboratoire (et plus largement partout hormis aux Etats-Unis).

Ainsi, nous avons dans un *premier temps* déterminé la seule condition expérimentale sans-artéfact (insert-earphones entièrement blindés dans une cage de Faraday), et à partir de là, nous avons apporté la preuve de l'**authenticité neurophysiologique du Speech ABR**. Jusqu'à présent, chaque équipe avait mis au point sa propre technique pour se prémunir des artéfacts, sans la décrire en détail, ni la pousser à une étude paramétrique (toutefois à approfondir), un peu à la façon des pionniers ('numerous checks'...) Fig. 1.13. D'autre part, le rapide coup d'oeil dans le monde de la surdité (avec les sourds unilatéraux chap.2.3) a donné des Speech ABR différents de ceux des normo-entendants (à audiogramme tonal équivalent), et suggère un fonctionnement distinct de leur génération dû à la surdité.

Ensuite, dans un *second temps*, et une explorations large-champ des **influences des paramètres acoustiques du stimulus sur les ABR**, nous avons montré la non-réciprocité du Speech ABR : le Speech ABR imite l'enveloppe temporelle du stimulus, mais à l'inverse, deux stimuli avec la même enveloppe temporelle ne donnent pas nécessairement le même Speech ABR. Nous avons également montré que le niveau sonore, tout comme la fréquence, avaient une influence sur le Speech ABR ; par contre, la nature du stimulus ne semble pas influencer. Enfin, en annexe C, nous examinons les perspectives de la représentation temps-fréquence pour contrer l'effet-jitter, conséquence de l'élimination des composantes hautes-fréquence du fait du moyennage.

Le *résultat-clé* de cette seconde partie est que **la structure fine temporelle a plus d'impact que l'enveloppe** du stimulus dans la création de la FFR. Ceci s'est illustré de deux façons : (i) certains analogues donnent une FFR (ceux spec-



tralement riches, et harmoniques), d'autres non (ceux appauvris spectralement), ou encore, pris sous un autre angle, (ii) nous avons obtenu de très bonnes FFR pour des sons complexes harmoniques, mais pas pour chacune des harmoniques prises indépendamment, et ceci, quelles qu'en soient les fréquences. Nous spéculons alors que le Speech ABR est à la fois le reflet de la complémentarité entre le *code neural temporel* (phase-locking sur l'enveloppe du stimulus), et le *code spatial* (utilisation de l'information provenant des différents filtres auditifs, en particulier dans les hautes-fréquences au delà de 500 Hz). Ce résultat peut-être rapproché, d'un côté aux observations que les informations de l'enveloppe temporelle sont disponibles par battement des harmoniques non-résolus dans les filtres auditifs centrées sur les hautes-fréquences [Gilbert 2007], et de l'autre, que la sortie des filtres auditifs en sortie de la base de la cochlée (où les hautes-fréquences sont codées) donnent lieu à une activité neurale mieux synchronisée qu'à l'apex, en raison de l'absence du délai de transmission de la vibration cochléaire [Dau 2003].

Les principales questions que nous nous sommes posés tout au long de l'implémentation de l'étude des Speech ABR sont reprises ci-dessous.

### **Les conditions de recueil peuvent-elles influencer la forme du Speech ABR ?**

Dans les *conditions de recueil*, on peut soit considérer les conditions physiques, liées au dispositif expérimental et à l'état musculaire du sujet, ou soit considérer les conditions plus intrinsèques liées à l'état d'attention du sujet.

Pour ce qui est des *conditions physiques*, la réponse est très fortement affirmative, comme l'atteste notre étude instrumentale paramétrique 2.2. Au niveau du dispositif expérimental, si toutes les précautions ne sont pas satisfaites, il y aura des forts risques d'artéfact ou de rapport signal sur bruit très défavorable. D'autre part, si le sujet a tendance à être très tendu musculairement (sans forcément en avoir conscience), le Speech ABR peut être très fortement masqué par l'activité des spikes musculaires, comme pour toutes les méthodes de recueil de potentiels électrophysiologiques.

Pour ce qui est des conditions liées à l'*attention* que le sujet porte sur le stimulus, les auteurs dans la littérature sont partagés. Il semblerait qu'étant donné l'origine neurophysiologique *bas-niveau* (tronc cérébral) du Speech ABR, et sachant que les mécanismes attentionnels sont plutôt attribués aux aires auditives plus intégrées (cortex), on aurait tendance à penser que l'attention n'a pas grande influence sur le Speech ABR. C'est également ce que nous pensons, étant donné le risque d'artéfact des données de Galbraith, et puisque les Speech ABR obtenus à l'état de sommeil ou éveillé étaient très similaires.

L'équipe de Picton, à propos des ASSR, pense qu'il existe une fréquence modulante limite de 40 Hz en dessous de laquelle les ASSR sont générées à un niveau cortical vu qu'il s'agit de la limite supérieure de phase-locking pour les neurones corticaux ; par conséquent pour ces basses fréquences, l'état d'attention influe significativement



[Ross 2004]. A l'inverse, ils pensent que pour les plus hautes fréquences modulantes, le sujet peut très bien somnoler sans perturber le recueil. Pour les équipes de Galbraith, Krishnan [Krishnan 2005] ou de Kraus, l'attention sélective a un effet sur l'amplitude des Speech ABR [Galbraith 1994], tout comme l'apprentissage à court terme [Song 2008]. Habituellement, pour maintenir un niveau d'attention fixe, un film est diffusé avec un niveau sonore faible. Craignant les sources de courant parasites, nous n'avons pas donné de consignes aux sujets, dont la plupart ont somnolé pendant les expérimentations sans pour autant que cela semble influencer sur la qualité de nos recueils. D'autre part, étant donné que nous avons appliqué la loi centrale normée aux Speech ABR afin d'harmoniser l'influence de chaque sujet sur la grande moyenne, nous n'avons jamais étudié l'amplitude des Speech ABR.

### **Le Speech ABR est-il sensible aux caractéristiques acoustiques du stimulus ? Y a-t-il une quelconque linéarité du Speech ABR ?**

L'exemple des sons purs et des sons complexes harmoniques de l'étude 3.2 dément clairement toute linéarité du Speech ABR : la réponse aux sons complexes harmoniques n'est en rien la somme des FFR aux harmoniques pris individuellement.

Les caractéristiques acoustiques du stimulus semblent avoir une influence capitale sur la synchronisation de la FFR, les hautes fréquences en particulier comme le montrent notre étude 3.2 et [Dau 2003], le cas analogue très-mauvais dans l'étude 3.2.2 ainsi que les effets des bruits en cours au laboratoire. Il semblerait donc que les sollicitations ayant lieu au sein de différentes bandes critiques se rejoignent (probablement dans les noyaux du tronc cérébral étant supposés générer le Speech ABR) et donnent lieu à cette activité qui est ensuite recueillie comme FFR. Ceci pourrait de même expliquer l'absence de FFR pour le son pur, même si cette observation contraste avec les belles FFR recueillies dans la littérature et au sein de l'équipe en réponse à des sons purs... avant que nous ne disposions de conditions assermentées sans-artefact. (Même si les latences rapportées plaident en leur faveur.)

La comparaison des deux types d'analogues, entre (i) ceux des études 3.2 & 3.2.2, et (ii) ceux de l'étude 3.4 est révélatrice, car la seconde a donné lieu à une FFR très bien définie, contrairement à la première. Le premier type d'analogue (i) était une somme de 5 sinus modulés par l'enveloppe complexe du son de parole. Le second type d'analogue (ii) avait par contre une structure harmonique complexe entière, mais aucune amplification locale de l'enveloppe spectrale n'a souligné la présence de formants (ceci étant la seule différence entre les stimuli /ka/ et leurs analogues). Cette observation, qu'un son avec une structure acoustique harmonique riche donne une FFR alors qu'une structure fine sporadique comme les 5 sinus n'en donne pas, appuie encore l'hypothèse que la FFR est issue d'un mécanisme de détection de coïncidences entre les différentes bandes critiques. Une investigation plus précise sur les effets respectifs des harmoniques résolus et non-résolus pourraient

permettre de préciser si la résolution d'harmoniques ou pas (et dans ce cas, des battements apparaissent à la fréquence fondamentale en sortie des filtres auditifs... comme dans la FFR) est importante dans la FFR. A l'heure actuelle, seules les études de [Greenberg 1987] et [Dau 2003] apportent quelques renseignements (imprécis) à ce sujet soit à cause du manque de sujets, soit à cause des imprécisions de modélisation. Les expériences de [Krishnan 2008], où il est montré que la FFR est sensible à la dynamique de  $F_0$  sont également à considérer.

Enfin, le cas particulier de l'analogue très-mauvais dans l'étude 3.2.2 semblerait mettre en doute notre idée que la FFR, quand elle existe, est synchronisée sur  $F_0$  (ici 200Hz), car dans ce cas, la synchronisation a eu lieu sur 400 Hz, qui pourrait être déduite des harmoniques présents (200 - 800 - 1200 - 2400 - 3200), qui correspondent aux harmoniques ( $\frac{1}{2}$ -2-3-8-11). Cette donnée isolée devrait être approfondie.

### **L'Onset Response est-elle une mesure objective du Voice Onset Time ?**

L'Onset Response est un proche parent des clic-ABR couramment utilisés en routine clinique dans la mesure des seuils auditifs. L'Onset Response semble bien être le résultat d'une réponse transitoire lié à une transition acoustique qui s'apparente à un clic (telle que le passage entre l'équivalent de la consonne à celui de la voyelle dans le cas de l'analogue /ba/). L'Onset Response a été décrite comme une potentielle mesure objective du VOT, qui est un indice acoustique primordial dans la perception de la parole pour discriminer par exemple le /ba/ (VOT = -55 ms) du /pa/ (aux alentours de 20 ms pour la langue Française). En revanche, dans cette étude, nous n'avons pas été à même d'observer de différences visibles entre les Onset Response engendrées par le /ba/ ou le /pa/, contrairement aux analogues pour des raisons que nous ne pouvons déterminer. Plusieurs ont émis l'hypothèse que l'Onset Response et la FFR sont issus de deux processus auditifs distincts [Kraus 2005], [Akhoun 2008c]. Ceci trouve un écho dans le fait que nous n'ayons jamais obtenu les deux en même temps, par exemple entre l'analogue /ba/ (Onset Response bien définie, pas de FFR) et le /ba/ (FFR bien définie, pas ou peu d'Onset Response).

### **Le Speech ABR est-il sensible à la nature du stimulus (parole – musique – autre) ?**

Nos expérimentations de l'étude 3.2, utilisant les analogues et les sons de parole, nous ont amenés à nous demander si le Speech ABR possède des composants différents, et en particulier une FFR mieux définie, pour des sons de parole (voire de musique) comparés à des sons non-signifiants.

Un numéro spécial de la revue *Philosophical Transaction of the Royal Society of London. Serie B – Biological Sciences* [Moore 2008] traite de la question de la

---

spécialisation du traitement de la parole. Les conclusions de ces articles semblent converger vers une spécialisation à un niveau cortical et un pré-traitement davantage *mécanique* (plutôt sensible aux caractéristiques acoustiques des stimuli qu'à la nature du son) dans le tronc cérébral. Ils s'appuient pour cela sur des études d'imagerie fonctionnelle.

Toutefois, les récentes études de l'équipe Kraus montrent clairement un traitement pré-cortical significatif (les linguistic-pitches ou les notes de violoncelles pour les musiciens). Les Speech ABR ainsi recueillis sont mieux définis pour les musiciens ([Wong 2007]; [Musacchia 2007]; [Kraus 2007]), et que ce prétraitement dans le tronc cérébral serait même sensible aux apprentissages à long-terme (langue maternelle : [Krishnan 2005]) et à court-terme ([Song 2008]). Ces travaux font donc bien état de mécanismes de traitement de l'information habituellement prêtées au traitement cortical. Les travaux de neurophysiologie multi-unitaire, présentés dans les grandes lignes en introduction chap.1.1, ne permettent pas vraiment de préciser cette question, étant donné qu'elles ont pour la plupart été réalisées sur des animaux anesthésiés. Toutefois, certaines cellules, telles que les cellules multistellate (dans le noyau cochléaire dorsal) sont décrites comme ayant la capacité de tirer profit de mécanismes inhibiteur (Lateral Inhibitory Network : comme ceux en jeu en vision pour détecter les coins) pour pouvoir extraire l'enveloppe spectrale des voyelles [Palmer 2004].

De notre étude 3.1, les latences importantes que nous avons obtenues indiquent la possibilité d'un rétrocontrôle cortical dans le tronc cérébral. Et d'autre part, nous avons observé une très bonne robustesse du Speech ABR en fonction de l'intensité de stimulation. Dès lors, rien n'empêcherait un flux particulier de se produire pour certains types de sons, ce qui irait dans le sens d'une explication des remarquables facultés de l'audition humaine de la parole... De plus, l'étude suivante 3.2.2 illustre une corrélation entre la qualité de perception d'un analogue (plus ou moins ressemblant au son de départ) et la qualité de synchronisation de la FFR. Mais il s'agit de fausses pistes, l'étude 3.4 ne montre pas de meilleure synchronisation de la FFR en réponse à des vrais sons de parole (ou de musique) par rapport aux analogues, et l'étude de l'étude 3.2 a de plus montré une excellente synchronisation de la FFR sur les sons complexes harmoniques (non-langagiers).

Dans leur ensemble, ces données montrent donc une grande variabilité de la sensibilité de la synchronisation de la FFR avec des analogues, qui peut soit être excellente ou inexistante sur des analogues. Nos différentes études ont donc servi à poser les bases de futures investigations sur notamment les effets du contexte (entendu comme étant lié à la nature du stimulus : stimuli de langue tonale ou IRN) sur la synchronisation de la FFR [Krishnan 2008], ou bien l'analyse simultanée des PEA corticaux et des Speech ABR [Galbraith 1994] pour observer de plus près ce supposé mécanisme de rétrocontrôle cortical. En parallèle, un choix méthodique de stimuli permettra de définir le plus précisément possible les caractéristiques

acoustiques qui influent sur le Speech ABR, et d'estimer quel est leur lien avec la perception de la parole (dans l'idée de l'étude 3.2.2).

### **Le Speech ABR est-il sensible à une fréquence de coupure, ou au nombre d'harmoniques ?**

Plutôt la fréquence de coupure, vu les variations de nombre d'harmoniques trouvés par exemple dans le chapitre 3.2. Cependant, dans nos études, nous avons considéré plusieurs types de hauteurs pour nos stimuli (les 3 /ba/ de l'étude préliminaire [Akhoun 2008a] présentée dans le proceeding POMA, puis les 3 /ka/ et les 3 sons de piano de l'étude 3.4, le /ba/ voix de femme et les deux sons complexes harmoniques utilisés dans l'étude 3.2. Dans toutes ces différentes études, le Speech ABR fut tout le temps passe-bas, avec une fréquence de coupure qui pouvait varier entre 500 Hz (estimée par FFT pour le /ba/ 120 Hz, dans l'étude 3.1) et 1200 Hz (en spectrogramme pour le /ba/ 200 Hz, dans l'étude 3.2).

Mais il semblerait qu'un problème de traitement du signal se profile. Est-ce que ces harmoniques ne seraient pas simplement dus à une distorsion de la synchronisation qui fait apparaître artificiellement des harmoniques supérieurs, comme c'est très bien visible dans les enregistrements d'artefact 2.2, Fig.2.5 ? Devrait-on donc sommer l'énergie de chaque composant et ne considérer que la synchronisation ainsi *sommée* sur la fréquence fondamentale ?

### **Est-ce que le Speech ABR est bien généré au niveau du tronc cérébral ? Y a-t-il un effet du contexte (linguistique) dans la présentation des stimuli ?**

Les latences observées pour le Speech ABR, ainsi que ses fréquences de coupure pointent vers des générateurs subcorticaux. Mais il existe probablement dans cette réponse une part corticale, qu'il serait très intéressant de mettre en évidence.

Grâce à la batterie de stimuli que nous avons utilisés, notamment grâce aux analogues, nous avons pu croiser nos données et celles de la littérature animale faite d'études in situ, afin de cibler les générateurs neurophysiologiques de la FFR. Ceux-ci correspondraient aux détecteurs de coïncidence du colliculus inférieur. Cette précision est importante, car les Speech ABR peuvent alors permettre un aperçu du bon-fonctionnement bien ciblé de cette partie du colliculus inférieur. Ceci serait en particulier très utile pour parfaire le diagnostic des surdités centrales ou des neuropathies auditives [Kraus 2000]. Ce centre nerveux est d'une importance capitale dans le traitement de l'information auditive, comme l'atteste la grande quantité d'afférences binaurales qui convergent dans le colliculus inférieur.

En reprenant la modélisation par AIM en Annexe (B) de l'activité neurale

---

périphérique, la quantité d'information neurale qui transite dans chaque bande de fréquence semble n'être pas distribuée de la même façon pour la voyelle du vrai son de parole par rapport à l'analogue : pour le vrai son de parole, chaque bande de fréquence possède des maxima d'activité neurale pour des écarts entre deux pics d'activité (abscisses) progressif, tandis que ces écarts sont les mêmes sur la plupart des bandes de fréquences dans le cas de l'analogue. De façon imagée, on peut dire que l'activité neurale en *crosses de hockey* telles qu'on peut les observer pour le vrai son de parole est plus à même de donner lieu à une forte synchronisation de la FFR que l'activité neurale avec des écarts fixes entre les pics d'activité des différents filtres auditifs. Les générateurs de la FFR sembleraient jouer le rôle d'intégrateurs de cette activité périphérique (les afférences doivent arriver de façon régulière plutôt que tout d'un coup), et qu'ils agiraient de préférence comme les détecteurs de coïncidences (nécessité d'avoir des afférences de différents filtres auditifs). Anatomiquement, ces détecteurs de coïncidences, situés dans le colliculus inférieur central, sont appelés cellules-disque (*disc-shape cells*), qui sont fonctionnellement qualifiées de cellules soutenues (*sustained cells*) [Rees 1997], [Voutsas 2005].

Nous avons également cherché à déterminer l'origine des Speech ABR en les enregistrant selon 2 voies de recueil bien choisies : sur la voie 1, entre le vertex et le nez, et sur la voie 2, entre la mastoïde et le nez (référence commune), avec le front comme masse commune. Le rationnel était que si le potentiel lié au Speech ABR était plutôt cortical, il atteindrait le vertex d'abord, le nez ensuite, puis la mastoïde ; ce qui entraînerait une opposition de phase entre les deux voies. Dans le cas où, au contraire, le Speech ABR serait davantage créé dans le tronc cérébral, il atteindrait approximativement le nez, le vertex et la mastoïde au même moment ; ce qui se traduirait par un enregistrement très comparable sur les 2 voies. Cependant, cette approche simple et très intéressante n'a pas été fructueuse, car les signaux de la voie 2 se sont avérés inexploitablement, car trop bruités.

L'effet du contexte de présentation n'est pas évident à investiguer vu les nombreuses répétitions nécessaires pour avoir une moyenne exploitable de Speech ABR. On pourrait toutefois mélanger aléatoirement les stimuli (au lieu de procéder à des séries de moyennages où le même stimulus est répété, puis le suivant, ...) et regrouper a posteriori les Speech ABR, puis les moyennner offline. Cette piste semble très intéressante à exploiter. En particulier, qu'advierait-il si les analogues non signifiants étaient inclus dans un contexte particulier et devenaient de ce fait signifiant ? Pour être encore plus rigoureux, les sine-wave speech [Remez 1981] seraient la façon optimale de répondre à cette question. Avec ces stimuli, des sons de parole (en général des phrases) sont reconstituées par des sommes de 3 sinusoïdes qui suivent les variations temporelles de 3 formants. Il est frappant de constater comme cette séquence de sinus peut être identifiée comme de la parole une fois que l'on connaît la phrase de départ. De cette façon, il serait idéal d'enregistrer les Speech ABR dans la condition où les sujets ne comprennent pas le stimulus, et lorsqu'ils connaissent le sens de la phrase. Pour l'instant, cette perspective est

limitée par la taille de la fenêtre d'enregistrement (inversement proportionnelle à la fréquence d'échantillonnage d'acquisition, qui doit être au moins 10 kHz), et par la nécessité de moyennner un grand nombre de traces pour avoir un rapport signal sur bruit acceptable. Une méthode pour se passer de moyennnage et pour enregistrer des Speech ABR sur des fenêtres aussi longues qu'on le souhaite ouvrirait des opportunités très grandes.

### **Est-ce que la FFR, quand elle existe, est nécessairement le reflet de l'enveloppe temporelle ?**

Toutes les données dont nous disposons vont dans ce sens... sauf une !

Le cas obtenu avec l'analogie très-mauvais (aucune information d'enveloppe et un poids spectral très fort sur les formants) montre que la synchronisation a lieu non pas sur le fondamental mais sur le premier harmonique à 400 Hz (voir flèche sur la Fig.3.20, en haut à gauche). Or, il s'avère que c'est le seul des 5 stimuli à avoir une structure pseudo-harmonique ; et il est composé des fréquences suivantes (avec chacune une amplitude comparable) : 200 - 800 - 1200 - 3200 - 4400 Hz Fig.3.16, c'est-à-dire les harmoniques suivant de 400 Hz : 1/2 - 2 - 3 - 8 - 11. Il s'avère que cette anecdote est lourde de sens, car dans ce cas, la FFR n'a pas été synchronisée sur le fondamental (200 Hz, comme l'indique le spectre 3.16), mais que la FFR a été semble-t-il le plus influencée par la structure harmonique dont la périodicité de 400 Hz a été extraite quelque part à travers les harmoniques 2-3-8 et 11 (par des phénomènes de battements dans les bandes critiques ?). Cette donnée isolée ne nous permet pas d'aller plus loin que ces quelques conjectures. Néanmoins, il serait très intéressant de procéder à une étude paramétrique faisant varier systématiquement le nombre d'harmoniques pour évaluer leur impact dans la formation de la FFR.

### **Comment se fait-il que la FFR soit enregistrée aussi clairement sur le scalp ?**

Il est légitime de se demander comment est-ce que la FFR parvient de façon aussi claire sur les électrodes placées sur le scalp, malgré la *cacophonie neurale* ambiante qui accompagne le trajet du message nerveux auditif ? En effet, les cellules dans tous les neurones du système auditif sont très spécialisés et ont notamment des caractéristiques fonctionnelles spécifiques tels que les taux de décharge spontanés, des fréquences de réponse, ainsi que des types de décharge (transitoires, périodiques synchronisées, périodiques mais à une périodicité intrinsèque, ...). Et pour compliquer encore la tâche d'un observateur en champ lointain de cette activité auditive (tel qu'un enregistrement de Speech ABR), chaque filtre auditif répond en décalé (en raison du délai de propagation de la vibration de la membrane basiliaire), et chaque noyaux cellulaires répondent également au fur et à mesure au son (avec des latences progressives d'une ms en sortie de la cochlée jusqu'à une dizaine de ms dans le col-



liculus inferieur). A priori, aucune réponse ne devrait parvenir aux électrodes sur le scalp, ou alors noyée dans cette cacophonie neurale. Eggermont [Eggermont 2007] a récemment décrit les phénomènes électriques qui font que le message nerveux peut parvenir jusqu'au scalp dans le cas des clic-ABR (ce qui peut être vraisemblablement étendu au cas des ABR tels que nous les avons enregistrés). La dynamique de transmission du message et l'exposition des générateurs neurophysiologiques sont importants. Afin d'être correctement recueilli, le message nerveux doit profiter des changements de conductivité du milieu biologique traversé (e.g. tissu / os), et dépend beaucoup de l'alignement des cellules. C'est au passage pourquoi il pense que le thalamus est peu enclin à donner des réponses enregistrables sur le scalp, car d'autres types de cellules masquent les cellules auditives, en les emprisonnant comme une cage de Faraday. Pour aller plus loin dans cette idée de dynamique de la source électrophysiologique nécessaire pour que le signal puisse bien se propager jusque sur les électrodes de scalp, le mode de code neural (dit rate Modulation Transfert Function) en jeu dans les cellules du colliculus inferieur est intéressant. Le principe du *rMTF* est de coder l'information de modulation de fréquence par des variations du taux de décharge proportionnels à cette fréquence modulante [Joris 2004], voir chap.1.1. Alors que dans les noyaux cellulaires précédents (en périphérie cochléaire), le type de codage n'est que temporal-MTF (tMTF) : les cellules déchargent chacune selon leurs caractéristiques spécifiques, mais avec une période correspondant à celle du stimulus. Intuitivement, on pourrait penser qu'un phénomène dynamique de variation périodique de courant synchronisé sur la périodicité de la modulante ( $F_0$ ), tel que le rMTF serait plus à même de résulter en une différence de potentiel enregistrable sur le scalp. L'aspect spéculatif du raisonnement est toutefois ici très fort.

### **Le Speech ABR peut-il être corrélé avec des mesures comportementales (psychoacoustiques) ?**

C'est le graal qui propulsera les Speech ABR dans le domaine clinique, avec l'identification précise des mécanismes neurophysiologiques sous-jacents [3.2].

Nous avons montré, dans la lignée de [Dau 2003], que la structure fine acoustique a une influence majeure dans la FFR (même si la FFR n'a aucune composante à de telles fréquences, elle en est en revanche très influencée). Or, de récentes études ont mis en avant le rôle capital de la structure fine acoustique dans la perception de la parole [Gilbert 2007], [Sheft 2007], et en particulier chez les malentendants [Lorenzi 2006]. Il apparait que les informations d'enveloppe qui sont primordiales dans la perception de la parole peuvent être extraites de la structure fine, dans les filtres auditifs ou les harmoniques ne sont pas résolus : deux ou plusieurs harmoniques étant analysés dans le même filtre auditif donnent lieu à des phénomènes de battements restituant l'enveloppe du signal [Gilbert 2006]. Nos résultats montrent que les informations contenues dans les filtres auditifs de hautes fréquences contri-

buent à la naissance de la FFR (dans le colliculus inferieur central). De plus amples investigations seraient semble-t-il les bienvenues pour évaluer la possibilité pour le Speech ABR d'être une mesure objective de la perception de la parole, ou plus précisément, de la correcte utilisation par le patient mal-entendant des indices acoustiques contenus dans la structure fine.

Greenberg a remarquablement montré le lien entre l'amplitude spectrale de la FFR et la sensation de hauteur tonale [Greenberg 1987]. Ensuite, ([Dimitrijevic A 2004]) ont tenté de relier les performances d'audiométrie vocale aux caractéristiques des ASSR, tandis que [Purcell 2004] ont montré la corrélation entre l'acuité temporelle (tMFT et gap détection) avec l'enveloppe following response à un son modulé en amplitude. Mikaël Ménard a quant à lui fait le lien entre l'amplitude des ASSR et les fonctions de sonie [Ménard 2008]. C'est certainement là une des grandes limitations de l'application des Speech ABR en clinique : un lien clair entre les Speech ABR et les paramètres psychoacoustiques manque encore.

#### **Quelle est l'influence de la surdité sur le Speech ABR ? En particulier, pourrait-on envisager un Speech eABR ?**

D'après les mesures que nous avons effectuées dans l'étude 2.3, il semblerait que la surdité ait un impact très important sur les Speech ABR. Ceux-ci semblent beaucoup moins bien définis chez les malentendants. De plus, est-ce que la fréquence de coupure du Speech ABR est sensible à l'état auditif du sujet dans ces fréquences ? Au delà ? Mais vu le manque d'information sur l'état fonctionnel des voies auditives, et le manque de connaissance précises sur les générateurs neurophysiologiques du Speech ABR, des expérimentation à large échelle devraient permettre de faire apparaître des données normatives, potentiellement corrélées aux données comportementales (seuils auditifs / perception de la parole). La seule étude à notre connaissances est celle de [Dimitrijevic A 2004].

L'objectif final serait d'utiliser les Speech ABR en marge du réglage des implants cochléaires (après l'avoir utilisé en dépistage). Mais l'obstacle de taille s'opposant à cette application, en plus du manque de connaissances sur le Speech ABR et la surdité, est lié à l'artefact engendré par la stimulation électrique.

# Conclusion générale

---

Le Speech ABR pourrait permettre de disposer d'une fenêtre d'observation de l'activité neurale issue à la fois du code temporel et spatial chez l'humain éveillé, en complémentarité de l'approche psychoacoustique, et en concrétisant les observations électrophysiologiques animalières réalisées sous anesthésie, avec en ligne de mire, le développement d'un outil pouvant assister l'exploration auditive clinique ou l'estimation de l'efficacité de la réhabilitation auditive (implant cochléaire, prothèse auditive). Dans ce travail, nous nous sommes assuré que nos recueils de Speech ABR étaient sans-artefact (??, [Akhoun 2008d]), et nous avons montré que la structure fine temporelle a plus d'impact que l'enveloppe du stimulus dans la création de la FFR.

Au regard de ces pré-études, il semblerait intéressant :

- D'observer paramétriquement les influences spectrales, au moyen de sons complexes harmoniques, selon la position des composantes spectrales, ou plus généralement, en reprenant les études de [Greenberg 1987] en détail.
- Puis, dans la continuité, de faire de même en filtrant des sons de parole dans certaines bandes de fréquence
- De compléter la comparaison entre les stimuli de différentes natures, pour évaluer l'effet de la nature du stimulus, avec une attention particulière aux sons de type *sine-wave speech*, [Remez 1981].
- De disposer de données normatives sur les effets de la perte auditive, ceci étant jusqu'à présent ignoré dans la littérature. En particulier, y a-t-il un lien entre la fréquence de coupure du Speech ABR et celle de l'audiogramme, dans le cas des surdités du type des indications pour l'implantation électro-acoustique.
- D'investiguer les effets de la réorganisation des voies auditives suite à assourdissement unilatéral, dans une étude longitudinale, avec comme paramètre principal la durée de surdité.
- D'aller plus loin dans l'authentification, ce coup-ci auditif, du Speech ABR, en montrant qu'il soit réellement insensible aux effets du complexe auditif/facial du réflexe stapédien.



# Calibration de la stimulation sonore avec les inserts-earphones

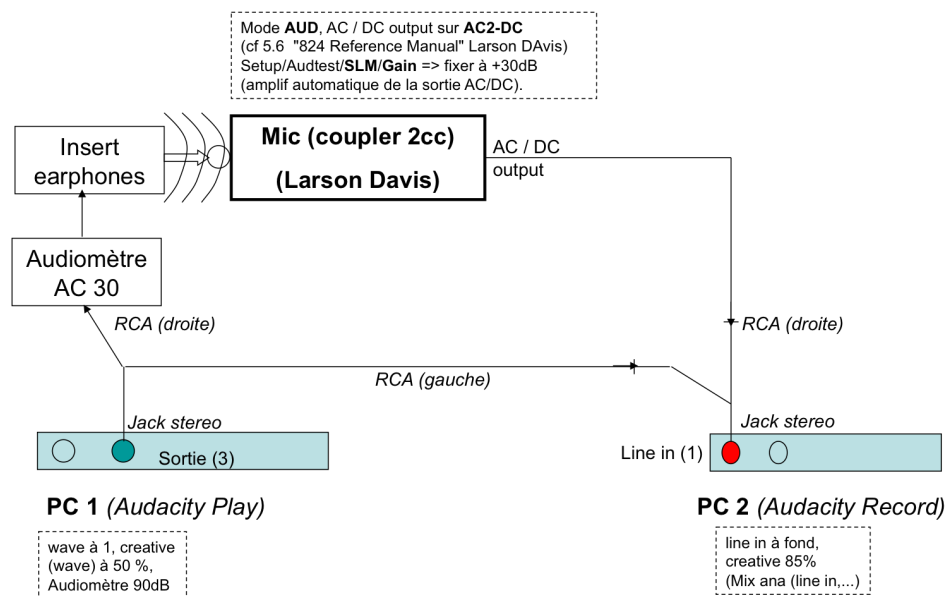


FIGURE A.1 – Synoptique de l'enregistrement du signal audio.

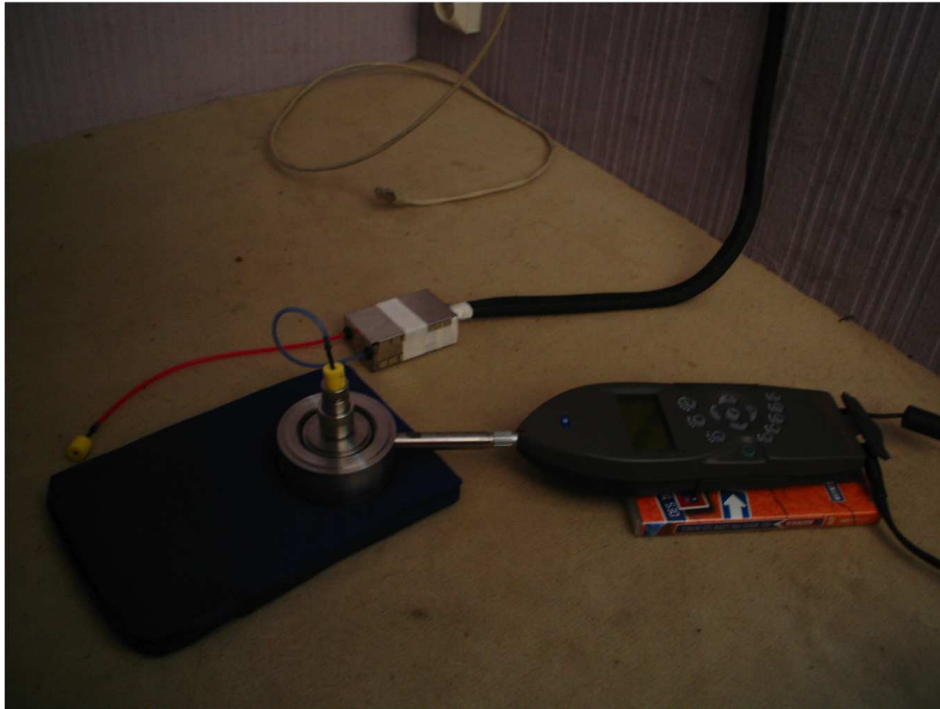


FIGURE A.2 – Aperçu de l'enregistrement du signal audio.

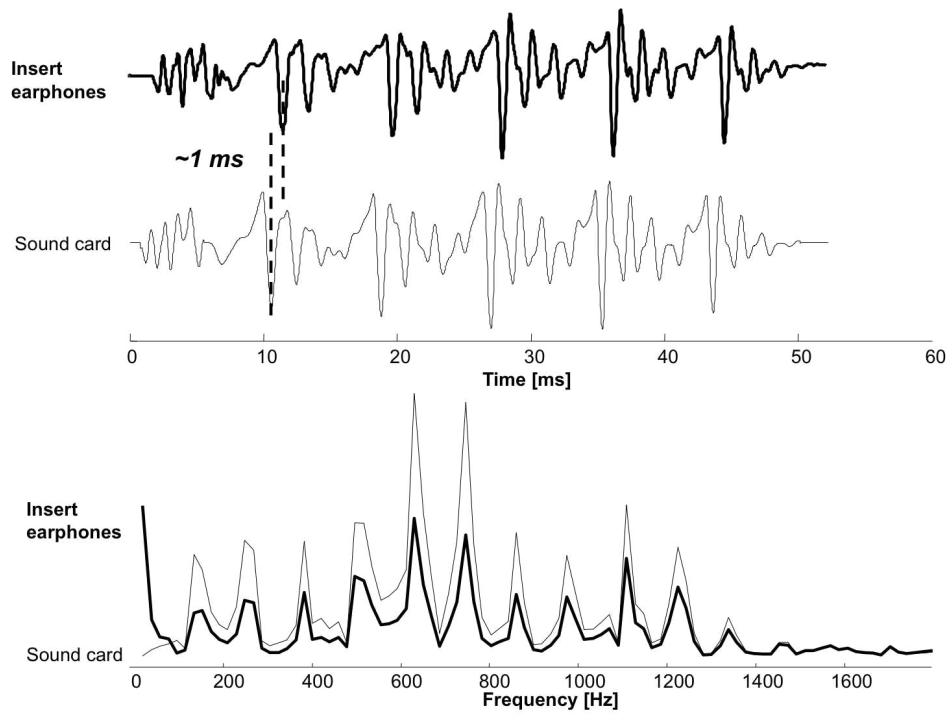


FIGURE A.3 – Comparaison du fichier son et du signal audio.



# Modélisation de l'activité neurale par AIM

Ci-dessous la modélisation par l'Auditory Image Model [Bleack 2004] des /ba/, /pa/, analogues, sons purs et sons complexes de l'étude 3.2.

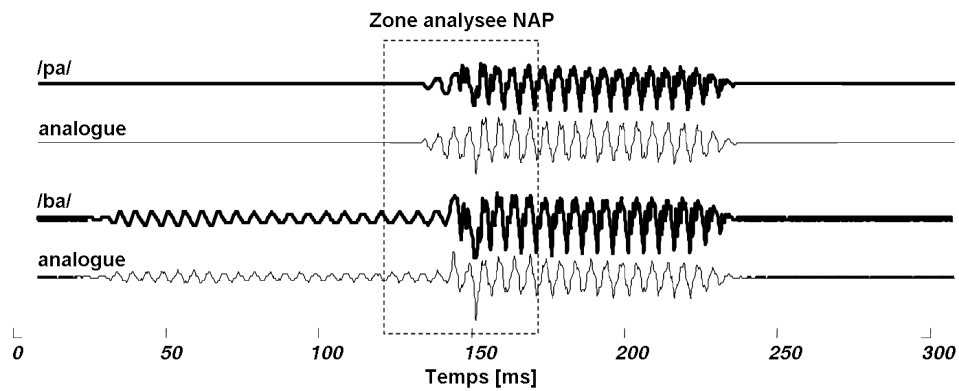
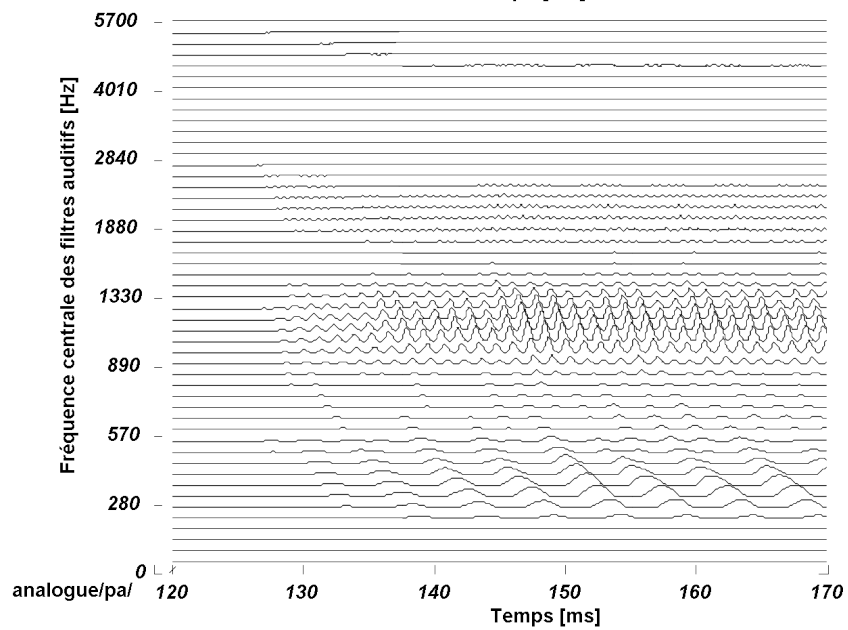
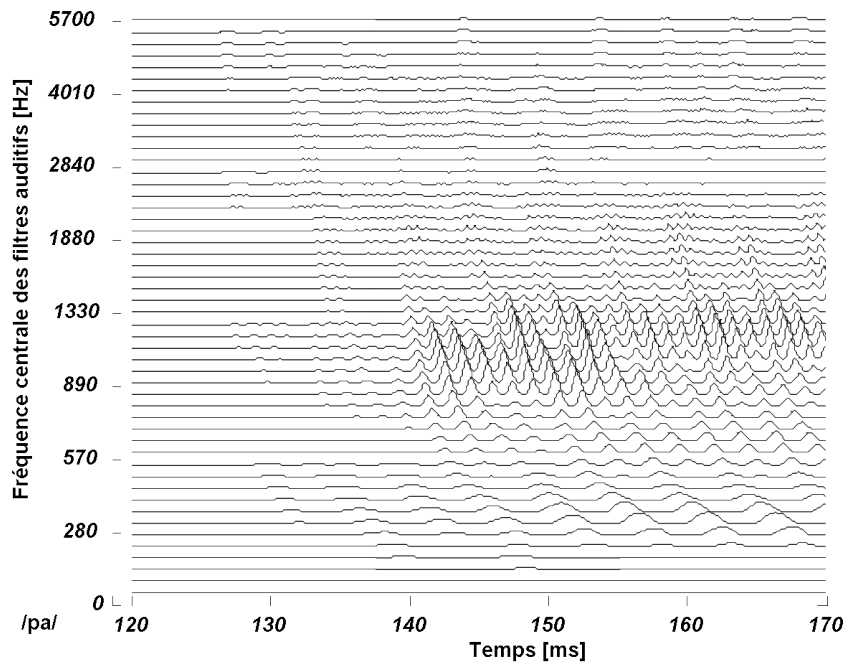
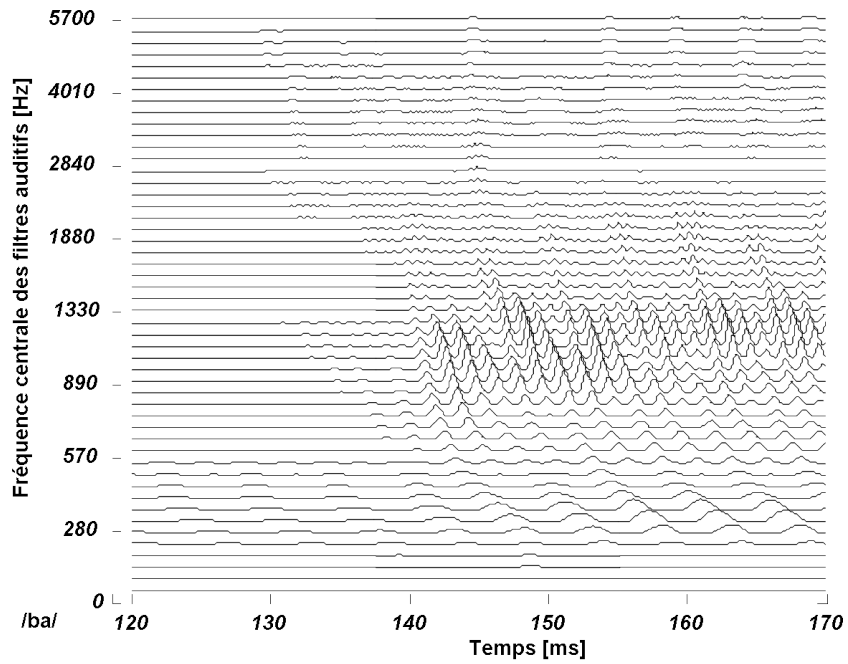


FIGURE B.1 – Zone des stimuli de parole et analogues analysée par AIM.



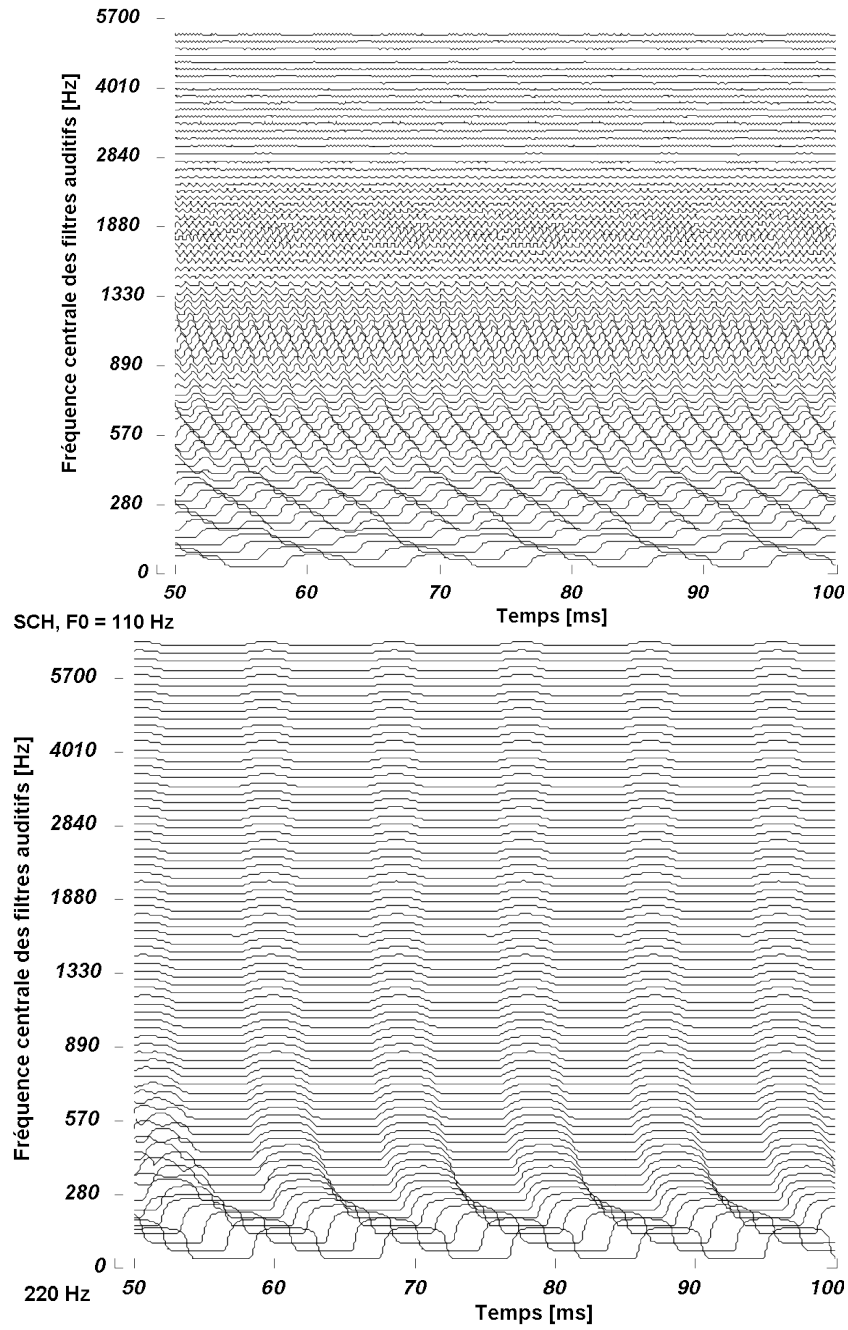


FIGURE B.3 – AIM des sons purs et sons complexes harmoniques.



# Speech ABR et analyse temps-fréquence. Effet jitter.

---

Nous présentons ici les résultats d'une étude comparant les avantages entre le moyennage temporel ou le moyennage des spectrogrammes.

## C.1 Annulation artificielle de la FFR (effet jitter)

En routine clinique, les ABR sont le plus souvent analysés d'après les traces temporelles. La FFR est quant à elle souvent analysée en fréquence, en raison de ses propriétés oscillatoires remarquables. Toutefois, dans une étude préliminaire de celle de l'étude (4.3), publiée dans [Akhoun 2008a], qui faisant intervenir des stimuli de hauteur tonale (fréquence fondamentale) assez élevées, nous avons constaté un problème directement lié à la technique de moyennage : la FFR s'annule au fur et à mesure des moyennages en raison de très petits décalages entre les pics et les creux des FFR entre les individus. Sur les Fig.C.2 et C.1, on remarque que la FFR est bien synchronisée à 575 Hz sur les Speech ABR individuels, mais cela n'est plus visible sur la grande moyenne. En effet, il peut exister un décalage (dit jitter) entre les pics de la FFR obtenus pour 2 sujets ; ce jitter entraîne un déphasage entre les FFR de ces sujets. Dans le cas limite où ce déphasage entre la FFR de deux sujets est de  $180^\circ$ , la grande moyenne de la FFR sur ces deux sujets est annulée artificiellement : ce qui semble se passer ici.

## C.2 Etude du compromis moyennage / cartes temps fréquences instantanées.

Pour lutter contre les soucis liés au jitter, et à l'atténuation artificielle de la FFR, le mieux serait de procéder au moyennage des spectrogrammes non seulement individuels, mais aussi des Speech ABR instantanés. Toutefois, trois problèmes semblent se poser : (i) nous comptons sur la méthode de moyennage pour atténuer au fur et à mesure l'artéfact lié au 50 Hz (EDF) en prenant des fenêtres d'acquisitions faisant  $(2k+1) \cdot 20$  ms, ensuite (ii) la polarité alternée habituellement utilisée pour supprimer l'artéfact (comme nous l'expliquons dans le chap. 1 précédent) ne marche plus avec le spectrogramme, et enfin (iii) ce processus est très gourmand

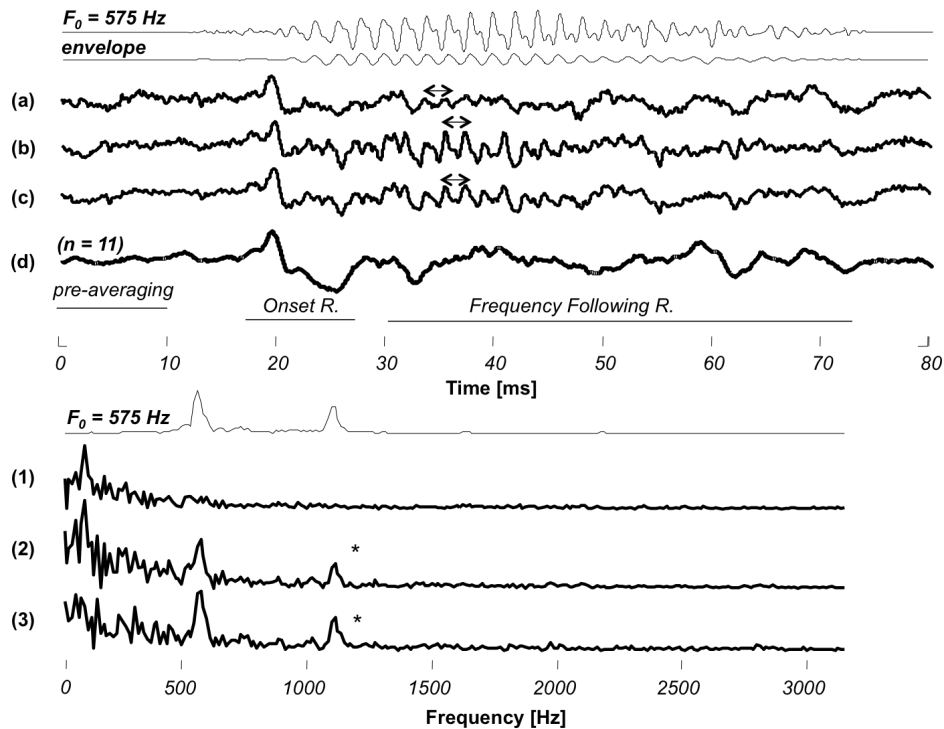


FIGURE C.1 – Annulation artificielle de la FFR par le moyennage - Speech ABR individuel (sujet 2) à 575 Hz. Contrairement à ce qui est suggéré sur la grande moyenne (d), certains sujets ont bien eu une FFR à 575 Hz. C'est particulièrement visible sur chaque polarité (a et b), même si la FFR est bien mieux définie en (b). La meilleure représentation de la FFR en polarité (a) comparée à la polarité (b) fait que la FFR est bien représentée en polarité alternée (c). Ceci se retrouve dans la représentation spectrale de la FFR (1, 2 et 3) juste en dessous du spectre du stimulus en trait fin ( $F_0 = 575$  Hz). Par conséquent, l'expression de la FFR s'est annulée lors du moyennage (des petits décalages entraînent des déphasages importants à de telles fréquences, ce qui n'était pas le cas à 120 Hz) : le passage à l'analyse temps-fréquence permet de s'affranchir de ce type d'annulation par effet-jitter. [Akhoun 2008c]



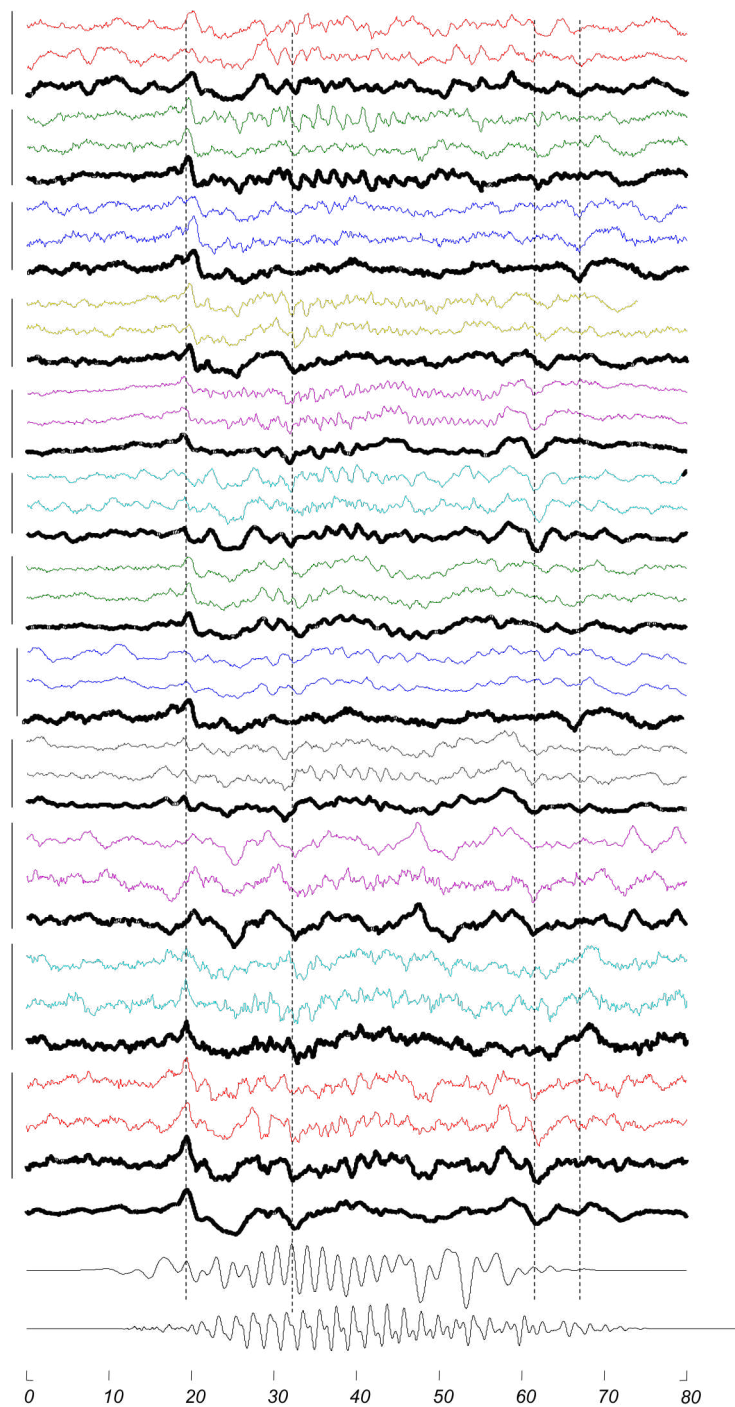


FIGURE C.2 – Speech ABR est bien défini sur les traces individuelles en réponse à /ba/ de  $f_0 = 575$  Hz.

en temps (une quarantaine de minutes ici). (i) Comme on pouvait s’y attendre, il ressort du moyennage des spectrogrammes successifs que la moyenne des 1500 traces successives aboutit à une énorme composante fréquentielle basse fréquence (qui englobe plus que simplement la bande correspondant au 50 Hz ainsi que des harmoniques à 900 et 1200 Hz), tandis qu’aucune FFR n’est visible, Fig. C.3. Notons que rajouter une étape de filtrage coupe-bande pourrait apporter une solution à ce problème, mais nous ne l’avons pas expérimenté.

(ii) La méthode de polarité alternée permet traditionnellement d’éliminer l’artéfact

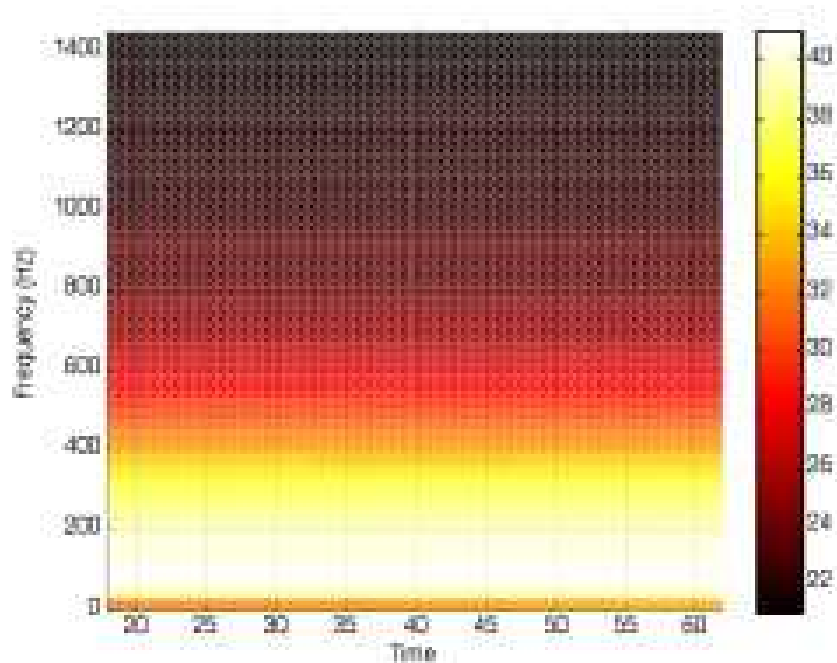


FIGURE C.3 – **Moyenne des 1500 cartes temps-fréquence successives de Speech ABR.** L’analyse temps-fréquence permet de ne plus avoir de problèmes de jitter dû à la suppression de composantes hautes-fréquences lors du moyennage temporel. Toutefois, dans ce cas, le bruit ambiant résulte en une bande basse-fréquence uniforme. D’où l’intérêt moyenner un certain nombre de traces temporelles avant de passer au spectrogramme.

électromagnétique résiduel ou le potentiel microphonique (en faisant l’hypothèse que les effets artéfactuels sont opposés sur les deux polarités, tandis que le Speech ABR est identique vu l’encodage identique des deux sons en opposition de phase Fig. 2.11 ou Fig. 3.3; mais elle n’est plus applicable avec les spectrogrammes d’amplitude qui utilisent des énergies, et de ce fait, sont toujours positives. Ainsi, la carte temps fréquence de la polarité alternée n’est pas égale à la somme des cartes temps-fréquence des deux polarités Fig. C.4. Les artéfacts sont visibles sur les deux bandes de fréquences qui sont plus énergétiques sur (3) que sur (4) ou (5).

Ces artéfact doivent donc au préalable être supprimés en temporel. Par conséquent, il convient de procéder en deux temps : tout d'abord sommer temporellement les deux Speech ABR (pour chaque sujets) dans chaque polarité afin de limiter les effets artéfactuels, puis ensuite moyennner les cartes temps-fréquences individuelles en polarité alternée ainsi obtenues.

(iii) Afin de limiter le temps de calcul, et également afin de réduire les effets du 50 Hz (cf. (i)), nous avons cherché à moyennner des groupes de traces temporelles, afin de trouver le meilleur compromis entre le nombre de moyennes de ces sous-groupes et la bonne visualisation de la FFR, avec le cas le plus concerné, pour le stimulus dont la fréquence fondamentale est de 575 Hz. On peut déduire que dans ce cas précis, le meilleur compromis a lieu pour le moyennnage des cartes temps fréquences calculées pour des groupes de 250 traces successives Fig. C.5.

Il s'agit ici d'une approche totalement empirique. Nous avons donc tenté de trouver le bon nombre de traces temporelles de Speech ABR instantanées à « pré-moyenner » (afin de limiter le 50 Hz en évitant au maximum la possibilité d'avoir des effets jitter) avant d'appliquer le spectrogramme pour passer à une représentation temps-fréquence, qui seront ensuite moyennnées. Par exemple, dans le cas « 100 », nous avons moyennnés  $1500 / 100 = 15$  sous-groupes de 100 traces instantanées, donc nous avons obtenus 15 cartes temps-fréquences que nous avons ensuite moyennnées. Dans le cas 250, nous avons moyennné 6 spectrogrammes de sous-groupes de 250 traces instantanées successives de Speech ABR... Le compromis à trouver est le suivant : si nous moyennnons trop de traces successives instantanées de Speech ABR (au maximum 1500, ici), nous aurons un risque de perdre des informations hautes-fréquences par effet-jitter (moyennnage de FFR instantanées (très légèrement) déphasées) tandis que si nous moyennnons des cartes temps-fréquences de Speech ABR insuffisamment moyennnées, nous risquerons fort d'obtenir une bande très fortement énergétique englobant les basses fréquences (résultant du bruit ambiant). Dans ce cas, le meilleur compromis semble se produire pour le cas « 250 ».

Dans cette étude, nous avons expérimenté une technique empirique de calculs de spectrogramme pour des groupes de Speech ABR successifs. Nous avons ainsi trouvé que le moyennnage temporel d'un assez grand nombre de traces successives (250) pouvait au préalable atténuer le 50 Hz sans que l'effet-jitter ne soit trop problématique (ce n'est toutefois qu'une hypothèse). En somme, le spectrogramme est un outil privilégié pour étudier la FFR, en se mettant à l'abri des problèmes de déphasages entre les FFR lors du moyennnage et de l'effet-jitter qui en découle. La plupart des études de FFR dans la littérature (e.g. [Wong 2007]) utilisent le logiciel Pratt, qui implémente le pitch-tracking (extraction de fondamental) par autocorrélation (de la même façon qu'avec les histogrammes poststimuli dans les études physiologiques ou les modélisations). Toutefois, nous avons choisi l'utilisation de cette technique du spectrogramme sous Matlab pour quantifier la synchronisation de la FFR dans le chapitre suivant. Notons qu'une investigation approfondie des différentes méthodes de représentations temps-fréquence (Wigner-Ville, ondelettes) pourrait permettre d'optimiser l'analyse des Speech ABR, et en particulier de mieux observer les varia-

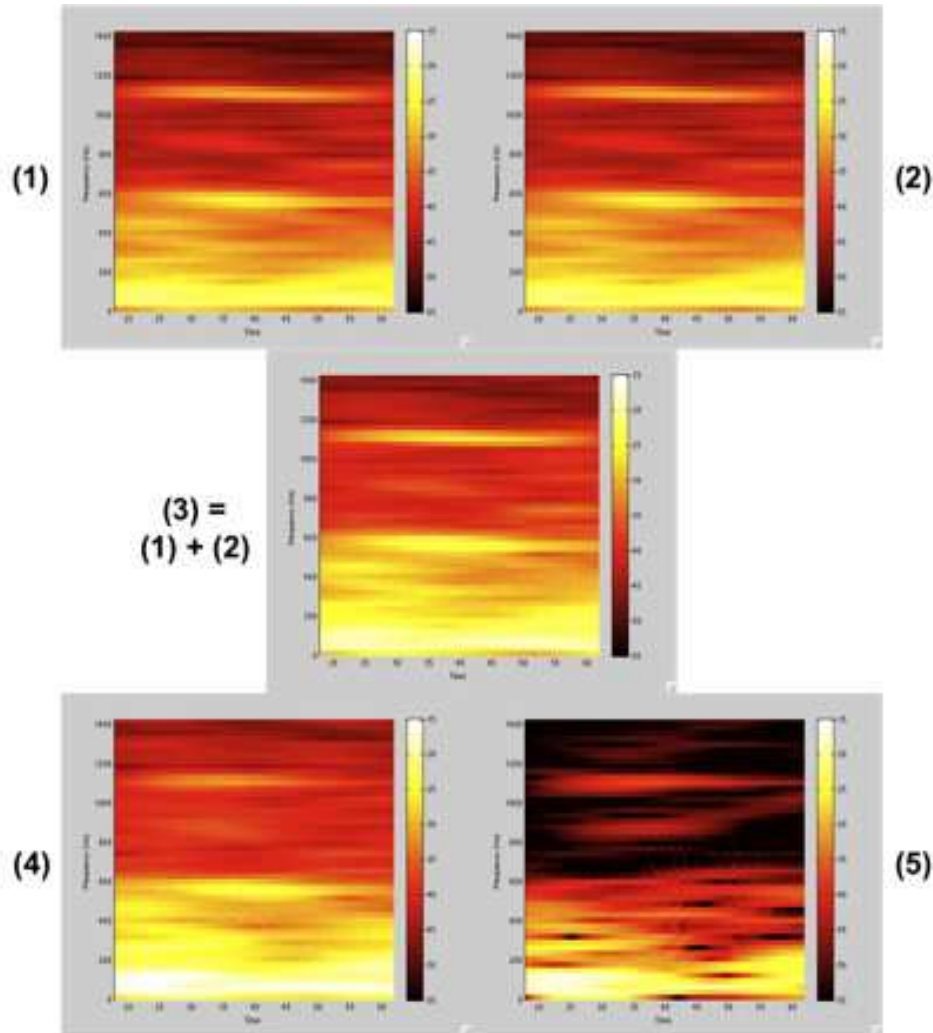


FIGURE C.4 – Pourquoi le spectrogramme ne peut-il plus supprimer les effets artéfactuels, comme la polarité alternée en temporel (pour un /ba/ à 575 Hz de fondamental) ? Les deux premiers spectrogrammes (1) et (2) sont ceux des grandes moyennes pour chaque polarité plus et moins. Le spectrogramme du milieu (3) est la somme des spectrogrammes précédents (pour chaque polarité). Les deux spectrogrammes du bas représentent respectivement la moyenne des 12 spectrogrammes de Speech ABR individuels en polarité alternée (4), et le spectrogramme de la grande moyenne temporelle (en polarité alternée) sur les 12 sujets (5).

tions de synchronisation [P.Flandrin 1998].

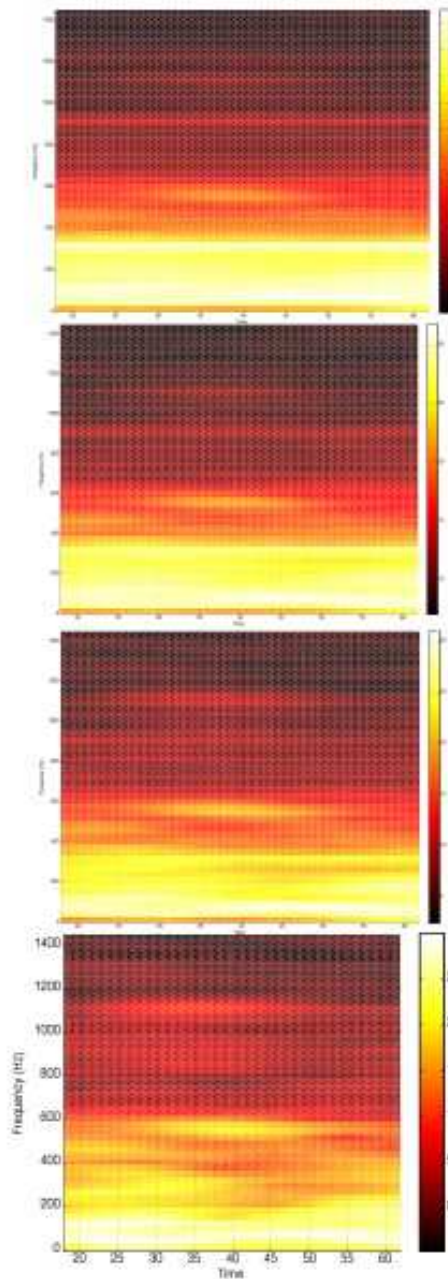


FIGURE C.5 – Etude du compromis de pré-moyennages temporel et de moyennages temps fréquence. Ici, 25, 50, 100, 250, 1500 traces instantanées successives de Speech ABR furent moyennées avant que les spectrogrammes de ces sous-moyennes soient eux-mêmes moyennés.





# Bibliographie

- [Abrams 2008] Daniel A Abrams, Trent Nicol, Steven Zecker et Nina Kraus. *Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech*. J Neurosci, vol. 28, no. 15, pages 3958–3965, Apr 2008. 26
- [Aiken 2006] Steven J Aiken et Terence W Picton. *Envelope following responses to natural vowels*. Audiol Neurootol, vol. 11, no. 4, pages 213–232, 2006. 34
- [Aiken 2008] Steven J Aiken et Terence W Picton. *Envelope and spectral frequency-following responses to vowel sounds*. Hear Res, Aug 2008. 34
- [Akhoun 2008a] I. Akhoun, B.Tillmann, A.Moulin, S.Gallégo, C.Berger-Vachon, L.Collet et H.Thai-Van. *Temporal, intensity and pitch relationship on Speech Auditory Brainstem Responses to phonemes /ba/ in Normal Hearing Adults*. Proceedings of Meetings on Acoustics, vol. 1, page Epub, 2008. iv, 81, 90, 123, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 166, 179
- [Akhoun 2008b] I. Akhoun, C.Richard, A.Moulin, S.Gallégo, C.Berger-Vachon, L.Collet et H.Thai-Van. *Comparison of speech vs speech-analogue temporal encoding in human : An auditory brainstem response and categorical perception assessment. Poster*. J Acous Soc Am, vol. 123, page 3863, 2008. 53, 58, 59, 60, 62, 64, 65
- [Akhoun 2008c] I. Akhoun, S. GallÉgo, A. Moulin, M. MÈnard, E. Veuillet, C. Berger-Vachon, L. Collet et H. Thai-Van. *The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults*. Clin Neurophysiol, vol. 119, no. 4, pages 922–933, Apr 2008. iv, 81, 82, 140, 164, 180
- [Akhoun 2008d] Idrick Akhoun, Annie Moulin, Arnaud Jeanvoine, Mikael MÈnard, FranÁois Buret, Christian Vollaire, Riccardo Scorretti, Evelyne Veuillet, Christian Berger-Vachon, Lionel Collet et Hung Thai-Van. *Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status An experimental parametric study*. J Neurosci Methods, vol. 175, no. 2, pages 196–205, Nov 2008. iii, 20, 35, 47, 63, 69, 71, 73, 75, 77, 79, 88, 171
- [Banai 2005] Karen Banai, Trent Nicol, Steven G Zecker et Nina Kraus. *Brainstem timing : implications for cortical processing and literacy*. J Neurosci, vol. 25, no. 43, pages 9850–9857, Oct 2005. 23, 26
- [Banai 2006] K. Banai et N.Kraus. Handbook of central auditory processing disorder : From science to practice., chapitre The neurobiology of central auditory processing disorder (CAPD), language impairment and learning disability., pages 100–120. Plural Publishing Inc. G.D. Chermak, F.E. Musiek (Eds.), USA, 2006. 21, 23, 44

- [Banai 2007] Karen Banai, Daniel Abrams et Nina Kraus. *Sensory-based learning disability : Insights from brainstem processing of speech sounds*. Int J Audiol, vol. 46, no. 9, pages 524–532, Sep 2007. 21, 23, 25, 44
- [Banai 2008] K. Banai et N.Kraus. The dynamic brainstem. implications for apd, chapitre Current Controversies in Central Auditory Processing Disorder., pages 269–289. Plural Publishing Inc. Mc Farland D et Cacace A. (Eds.), San Diego, Californie, 2008. 20
- [Bleeck 2004] S. Bleeck, T. Ives et R.Patterson. *Aim-mat : the Auditory Image Model in MATLAB*. Acta Acustica, vol. 90, pages 781–788, 2004. 91, 175
- [Brown 2003] Carolyn J Brown. *Clinical uses of electrically evoked auditory nerve and brainstem responses*. Curr Opin Otolaryngol Head Neck Surg, vol. 11, no. 5, pages 383–387, Oct 2003. 21
- [Bélanger 2007] O Bélanger, Traube C et J. Piché. *Designing and controlling a source-filter model for naturalistic and expressive singing voice synthesis*. Proceedings of the International Computer Music Conference (ICMC'07), Copenhagen, Denmark, 2007. 123
- [Clark 2003] G.M. Clark. Cochlear implants. fundamentals and applications, modern acoustic and signal processing. Springer, Melbourne, 2003. 4, 8, 14
- [Clark 2006] Graeme M Clark. *The multiple-channel cochlear implant : the interface between sound and the central nervous system for hearing, speech, and language in deaf people-a personal perspective*. Philos Trans R Soc Lond B Biol Sci, vol. 361, no. 1469, pages 791–810, May 2006. 14
- [Cunningham 2002] Jenna Cunningham, Trent Nicol, Cynthia King, Steven G Zecker et Nina Kraus. *Effects of noise and cue enhancement on neural responses to speech in auditory midbrain, thalamus and cortex*. Hear Res, vol. 169, no. 1-2, pages 97–111, Jul 2002. 39
- [Dajani 2005] Hilmi R Dajani, David Purcell, Willy Wong, Hans Kunov et Terence W Picton. *Recording human evoked potentials that follow the pitch contour of a natural vowel*. IEEE Trans Biomed Eng, vol. 52, no. 9, pages 1614–1618, Sep 2005. 34
- [Dau 2003] Torsten Dau. *The importance of cochlear processing for the formation of auditory brainstem and frequency following responses*. J Acoust Soc Am, vol. 113, no. 2, pages 936–950, Feb 2003. 34, 35, 36, 39, 40, 44, 110, 111, 162, 163, 164, 169
- [Delorme 2004] Arnaud Delorme et Scott Makeig. *EEGLAB : an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis*. J Neurosci Methods, vol. 134, no. 1, pages 9–21, Mar 2004. 52
- [Dimitrijevic A 2004] Picton TW. Dimitrijevic A John MS. *Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults*. Ear and Hearing, vol. 25, pages 68–84, 2004. 28, 44, 170
- [Eggermont 2001] J. J. Eggermont. *Between sound and perception : reviewing the search for a neural code*. Hear Res, vol. 157, no. 1-2, pages 1–42, Jul 2001. 1

- [Eggermont 2004] J. J. Eggermont. Evoked potentials manual., chapitre The Auditory Brainstem Response., pages 150–200. Kluwer Academic. Colon EJ, Vissel SL (Eds.), Amsterdam, 2004. 39
- [Eggermont 2007] J. J. Eggermont. 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 Willkins. Burkard R, Don M, Eggermont JJ (Eds.), Philadelphia, Baltimore, New-York, London, Buenos-Aires, Hong-Kong, Sydney et Tokyo, 2007. 21, 44, 169
- [Galbraith 1993] G. C. Galbraith et C. Arroyo. *Selective attention and brainstem frequency-following responses*. Biol Psychol, vol. 37, no. 1, pages 3–22, Oct 1993. 30
- [Galbraith 1994] G. C. Galbraith. *Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli*. Electroencephalogr Clin Neurophysiol, vol. 92, no. 4, pages 321–330, Jul 1994. 42, 163, 165
- [Galbraith 1995a] G. C. Galbraith, P. W. Arbagey, R. Branski, N. Comerci et P. M. Rector. *Intelligible speech encoded in the human brain stem frequency-following response*. Neuroreport, vol. 6, no. 17, pages 2363–2367, Nov 1995. 23, 35, 42, 43
- [Galbraith 1995b] G. C. Galbraith et B. Q. Doan. *Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli*. Int J Psychophysiol, vol. 19, no. 3, pages 203–214, Apr 1995. 35
- [Galbraith 1997] G. C. Galbraith, S. P. Jhaveri et J. Kuo. *Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition*. Electroencephalogr Clin Neurophysiol, vol. 102, no. 1, pages 46–53, Jan 1997. 30
- [Galbraith 2000] GC Galbraith, DM Olfman et TF Ullman. *Putative measure of peripheral and brainstem frequency-following in humans*. Neuroscience letters, vol. 292, pages 123–127, 2000. 35
- [Galbraith 2001] G. C. Galbraith. *From brainstem to cortex : neurobiologic research provides keys to the riddles, mysteries and enigmas of brain dysfunction*. Clin Neurophysiol, vol. 112, no. 5, pages 721–723, May 2001. 21
- [Galbraith 2003] GC Galbraith, DM Olfman et TF Ullman. *Selective attention affects human brain stem frequency-following response*. Neuroreport, vol. 14, pages 735–738, 2003. 30
- [Galbraith 2004] Gary C Galbraith, Elizabeth M Amaya, Jacinta M Diaz de Rivera, Namee M Donan, Mylien T Duong, Jeffrey N Hsu, Kim Tran et Lian P Tsang. *Brain stem evoked response to forward and reversed speech in humans*. Neuroreport, vol. 15, no. 13, pages 2057–2060, Sep 2004. 30

- [Gallégo 1998] S. Gallégo, B. Frachet, C. Micheyl, E. Truy et L. Collet. *Cochlear implant performance and electrically-evoked auditory brain-stem response characteristics*. *Electroencephalogr Clin Neurophysiol*, vol. 108, no. 6, pages 521–525, Nov 1998. 20, 21
- [Gaumond 1982] R. P. Gaumond, C. E. Molnar et D. O. Kim. *Stimulus and recovery dependence of cat cochlear nerve fiber spike discharge probability*. *J Neurophysiol*, vol. 48, no. 3, pages 856–873, Sep 1982. 91
- [Gilbert 2006] GaËtan Gilbert et Christian Lorenzi. *The ability of listeners to use recovered envelope cues from speech fine structure*. *J Acoust Soc Am*, vol. 119, no. 4, pages 2438–2444, Apr 2006. 169
- [Gilbert 2007] GaËtan Gilbert, Isabelle Bergeras, DorothÈe Voillery et Christian Lorenzi. *Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues*. *J Acoust Soc Am*, vol. 122, no. 3, page 1336, Sep 2007. 44, 162, 169
- [Glasberg 1990] B. R. Glasberg et B. C. Moore. *Derivation of auditory filter shapes from notched-noise data*. *Hear Res*, vol. 47, no. 1-2, pages 103–138, Aug 1990. 3
- [Greenberg 1987] S. Greenberg, J. T. Marsh, W. S. Brown et J. C. Smith. *Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones*. *Hear Res*, vol. 25, no. 2-3, pages 91–114, 1987. 22, 32, 33, 35, 42, 44, 110, 120, 122, 164, 170, 171
- [Grose 2001] J. H. Grose, J. W. Hall et E. Buss. *Gap duration discrimination in listeners with cochlear hearing loss : effects of gap and marker duration, frequency separation, and mode of presentation*. *J Assoc Res Otolaryngol*, vol. 2, no. 4, pages 388–398, Dec 2001. 13
- [Guiraud 2007] Jeanne Guiraud, StÈphane Gallego, Laure Arnold, Patrick Boyle, Eric Truy et Lionel Collet. *Effects of auditory pathway anatomy and deafness characteristics ? (1) : On electrically evoked auditory brainstem responses*. *Hear Res*, vol. 223, no. 1-2, pages 48–60, Jan 2007. 20
- [Hall 1979] J. W. Hall. *Auditory brainstem frequency following responses to waveform envelope periodicity*. *Science*, vol. 205, no. 4412, pages 1297–1299, Sep 1979. 41
- [Heinz 2001] M. G. Heinz, H. S. Colburn et L. H. Carney. *Evaluating auditory performance limits : i. one-parameter discrimination using a computational model for the auditory nerve*. *Neural Comput*, vol. 13, no. 10, pages 2273–2316, Oct 2001. 38
- [Hewitt 1991] M. J. Hewitt et R. Meddis. *An evaluation of eight computer models of mammalian inner hair-cell function*. *J Acoust Soc Am*, vol. 90, no. 2 Pt 1, pages 904–917, Aug 1991. 84, 91, 92
- [Jeanvoine 2008] A. Jeanvoine. Rapport de stage (master 2 recherche génie biomédical), chapitre Etude de l'artéfact électromagnétique sur les enregistrements

- de Speech ABR. Université de Lyon. Université Claude Bernard Lyon 1, 2008. 54, 55, 56, 61
- [Jewett 1970] D. L. Jewett, M. N. Romano et J. S. Williston. *Human auditory evoked potentials : possible brain stem components detected on the scalp*. Science, vol. 167, no. 924, pages 1517–1518, Mar 1970. 21, 22
- [John 2000] M. S. John et T. W. Picton. *Human auditory steady-state responses to amplitude-modulated tones : phase and latency measurements*. Hear Res, vol. 141, no. 1-2, pages 57–79, Mar 2000. 36, 87
- [Johnson 2005] Krista L Johnson, Trent G Nicol et Nina Kraus. *Brain stem response to speech : a biological marker of auditory processing*. Ear Hear, vol. 26, no. 5, pages 424–434, Oct 2005. 23, 84
- [Johnson 2008] Krista L Johnson, Trent Nicol, Steven G Zecker, Ann R Bradlow, Erika Skoe et Nina Kraus. *Brainstem encoding of voiced consonant-vowel stop syllables*. Clin Neurophysiol, Sep 2008. 35, 121
- [Joris 2004] P. X. Joris, C. E. Schreiner et A. Rees. *Neural processing of amplitude-modulated sounds*. Physiol Rev, vol. 84, no. 2, pages 541–577, Apr 2004. 1, 7, 8, 22, 169
- [Khalifa 2001] S. Khalifa, R. Bougeard, N. Morand, E. Veuillet, J. Isnard, M. Guenot, P. Ryvlin, C. Fischer et L. Collet. *Evidence of peripheral auditory activity modulation by the auditory cortex in humans*. Neuroscience, vol. 104, no. 2, pages 347–358, 2001. 6
- [Khanna 1989] S. M. Khanna et M. C. Teich. *Spectral characteristics of the responses of primary auditory-nerve fibers to amplitude-modulated signals*. Hear Res, vol. 39, no. 1-2, pages 143–157, May 1989. 3
- [Kraus 1996] N. Kraus, T. J. McGee, T. D. Carrell, S. G. Zecker, T. G. Nicol et D. B. Koch. *Auditory neurophysiologic responses and discrimination deficits in children with learning problems*. Science, vol. 273, no. 5277, pages 971–973, Aug 1996. 23
- [Kraus 2000] N. Kraus, A. R. Bradlow, M. A. Cheatham, J. Cunningham, C. D. King, D. B. Koch, T. G. Nicol, T. J. McGee, L. K. Stein et B. A. Wright. *Consequences of neural asynchrony : a case of auditory neuropathy*. J Assoc Res Otolaryngol, vol. 1, no. 1, pages 33–45, Aug 2000. 166
- [Kraus 2005] Nina Kraus et Trent Nicol. *Brainstem origins for cortical 'what' and 'where' pathways in the auditory system*. Trends Neurosci, vol. 28, no. 4, pages 176–181, Apr 2005. 26, 164
- [Kraus 2007] N. Kraus et K. Banai. *Auditory-Processing malleability. Focus on Language and Music*. Current directions in psychological sciences., vol. 16, pages 105–110, 2007. 165
- [Krishnan 2004] Ananthanarayan Krishnan, Yisheng Xu, Jackson T Gandour et Peter A Cariani. *Human frequency-following response : representation of pitch contours in Chinese tones*. Hear Res, vol. 189, no. 1-2, pages 1–12, Mar 2004. 28, 34, 35

- [Krishnan 2005] Ananthanarayan Krishnan, Yisheng Xu, Jackson Gandour et Peter Cariani. *Encoding of pitch in the human brainstem is sensitive to language experience*. *Brain Res Cogn Brain Res*, vol. 25, no. 1, pages 161–168, Sep 2005. 28, 163, 165
- [Krishnan 2008] Ananthanarayan Krishnan, Jayaganesh Swaminathan et Jackson T Gandour. *Experience-dependent Enhancement of Linguistic Pitch Representation in the Brainstem Is Not Specific to a Speech Context*. *J Cogn Neurosci*, Aug 2008. 31, 34, 42, 111, 135, 164, 165
- [Kuhl 1991] P. K. Kuhl. *Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not*. *Percept Psychophys*, vol. 50, no. 2, pages 93–107, Aug 1991. 104
- [Loizou 1999] P. C. Loizou. *Introduction to cochlear implants*. *IEEE Eng Med Biol Mag*, vol. 18, no. 1, pages 32–42, 1999. 14
- [Lorenzi 2006] Christian Lorenzi, GaËtan Gilbert, HÈloÛse Carn, StÈphane Garnier et Brian C J Moore. *Speech perception problems of the hearing impaired reflect inability to use temporal fine structure*. *Proc Natl Acad Sci U S A*, vol. 103, no. 49, pages 18866–18869, Dec 2006. 10, 169
- [Lyon 1996] Richard Lyon et Shihab Shamma. *Springer handbook of auditory research (vol. 6). auditory computation.*, chapitre *Auditory representation of pitch and timbre.*, pages 150–200. Springer. Hawkins HL, McMullen T, Fay RR, Popper AN (Eds.), New-York, 1996. 92
- [Marsh 1970] J. T. Marsh, F. G. Worden et J. C. Smith. *Auditory frequency-following response : neural or artifact?* *Science*, vol. 169, no. 951, pages 1222–1223, Sep 1970. 52
- [Marsh 1975] J. T. Marsh, W. S. Brown et J. C. Smith. *Far-field recorded frequency-following responses : correlates of low pitch auditory perception in humans*. *Electroencephalogr Clin Neurophysiol*, vol. 38, no. 2, pages 113–119, Feb 1975. 22
- [Martin 2008] BA Martin, KL Tremblay et P Korczac. *Speech Auditory Evoked Potential : from the laboratory to the clinic*. *Ear and Hearing*, vol. 29, pages 285–313, 2008. 21
- [McDermott 2004] Hugh J McDermott. *Music perception with cochlear implants : a review*. *Trends Amplif*, vol. 8, no. 2, pages 49–82, 2004. 14
- [McKay 2005] Colette McKay, Leonie Fewster et Pam Dawson. *A different approach to using neural response telemetry for automated cochlear implant processor programming*. *Ear Hear*, vol. 26, no. 4 Suppl, pages 38S–44S, Aug 2005. 23
- [Melcher 1996] J. R. Melcher et N. Y. Kiang. *Generators of the brainstem auditory evoked potential in cat. III : Identified cell populations*. *Hear Res*, vol. 93, no. 1-2, pages 52–71, Apr 1996. 21, 38, 39
- [Moore 2003] Brian C J Moore et Brian R Glasberg. *Behavioural measurement of level-dependent shifts in the vibration pattern on the basilar membrane at 1 and 2 kHz*. *Hear Res*, vol. 175, no. 1-2, pages 66–74, Jan 2003. 91



- [Moore 2005] B. C. Moore. An introduction to the psychology of hearing, volume 1. Emerald Group Publishing, Bradford, West-Yorkshire, 5th édition, 2005. 5, 15, 22
- [Moore 2008] Brian C J Moore, Lorraine K Tyler et William Marslen-Wilson. *Introduction. The perception of speech : from sound to meaning*. Philos Trans R Soc Lond B Biol Sci, vol. 363, no. 1493, pages 917–921, Mar 2008. 22, 164
- [Musacchia 2006] Gabriella Musacchia, Mikko Sams, Trent Nicol et Nina Kraus. *Seeing speech affects acoustic information processing in the human brainstem*. Exp Brain Res, vol. 168, no. 1-2, pages 1–10, Jan 2006. 31, 32, 33
- [Musacchia 2007] Gabriella Musacchia, Mikko Sams, Erika Skoe et Nina Kraus. *Musicians have enhanced subcortical auditory and audiovisual processing of speech and music*. Proc Natl Acad Sci U S A, vol. 104, no. 40, pages 15894–15898, Oct 2007. 28, 32, 34, 165
- [Musiek 2007] Frank Musiek, Jenny Shinn et Bob Jirsa. Auditory evoked potentials. principles and clinical applications., chapitre AuditoryBrainstemResponses in Auditory Nerve and Brainstem Responses., pages 291–312. Lippincott, Williams et Willkins. Burkard R, Don M, Eggermont JJ (Eds.), Philadelphia, Baltimore, New-York, London, Buenos-Aires, Hong-Kong, Sydney et Tokyo, 2007. 21
- [Ménard 2004] Mikael Ménard, StÉphane Gallego, Eric Truy, Christian Berger-Vachon, John D Durrant et Lionel Collet. *Auditory steady-state response evaluation of auditory thresholds in cochlear implant patients*. Int J Audiol, vol. 43 Suppl 1, pages S39–S43, Dec 2004. 20
- [Ménard 2008] Mikael Ménard, StÉphane GallÉgo, Christian Berger-Vachon, Lionel Collet et Hung Thai-Van. *Relationship between loudness growth function and auditory steady-state response in normal-hearing subjects*. Hear Res, vol. 235, no. 1-2, pages 105–113, Jan 2008. 35, 170
- [Palmer 2004] Alan Palmer et Shihab Shamma. Springer handbook of auditory research (vol. 18). speech processing in the auditory system., chapitre Physiological representations of speech., pages 150–200. Springer. Greenberg S, Ainsworth AW, Fay RR, Popper AN (Eds.), New-York, 2004. 1, 165
- [Patterson 1995] R. D. Patterson, M. H. Allerhand et C. GiguÈre. *Time-domain modeling of peripheral auditory processing : a modular architecture and a software platform*. J Acoust Soc Am, vol. 98, no. 4, pages 1890–1894, Oct 1995. 3, 91
- [Perrot 2006] Xavier Perrot, Philippe Ryvlin, Jean Isnard, Marc GuÈnot, HÈlÈne Catenoux, Catherine Fischer, FranÁois MauguiÈre et Lionel Collet. *Evidence for corticofugal modulation of peripheral auditory activity in humans*. Cereb Cortex, vol. 16, no. 7, pages 941–948, Jul 2006. 6
- [P.Flandrin 1998] P.Flandrin. *Localisation dans le plan Temps-FrÉquence*. Traitement du Signal, vol. 15, pages 483–492, 1998. 185

- [Picton 1974] T. W. Picton, S. A. Hillyard, H. I. Krausz et R. Galambos. *Human auditory evoked potentials. I. Evaluation of components*. *Electroencephalogr Clin Neurophysiol*, vol. 36, no. 2, pages 179–190, Feb 1974. 21
- [Picton 2003a] Terence W Picton, M. Sasha John, Andrew Dimitrijevic et David Purcell. *Human auditory steady-state responses*. *Int J Audiol*, vol. 42, no. 4, pages 177–219, Jun 2003. 20, 55, 90
- [Picton 2003b] Terence W Picton, M. Sasha John, David W Purcell et Gilles Plourde. *Human auditory steady-state responses : the effects of recording technique and state of arousal*. *Anesth Analg*, vol. 97, no. 5, pages 1396–1402, Nov 2003. 30
- [Picton 2004] Terence W Picton et M. Sasha John. *Avoiding electromagnetic artifacts when recording auditory steady-state responses*. *J Am Acad Audiol*, vol. 15, no. 8, pages 541–554, Sep 2004. 26
- [Picton 2007] Terry Picton. *Auditory evoked potentials. principles and clinical applications.*, chapitre *Audiometry using Auditory Brainstem Responses*, pages 441–462. Lippincott, Williams et Willkins. Burkard R, Don M, Eggermont JJ (Eds.), Philadelphia, Baltimore, New-York, London, Buenos-Aires, Hong-Kong, Sydney et Tokyo, 2007. 36, 37
- [Purcell 2004] David W Purcell, Sasha M John, Bruce A Schneider et Terence W Picton. *Human temporal auditory acuity as assessed by envelope following responses*. *J Acoust Soc Am*, vol. 116, no. 6, pages 3581–3593, Dec 2004. 22, 36, 44, 87, 170
- [Rees 1997] A. Rees, A. Sarbaz, M. S. Malmierca et F. E. Le Beau. *Regularity of firing of neurons in the inferior colliculus*. *J Neurophysiol*, vol. 77, no. 6, pages 2945–2965, Jun 1997. 167
- [Regan 1966] D. Regan. *Some characteristics of average steady-state and transient responses evoked by modulated light*. *Electroencephalogr Clin Neurophysiol*, vol. 20, no. 3, pages 238–248, Mar 1966. 36
- [Remez 1981] R Remez. *Speech perception without traditional speech cues*. *Science*, vol. 22, pages 947–949, 1981. 105, 167, 171
- [Ross 2004] B. Ross, T. W. Picton, A. T. Herdman et C. Pantev. *The effect of attention on the auditory steady-state response*. *Neurol Clin Neurophysiol*, vol. 2004, page 22, 2004. 30, 162
- [Russo 2004] Nicole Russo, Trent Nicol, Gabriella Musacchia et Nina Kraus. *Brainstem responses to speech syllables*. *Clin Neurophysiol*, vol. 115, no. 9, pages 2021–2030, Sep 2004. 23, 26, 30, 84, 136
- [Russo 2008] Nicole Russo, Charles Larson et Nina Kraus. *Audio-vocal system regulation in children with autism spectrum disorders*. *Exp Brain Res*, vol. 188, no. 1, pages 111–124, Jun 2008. 26
- [Shannon 2002] Robert V Shannon. *The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design*. *Am J Audiol*, vol. 11, no. 2, pages 124–127, Dec 2002. 10

- [Sheft 2007] S. Sheft, M. Ardoïnt et C. Lorenzi. *Speech identification based on temporal fine structure cues*. J Acous Soc Am, vol. 124, pages 562–575, 2007. 169
- [Smith 1975] J. C. Smith, J. T. Marsh et W. S. Brown. *Far-field recorded frequency-following responses : evidence for the locus of brainstem sources*. Electroencephalogr Clin Neurophysiol, vol. 39, no. 5, pages 465–472, Nov 1975. 22
- [Song 2008] Judy H Song, Erika Skoe, Patrick C M Wong et Nina Kraus. *Plasticity in the adult human auditory brainstem following short-term linguistic training*. J Cogn Neurosci, vol. 20, no. 10, pages 1892–1902, Oct 2008. 30, 163, 165
- [Swaminathan 2008] Jayaganesh Swaminathan, Ananthanarayan Krishnan, Jackson T Gandour et Yisheng Xu. *Applications of static and dynamic iterated rippled noise to evaluate pitch encoding in the human auditory brainstem*. IEEE Trans Biomed Eng, vol. 55, no. 1, pages 281–287, Jan 2008. 31
- [Thai-Van 2004] Hung Thai-Van, Eric Truy, Basile Charasse, Florent Boutitie, Jean-Marc Chanal, Nadine Cochard, Jean-Pierre Piron, SÈbastien Ribas, Olivier Deguine, Bernard Fraysse, Michel Mondain, Alain Uziel et Lionel Collet. *Modeling the relationship between psychophysical perception and electrically evoked compound action potential threshold in young cochlear implant recipients : clinical implications for implant fitting*. Clin Neurophysiol, vol. 115, no. 12, pages 2811–2824, Dec 2004. 23
- [Tillmann 2008] B Tillmann, Rusconi E, Peretz I, Traube C, Butterworth B et Umiltà C. Music, language and the mind, chapitre Pitch discrimination differs in speech and music : Evidence from a case of congenital amusia., page à paraître. Tufts University, Boston, Massachusetts, Etats-Unis, 2008. 123
- [Tomita 2004] Masahiko Tomita, Arnaud J Noreña et Jos J Eggermont. *Effects of an acute acoustic trauma on the representation of a voice onset time continuum in cat primary auditory cortex*. Hear Res, vol. 193, no. 1-2, pages 39–50, Jul 2004. 82
- [van den Honert 1987] C. van den Honert et P. H. Stypulkowski. *Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli*. Hear Res, vol. 29, no. 2-3, pages 207–222, 1987. 16
- [Verpy 2008] Elisabeth Verpy, Dominique Weil, Michel Leibovici, Richard J Goodyear, Ghislaine Hamard, Carine Houdon, Gaëlle M LefÈvre, Jean-Pierre Hardelin, Guy P Richardson, Paul Avan et Christine Petit. *Stereocilin-deficient mice reveal the origin of cochlear waveform distortions*. Nature, Oct 2008. 90
- [VeUILlet 2007] Evelyne VeUILlet, Annie Magnan, Jean Ecalle, Hung Thai-Van et Lionel Collet. *Auditory processing disorder in children with reading disabilities : effect of audiovisual training*. Brain, vol. 130, no. Pt 11, pages 2915–2928, Nov 2007. 87

- [Voutsas 2005] Kyriakos Voutsas, Gerald Langner, Jürgen Adamy et Michael Ochse. *A brain-like neural network for periodicity analysis*. IEEE Trans Syst Man Cybern B Cybern, vol. 35, no. 1, pages 12–22, Feb 2005. 7, 9, 11, 12, 167
- [Wever 1930] Ernest Glen Wever et Charles W Bray. *AUDITORY NERVE IMPULSES*. Science, vol. 71, no. 1834, page 215, Feb 1930. 22, 24, 35
- [Wible 2005] Brad Wible, Trent Nicol et Nina Kraus. *Correlation between brainstem and cortical auditory processes in normal and language-impaired children*. Brain, vol. 128, no. Pt 2, pages 417–423, Feb 2005. 23, 30
- [Wong 2007] Patrick C M Wong, Erika Skoe, Nicole M Russo, Tasha Dees et Nina Kraus. *Musical experience shapes human brainstem encoding of linguistic pitch patterns*. Nat Neurosci, vol. 10, no. 4, pages 420–422, Apr 2007. 28, 29, 34, 165, 183
- [Worden 1968] F. G. Worden et J. T. Marsh. *Frequency-following (microphonic-like) neural responses evoked by sound*. Electroencephalogr Clin Neurophysiol, vol. 25, no. 1, pages 42–52, Jul 1968. 22
- [Xu 1997] Z. M. Xu, E. De Vel, B. Vinck et P. B. van Cauwenberge. *Choice of a tone-pip envelope for frequency-specific threshold evaluations by means of the middle-latency response : normally hearing subjects and slope of sensorineural hearing loss*. Auris Nasus Larynx, vol. 24, no. 4, pages 333–340, Oct 1997. 121
- [Xu 2006] Yisheng Xu, Jackson T Gandour et Alexander L Francis. *Effects of language experience and stimulus complexity on the categorical perception of pitch direction*. J Acoust Soc Am, vol. 120, no. 2, pages 1063–1074, Aug 2006. 28, 31, 34, 111
- [Zatorre 2008] R Zatorre et JT Gandour. *Specialization of speech processing, moving beyond the dichotomies*. Philos Trans R Soc Lond B Biol Sci, vol. 363, no. 1493, pages 947–963, Mar 2008. 31, 123
- [Zeng 2004] Fan-Gang Zeng. *Trends in cochlear implants*. Trends Amplif, vol. 8, no. 1, pages 1–34, 2004. 14, 16

---

**Résumé :** Les ABR sont un outil prometteur d'exploration fonctionnelle auditive, qui pourrait à terme être implémenté en routine clinique afin de dépister la surdité et d'évaluer l'efficacité des techniques de réhabilitation auditive. Le contexte scientifique et une présentation de l'état de l'art est proposée dans le *premier chapitre*. Dans le *second chapitre*, nous la réponse Speech ABR est authentifiée neurophysiologiquement au moyen (i) d'une étude paramétrique quantifiant les effets de l'artefact électromagnétique, et (ii) en montrant que le Speech ABR n'existe que si le stimulus a bien été entendu. Puis, dans le *troisième chapitre* nous discutons les indices acoustiques influençant les Speech ABR. L'étude pilote décrit le Speech ABR à /ba/ comme une succession d'une réponse transitoire et soutenue (Frequency Following Response, FFR), puis nous montrons que l'enveloppe temporelle a moins d'impact que la structure fine (et en particulier la structure harmonique) dans la génération de la FFR. Nos données (préliminaires) ne semblent pas suggérer de sensibilité à la nature du stimulus. En revanche, le niveau sonore et la fréquence du stimulus ont des conséquences directes sur le Speech ABR.

**Mots clés :** Electrophysiologie, exploration audiolgique objective, tronc cérébral.

---

### Speech auditory brainstem response (Speech ABR)

**Abstract :** Here, we present a two-fold study of Speech Auditory Brainstem Response : first, in *chapter two*, it is shown that Speech ABR is a genuine neural response, and second, in *chapter 3*, we investigate the different acoustic impact of the stimuli (temporal envelope and fine structure, harmonic content, intensity, frequency, nature of the stimulus) on Speech ABR. Before that, the scientific context, and an overview of existing studies is given in *chapter 1*.

**Keywords :** Electrophysiology, Brainstem, electromagnetic artefact, temporal and spatial neural code, neurophysiologic generators, unilateral deafness.

---

---