AIX-MARSEILLE UNIVERSITE

FACULTE DE MÉDECINE DE MARSEILLE

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE – ED 62

T H È S E

Présentée et publiquement soutenue devant

LA FACULTÉ DE MÉDECINE DE MARSEILLE

Le 20 Septembre 2016

Par Bastien INTARTAGLIA

Effet de l'expertise et de l'attention sur l'encodage neuronal des sons de parole

Pour l'obtention du DOCTORAT d'AIX-MARSEILLE UNIVERSITÉ

Spécialité : NEUROSCIENCES

Réalisé à l'Institut de Neurosciences des Systèmes

sous la direction de

Daniele Schön

Membres du Jury de la Thèse :

Anne Caclin, Chargée de recherche, Université de Lyon	Rapporteur
Simone Dalla-Bella, Professeur, Université de Montpellier	Rapporteur
Pascal Belin, Professeur, Aix-Marseille Université	Examinateur
Daniele Schön, Directeur de recherche, Aix-Marseille Université	Directeur de thèse

Résumé

La perception auditive, longtemps considérée comme un phénomène passif, apparaît aujourd'hui comme un processus dynamique caractérisé par des interactions continues à double sens entre les systèmes corticaux et sous-corticaux, entre les systèmes cognitifs et sensoriels. Les récentes découvertes suggèrent que l'engagement actif pendant la perception des sons entraîne des réorganisations neuronales à plusieurs niveaux du système auditif et à plusieurs échelles temporelles. Les travaux réalisés durant cette thèse se sont focalisés sur les effets de plasticité induits au niveau sous-cortical : à long terme, par l'étude des effets de l'expertise (linguistique et musicale) et à court terme par l'étude des effets attentionnels. Notre hypothèse était que l'expertise linguistique renforce l'encodage automatique des sons natifs au niveau des indices linguistiques les plus importants.

Les réponses complexes du tronc cérébral (*complex Auditory Brainstem Responses* ou *cABRs*) se sont avérées plus robustes pour les sons natifs avec un renforcement spécifique de l'encodage des fréquences formantiques qui jouent un rôle important dans la discrimination des phonèmes, ce qui confirme l'hypothèse selon laquelle l'expertise linguistique induit une réorganisation des structures sous-corticales et que cette plasticité se produit en fonction de ce qui est important pour le locuteur. Un corpus d'études important montre des effets de transfert entre musique et langage. Nous avons montré que les différences d'expertise linguistique sont compensées par la pratique musicale grâce à un renforcement de l'encodage neuronal des indices linguistiques importants (formants). Puisque ces effets de plasticité se produisent de manière spécifique au niveau des dimensions acoustiques importantes, nous avons postulé qu'à plus court terme, des processus attentionnels *top-down* moduleraient les réponses sous-corticales de manière à induire un renforcement sélectif du codage neuronal. La comparaison des réponses neuronales recueillies en condition d'écoute passive et active (*tapping*) a révélé une représentation de la fréquence fondamentale plus robuste lors de l'écoute active des sons, confirmant l'influence des processus *top-down* sur l'encodage sous-cortical des sons de parole. Le design expérimental de cette étude nous a aussi permis d'étudier les modulations dynamiques et rapides de l'attention. Plus précisément, en entraînant l'attention au rythme d'un stimulus sur deux via une tâche de *tapping*, nous avons observé une modulation significative de la représentation du son en fonction de la phase du *tapping*, révélant ainsi les effets d'une dynamique oscillatoire (attentionnelle) au niveau de la voie auditive sous-corticale.

En conclusion, ce travail de thèse a permis de révéler la nature dynamique de la perception auditive, en montrant une modulation de l'encodage des sons au niveau sous-cortical à plusieurs échelles temporelles : à long terme via les effets de l'expertise et à court terme via les effets de l'attention. Les réponses neuronales sous-corticales sont probablement modulées via des boucles de rétrocontrôle entre le cortex et les structures sous-corticales. La communication entre ces deux niveaux de traitement pourrait se faire par des mécanismes de couplage entre la phase des ondes corticales lentes et l'amplitude des réponses du tronc cérébral. Néanmoins, de futures études sont nécessaires pour explorer cette hypothèse.

Abstract

Auditory processing, initially viewed as a passive phenomenon, is now considered as a dynamic process characterized by continuous interactions between cortical and subcortical structures, between cognitive and sensory systems. Recent findings suggest that the active engagement with sounds shapes the auditory system at different levels and timescales. The work of this thesis focused on the effects of plasticity at the brainstem level: over a long-timescale, by studying experience-dependent plasticity (linguistic and musical experience) and over a short-timescale, by studying the effect of attention. We hypothesized that linguistic experience would strengthen the automatic processing of native sounds especially on linguistically-relevant dimensions.

We recorded complex auditory brainstem responses (cABRs) in American English and French native speakers, listening to natural syllables of both languages. Neural representations were more robust for the native sounds, with a specific neural enhancement of formant frequencies that play an important role in phonemes discrimination. Furthermore, American participants exhibited greater neural representation of the fundamental frequency, consistent with its importance in English to convey stress patterns. These findings support the hypothesis that language experience shapes subcortical processing of speech and that this plasticity is sensitive to what is relevant for the listener. The transfer effects between music and speech has been extensively studied. Therefore, in a subsequent study we investigated the effect of musical expertise on subcortical processing of foreign speech sounds by comparing subcortical responses to an English syllable between three groups: American nonmusicians, French nonmusicians, and French musicians. Results showed that language differences are compensated by musical expertise in non-native musicians by strengthening the encoding of the fundamental frequency and the formants. Because experience-dependent plasticity occurs on specific acoustic dimensions that are relevant for the listener, we proposed that, on a shorter-timescale, attentional top-down processes may modulate subcortical encoding of speech sounds in order to induce selective enhancements. Comparing neural responses recorded under passive and active (tapping) listening conditions revealed an enhancement of the fundamental frequency during active listening, which is consistent with our prediction that subcortical processing is modulated by top-down processes. Moreover, by entraining the attentional cycles with the phase of the tapping (one out two stimuli), we observed that sound representations were significantly modulated at the same pace, revealing an effect of a fast oscillatory (attentional) dynamic over subcortical auditory processing.

To conclude, this work revealed the dynamic nature of auditory perception, showing that subcortical encoding of speech sounds is modulated at different timescales : on a long-term by experience-dependent plasticity, and on a short-term by attentional modulations. Subcortical responses are probably modulated via feedforward loops from cortical to subcortical structures. Phase-amplitude coupling could be the neural mechanism of communication between these distant neural structures. However, further studies are needed to explore this hypothesis.

Remerciements

A travers ces quatre années de thèse, j'ai rencontré de nombreuses personnes qui m'ont accompagné et aidé à réaliser ce travail, mais surtout qui ont su enrichir ma vie sur tous ses aspects, et je tiens ici à les remercier sincèrement.

Je voudrais d'abord remercier chaleureusement mon directeur de thèse Daniele Schön, pour m'avoir fait confiance tout au long de cette thèse, pour m'avoir transmis certaines de ses qualités professionnelles et humaines (pas toutes restons humbles), pour son soutien inconditionnel, et pour sa bonne humeur sans faille qui a rendu la réalisation de ce travail bien plus agréable.

Je remercie également Catherine Liégeois-Chauvel qui a toujours gardé un œil bienveillant sur moi dès mon arrivée dans le laboratoire, et sans qui ce travail de thèse n'aurait sûrement pas pu aboutir.

Je remercie les membres de mon jury de thèse d'avoir accepté d'évaluer ce travail de recherche et de venir sur Marseille pour apporter leurs jugements scientifiques, de qualité, sur ce projet.

Je voudrais adresser mes remerciements à l'ensemble des membres du laboratoire de l'Institut de Neurosciences des Systèmes. A Agnès qui a su m'initier au monde de la recherche, qui m'a encadré pendant de nombreuses années, et dont j'ai pu apprécier les qualités humaines et professionnelles. Patrick Marquis allias PPM pour tes compétences inégalables en électrophysiologie qui ont apporté une aide précieuse dans la réalisation de mes expériences, pour toutes les discussions diverses et variées que nous avons pu partager au cours de ces années. Une mention spéciale à Anaïs, que j'ai toujours considéré comme ma grande sœur de thèse, qui m'a vu « grandir » intérieurement et extérieurement pendant toutes ces années, qui m'a guidé dans mes choix de vie et avec qui j'ai pu développer une amitié sincère. Merci à Mina pour ta bonne humeur quotidienne et pour ta personnalité hors du commun. Un merci général à tous ceux que j'ai pu croiser dans ce laboratoire, Hester, Rohit, Pascale, Clément, Julien, Céline, Mylène, Nia, Sophie, Marie, Rachel, Alexia, et à tous ceux que j'oublie ! Merci aussi à Véronique pour m'avoir aidé dans toutes les démarches administratives et pour ta gentillesse.

Je voudrais remercier aussi tous les membres de l'Auditory Neuroscience Lab qui m'ont accueilli pendant mon séjour à Chicago et avec qui j'ai pu développer une collaboration fructueuse.

Merci aux infirmières du service de neurophysiologie clinique pour leur accueil et pour leur sympathie.

Un grand merci à tous les participants qui ont su résister à la torture physique et psychologique de mes expériences.

Je ne peux pas oublier tous mes amis avec qui j'ai pu partager les hauts et les bas de ma thèse : JB, Nico, Rakib, Frédo, Morgan, Serge, Sahra et tout le reste de la promo du Master Neuro. Cécile je n'oublierai pas ton soutien pendant toutes ces années que nous avons partagées.

Enfin, un immense merci à mes parents Geneviève et Daniel ainsi qu'à ma sœur Marine, pour leur soutien et leur amour inconditionnel. Et pour finir en beauté, au sens propre comme au figuré, merci à Manto pour tout le bonheur que tu m'as apporté.

Table des matières

Première partie - INTRODUCTION, OBJECTIFS et HYPOTHESES 1

Introduction 3

1. Les réponses évoquées du tronc cérébral auditif (*ABRs*) 3

1.1. ABRs évoquées par des clics (*click-ABRs*) 3

1.2. ABRs évoquées par des sons complexes (*complex-ABRs* ou *cABRs*) 7

1.2.1. Similarité stimulus-réponse 8

1.2.2. Réponses transitoires et soutenues 10

2. Plasticité du système auditif 13

2.1. Cadre théorique 13

2.1.1. Une plasticité à tous les niveaux 15

2.1.2. Une réorganisation à plusieurs vitesses 21

2.2. Plasticité dépendante de l'expérience 21

2.2.1. Expérience musicale 22

2.2.2. Expérience linguistique 34

2.3. Plasticité dépendante du contexte 43

2.3.1. Effets du contexte attentionnel 43

2.3.2. *Active sensing* et *Dynamic Attending Theory* 45

Objectifs et hypothèses 51

Deuxième partie - MATERIELS et METHODES 53

Matériels et méthodes 55

3. Choix du stimulus 55

3.1. Sons artificiels ou naturels 55

3.2. Sons de parole et de musique 57

3.3. Composantes acoustiques transitoires et continues 58

4. Présentation du stimulus 59

4.1. Matériel de stimulation 59

4.2. Intensité 61

4.3. Cadence de stimulation 62

4.4. Polarité du stimulus 63

4.5. Stimulation monaurale et binaurale 65

4.6. Transducteurs ... 66
5. Enregistrements électrophysiologiques ... 67
5.1. Montage ... 67
5.2. Fréquence d'échantillonnage ... 69
5.3. Nombre d'essais ... 69
5.4. Fenêtre d'enregistrement ... 70
5.5. Conditions d'enregistrement passives et actives ... 71
6. Détection et réduction des artéfacts ... 72
6.1. Courant de secteur (50 ou 60 Hz) ... 72
6.2. Artéfact du stimulus ... 73
6.3. Microphone cochléaire ... 75
6.4. Artéfacts musculaires ... 75
7. Traitement des données ... 76
7.1. Données brutes ... 76
7.2. Rejet des essais bruités ... 77
7.3. Filtrage ... 78
7.4. Moyennage ... 78
7.5. Rapport signal/bruit (*SNR*) ... 79
8. Analyses des données ... 80
8.1. Réponses transitoires (*onset, offset,...*) ... 80
8.1.1. Latence et amplitude des pics ... 80
8.1.2. Cross-phaseogramme ... 81
8.2. Réponses continues (*FFR*) ... 83
8.2.1. Corrélation stimulus-réponse ... 83
8.2.2. Autocorrélation ... 84
8.2.3. Transformée de Fourier (FFT) ... 87
8.2.4. Consistance temporelle des *cABRs* (*neural response consistency*) ... 90
8.2.5. Cohérence de phase inter-essais (*inter-trial phase-coherence*) ... 91
9. Forces et limites de la technique des *cABRs* ... 93

Troisième partie - TRAVAUX EXPERIMENTAUX ... 95

Axe « Effet de l'expertise linguistique et musicale » ... 97

10. Effet de l'expertise linguistique ... 98
11. Effet de transfert entre musique et langue étrangère ... 125

Axe « Effet de l'attention » 147

Quatrième partie - DISCUSSION et PERSPECTIVES 179

Discussion générale et perspectives 181

12. Résumé des résultats 181

13. Discussion 182

14. Perspectives 186

14.1. Tentative d'unification des effets attentionnels à court terme et des effets d'expertises à long terme 186

14.2. Mécanismes neuronaux potentiels des modulations attentionnelles *top-down* 189

Conclusion 193

Références bibliographiques 195

Première partie - INTRODUCTION, OBJECTIFS et HYPOTHESES

Introduction

1. Les réponses évoquées du tronc cérébral auditif (*ABRs*)

1.1. ABRs évoquées par des clics (*click-ABRs*)

En 1971, un article paru dans le journal *Brain* a révolutionné le monde clinique de l'évaluation audiométrique (Jewett and Williston, 1971). A l'aide de trois électrodes placées sur le scalp des participants, Don Jewett at John Williston ont enregistré pour la première fois, chez l'Homme, des *potentiels évoquées auditifs* (*PEA*) précoces (<10 ms) en réponse à des clics. Ces PEA précoces, aussi appelés *click-ABRs*, sont une série de 7 ondes, numérotées de I à VII avec des latences allant de 1.4 à environ 10 ms. En comparant la forme et la latence de ces ondes à celles déjà reportées dans la littérature chez l'animal, les auteurs ont conclu que l'onde I reflétait l'activité du nerf auditif, tandis que les ondes II à VII reflétaient l'activation séquentielle des noyaux du tronc cérébral auditif par lesquels est relayé l'influx nerveux avant d'atteindre le cortex. Cette étude a montré de ce fait que l'on pouvait évaluer de manière non-invasive (électrodes de scalp) et objective (i.e. sans la participation active sur sujet), l'intégrité fonctionnelle des voies sous-corticales du système auditif. Suite à ces travaux, cette méthode d'évaluation auditive a été intégrée comme routine par les praticiens ORL afin de réaliser des diagnostics de surdité précoce, en particulier chez les nouveau-nés qui sont encore incapables de communiquer.

Afin d'enregistrer les réponses du tronc cérébral auditif aux clics ou *click-ABRs*, le clinicien pose trois électrodes de scalp (vertex, Fz et lobe de l'oreille ou mastoïde). Dans un montage vertical, c'est le différentiel d'activité entre le vertex et l'oreille qui permet de recueillir les réponses provenant du tronc cérébral auditif. Les *click-ABRs* sont de très faible amplitude (<1 μV) puisqu'ils sont enregistrés à distance de leurs générateurs neuronaux, il faut donc moyenner un grand nombre d'essais (de quelques centaines à quelques milliers) afin de les faire émerger du bruit de fond. En effet, l'activité électrique corticale, musculaire ou provenant des appareils électroniques environnant (téléphones, ordinateurs, prises électriques,...) peuvent parasiter le signal enregistré par l'électrode. Néanmoins, cette activité parasite survient de manière aléatoire

par rapport à la présentation du stimulus, à l'inverse de l'activité des relais sous-corticaux qui est en phase avec la présentation du stimulus. En moyennant un grand nombre d'essais, l'activité parasite s'annule tandis que l'activité sous-corticale émerge du bruit de fond, ce qui permet d'observer l'activité sous-corticale, bien que celle-ci soit de faible amplitude.

Chez un même sujet, les *click-ABRs* sont très reproductibles, en particulier l'onde V qui semble la plus robuste aux différents paramètres de stimulation (cadence de stimulation, nombre d'essais moyennés, etc...). La **figure 1a** est un exemple de mesure effectuée chez un normo-entendant, sur laquelle les cinq ondes, numérotées en chiffre romains de I à V, sont facilement identifiables. En évaluant la latence de l'onde V en fonction de l'intensité de stimulation, et en comparant les valeurs obtenues à des valeurs normatives, le praticien ORL est donc en mesure de réaliser une évaluation audiométrique de manière objective. Ainsi, un patient sera considéré comme normo-entendant si la latence de l'onde V se situe à l'intérieur des limites normales, alors qu'un patient présentant des pertes auditives aura un retard ou une absence d'onde V pour une intensité donnée (**Figure 1b**). Les ondes plus précoces, en particulier l'onde I et III permettent un diagnostic plus fin. La présence ou l'absence des ces ondes, ainsi que l'intervalle de temps entre l'onde I et III, III et V, et I et V donnent des indications sur l'origine possible des pertes auditives, ce qui permet par exemple de distinguer les pathologies provenant de l'oreille interne de celles provenant de l'oreille moyenne.

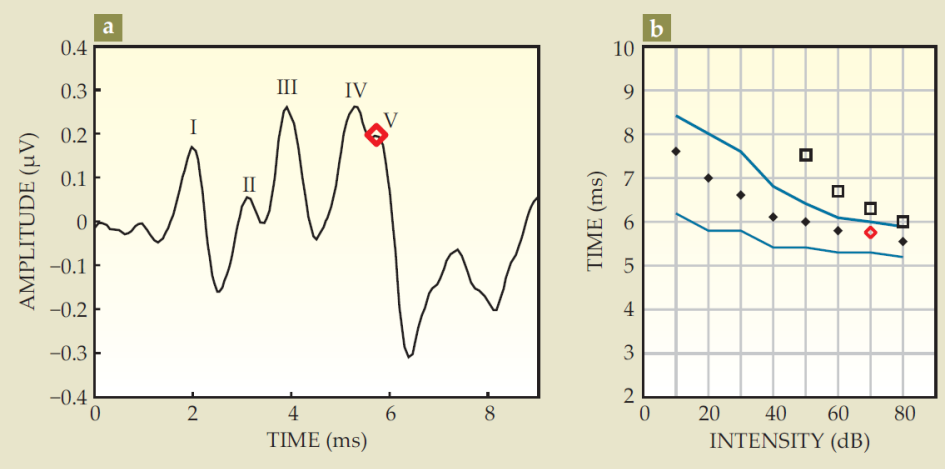


Figure 1. Potentiels évoqués auditifs du tronc cérébral en réponse à des clics. (**a**) Mesure effectuée chez un normo-entendant, avec un clic présenté à l'oreille droite à 70 décibels (dB). Les ondes sous-corticales sont numérotées en chiffre romain de I à V. Le losange rouge indique le pic de l'onde V à 5.75 ms après la présentation du son. (**b**) Latence de l'onde V en fonction de l'intensité de stimulation chez un normo-entendant (losange noir) et chez un patients présentant des pertes auditives (carrés noirs). Le losange rouge correspond à la mesure du panel a. Les lignes bleues représentent l'intervalle temporel pour les individus normo-entendants. En dessous de 50 dB, l'onde V n'est plus visible chez le patient malentendant (Kraus, 2011).

La technique des *click-ABRs* est de nos jours très répandue en clinique puisqu'elle présente de nombreux avantages : les réponses neuronales sont faciles et rapides à obtenir (quelques minutes seulement). C'est une technique non-invasive (électrodes de scalp) qui peut être réalisée même durant le sommeil, ce qui permet de réaliser facilement un diagnostic chez les nouveau-nés par exemple. Les potentiels évoqués auditifs ont été identifiés comme marqueurs biologiques de la maturation du système auditif (Eggermont, 1995), et sont aussi sensibles aux troubles auditifs (Tillein et al., 2012). En enregistrant les ondes I à V et les potentiels corticaux P2 et N2 à différents âges, Eggermont (1995) a montré que les nouveaux nés présentent des latences plus longues par rapport aux adultes, suggérant une maturation et un développement des fonctions auditives durant la période post-natale. En effet, la latence de ces différentes composantes neuronales diminue en fonction de l'âge avec une durée de maturation s'allongeant de manière graduelle de la périphérie vers les voies centrales. La latence de l'onde I d'un nouveau-né étant proche de celle d'un adulte alors que l'onde V et les composantes plus tardives mettront environ 6 ans pour atteindre un état de maturité.

Bien que la méthode des *click-ABRs* soit utile dans l'évaluation des fonctions auditives, elle ne fournit néanmoins qu'une information partielle de type *on-off* sur le traitement des sons au niveau sous-cortical. Or, les troubles auditifs touchent souvent une partie restreinte du spectre audible (e.g. les hautes fréquences dans la presbyacousie), ce qui reste difficile à mettre en évidence avec les clics dont le contenu acoustique est pauvre.

Par la suite, les chercheurs ont donc utilisé des sons plus complexes tels que des ondes sinusoïdales pures de différentes fréquences afin d'étudier plus finement l'encodage spectral sous-cortical. En utilisant des sons purs de fréquences allant de 250 Hz à 2 kHz, Moushegian et al. (1973) ont montré que l'activité des neurones sous-corticaux se synchronise sur la fréquence du stimulus. En raison de cette propriété de suivi de fréquence, les ABR évoqués par des sons sinusoïdaux sont appelés *Frequency-Following Responses* (*FFR*) ou *Auditory Steady-State Responses* (*ASSR*). La réponse sous-corticale à un son sinusoïdale est aussi une sinusoïde ou quasi-sinusoïde (dû à la non-linéarité de l'encodage). Ainsi, la transformée de Fourier de la *FFR* révèlera des composantes fréquentielles très similaires à celles du stimulus, telles que la fréquence fondamentale (*F*0) et ses harmoniques (**Figure 2**).

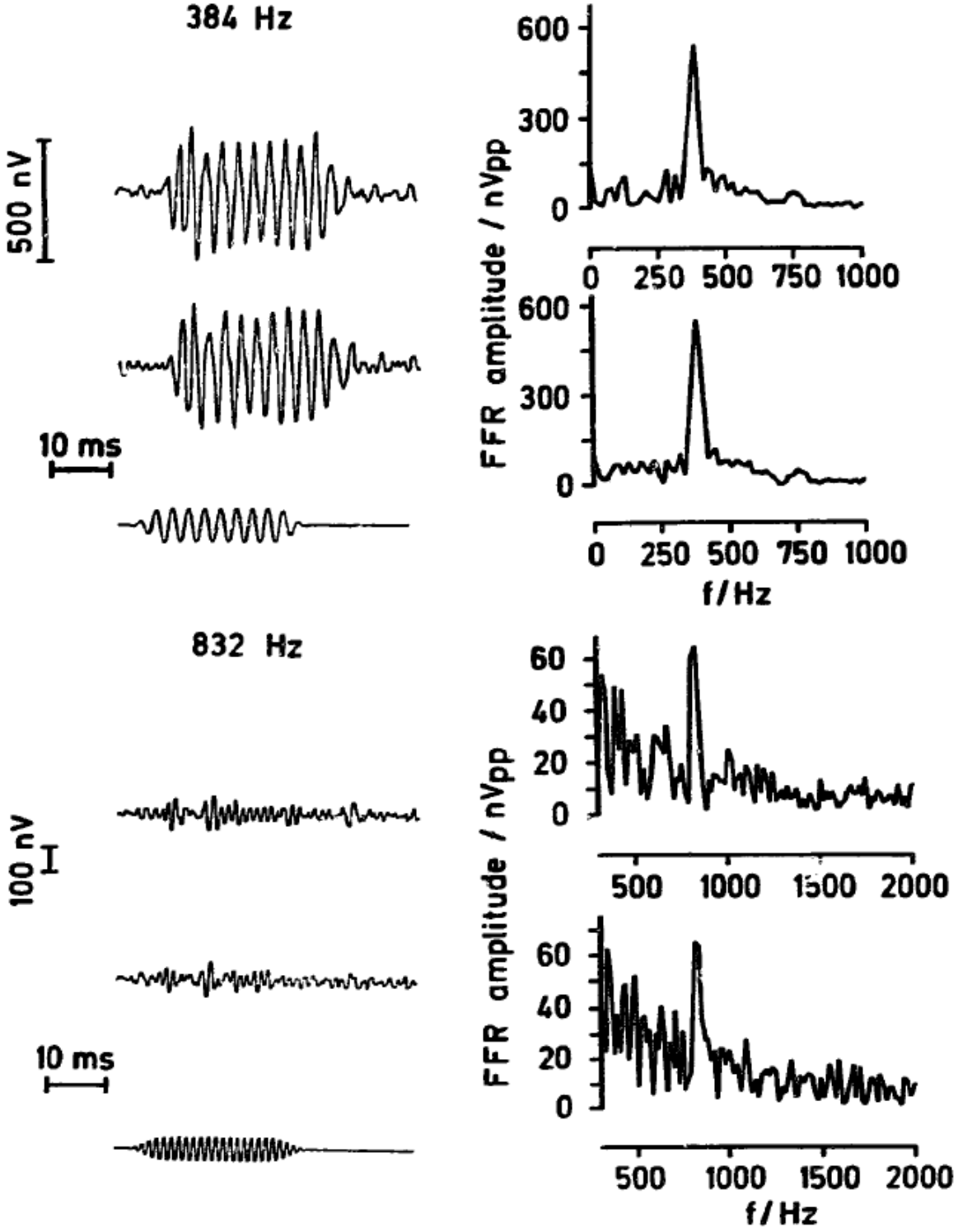


Figure 2. FFR évoquées par un tone-burst à 384 Hz (panels hauts) et à 832 Hz (panels bas) dans le plan temporel (panels gauches) et spectral (panels droits) pour un sujet sur deux enregistrements successifs. Le troisième tracé de chaque panel de gauche représente le stimulus (Hoormann et al., 1992).

La **figure 2** montre en effet que la réponse est très similaire à l'onde sonore qui l'évoque, et la FFT de la réponse révèle un pic correspondant à la fréquence fondamentale présente dans le stimulus. La *FFR* est une mesure robuste comme le montrent les deux tracés et FFT enregistrés pour un même sujet. Cette approche fût donc adoptée en complément des *clicks-ABRs* afin d'évaluer les seuils auditifs de manière plus spécifiques, et mettre ainsi en évidence des troubles auditifs spécifiques à certaines fréquences (Rance and Rickards, 2002). Bien que les sons sinusoïdaux et tone-bursts soient utilisés depuis longtemps afin d'évaluer le seuil auditifs, ces stimuli restent néanmoins peu représentatifs des sons naturels auxquels nous sommes confrontés au quotidien, ils ne permettent pas d'évaluer le traitement fin des sons au niveau du tronc cérébral tel que celui qui se produit dans un environnement sonore naturellement plus complexe (sons de parole, musique, sons de l'environnement). La communauté scientifique et clinique s'est donc peu à peu tournée vers des stimuli de plus en plus complexes et naturels qui semblent offrir de nouvelles perspectives.

1.2. ABRs évoquées par des sons complexes (*complex-ABRs* ou *cABRs*)

L'utilisation de sons complexes afin d'évoquer les *ABRs* a ouvert la voie à de nouveaux champs d'étude et d'analyse qui ont considérablement enrichi nos connaissances du système auditif. Pour se rapprocher au mieux des conditions d'écoute écologiques, les sons utilisés pour évoquer les réponses du tronc cérébral se sont complexifiés : des ondes sinusoïdales pures, en passant par les tones-burst (hybride entre un clic et un son pur), pour arriver aux sons verbaux (consonnes, voyelles, syllabes, mots), musicaux (notes, courtes mélodie), et même aux sons de l'environnement (Krishnan, 2002, 1999; Musacchia et al., 2007; Strait et al., 2009). Les réponses évoquées du tronc cérébral aux sons complexes ou *complex ABRs* (*cABRs*), fournissent de nombreuses informations sur la qualité de l'encodage temporel et spectral des sons au niveau du tronc cérébral.

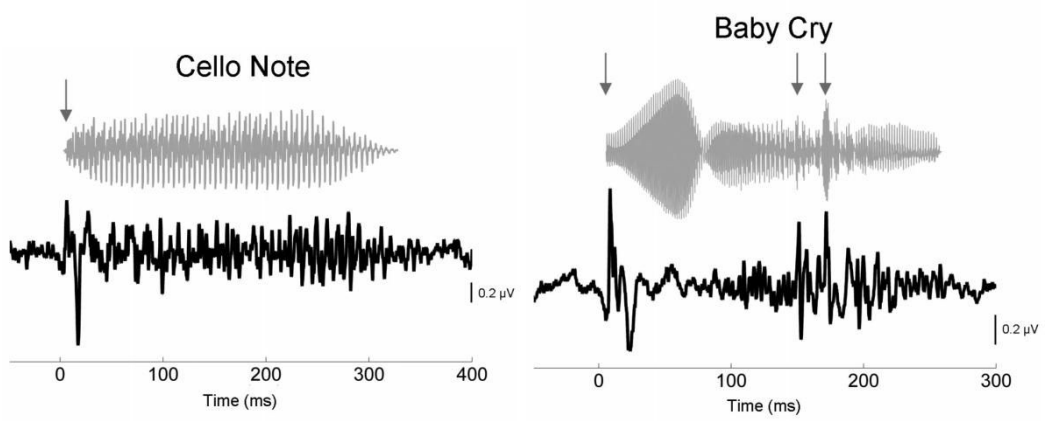


Figure 3. Exemples de *cABRs* (tracés noirs) évoqués par des sons complexes (tracés gris). Afin de maximiser la ressemblance visuelle entre stimulus et réponse, le tracé des stimuli a été décalé dans le temps pour débuter à l'*onset* de la réponse. Les flèches indiquent les principales caractéristiques transitoires qui sont représentées sous forme de pics amples dans les réponses neuronales. **Gauche** : *cABR* évoquée par un note de violoncelle (G2, 100 Hz). Le contact entre l'archet et les cordes du violoncelle déclenche une réponse de type *onset*, alors que la vibration périodique de la corde qui suit entraîne une réponse soutenue. Aucune réponse *offset* n'est observée du fait de la diminution progressive de l'intensité sonore. **Droite** : les variations spectro-temporales abruptes entraînent des réponses de type transitoires, similaires aux réponses *onset*. Adaptée à partir de (Skoe and Kraus, 2010).

1.2.1. Similarité stimulus-réponse

Une des propriétés les plus étonnantes des *cABRs* est leur grande fidélité au stimulus, c'est-à-dire que la forme de l'onde neuronale est très similaire à celle du signal acoustique qui l'évoque. L'activité des neurones sous-corticaux reproduit avec précision les variations temporelles et spectrales présentent dans le stimulus. Lorsque l'on superpose les deux ondes, avec un décalage temporel correspondant aux délais de conduction neuronal entre la cochlée et le tronc cérébral, la ressemblance entre les deux ondes est encore plus flagrante (**Figure 4**). Cette similarité entre stimulus et réponse neuronale est telle que, lorsque l'onde neuronale est convertie en onde sonore, il est possible de reconnaître le stimulus qui l'a évoqué (Galbraith et al., 1995). De ce fait, il est possible d'étudier les *cABR* avec des outils d'analyse proches de ceux utilisés dans le traitement du signal sonore.

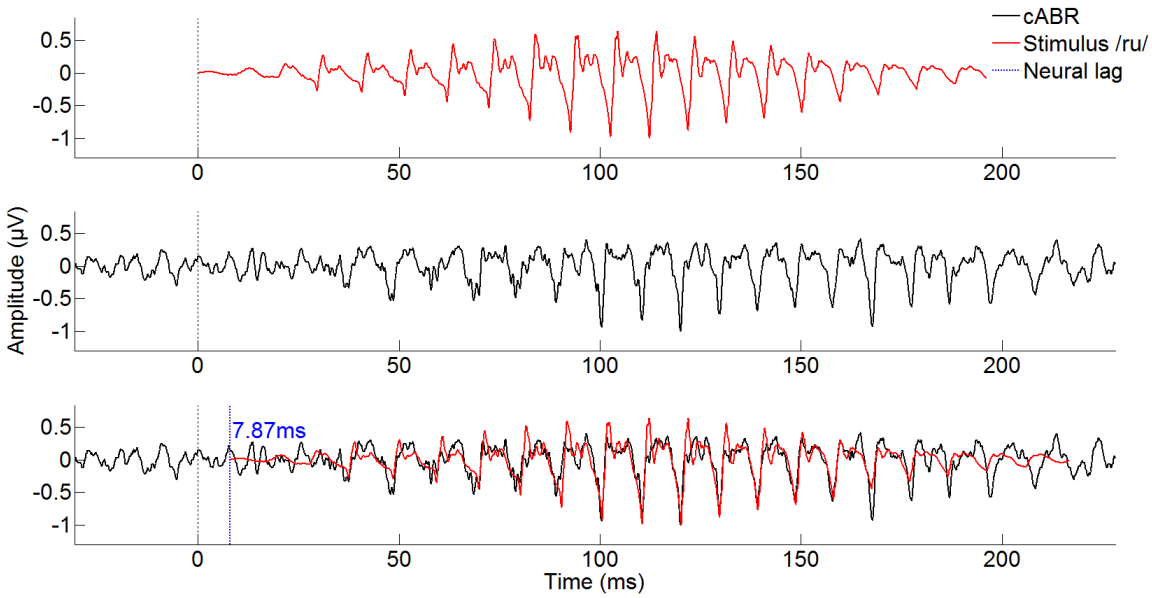


Figure 4. Similarité temporelle entre stimulus et réponse. **Haut** : signal acoustique de la syllabe naturelle /ru/. **Milieu** : *cABR* évoquée par le stimulus /ru/ chez un sujet français issu d'Intartaglia et al. (2016). **Bas** : superposition du stimulus (tracé rouge) et de la réponse (tracé noir). Afin de maximiser la ressemblance entre les deux signaux, le tracé du stimulus a été décalé dans le temps d'un délai de 7.87 ms, correspondant au délai neuronal entre l'onset du stimulus et l'onset de la réponse estimé à partir du niveau de corrélation maximum entre les deux tracés. L'amplitude du stimulus a été normalisée en fonction de l'amplitude maximale de la réponse. Une réponse de type *onset* est observée à 7.87 ms, suivie d'une réponse soutenue, de type *FFR* correspondant à fréquence fondamentale présente dans le stimulus (~100 Hz).

Dans le plan temporel, il est par exemple possible de mesurer l'indice de corrélation stimulus-réponse qui renseigne sur la fidélité de la réponse par rapport au stimulus. Dans le plan spectral, le calcul de la transformée de Fourier (FFT) du stimulus et de la réponse neuronale permet d'apprécier la qualité de l'encodage neuronal des différentes fréquences contenues dans le stimulus (**Figure 5**). Cette dernière analyse est particulièrement bien adaptée à l'analyse de la *FFR* dont le contenu spectral est proche de celui du stimulus, et pour laquelle le calcul de la FFT est plus robuste car se faisant sur la partie stationnaire du signal.

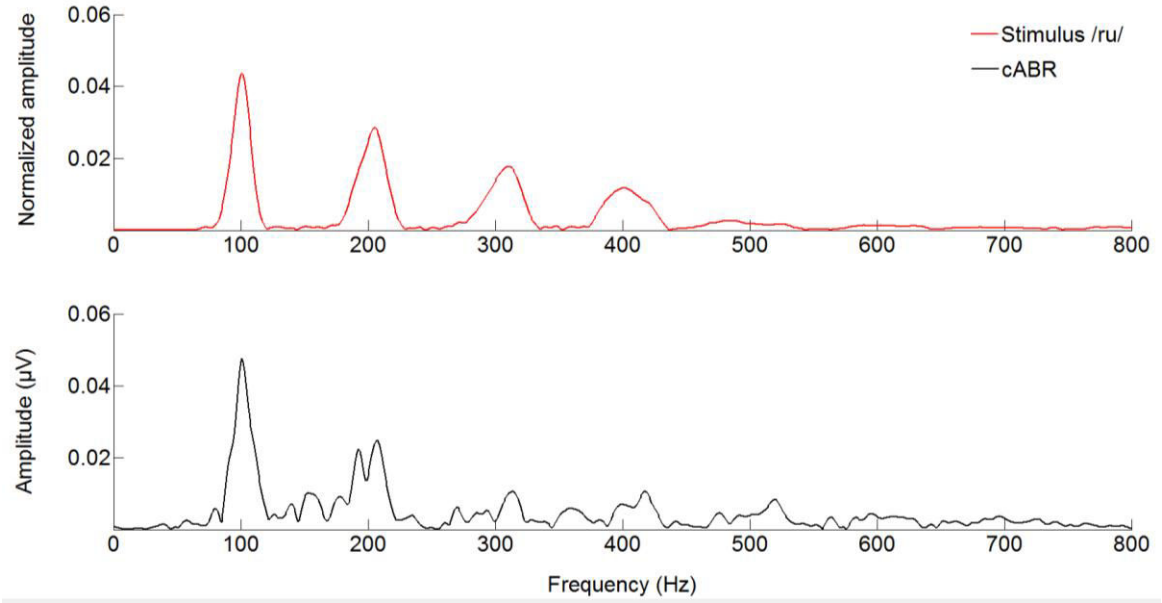


Figure 5. Transformées de Fourier (FFT) du stimulus /ru/ (tracé rouge) et de la réponse sous-corticale évoquée (tracé noir) chez le même sujet que celui de la figure 4 (Intartaglia et al., 2016). La FFT a été calculée sur la durée totale du stimulus, et sur la même durée dans la réponse en incluant un délai neuronal de 7.87 ms. L'amplitude du stimulus a été normalisée en fonction de l'amplitude maximale de la réponse. L'analyse spectrale révèle une grande similarité fréquentielle entre stimulus et réponse neuronale.

1.2.2. Réponses transitoires et soutenues

Les *cABR* comportent deux classes de réponses : les réponses transitoires et soutenues. Les réponses transitoires sont évoquées par des stimuli apériodiques et de courte durée. Par exemple, le début et la fin d'un son peuvent engendrer des réponses transitoires de type *onset* et *offset* respectivement (voir **Matériels et Méthodes Figure 25** pour plus de détails). La transition spectro-temporelle rapide entre une consonne et une voyelle peuvent également engendrer des réponses transitoires. Ce sont les caractéristiques d'attaque temporelle qui vont déterminer la forme et la robustesse des réponses transitoires des *cABRs*. Ainsi, les sons présentant une augmentation abrupte d'intensité sonore, tel que les clics, les consonnes occlusives (e.g. /d/ ou /t/), ou les notes d'instruments à percussion (e.g. piano, batterie), vont engendrer une stimulation plus synchrone de la cochlée, ce qui va déclencher la décharge synchrone d'une large population de neurones. Cela évoque une réponse transitoire plus ample et plus précoce que des sons présentant des caractéristiques d'attaque plus lentes, tels que les consonnes nasales (e.g. /m/, /n/) ou fricatives (e.g. /r/, /f/), ou les notes d'instruments à cordes frottées (e.g. violon,

violoncelle). Les flèches de la **figure 3** indiquent les caractéristiques transitoires de deux sons complexes qui se traduisent par des pics de haute amplitude dans les réponses sous-corticales.

A l'inverse, les réponses soutenues sont évoquées par des sons périodiques et de longue durée, comme les voyelles, les notes de musique soutenues et les sons sinusoïdaux. Comme dans le cas des sons purs décrits précédemment, l'activité des neurones du tronc cérébral va se synchroniser sur les mêmes fréquences que celles présentes dans le stimulus. Cette propriété, appelée verrouillage de phase ou *phase-locking* confère aux réponses du tronc cérébral une grande similarité avec le stimulus qui l'évoque. Il est cependant important de noter qu'il existe une limite au-delà de laquelle les neurones ne parviennent plus à suivre la cadence imposée par le stimulus, limite qui se situe aux alentours de 1.5-2 kHz (Aiken and Picton, 2008; Krishnan, 2002; Moushegian et al., 1973). Le tronc cérébral se comporte comme un filtre passe-bas, qui va amplifier les basses fréquences, en particulier la fréquence fondamentale, au détriment des hautes fréquences. Néanmoins, au-delà de cette limite de *phase-locking*, le codage des fréquences reposerait sur un *place-code*, c'est-à-dire que la distinction des hautes fréquences se ferait selon le lieu de stimulation de la cochlée qui va engendrer des décalages temporels différents des réponses neuronales selon la fréquence (Chandrasekaran and Kraus, 2010). En effet, la cochlée est organisée de manière tonotopique : les hautes fréquences sont traitées à la base de la cochlée par laquelle arrivent les stimulations sensorielles, tandis que les basses fréquences sont traitées à l'apex. Le temps de propagation pour atteindre les cellules ciliées de la cochlée seront donc plus court pour les hautes fréquences et plus long pour les basses fréquences. Des différences subtiles de fréquences peuvent donc être transcrites sous forme de décalages temporels au niveau des réponses neuronales. Par exemple, les différences de trajectoire des formants entre les syllabes /ba/, /da/ et /ga/ vont se traduire par des différences de latences dans les *cABRs* (Krista L. Johnson et al., 2008). La syllabe /ga/ qui présente des fréquences formantiques plus élevées engendre les réponses les plus précoces, suivit par la syllabe /da/ puis /ba/.

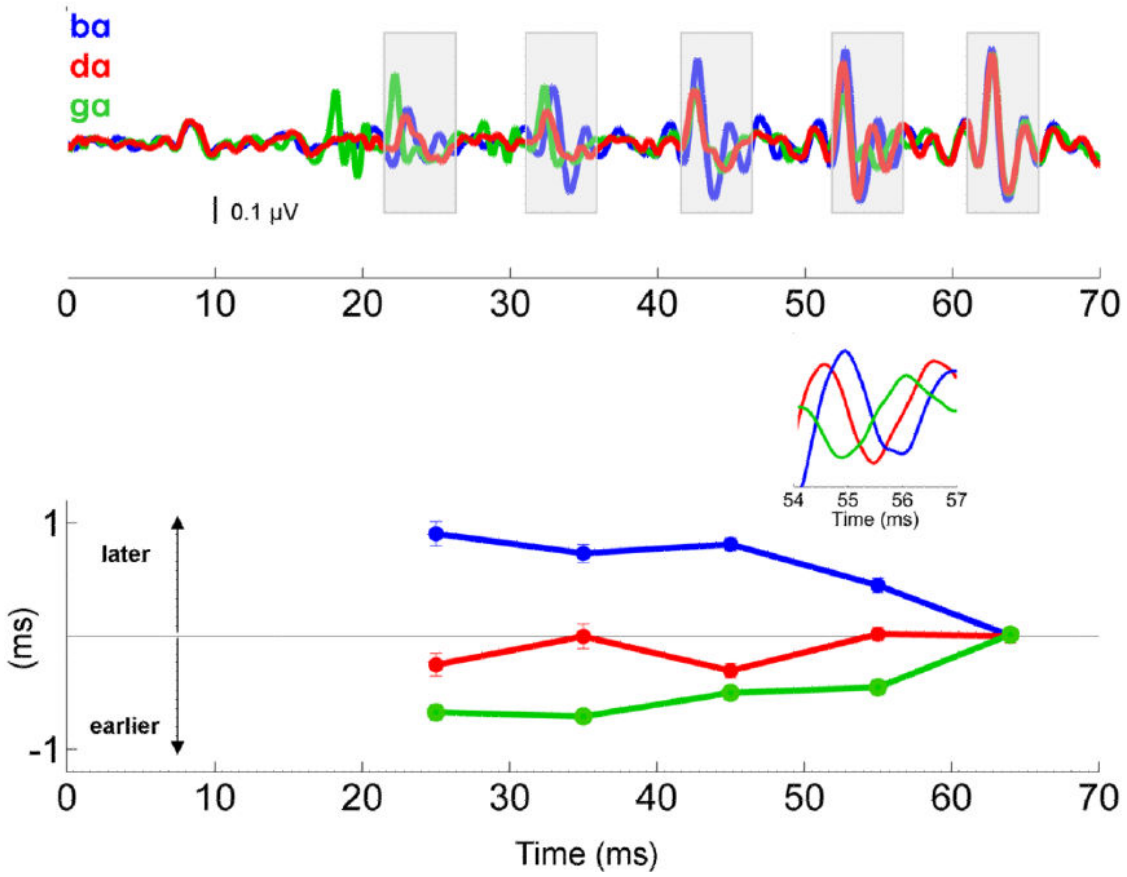


Figure 6. *Place code* dans les réponses sous-corticales. Les différences de trajectoires formantiques entre les syllabe /ba/, /da/ et /ga/ se traduisent par des différences de latences dans les *cABRs*. **Haut :** *cABRs* évoquées par les trois syllabes. Les différences de latences sont les plus évidentes à certaines fenêtres temporelles (encadrés gris). L'encadré représente un agrandissement de la fenêtre temporelle à ~55 ms. **Bas :** Différences de latences normalisées entre chaque pic et la moyenne des trois pics en fonction du temps. La syllabe /ga/ qui présente des fréquences formantiques plus élevées déclenchent les réponses les plus précoces, suivit par la syllabe /da/ puis du /ba/. Adaptée à partir de Johnson et al. (2008) et Skoe and Kraus (2010).

Nous venons de voir les propriétés qui font des *cABR* des réponses neuronales si particulières et qui rendent possible l'utilisation d'outils d'analyse avancés. Les *cABR* permettent donc une évaluation objective de la qualité de l'encodage des sons de parole au niveau sous-cortical. Dans le prochain chapitre, nous verrons que le tronc cérébral n'est pas un simple relais passif du système auditif, mais plutôt un lieu convergence entre les informations sensorielles provenant de la périphérie (*bottom-up*) et les modulations d'origine corticale (*top-down*)., au niveau duquel se produisent des effets de plasticité.

2. Plasticité du système auditif

2.1. Cadre théorique

Le système nerveux auditif est généralement vu comme une série de relais neuronaux ayant chacun une fonction bien définie (Winer and Schreiner, 2005). Selon cette approche, la perception d'un son va mette en jeu successivement le nerf auditif, les noyaux cochléaires, le tronc cérébral et le thalamus avant d'atteindre le cortex auditif primaire puis associatif (**Figure 7**). A mesure que l'information auditive remonte les voies afférentes de la périphérie vers les centres, la sensibilité au contenu acoustique des structures neuronales décroît alors que parallèlement la sensibilité aux caractéristiques abstraites (syllabes, mots, intelligibilité) augmente (Okada et al., 2010). Cette organisation hiérarchique du système auditif résulte en partie des limites physiologiques de *phase-locking* inhérentes aux structures neuronales : les neurones du tronc cérébral sont capables d'encoder les fréquences allant jusqu'à 1.5-2 kHz alors que le cortex a une limite de *phase-locking* située aux alentours de 100 Hz (Steinschneider et al., 2013).

Cette approche analytique du système auditif ne peut néanmoins expliquer les dernières découvertes montrant une influence du système cortical sur le traitement sous-cortical des sons (Kral and Eggermont, 2007). En effet, un système est par définition composé d'éléments en interaction les uns avec les autres, et même si les modèles développés jusqu'à présent prennent en compte ces interactions, les recherches se sont focalisées sur la caractérisation des fonctions spécifiques à chaque niveau, indépendamment les uns des autres. Néanmoins, les dernières années de recherche en neurosciences auditives ont vu l'émergence d'une nouvelle vision du système auditif, à la fois distribué mais intégré, où les fonctions de chaque relai seraient le résultat d'une interaction entre tous les niveaux du système (Bajo and King, 2013; Kraus and White-Schwoch, 2015; Tzounopoulos and Kraus, 2009). Cette approche dynamique conçoit la perception auditive comme une interaction à double sens entre les systèmes corticaux et sous-corticaux, entre les systèmes cognitifs et sensoriels (**Figure 8**). Au niveau anatomique, les flux d'information ascendants (*bottom-up*) et descendants (*top-down*) sont rendus possibles grâce aux multiples projections neuronales afférentes, de la cochlée au cortex (Harrison and Howe, 1974a), et efférentes, du cortex à la cochlée (Harrison and Howe, 1974b). Dans ce système, le tronc cérébral occupe une position clé puisqu'il se trouve au point de convergence entre les voies

ascendantes et descendantes, à mi-chemin entre systèmes sensoriels et cognitifs. L'étude du tronc cérébral auditif présente donc un intérêt majeur lorsqu'il s'agit de comprendre le fonctionnement du système auditif de manière intégrée.

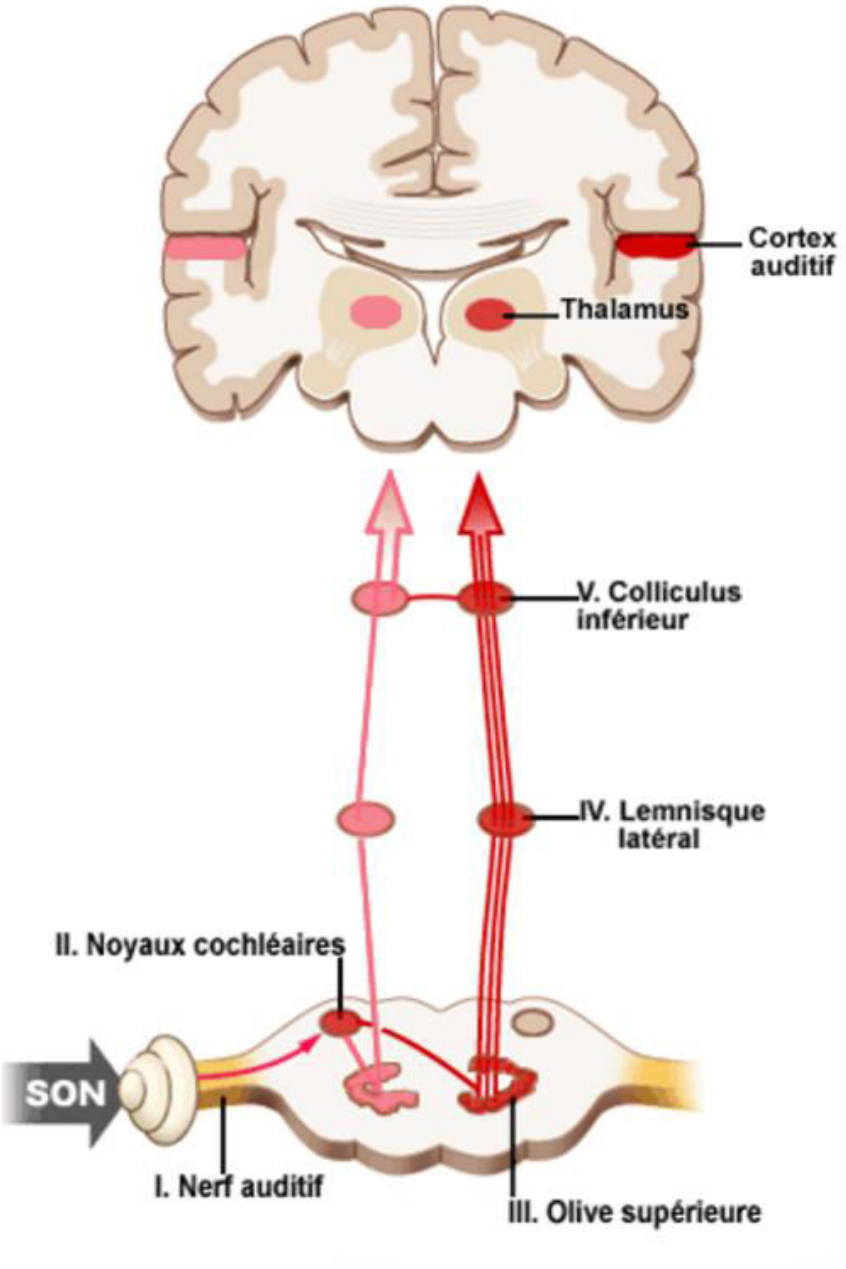


Figure 7. Vue *bottom-up* classique du système auditif. Le son arrive au niveau de la cochlée, ce qui entraîne la décharge de neurones au niveau du nerf auditif qui est ensuite relayée par plusieurs noyaux sous-corticaux avant d'atteindre le cortex auditif primaire (image issue du site http://www.cochlea.eu/).

2.1.1. Une plasticité à tous les niveaux

La capacité du cerveau à se réorganiser en fonction de son environnement, nommée *plasticité cérébrale*, se produit tout au long de la vie, et s'observe à tous les niveaux d'organisation du vivant : de l'altération de l'expression de certains gènes, en passant par la modification des connexions qui relient les neurones jusqu'au comportement. Les phénomènes de plasticité neuronale sont à la base de tout apprentissage, et au niveau du système auditif, ils ont été observés de la cochlée jusqu'au cortex (Edeline et al., 2011; Gao and Suga, 2000; Perrot et al., 2006). Afin de mieux comprendre comment ces différents niveaux d'organisation interagissent entre eux lors des phénomènes de réorganisation neuronale, il est intéressant d'utiliser deux dichotomies : les projections afférentes vs. efférentes et les régions primaires vs. secondaires.

Projections afférentes et efférentes

Les voies afférentes ou *bottom-up* transmettent l'information de la périphérie vers les centres, c'est-à-dire de la cochlée, en passant par les différents relais sous-corticaux, avant d'atteindre le cortex auditif primaire, puis secondaire et enfin associatif. A l'inverse, les projections efférentes ou *top-down* transmettent les signaux dans le sens contraire, du centre vers la périphérie, du cortex vers la cochlée (**Figure 8**). Selon les dernières découvertes, les voies efférentes, appelées aussi système corticofuge, joueraient un rôle important dans la réorganisation des circuits neuronaux du système auditif (Kral and Eggermont, 2007).

Le système corticofuge forme de nombreuses boucles de rétroaction entre les neurones corticaux et sous-corticaux. La plus courte d'entres elles est la boucle corticothalamique et la plus longue est celle entre le cortex et les cellules ciliées de la cochlée (Huffman and Henson Jr., 1990). De nombreuses études chez l'animal, et en particulier chez la chauve-souris, ont montré que le système corticofuge est capable d'affiner ou de déplacer les accords en fréquence ou *frequency tuning* des neurones sous-corticaux, dans le domaine spectral, temporel et spatial (Suga, 2008). Il jouerait donc un rôle majeur dans la réorganisation du système auditif en fonction de l'expérience. Par exemple, la stimulation électrique de neurones du cortex auditif primaire entraîne soit une facilitation soit une inhibition des réponses des neurones thalamiques et colliculaires.

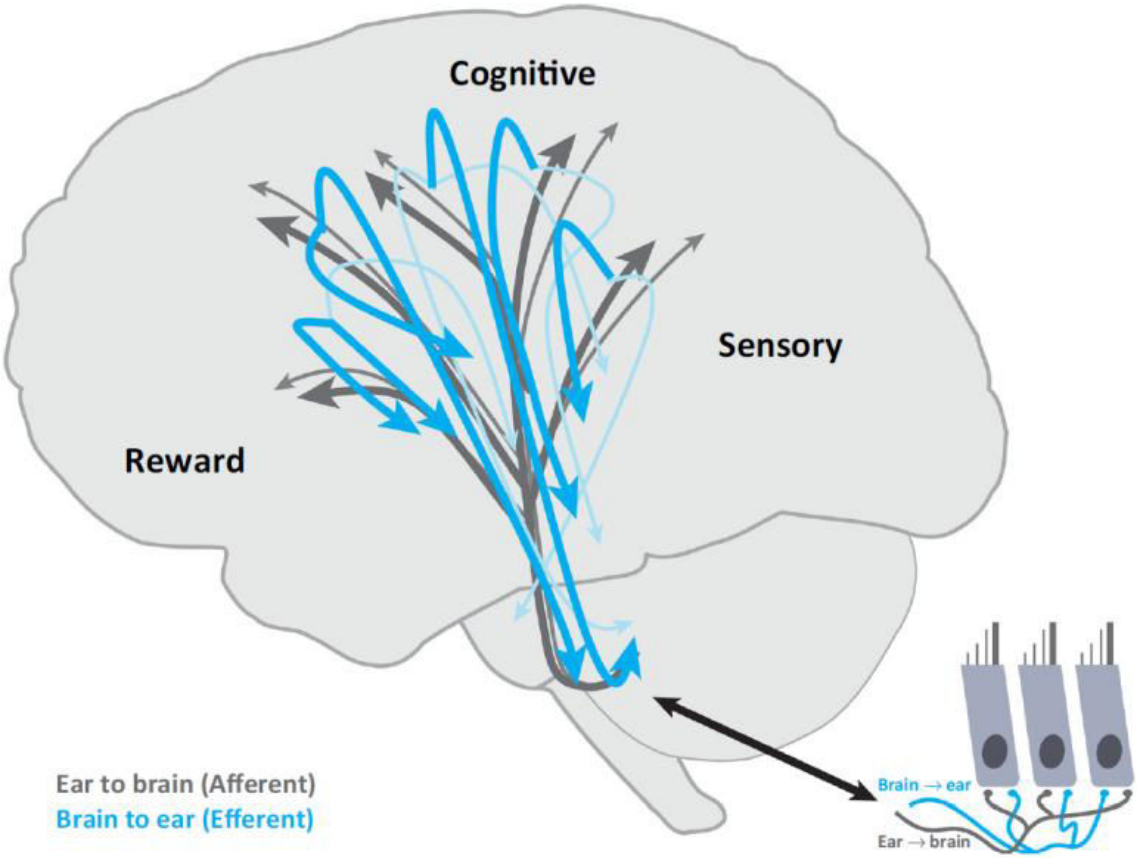


Figure 8. Le système auditif est un circuit à la fois distribué et intégré qui met en jeu les systèmes de récompense, cognitifs, et sensoriels. L'interaction entre ces systèmes se fait par des voies afférentes de la cochlée vers le cortex (flèches grises) et efférentes du cortex vers la cochlée (flèches bleues). Ces projections neuronales contiennent les divisions primaires (couleurs foncées) et non-primaires (couleurs claires) du système auditif et constituent les bases structurales du traitement des sons et de la plasticité neuronale (issue de (Kraus and White-Schwoch, 2015).

L'effet d'inhibition ou de facilitation dépend de la fréquence du son pour laquelle les neurones répondent, et de la relation de *frequency tuning* entre les neurones corticaux stimulés et les neurones sous-corticaux enregistrés. En effet, le seuil de réponse d'un neurone est généralement plus bas pour une fréquence donnée. Cette fréquence est définie comme la meilleure fréquence ou *best frequency* (*BF*) du neurone. Si la *BF* du neurone sous-cortical enregistré correspond à celle du neurone cortical stimulé, la réponse du neurone sous-cortical est amplifiée au niveau de sa *BF* et inhibée au niveau des fréquences plus basses ou plus hautes que la *BF*. De ce fait, la *frequency tuning* est affinée. Inversement, si la *BF* du neurone sous-cortical enregistré est différente de celle du neurone cortical stimulé, la réponse du neurone sous-cortical est inhibée au niveau de sa *BF* et amplifiée au niveau des autres fréquences. De ce fait, la *frequency tuning* est décalée dans l'espace fréquentiel, ce qui modifie la *BF* des neurones sous-corticaux en fonction de la fréquence du son (Suga, 2008). Ces phénomènes de réorganisation vont même se produire au niveau des structures les plus périphériques, telles que la cochlée. Par exemple, chez la chauve-souris moustachue qui émet des ultrasons dont la fréquence est comprise entre 60 et 62 kHz, le microphone cochléaire répondra préférentiellement aux alentours de 61 kHz. Or, la stimulation électrique des neurones corticaux entraîne un déplacement de la *BF* du microphone cochléaire controlatéral à la stimulation d'environ 0.25 kHz autour de 61 kHz, en fonction de la *BF* des neurones corticaux stimulés (Xiao and Suga, 2002). Cependant, la stimulation électrique des neurones corticaux est loin d'être la manière la plus écologique d'étudier les réorganisations neuronales liées au système corticofuge. Nous pouvons donc nous demander si de tels résultats sont applicables dans des conditions naturelles d'écoute, voire même chez l'Humain. Afin de répondre à cette interrogation, cette même équipe de recherche a utilisé une stimulation acoustique à la place d'une stimulation électrique directe du cortex auditif primaire. Ils ont montré qu'une stimulation acoustique répétitive de courte durée (i.e. un tone-burst) entraîne elle aussi un déplacement de la *BF* des neurones corticaux ou sous-corticaux (Yan and Suga, 1998). Ces résultats suggèrent que le rôle du système corticofuge serait de réorganiser les cartes sensorielles sous-corticales en fonction de l'expérience auditive. Néanmoins, les auteurs ont observé que lorsque la stimulation acoustique est appliquée seule, le déplacement de la *BF* des neurones est beaucoup moins importante que celle observée par stimulation électrique. De manière intéressante, une étude plus récente a montré que lorsque la stimulation acoustique est couplée à un stimulus conditionnant (i.e. une stimulation électrique d'un membre inférieur), ils observaient une réorganisation des neurones corticaux et sous-corticaux tout aussi importante que celle observée par stimulation électrique du cortex auditif. De plus, le déplacement de la *BF* des neurones dépendait de la fréquence de la stimulation acoustique. Ces résultats montrent d'une part que ces réorganisations sont dirigées

par le système corticofuge, et d'autre part qu'elles sont sensibles à l'importance accordée aux stimulations acoustiques par le sujet. Autrement dit, le système auditif réorganise ses propriétés de réponse neuronale, à la fois en fonction des propriétés acoustiques des stimulations auxquelles il est exposé, mais aussi en fonction de l'importance attribuée à ces stimulations. Ces découvertes amènent à une autre question : si les phénomènes d'adaptation neuronale sont sensibles à la valence comportementale et/ou émotionnelle attribuée aux sons, d'autres systèmes autres que le système auditif *per se* seraient-ils impliqués dans ces phénomènes de plasticité ?

Afin de répondre à cette question, l'équipe de Gao et Suga a étudié l'effet de l'inactivation bilatérale du cortex somatosensoriel sur le déplacement de la *BF* des neurones du cortex auditif primaire et colliculaire. Les résultats montrent que l'inactivation seule du système somatosensoriel n'entraîne pas de perturbation des réponses des neurones corticaux ou colliculaires, mais va sélectivement entraver les changements des réponses neuronales normalement observés lors d'un conditionnement (i.e. stimulus acoustique + simulation électrique d'un membre inférieur) (Gao and Suga, 2000, 1998). Ces résultats suggèrent que lorsque les stimulations acoustiques présentent un intérêt comportemental (e.g. par conditionnement par la peur), les réorganisations du système auditif vont dépendre d'autres systèmes, tels que le système somatosensoriel qui pourrait jouer un rôle dans l'attribution de la valence comportementale et/ou émotionnelle aux stimuli. Les régions non-primaires permettraient donc aux régions primaires d'adapter leurs réponses en fonction du contexte et de la valence des stimuli. Nous allons maintenant voir comment la relation entre les régions primaires, secondaires et associatives permet d'assurer les réorganisations neuronales du système auditif en fonction de l'expérience sensorielle.

Régions primaires et secondaires

Comme nous l'avons vu précédemment, la perception auditive est un processus d'interaction dynamique entre plusieurs systèmes cognitifs et sensoriels, et ces interactions se font par les voies afférentes et efférentes du système. Il est donc probable que ces interactions puissent jouer un rôle dans les phénomènes de plasticité du système auditif. La dichotomie entre les régions primaires et secondaires va nous aider à mieux comprendre le rôle joué par les voies non-primaires et comment elles interagissent avec les voies primaires dans les réorganisations neuronales du système auditif. La distinction entre régions primaires et secondaires peut s'observer dans la sensibilité des neurones au contenu acoustique des stimuli, mais aussi dans la capacité de réorganisation neuronale de ces régions. La plasticité induite par l'apprentissage est généralement associée aux régions non-primaires, dites de "haut" niveau, où la mémoire des

expériences passées serait stockée. Néanmoins, les études récentes en neurophysiologie ont apporté des preuves solides que les régions primaires ou sensorielles sont aussi le siège de nombreux phénomènes de plasticité dépendants de l'expérience (Weinberger, 2004). De la cochlée vers le cortex auditif primaire, secondaire et associatif, les régions sont de moins en moins sensibles au contenu acoustique, et de plus en plus malléables en fonction de l'expérience et du contexte cognitif (e.g. attentionnel), et inversement du centre vers la cochlée. Il existe ainsi une organisation parallèle en gradient inverse des voies primaires et non-primaires à travers le système auditif (**Figure 9**). Par exemple, il a été montré qu'en condition d'écoute active l'activité des neurones du cortex frontal, du cortex auditif secondaire et primaire était amplifiée. Du cortex frontal vers le cortex auditif primaire, l'amplification des réponses neuronales suivait un gradient décroissant, suggérant qu'à mesure que l'information est transmise vers les aires de plus haut niveau cognitif, la sensibilité au contenu acoustique diminue alors que la sensibilité au contexte attentionnel augmente (Atiani et al., 2014). Cette organisation permettrait au système d'être à la fois assez souple pour s'adapter rapidement à des changements de contexte, comme par exemple lorsqu'il s'agit de porter attention de manière sélective à un locuteur (i.e. l'effet *cocktail party*), tout en maintenant assez de stabilité afin de tirer parti des expériences passées (langage, musique, mémoire, etc...). Les voies non-primaires seraient capables de s'adapter "en ligne", tandis que les régions primaires seraient plus sensibles aux changements répétées sur le long-terme. Au plus une expérience auditive est répétée, au plus les adaptations neuronales qui en résultent se transmettraient aux voies primaires, qui deviennent alors un miroir des expériences auditives passées à travers les changements de propriétés de réponse neuronale (Kraus and White-Schwoch, 2015). Ce gradient inverse de sensibilité acoustique et de labilité entre voies primaires et secondaires, voies efférentes et afférentes suggère aussi qu'il existerait une plasticité à plusieurs vitesses dans le système auditif.

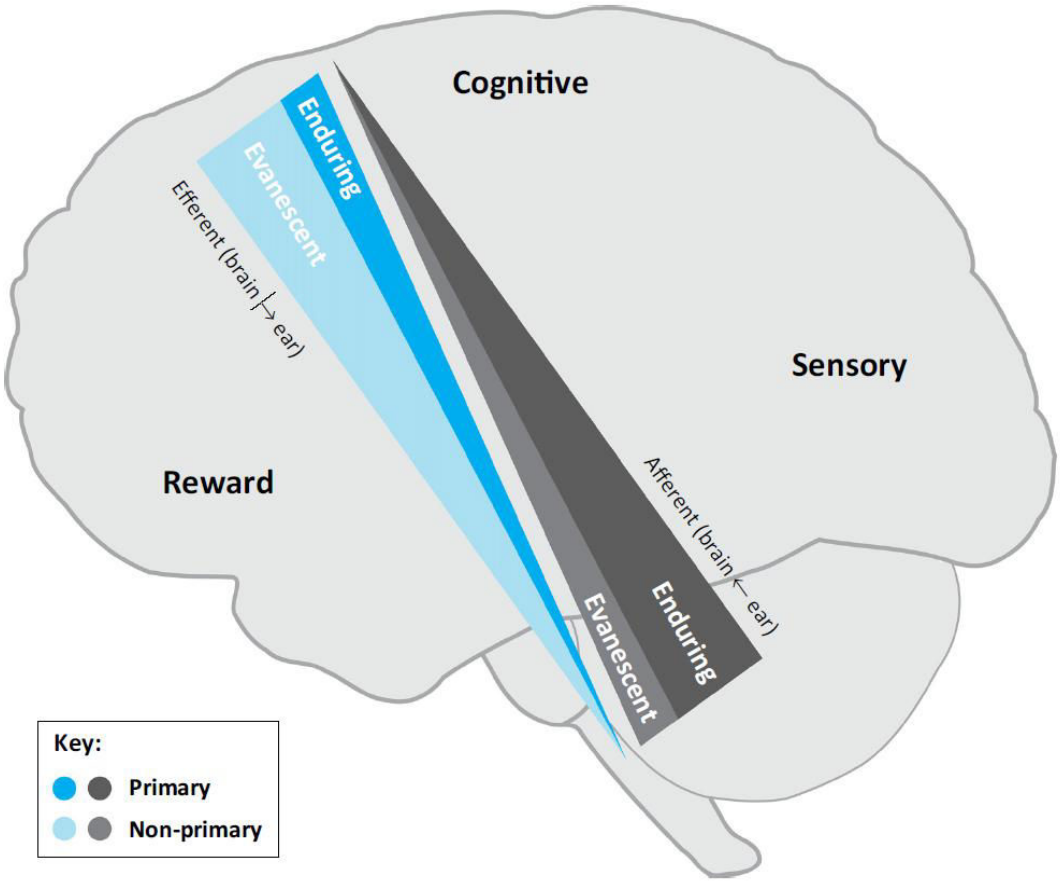


Figure 9. Organisation parallèle en gradient inverse des voies primaires et non-primaires à travers le système auditif. Le triangle plus large de la périphérie vers le centre (voie afférente primaire en gris foncé) illustre la prédominance de la stabilité des voies primaires, tandis que le triangle plus étroit (voie afférente non-primaire en gris clair) représente le caractère peu labile et plastique des voies non-primaires de la cochlée vers le cortex. A l'inverse, du centre vers la périphérie, le triangle plus étroit (voie efférente primaire, en bleu foncé) représente la moins grande stabilité du système primaire, tandis que le triangle plus large (voie efférente secondaire, en bleu clair) illustre la plus grande malléabilité des régions non-primaires. La balance entre stabilité et labilité entre les voies afférentes et efférentes pourrait sous-tendre la capacité du système à maintenir suffisamment de plasticité afin de s'adapter aux nouvelles situations tout en maintenant assez de stabilité afin de tirer parti des expériences passées (langage, musique, etc...). Au plus une expérience auditive est répétée, au plus elle va se transmettre aux voies primaires, qui deviennent alors un miroir des expériences auditives passées grâce aux changements de leurs propriétés de réponse neuronale. Adapté de Kraus & White-Schwoch, (2015).

2.1.2. Une réorganisation à plusieurs vitesses

Classiquement, la plasticité neuronale est vue comme se produisant à deux vitesses, rapide et lente, ou à court et long terme, mettant en jeu des mécanismes neurophysiologiques différents. Certaines adaptations neuronales sont très rapides, comme celles qui se produisent quelques minutes après une modification des demandes attentionnelles requises pour une tâche (Fritz et al., 2005). A l'inverse, d'autres études ont mis en évidence des effets de plasticité corticale sur de plus longues périodes (Weinberger, 2004), on parle alors de plasticité dépendante de l'expérience. Une hypothèse alternative proposée par Fritz et collaborateurs suggère que les adaptations neuronales à court et long terme ne seraient que les deux extrêmes d'un même continuum. Les réorganisations neuronale rapides (de la seconde à la minute) et évanescentes, seraient le résultat d'une adaptation au contexte comportemental (e.g. lorsqu'un musicien apprend à discriminer des notes en portant son attention sur les notes qu'il joue). Lorsque ces changements rapides sont renforcés par répétition (e.g. la pratique répétée d'un instrument à travers les jours, mois et années), les adaptations neuronales pourraient se cumuler et s'ancrer de manière permanente dans le système auditif, donnant ainsi lieu aux effets de plasticité dépendants de l'expérience à long-terme (e.g. les musiciens ont de meilleures capacités de discrimination fréquentielle au niveau comportemental et neuronal).

Dans cette thèse nous nous sommes intéressés aux deux extrêmes de ce continuum, en investiguant les effets du contexte attentionnel, et les effets d'expertise sur le long terme, sur l'encodage des sons de parole au niveau sous-cortical. Bien que cette thèse s'articule autour des effets de plasticité au niveau sous-cortical, nous ne pouvons passer à côté de l'ampleur des études ayant déjà été réalisées sur la plasticité au niveau cortical. Nous allons dans les sous-chapitres suivants faire une brève revue des études portant sur la plasticité à long terme ou dépendante de l'expérience, puis celles portant sur la plasticité à court terme ou dépendante du contexte.

2.2. Plasticité dépendante de l'expérience

Les interactions constantes entre le système auditif et l'environnement vont jouer un rôle majeur dans le développement et la maturation du système nerveux auditif (Benasich et al., 2014; Dehaene-Lambertz and Spelke, 2015; K. L. Johnson et al., 2008). Les différents relais du système auditif vont se réorganiser en fonction de la qualité mais aussi de la quantité de ces interactions. Chez les patients souffrant de troubles auditifs, ces interactions sont réduites ou appauvries, et à

l'inverse chez les musiciens ou les bilingues, ces interactions sont enrichies. Ainsi, l'étude des fonctions auditives chez ces populations nous aide à mieux comprendre le rôle que joue l'expérience dans la réorganisation du système auditif (Kilgard, 2012).

Toutes les expériences que nous vivons peuvent modifier notre architecture cérébrale, mais nous allons ici nous focaliser sur deux expériences en particulier : l'expérience musicale et linguistique. L'étude de la plasticité induite par la pratique musicale ou d'une langue présente trois intérêts majeurs. Les stimuli (sons musicaux ou verbaux) sont complexes, ce qui permet d'étudier l'influence de l'expérience sur le traitement fin des sons. Deuxièmement, la durée d'exposition à ces stimuli permet d'établir un lien entre la taille des effets observés et la durée de l'expérience (e.g. durée de la pratique musicale ou d'une langue). Enfin, les stimuli complexes qui entrent en jeu dans ce type d'expérience sont porteurs de sens, ce qui comme nous le verrons, est un facteur majeur dans les effets de plasticité neuronale.

2.2.1. Expérience musicale

Jouer de la musique à un niveau professionnel relève d'un comportement d'une extrême complexité à la fois au niveau cognitif et comportemental. Ainsi, l'apprentissage d'un instrument de musique améliore, à travers de nombreuses années de pratique, la perception de sons musicaux (Besson et al., 2007; Pantev et al., 1998; Schneider et al., 2002; Tierney and Kraus, 2014) et développe les facultés cognitives et sensori-motrices des musiciens (Krause et al., 2010; Patel, 2014, 2012, 2011; Strait and Kraus, 2011; Tierney and Kraus, 2014). Les recherches menées sur le traitement des sons chez les musiciens ont montré qu'il existe des corrélats anatomiques et fonctionnels aux bénéfices induits par la pratique musicale. Par exemple, les musiciens présentent un seuil de discrimination fréquentielle significativement plus bas que les non-musiciens (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Parbery-Clark et al., 2011; Strait et al., 2010). Au niveau cortical, cela se traduit par une plus grande différence des potentiels évoqués auditifs (PEA) lors de la perception de deux sons proches en fréquence (Bermudez et al., 2009; Gaser and Schlaug, 2003; Schneider et al., 2002). L'amélioration des performances motrices se reflète au niveau du cortex moteur par un recrutement neuronal plus important au niveau des représentations corticales associées à la pratique de l'instrument. Par exemple, les violonistes doivent développer une grande dextérité des doigts de la main gauche s'ils veulent jouer des pièces de musique complexes. Cette sollicitation intense et fine développe chez ces musiciens une

représentation sensori-motrice des doigts de la main gauche plus importante par rapport à celle de non-musiciens. A l'inverse, la main droite qui manipule l'archet, réalise une tâche beaucoup moins complexe, ne nécessitant pas de mouvements fins des doigts, ce qui se traduit au niveau cortical par des représentations sensori-motrices équivalentes à celles des non-musiciens (Elbert et al., 1995).

Le développement des techniques d'imagerie cérébrale telles que l'*Imagerie par Résonance Magnétique* (*IRM*) à haute résolution ont permis d'étudier avec précision les modifications anatomiques induites par l'expérience musicale. Par exemple, lors de l'écoute de sons sinusoïdaux, Schneider et collaborateurs (2002) ont montré qu'il existait une forte corrélation entre l'amplitude des PEA du cortex auditif primaire, le volume de matière grise au niveau du gyrus de Heschl et le niveau d'expertise musicale des sujets (Schneider et al., 2002). Parmi les musiciens, certains présentent des capacités exceptionnelles telles que l'*oreille absolue.* Cette aptitude à reconnaître, lors de l'écoute isolée d'un son, la note de musique correspondante a longtemps été considérée comme innée, mais il s'avère qu'elle repose à la fois sur des prédispositions génétiques et sur l'entraînement musical (Gregersen et al., 1999). Cet avantage reposerait sur une particularité anatomique au niveau du planum temporal, caractérisée par une asymétrie latéralisée à gauche plus importante chez ces musiciens par rapport aux non-musiciens (G. Schlaug et al., 1995). Les différences anatomiques entre musiciens et non-musiciens ne se limitent pas au cortex auditif mais se retrouvent dans toutes les régions du cerveau sollicitées par la pratique musicale. Ainsi, les pianistes et violonistes professionnels, qui possèdent une grande dextérité, ont une représentation corticale motrice de la main plus importante au niveau des deux hémisphères par rapport aux non-musiciens, ainsi qu'une meilleure symétrie interhémisphérique. De plus, chez les musiciens le volume des aires motrices de la main est inversement proportionnel à l'âge de début de la pratique musicale, ce qui suggère que le cortex moteur peut montrer des adaptations fonctionnelles et structurelles induites par l'expérience musicale sur le long terme (Amunts et al., 1997). Le cervelet, aussi connu pour son rôle dans les apprentissages moteurs, est mis en jeu dans des tâches nécessitant une grande précision temporelle (Anderson et al., 2002, 1994; Kleim et al., 1997). En mesurant le volume cérébelleux de musiciens experts et non-musiciens, Hutchinson et collaborateurs ont montré que les pianistes possèdent non seulement un cervelet plus volumineux, mais qu'il existe également une forte corrélation entre le volume cérébelleux et le niveau de pratique musical (Hutchinson et al., 2003). Ces résultats supportent une fois de plus l'hypothèse selon laquelle les différences structurelles observées seraient induites par la pratique musicale intense et répétée plutôt qu'à des prédispositions innées. En outre, la pratique d'un instrument de musique nécessite une grande coordination motrice des mains. Le volume du corps

calleux, qui correspond au principal passage des fibres connectant les deux hémisphères, serait un bon indicateur de la communication interhémisphérique et donc de la coordination bimanuelle. En mesurant le volume de la région sagittale antérieure du corps calleux, chez des joueurs professionnels d'instruments à cordes ou de claviers, Schlaug et collaborateurs ont montré que la grande coordination motrice acquise par la pratique précoce et intense de la musique est associée à un nombre de connexions interhémisphériques plus important, ce qui suggère une meilleure interaction entre les deux hémisphères chez les musiciens, nécessaire à la coordination motrice des deux mains (Schlaug et al., 1995).

Plus récemment, le développement de l'IRM en Tenseur de Diffusion a permis l'étude des modifications structurales fines induites par la pratique musicale à long terme au niveau de la matière blanche. Grâce à cette technique, Oechslin et collaborateurs ont étudié, chez un groupe de musiciens et de non-musiciens, l'anisotropie fractionnelle du faisceau longitudinal supérieur (FLS) qui relie des structures mises en jeu dans la perception et la production de la musique et de la parole. Les musiciens ayant l'*oreille absolue* présentent une plus grande asymétrie latéralisée à gauche à la différence des musiciens qui ne possèdent pas cette faculté ou des non-musiciens. De plus, cet effet est fortement corrélé avec le degré d'expertise musicale, ce qui suggère que le niveau d'organisation architectural des fibres nerveuses est aussi un bon indicateur anatomique du niveau d'expertise musicale (Oechslin et al., 2010). D'autres faisceaux de neurones sont aussi le siège de réorganisations liées à la pratique d'un instrument de musique. Par exemple, Imfeld et collaborateurs ont montré chez les musiciens une augmentation de la diffusivité du faisceau cortico-spinal, qui désigne l'ensemble des fibres qui transmettent la commande motrice du cortex moteur aux motoneurones de la moelle épinière, et qui joue un rôle primordial dans le contrôle moteur volontaire. De plus, cette augmentation est inversement proportionnelle à l'âge de début de la pratique musicale (Imfeld et al., 2009). Ces effets de plasticité dépendants de l'expertise musicale se retrouvent aussi au niveau du cortex cérébelleux. Le cervelet droit des musiciens présentent non seulement un plus grand volume de matière grise, mais aussi de matière blanche par rapport aux sujets contrôles non-musiciens (Abdul-Kareem et al., 2011). L'ensemble de ces études utilisant l'IRM en tenseur de diffusion ont ainsi mis en évidence que les performances comportementales des musiciens sont non seulement associées à un plus grand volume de matière grise au niveau des régions mises en jeu lors de la pratique musicale, mais également à une meilleure connectivité structurelle entre ces régions.

Génétique ou épigénétique ?

La plupart des études décrites ci-dessus montrent que les différences anatomiques, fonctionnelles et comportementales entre musiciens et non-musiciens corrèlent fortement avec le degré d'expertise musicale, ce qui suggère que ces différences pourraient être le résultat d'une acquisition sur le long terme plutôt qu'à des prédisposition innées (Bermudez et al., 2009; Gaser and Schlaug, 2003). En revanche, seules les études longitudinales permettent de déterminer si effets observés ont une base génétique ou épigénétique. Afin de lever le voile sur cette ambiguïté, des enfants ont été assignés de manière aléatoire à un entraînement musical ou artistique (e.g. peinture) pendant une durée donnée. Des mesures structurelles, fonctionnelles et comportementales ont été réalisées avant, pendant et après entraînement. Après seulement six mois de pratique, les enfants assignés à l'entraînement musical montraient de meilleures performances dans une tâche de discrimination de hauteur tonale, reflétées au niveau cortical par une plus grande différence des réponses évoquées par des variations subtiles de la tonalité des sons, alors que cet effet était beaucoup moins important chez les enfants ayant suivi des cours de peinture (Moreno et al., 2009). Une autre étude longitudinale réalisée par le même groupe, s'étalant cette fois sur une période de 12 mois, a montré qu'après une telle période d'entraînement, les bénéfices s'observaient au niveau comportemental par une amélioration des capacités de traitement pré-attentives de la durée des syllabes et du VOT, et au niveau neuronal par une plus grande MMN post-entraînement, uniquement chez les enfants ayant suivi l'entraînement musical (Chobert et al., 2014). De même, après quinze mois d'entraînement musical, les enfants ont montré une augmentation du volume des régions impliquées dans la pratique de la musique comme le cortex auditif et moteur primaires ainsi qu'au niveau du corps calleux (Hyde et al., 2009). Ces modifications structurales étaient en outre associées à une amélioration des performances auditives (rythme et mélodie) et motrices (dextérité).

En résumé, bien que les d'études longitudinales soient encore peu nombreuses, elles apportent néanmoins l'évidence directe que les différences observées entre musiciens et non-musiciens chez les adultes ne reposent pas sur des différences génétiques, mais que ces différences s'acquièrent par la pratique musicale répétée tout au long de la vie d'un individu.

Effets de transfert entre expertise musicale et langage

Les bénéfices induits par la pratique musicale peuvent être hautement spécifiques. Par exemple, chez les musiciens (trompettistes et violonistes) les potentiels de champs magnétiques

sont plus importants pour des sons musicaux que des sons purs. De plus, au sein des sons musicaux, plus le timbre du son se rapproche de celui de l'instrument pratiqué par le musicien, plus la représentation corticale sera importante (Pantev et al., 2003, 2001, 1998). A l'inverse, certains avantages induits par la pratique musicale peuvent se transférer à d'autres domaines de compétences, et en particulier au langage qui partage de nombreuses caractéristiques avec la musique. De nombreuses données électrophysiologiques montrent que la pratique musicale améliore l'encodage neuronal des sons de parole . L'équipe de Kraus et collaborateurs a montré à de nombreuses reprises que l'encodage sous-cortical des sons de parole est amélioré chez les individus ayant suivi une pratique musicale (Kraus and Chandrasekaran, 2010). De plus, le degré d'amélioration de l'encodage neuronal est corrélé avec le nombre d'années de pratique musicale, ce qui suggère que ces effets de transfert sont dépendants de l'expérience musicale (Musacchia et al., 2007; Strait et al., 2009; Wong et al., 2007). Par exemple, les représentations sous-corticales des sons de parole sont plus robustes à la présence de bruit chez les musiciens (Parbery-Clark et al., 2009a). La pratique musicale permet aussi d'améliorer la représentation des composantes fréquentielles importantes dans la discrimination des syllabes, à savoir les trajectoires formantiques (Parbery-Clark et al., 2012). De plus, ces bénéfices ont été associés à une amélioration des compétences linguistiques telles que la capacité à percevoir la parole dans du bruit, les compétences en lecture, la mémoire de travail auditive (Tierney and Kraus, 2013), et même la capacité à apprendre de nouveaux patterns langagiers via l'amélioration de la segmentation implicite de nouveaux mots contenus dans un flux de parole continu (François et al., 2013).

Pourquoi existe-t-il des effets de transfert entre musique et langage ?

Les premières recherches, principalement issues de la neurologie et de la neuropsychologie, ont véhiculé l'idée que la musique et le langage étaient deux phénomènes bien distincts, ayant des corrélats anatomo-fonctionnels différents : la musique étant latéralisée au niveau de l'hémisphère droit, alors que le langage mettrait en jeu l'hémisphère gauche. Le développement des techniques d'imagerie cérébrale a renversé cette idée en révélant que la musique et le langage mettent en jeu des réseaux neuronaux beaucoup plus distribués au sein du cerveau, et surtout que ces deux fonctions partagent des ressources neuronales communes (e.g., Abrams et al., 2011; Levitin and Menon, 2003; Maess et al., 2001). L'hypothèse OPERA, proposée par Patel (2011) suggère que les compétences acquises par la pratique de la musique peuvent se transférer au langage lorsque cinq conditions sont réunies : (1) ***O****verlap* : il existe un

recouvrement des réseaux neuronaux mis en jeu lors du traitement de caractéristiques acoustiques importantes à la fois en musique et en langage (e.g. la hauteur tonale), (2) ***P****recision* : au niveau des réseaux neuronaux partagés, le traitement des caractéristiques acoustiques communes demande une précision plus importante en musique qu'en langage, (3) ***E****motion* : la mise en jeu de ces réseaux par la pratique musicale véhicule une émotion positive forte, (4) ***R****epetition* : les activités musicales qui mettent en jeu ces réseaux doivent être fréquemment répétées, et (5) ***A****ttention* : la mise en jeu de ces réseaux neuronaux par la pratique musicale est associée à une attention ciblée. Selon l'hypothèse OPERA lorsque ces cinq conditions sont réunies, la plasticité neuronale induite par la pratique musicale conduit à un fonctionnement plus précis des circuits neuronaux partagés par la musique et le langage. De ce fait, le traitement des sons de parole bénéficie des avantages acquis par la pratique musicale.

Overlap

La mise en jeu de ressources neuronales communes lors du traitement des sons de musique ou de parole n'est pas si surprenante puisque la musique et la parole partagent de nombreuses similarités. La musique et la parole nécessitent toutes deux la perception et la production de séquences sonores complexes. Ces séquences sont organisées de manière hiérarchique (i.e. en syllabes, mots, et phrases pour la parole, et en notes, rythmes et métriques pour la musique) et leur décodage nécessite une représentation efficiente des variations spectro-temporelles rapides qui les caractérisent. De plus, elles sont basées sur les mêmes caractéristiques acoustiques, telles que la hauteur tonale, le timbre, la durée et l'intensité des sons (Besson et al., 2011). En sollicitant de manière intense et répétée ces réseaux communs, la pratique musicale va améliorer les capacités de traitement de ces caractéristiques acoustiques via la plasticité neuronale, ce qui aura une influence positive sur la perception des sons langagiers. En pratique, lors de l'apprentissage de la musique les musiciens apprennent à porter leur attention aux fines variations acoustiques des sons. Par exemple, le musicien doit être capable de jouer des pièces de musique complexes, et souvent dans des environnements bruyants, ce qui va l'amener à porter une attention particulière à la hauteur tonale des sons afin de jouer la bonne note, au décours temporel, afin de jouer la bonne note au bon moment, et au timbre, afin de différencier les sons produits par son instrument du bruit environnant (autres instruments de musique, bruit d'une salle de concert, etc...). Ces trois propriétés acoustiques sont autant fondamentales en musique qu'en langage (Kraus et al., 2009).

De nombreuses études rapportent que les musiciens ont de meilleures capacités à détecter les fines variations de la hauteur tonale de sons purs (Kishon-Rabin et al., 2001), de sons musicaux (Fujioka et al., 2004; Trainor et al., 1999) mais aussi verbaux (Schön et al., 2004), et ce à plusieurs niveaux du traitement de l'information. Par exemple, en demandant explicitement à un groupe de musiciens experts et de non-musiciens de détecter les variations de *pitch* à la fin d'une séquence musicale ou d'une phrase, Schön et collaborateurs (2004) ont montré non seulement des temps de réponse plus courts et des réponses neuronales plus précoces chez les musiciens lorsqu'il s'agissait du matériel musical, mais de manière intéressante cet avantage s'observait aussi pour le matériel langagier (Schön et al., 2004). Ces avantages induits par la pratique musicale s'observent également à une étape plus précoce du traitement de la parole, au niveau des réponses pré-attentives du tronc cérébral. La fréquence fondamentale (F0), qui est le principal corrélat acoustique de la hauteur perçue, est représentée de manière fidèle au niveau sous-cortical par la FFR. Afin de déterminer si la pratique musicale modifie le traitement automatique des sons de parole, Wong et al. (2007) ont comparé, chez un groupe de musiciens et de non-musiciens anglais, les réponses sous-corticales évoquées par les trois patterns de hauteur tonale de la syllabe chinoise /mi/. En évaluant la corrélation entre la trajectoire de la FFR et celle de la F0 présente dans le stimulus, ils ont montré que les musiciens avaient une représentation sous-corticale plus fidèle et plus robuste des variations fines et rapides de la F0 des stimuli Mandarins par rapport à des non-musiciens (**Figure 10**). De plus, le niveau de corrélation stimulus-réponse était corrélé avec le nombre d'années et l'âge de début de la pratique la pratique musicale. Ces résultats révèlent d'une part que l'encodage automatique du *pitch* des sons de parole est amélioré chez les musiciens, et d'autre part que ces effets d'expertise se transfèrent aussi aux sons verbaux étrangers, puisque les participants de cette étude étaient de langue maternelle anglaise.

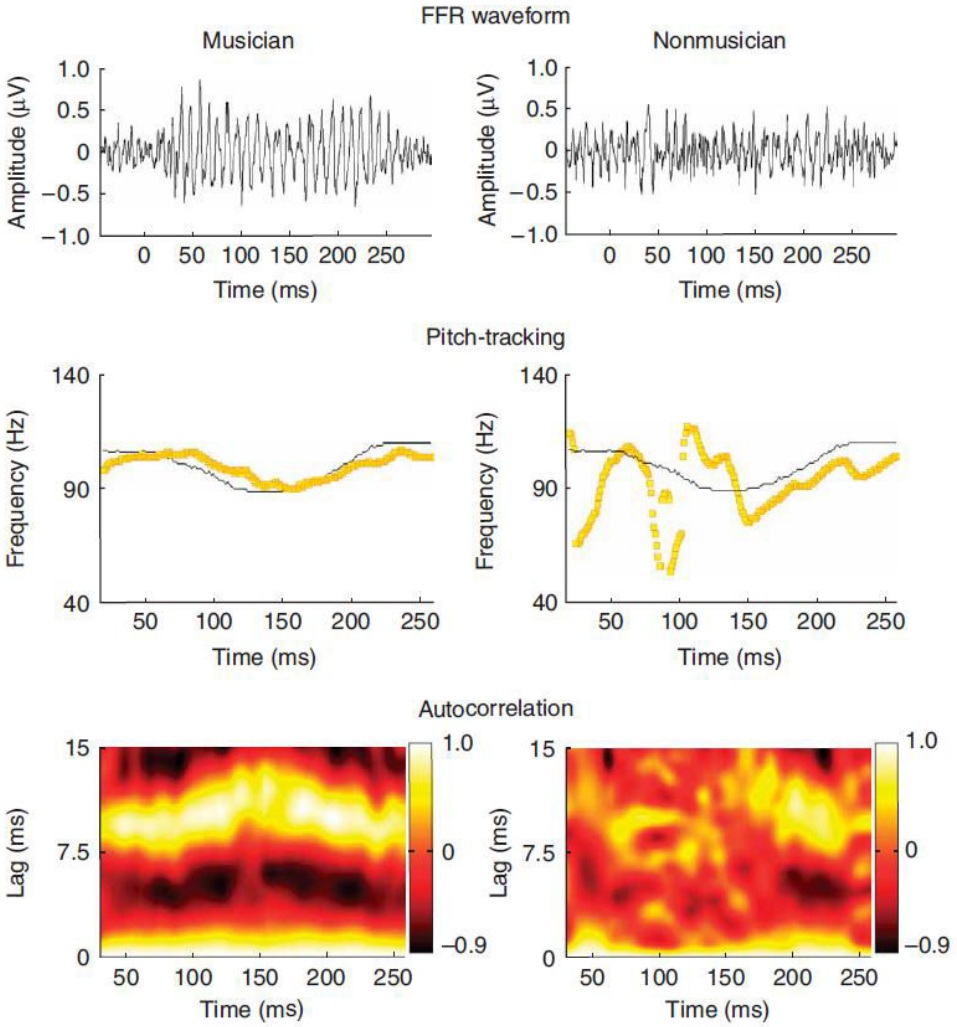


Figure 10. FFR évoquées par une syllabe tonale du chinois Mandarin. **Haut**, réponses neuronales enregistrées chez un musicien (gauche) et un non-musicien (droite). **Milieu**, la ligne noire représente la trajectoire de la fréquence fondamentale du stimulus, tandis que la ligne jaune représente celle contenue dans la FFR du sujet. **Bas**, autocorrélogrammes des réponses neuronales. La couleur indique le niveau de corrélation, les niveaux plus clairs indiquant de plus hauts niveaux de corrélation. Issue de Wong et al. (2007).

Le timbre permet de différencier deux sons ayant le même *pitch*. Le *timbre* est une propriété multidimensionnelle des sons qui émerge de l'interaction entre les variations spectrales, temporelles, du profil des harmoniques (certaines comme les formants étant plus amplifiées et d'autres atténuées), et des temps d'attaque et de descente des sons. En musique, le *timbre* permet de différencier deux instruments de musique jouant la même note, et dans la parole il permet de différencier le locuteur des autres personnes ou du bruit environnant (e.g. dans l'effet *cocktail party*). Les musiciens ont de meilleures représentations corticales des sons dont le timbre se rapproche de l'instrument pratiqué (Pantev et al., 2001). Néanmoins, cet avantage n'est pas

spécifique aux sons musicaux puisque de nombreuses études rapportent que les musiciens ont également de meilleures capacités à reconnaître des mots présentés dans du bruit de parole (Parbery-Clark et al., 2009b), une tâche se basant fortement sur l'encodage du timbre de la voix du locuteur. Au niveau sous-cortical, les musiciens présentent des réponses neuronales aux sons de parole plus robustes au bruit comparés aux non-musiciens (Parbery-Clark et al., 2009a).

Une autre caractéristique acoustique importante à la fois dans le décodage de la musique et de la parole est le décours temporel ou *timing*, véhiculé principalement par l'enveloppe temporelle des sons. En parole, les variations temporelles de l'enveloppe résultent de l'interaction entre la source, c'est-à-dire les cordes vocales, et des filtres, c'est-à-dire les articulateurs. Cette interaction permet notamment de véhiculer des indices suprasegmentaux tels que les pauses, le début et la fin des phrases, et à moindre échelle les trajectoires formantiques qui sont importantes pour distinguer les consonnes et voyelles. En musique, le *timing* participe aussi à la segmentation des sons, et à une échelle plus large à la perception du rythme. Afin de déterminer si l'amélioration du traitement du *timing* peut se transférer de la musique vers le langage, Chobert et collaborateurs (2014) ont assigné de manière aléatoire des enfants à un entraînement musical ou de peinture, et ont mesuré leurs réponses corticales évoquées par de fines variations temporelles sur du matériel verbal, tel que la durée des syllabes ou le temps de voisement (VOT). Les résultats ont montré qu'après 12 mois de pratique, seuls les enfants ayant suivi un apprentissage musical présentaient une amélioration significative du traitement de la durée et du VOT, et ce de manière pré-attentive puisque ces effets se reflétaient au niveau de la MMN, une réponse corticale reflétant la détection automatique d'incongruences, ici temporelles (Chobert et al., 2014; Picton et al., 2000). Une autre étude portant cette fois chez le musicien adulte professionnel a montré que lorsque un son est omis durant une séquence sonore isochrone (i.e. avec des intervalles inter-stimuli réguliers, ISI), une réponse MMN est observée chez les participants qu'ils soient musiciens ou non (Rüsseler et al., 2001). En revanche, lorsque la cadence de présentation est ralentie (ISI>120 ms), les non-musiciens voient leur MMN réduite voire même absente pour des ISI de 220 ms. De plus, lorsque des sons surviennent avec des décalages temporels inférieurs à 50 ms, seuls les musiciens présentent encore une MMN. Ces résultats révèlent la plus grande sensibilité des musiciens aux fines variations temporelles des sons comparés aux non-musiciens.

L'ensemble de ces études montrent que l'entraînement musical affecte le traitement sensoriel cortical et sous-cortical des sons, sur plusieurs dimensions acoustiques (*timbre, pitch, timing*) partagées par la musique et la parole, ce qui expliquerait les effets de transfert entre ces deux domaines. Le partage de ressources cognitives et neuronales (*Overlap*) dans le traitement des

caractéristiques acoustiques communes à la musique et à la parole serait donc une condition nécessaire aux effets de transfert comme le suggère l'hypothèse OPERA.

Precision

Une fois la condition *Overlap* remplie, il faudrait que la sollicitation des réseaux neuronaux communs par la pratique musicale exige une précision plus importante que celle nécessaire au traitement des sons de parole. Selon Patel (2011), ce serait le cas puisque les musiciens sont amenés à jouer des pièces de musique complexes, où les variations de *pitch* aussi fines qu'un demi-ton sont structurellement importantes (Eerola et al., 2006), sans parler de la finesse des ajustements nécessaires chez les musiciens qui jouent un instrument à justesse variable (e.g. violon, trompette, flute, ...). En revanche, la perception et la production des sons de parole ne seraient pas aussi exigeante en termes de précision, même dans les langues tonales pour lesquelles les variations du *pitch* sont fondamentales, puisqu'elles permettent de distinguer le sens des mots monosyllabiques. En effet, Patel et al. (2010) ont comparé l'intelligibilité de phrases chinoises naturelles ou dont les variations de la F0 ont été artificiellement supprimées. Même sans les variations de la F0, les locuteurs natifs de Chinois Mandarin ont trouvé les phrases parfaitement compréhensibles. Les locuteurs pourraient se passer de cette information, en intégrant d'autres informations contextuelles (sémantiques, syntaxe...) lorsque celles-ci sont disponibles afin de déduire les mots qui sont prononcés. Ces résultats confirmeraient l'idée selon laquelle, même si le langage impose une certaine précision dans le traitement des caractéristiques acoustiques des sons, celle-ci n'est pas aussi importante qu'en musique puisque dans la parole de nombreux indices acoustiques ont des fonctions linguistiques redondantes. Par exemple, la distinction entre la syllabe /b/ et /p/ repose à la fois sur le temps de voisement (VOT), la durée de la voyelle, les variations de la F0 et les trajectoires des premiers et deuxièmes formants. Prises séparément, ces informations peuvent être ambigües, mais lorsqu'elles sont intégrées les unes aux autres, en plus des informations contextuelles contenues dans la phrase, la catégorie phonologique du son se précise sans ambiguïté (Toscano and McMurray, 2010). Cette redondance d'indices phonologiques permettrait donc de diminuer le niveau d'exigence requis pour le traitement des caractéristiques acoustiques des sons de parole, alors qu'en musique le niveau de précision est bien plus important puisqu'une variation aussi fine qu'un demi-ton peut complètement changer l'appréciation de la performance musicale.

Emotion, Repetition and Attention

Selon l'hypothèse OPERA, les conditions *Overlap* et *Precision* ne seraient pas suffisantes pour que la pratique musicale induise des effets de plasticité neuronale favorable à l'encodage de sons de parole. Il faudrait de plus que la pratique musicale véhicule une émotion positive (*Emotion*), qu'elle soit fréquemment répétée (*Repetition*) et qu'elle soit associée à une attention ciblée (*Attention*) sur les caractéristiques acoustiques des sons. La dimension émotionnelle est très importante dans tout apprentissage. Il en est de même pour l'apprentissage d'un instrument de musique. Si l'enfant ne prend aucun plaisir à jouer, il y a de fortes chances pour qu'il n'excelle pas dans le domaine. En effet, écouter ou jouer de la musique ne met pas uniquement en jeu le système auditif, mais tout un réseau cérébral distribué dans lequel se produisent notamment des interactions entre les cortex auditifs et les régions limbiques (Blood et al., 1999; Salimpoor et al., 2013).

Pour que des adaptations neuronales se produisent, il serait nécessaire que la musique soit pratiquée de manière répétée sur le long terme. La majorité des études portant sur les effets de plasticité induits par l'expérience musicale montre un lien étroit entre le nombre d'années de pratique et la taille des effets observés (Strait and Kraus, 2014). Par exemple, l'amélioration de la capacité à percevoir la parole dans du bruit est étroitement dépendante du nombre d'années de pratique musicale, mais aussi du nombre d'heures de pratique de l'instrument de musique (Slater et al., 2015)

L'attention portée sur les indices acoustiques durant ces heures d'entraînement musical jouerait également un rôle important dans ces effets de plasticité neuronale. Les études chez l'animal ont montré que les champs récepteurs spectro-temporels du cortex auditif primaire chez le furet s'adaptent rapidement en fonction des propriétés acoustiques du stimulus et de la tâche réalisée (e.g. discrimination fréquentielle ou temporelle). De plus, ces adaptations neuronales sont d'autant plus importantes que l'animal porte attention aux stimuli (Fritz et al., 2007, 2005). Dans la pratique musicale, l'attention focalisée sur certains indices acoustiques (*pitch*, *timbre*, *timing*...) permettrait de renforcer leur encodage neuronal, et cette attention répétée sur le long terme entraînerait des adaptations neuronales qui viendraient s'ancrer de manière permanente dans les réseaux neuronaux.

En résumé, la musique et le langage sollicitent des réseaux neuronaux communs lorsqu'il s'agit de traiter des informations acoustiques similaires (*pitch*, *timing*, *timbre*, enveloppe...). Lors de

la pratique d'un instrument de musique, ces réseaux seraient sollicités d'une manière plus importante que lors de la production ou de la perception du langage, car en musique des variations acoustiques subtiles peuvent modifier drastiquement l'appréciation de la mélodie jouée. En revanche, dans la parole il existe une redondance des indices acoustiques servant à discriminer les catégories phonologiques, ce qui permet de réduire l'exigence nécessaire au traitement acoustique des sons. La pratique répétée et intensive de la musique entraînerait donc des adaptations neuronales aux niveaux de ces réseaux communs, ce qui en retour induirait des bénéfices dans le traitement des sons de parole.

Expertise musicale et langue étrangère

Il est généralement admis que les musiciens ont plus de facilité à apprendre une langue étrangère comparés aux non-musiciens. Si, comme nous l'avons vu précédemment, la perception des sons musicaux et langagiers sollicitent des réseaux neuronaux communs, nous pouvons supposer que les sons d'une langue étrangère utilisent au moins en partie les mêmes circuits, et donc que la pratique musicale pourrait avoir une influence sur leur traitement. Cette idée, plutôt intuitive, que les personnes ayant une oreille "plus fine" seraient plus à même d'apprendre une nouvelle langue n'est pas dénuée de preuves scientifiques (Slevc and Miyake, 2006).

La parole est un signal acoustique complexe qui contient de nombreux indices acoustiques et phonologiques importants pour comprendre le message véhiculé par le locuteur. Les effets de la pratique musicale sur la perception des langues étrangères ont donc été surtout étudiés au niveau de ces indices. L'un d'entre eux, le *pitch*, fut le sujet de nombreuses études, probablement car cette caractéristique acoustique véhicule des informations importantes à la fois dans la musique et dans certaines langues (Chinois Mandarins, Portugais, Espagnol,...), ce qui réunit des conditions favorables aux effets de transfert tels que le prédit l'hypothèse OPERA (Patel, 2011). Il a été montré, par exemple, que les français musiciens ont de meilleures capacités à discriminer les fines variations d'intonations (i.e. du *pitch*) en Portugais comparés aux non-musiciens (Marques et al., 2007). De plus, cet avantage se reflète au niveau cortical par des potentiels évoqués plus précoces (P300), révélant ainsi l'influence positive de l'expertise musicale sur le traitement de la prosodie des langues étrangères. Les variations du *pitch* sont encore plus fines et porteuses de sens dans les langues tonales puisqu'elles permettent de discriminer des mots monosyllabiques. Delogu et al. (2006) ont montré que chez les italiens, pour qui des variations aussi subtiles du *pitch* ne sont pas phonologiquement importantes, l'expertise musicale

améliorait leurs capacités à discriminer des patterns de hauteur tonale en Chinois Mandarin. Cet avantage à discriminer les tons se reflète au niveau cortical par une MMN de plus grande amplitude chez les musiciens par rapport aux non-musiciens (Martínez-Montes et al., 2013). Au niveau sous-cortical, nous avons vu précédemment avec l'étude de Wong et collaborateurs (2007) que les musiciens anglais avaient un encodage sous-cortical de la F0 des syllabes chinoises plus robuste comparés aux non-musiciens anglais, suggérant ainsi que l'exposition et la manipulation intensive du *pitch* des sons durant la pratique musicale influence de manière positive l'encodage des syllabes tonales Chinoises à un niveau de traitement précoce. Ces études montrent que les effets de plasticité induits par la pratique musicale n'affectent pas uniquement le traitement des sons de la langue maternelle, mais peuvent se généraliser aussi aux sons d'une langue étrangère.

L'ensemble des travaux portant sur le lien entre musique et langage indique clairement que les musiciens ont un encodage neuronal des sons plus fin et plus robuste, ce qui leurs confère des avantages perceptifs comparés aux non-musiciens. Les bénéfices induits par la pratique musicale se transfèrent aussi à d'autres domaines de compétences utilisant les mêmes réseaux neuronaux, comme le langage, qu'il soit natif ou étranger. Les études abordant les effets de transfert entre musique et langue étrangère à un niveau pré-attentif du traitement, c'est-à-dire au niveau sous-cortical, restent encore rares, et se sont principalement focalisées sur les avantages des musiciens dans l'encodage des syllabes tonales du Chinois Mandarin (c.f. Wong et al., 2007). Il nous a semblé intéressant de déterminer si, au niveau sous-cortical, l'amélioration du codage des sons étrangers pouvait être observée pour des langues beaucoup plus proches au niveau phonétique. Nous avons tenté de répondre à cette question dans la troisième étude de cette thèse.

Etant donné que l'exposition et la pratique intensive de la musique semblent induire une réorganisation du système auditif et améliorer l'efficience du traitement des sons, nous pouvons nous demander si, notre langue maternelle, à laquelle nous sommes constamment exposés depuis notre naissance et même avant, peut sculpter de la même manière le système auditif et améliorer spécifiquement l'encodage des sons appartenant au répertoire phonémique de notre langue.

2.2.2. Expérience linguistique

L'exposition à la langue maternelle commence très tôt dans le développement d'un individu. Durant l'état fœtal, l'enfant en gestation est déjà exposé aux sons émis par sa mère et par l'environnement. Bien que seule une partie du spectre audible puisse atteindre les oreilles de

l'enfant (principalement les basses fréquences), ce dernier est déjà exposé fréquemment à sa langue maternelle. Bien avant la capacité à produire des mots, l'enfant est déjà sensible aux sons vocaux et aux différentes combinaisons possibles pour créer des mots et des phrases dans sa langue maternelle. Dès la naissance, l'enfant peut reconnaître la prosodie de sa langue maternelle (Mehler et al., 1988). Vient ensuite la construction de son répertoire phonétique pendant sa première année de vie : la sensibilité aux voyelles de sa langue maternelle augmente (Kuhl et al., 1992), puis celle aux consonnes (Werker and Tees, 2002), puis finalement l'enfant devient sensible aux combinaisons de phonèmes qui sont possibles dans sa langue maternelle (Jusczyk and Luce, 1994). A mesure que l'enfant grandit, il devient de plus en plus sensible aux caractéristiques abstraites des sons de sa langue, telles que les mots (Shi et al., 2006), la grammaire (Shi, 2014) et la syntaxe (Bernal et al., 2010). Il est intéressant de noter qu'à mesure que l'enfant devient "expert" dans le traitement des sons de sa langue maternelle, il perd en contrepartie sa capacité à discriminer les sons verbaux des langues qui lui sont étrangères, phénomène décrit sous le nom de « perceptual narrowing » (Cheour et al., 1998; Kuhl et al., 2006). Par exemple, à l'âge de 6-8 mois un enfant anglais est encore sensible aux contrastes phonologiques présents dans l'Hindi ou le Salish, avec des performances aussi bonnes que des enfants natifs, alors que déjà à l'âge de 10-12 mois, leurs performances chutent pour rejoindre celles d'adultes anglais (Werker and Tees, 2002).

Les réseaux de neurones vont, à mesure de l'exposition croissante aux sons de la langue maternelle, se réorganiser afin de traiter les sons natifs de manière plus fine et plus robuste comparés aux sons des langues étrangères. Au niveau cortical, Cheour et al. (1998) ont montré que de l'âge de 6 à 12 mois, l'amplitude des réponses MMN diminue drastiquement pour les contrastes phonémiques de sons verbaux non-natifs. De même, à l'âge de 6 mois, les enfants montrent des variations de l'activité gamma plus importantes lorsqu'il s'agit de contrastes phonémiques natifs comparés aux contrastes non-natifs (Ortiz-Mantilla et al., 2013). Chez l'adulte, cette perte de sensibilité aux contrastes phonémiques non-natifs perdure puisque chez des japonais, des phonèmes français très proches au niveau acoustique vont engendrer une MMN très réduite en amplitude voire même absente comparés aux français adultes (Dehaene-Lambertz et al., 2000). Les données métaboliques apportent un autre point de vue sur ces effets de plasticité dépendants de l'expertise linguistique. Raizada et al. (2010) ont par exemple enregistré les patterns d'activité métabolique chez les japonais lors de l'écoute des syllabes anglaises /ra/ et /la/, pour qui le contraste phonémique entre les consonnes /r/ et /l/ n'existe pas. Les résultats ont montré que la séparabilité des patterns d'activité des deux syllabes prédisait les performances des sujets à discriminer les deux syllabes (Raizada et al., 2010).

Ces données montrent que le réseau cortical se réorganise durant le développement en fonction des contrastes phonémiques présents dans la langue maternelle, ce qui en contrepartie affecte négativement l'encodage neuronal des sons verbaux étrangers.

Hypothèses bottom-up et top-down

Deux principaux mécanismes ont été proposés afin d'expliquer cette sensibilité précoce aux sons de la langue maternelle : l'apprentissage statistique (*statistical learning*) et la sensibilité aux patterns abstraits de la parole. L'apprentissage statistique est un processus *bottom-up* où la force d'encodage des informations présentes dans l'environnement du sujet dépendra de leur distribution statistique. Ce mécanisme permet aux enfants d'apprendre quels sont les phonèmes ayant le plus de probabilité de se suivre pour former des syllabes ou des mots (Kuhl, 2004). En effet, la parole est un flux acoustique continu qu'il faut savoir segmenter afin de séparer les mots et de les associer à des représentations conceptuelles. En général, les syllabes qui appartiennent au même mot ont plus de chance de se suivre, à l'inverse des syllabes qui recouvrent les limites des mots, c'est-à-dire celles en début et fin des mots. Par exemple, en français les syllabes /ba/ et /ton/ ont plus de chances de se suivre dans un flux de parole continu puisqu'elles sont associées à des mots présents dans la langue française (e.g. /bâton/, /bâtonnet/), comparées aux syllabes /ba/ et /mi/. En général, ce mécanisme est mis en évidence en exposant de manière passive un enfant à un flux de parole synthétique et artificiel (i.e. avec des syllabes et des mots n'existant pas dans sa langue). L'idée est que les seuls indices nécessaires pour segmenter le flux de parole en mots sont les probabilités statistiques entre pairs de syllabes (haute probabilité au sein d'un mot, et faible entre les mots). A posteriori, l'apprentissage de l'enfant est évalué dans un test de choix forcé entre deux items, l'un étant un mot faisant partie du langage artificiel, l'autre non. Dans l'étude de Maye et al. (2002), deux groupes d'enfants ont été exposés à un continuum phonétique entre les syllabes /da/ et /ta/. Pour un groupe, le continuum présentait une distribution bimodale, c'est-à-dire que les extrêmes de ce continuum /da/ et /ta/ étaient plus fréquents que les syllabes hybrides /dta/ ou /tda/ se trouvant au milieu du continuum (voir **Figure 11**). A l'inverse pour l'autre groupe, les syllabes hybrides étaient plus fréquentes. Seuls les enfants exposés à la distribution bimodale ont été par la suite capables de discriminer les syllabes /da/ et /ta/, révélant ainsi l'influence de la distribution statistique des phonèmes dans l'apprentissage des langues.

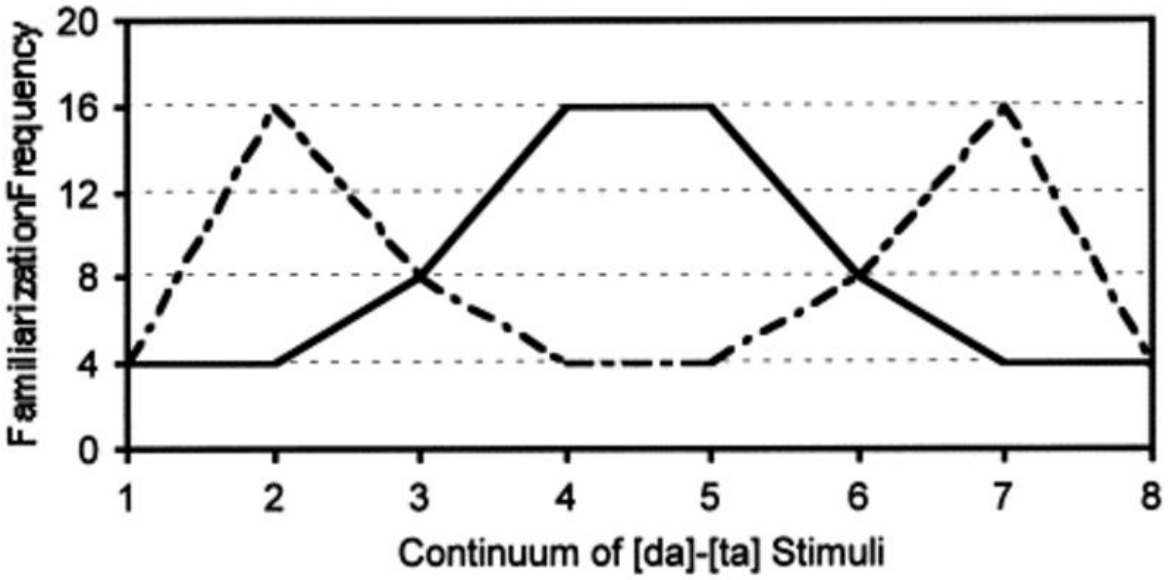


Figure 11. Distribution statistique bimodale (tracé en pointillés) et unimodale (tracé plein) du continuum /da/-/ta/ auquel ont été exposés les deux groupes d'enfants durant la phase de familiarisation dans l'étude de Maye et al. (2002).

Pendant longtemps l'approche *bottom-up* des processus d'apprentissage de la langue a prédominé, mais elle présente certaines limites. Il existe en effet un recouvrement important des catégories phonémiques dans le langage naturel (**Figure 12**). Ainsi, certains phonèmes, bien que distincts phonétiquement parlant, peuvent présenter une similarité acoustique importante selon le locuteur, la vitesse d'élocution ou selon la place du phonème dans un mot ou une phrase. Ces facteurs vont induire une variabilité acoustique importante pour un même phonème. Par exemple, les voyelles anglaises /ae/ et /ɛ/ sont très similaires au niveau acoustique, avec des rapports formantiques proches. La variabilité dans la prononciation de ces voyelles se traduit par des distributions qui se recouvrent de manière importante dans l'espace vocalique. De ce fait, le *statistical learning*, et plus largement les hypothèses *bottom-up*, ne peuvent pas expliquer complètement le développement du répertoire phonémique de la langue maternelle. D'autres hypothèses ont donc été proposées, ayant cette fois une approche *top-down*, dans lesquelles les représentations de haut niveau auraient une influence sur le traitement et la catégorisation des informations acoustiques entrantes au niveau des structures sensorielles.

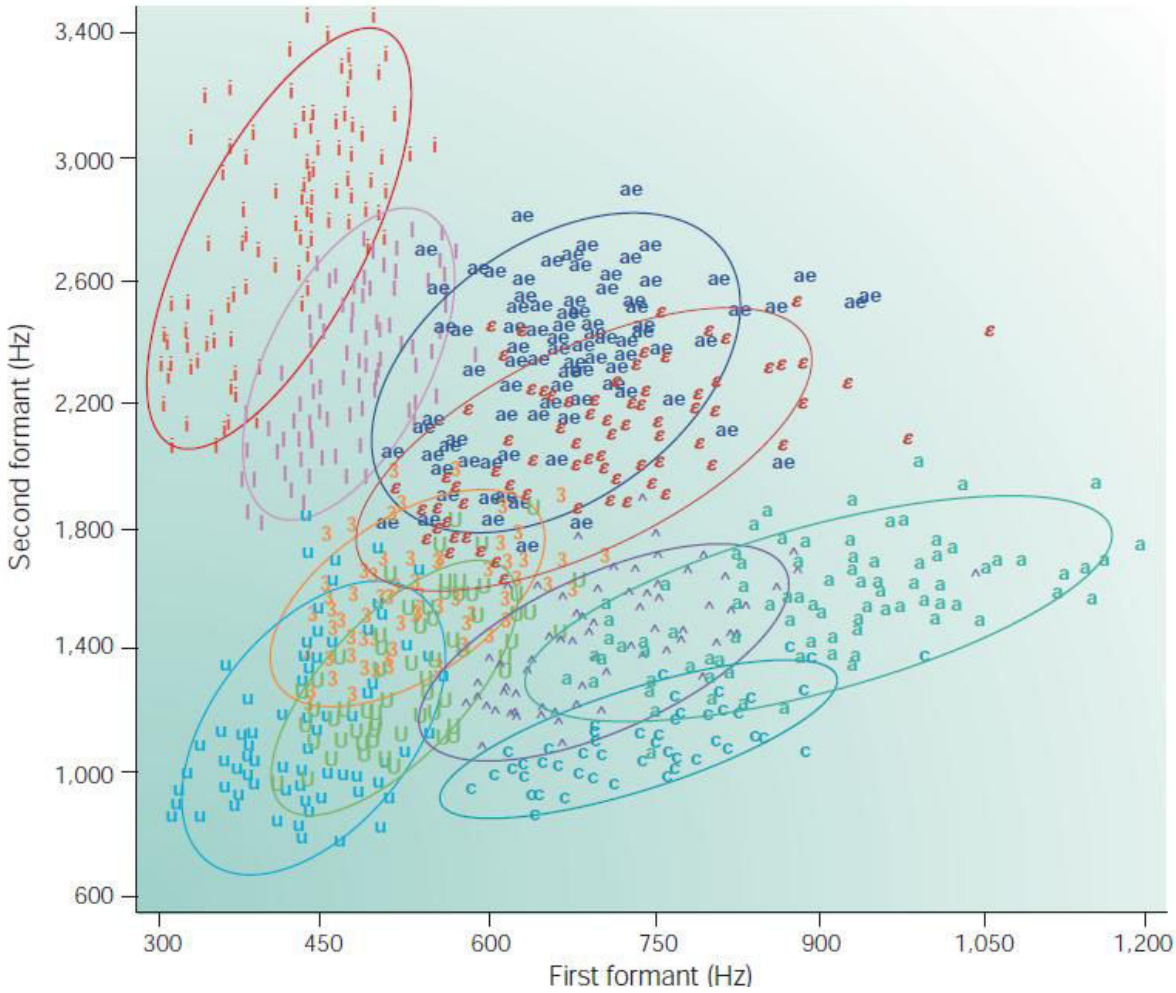


Figure 12. Illustration de la variabilité acoustique des phonèmes anglais dans l'espace vocalique. Valeurs moyennes des premiers (abscisse) et seconds formants (ordonnée) pour 12 voyelles anglaises prononcées par des hommes, femmes et enfants (issue de Kuhl et al. 2004).

L'une de ces hypothèses, nommée *Native Language Neural Commitment* (*NLNC*), propose que l'exposition à la langue maternelle forme des réseaux de neurones dédiés à l'encodage des patterns sonores présents dans notre langue, ce qui va influencer l'apprentissage futur de nouvelles langues (Kuhl, 2000). Des évidences provenant de la magnétoencéphalographie (MEG) viennent appuyer cette hypothèse : le traitement des sons étrangers entraîne une activation cérébrale plus étendue et plus longue contrairement aux sons de la langue maternelle (Zhang et al., 2005). Cette étude révèle que l'exposition précoce à la langue maternelle va induire un "engagement neuronal" (*neural commitment*) qui va interférer avec le traitement des sons verbaux étrangers, le rendant ainsi moins efficient. De ce fait, les connaissances linguistiques *a priori* ont une influence sur l'encodage des sons verbaux perçus *a posteriori*, ce qui suggère la mise en jeu de processus *top-down* venant moduler l'encodage de nouvelles informations acoustiques.

Une autre hypothèse, initialement développée pour le système visuel, nommée *The Reverse Hierarchy Theory*, propose que l'exposition sur le long terme à certains contextes sensoriels induit une réorganisation neuronale qui débuterait au niveau des structures de haut niveau puis se propagerait progressivement aux structures de bas niveau (Ahissar and Hochstein, 2004). Cette hypothèse a été récemment étendue au système auditif suite à des études ayant mis en évidence l'influence des réorganisations corticales sur le traitement des sons par les neurones sous-corticaux (Gutschalk et al., 2008; Suga, 2008; Gao and Suga, 2000).

L'ensemble de ces travaux renforce l'hypothèse selon laquelle les effets de plasticité dépendants de l'expertise linguistique sont le résultat d'une interaction continue à double sens entre les structures associatives de haut niveau et les structures sensorielles de bas niveau.

Plasticité expérience-dépendante au niveau du tronc cérébral : état des lieux

Dans le but de tester comment l'expérience linguistique peut avoir une influence sur le traitement des sons à une étape plus précoce du traitement, Krishnan et collaborateurs ont comparé les réponses du tronc cérébral évoquées par des syllabes tonales du Chinois Mandarin chez des sujets natifs ou anglais (Krishnan et al., 2005). Comme nous l'avons vu précédemment, en Chinois Mandarin les variations de la hauteur tonale constituent des informations linguistiques importantes pour les locuteurs natifs puisqu'elles permettent de faire la distinction ente des mots monosyllabiques. Sachant que le principal corrélat acoustique de la hauteur tonale est la fréquence fondamentale (F0), leurs analyses se sont focalisées sur la qualité et la robustesse de l'encodage de cette variable acoustique chez les deux groupes de sujets. Les résultats montrent que les Chinois ont une FFR qui suit de manière plus fidèle les variations de la F0 des stimuli, en particulier lors des variations rapides (voir **Figure 13**). Ces résultats montrent qu'une exposition plus intense à des patterns de hauteur tonale complexes et variés va renforcer l'encodage de la F0 à un niveau précoce du traitement.

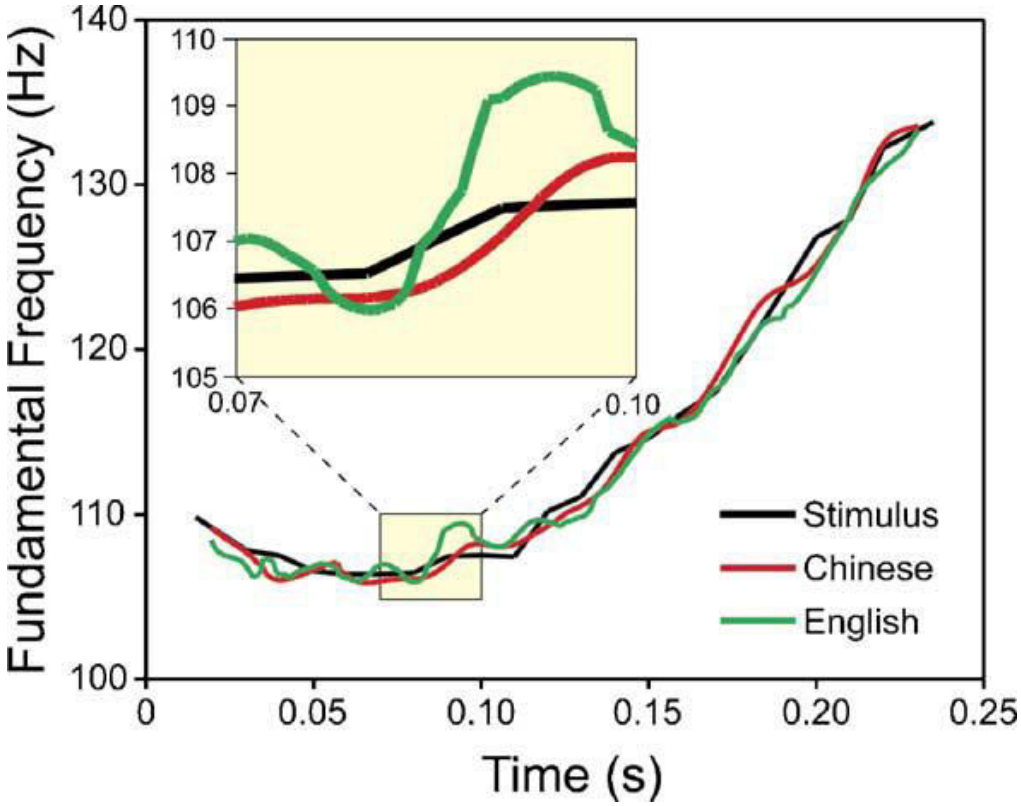


Figure 13. Trajectoires moyennes de la F0 extraites des FFRs des chinois (rouge) et des anglais (vert). La courbe noire indique la trajectoire de la F0 du stimulus (syllabe /yi/) utilisé pour évoquer les FFRs. L'encadré montre que les chinois ont une FFR qui suit de manière plus fidèle les variations de F0 du stimulus comparée à celle des anglais. Issue de Krishnan et al. (2005).

Afin de déterminer si ces différences d'encodage de la F0 étaient liés à des effets purement linguistiques ou bien à une meilleure capacité globale à encoder les trajectoires de la F0 chez les Chinois, cette même équipe a utilisé dans une étude ultérieure des sons Mandarins dont la F0 avait une direction similaire (i.e. montante ou descendante) aux sons utilisés dans la première expérience, mais dont les trajectoires étaient différentes, cette fois-ci linéaires au lieu d'être curvilinéaires (voir **Figure 14**). Aucune différence entre les deux groupes de participants n'a été observée, ce qui signifie que les Chinois présentent un avantage dans l'encodage sous-cortical de la F0 uniquement lorsque les stimuli ont des trajectoires de hauteur tonale présentes naturellement dans leur langue (Xu et al., 2006). Les auteurs sont allés encore plus loin dans leurs investigations en posant la question de savoir si les effets d'expertise linguistique seraient spécifiques à certaines dimensions acoustiques, indépendamment de tout contexte langagier. Ils ont pour cela utilisé des sons Mandarins dépourvus de tout attribut linguistique, en utilisant du bruit blanc (*iterated rippled noise*) dont les variations de la F0 suivaient celles présentes dans les stimuli Mandarins naturels. Les résultats ont révélé un encodage plus robuste des trajectoires fréquentielles des sons, et plus spécifiquement au niveau des variations rapides de la F0 et

seulement pour les sons ayant les trajectoires fréquentielles les plus complexes (Krishnan et al., 2009). De ce fait, les avantages des Chinois dans l'encodage des sons Mandarins ne peuvent s'expliquer uniquement par une meilleure capacité globale d'encodage de la hauteur tonale, sinon les auteurs auraient dû observer un encodage plus robuste au niveau de toutes les portions du stimulus, et en outre pour tous les stimuli utilisés. D'après ces études les réorganisations du tronc cérébral induites par l'expérience linguistique sont dépendantes de certaines dimensions des trajectoires fréquentielles présentes dans la langue Chinoise. Ces découvertes révèlent aussi que l'expérience linguistique chez les Chinois a modifié la façon dont leurs neurones sous-corticaux encodent les patterns de *pitch* indépendamment du contexte (langagier ou non) dans lequel ils se trouvent.

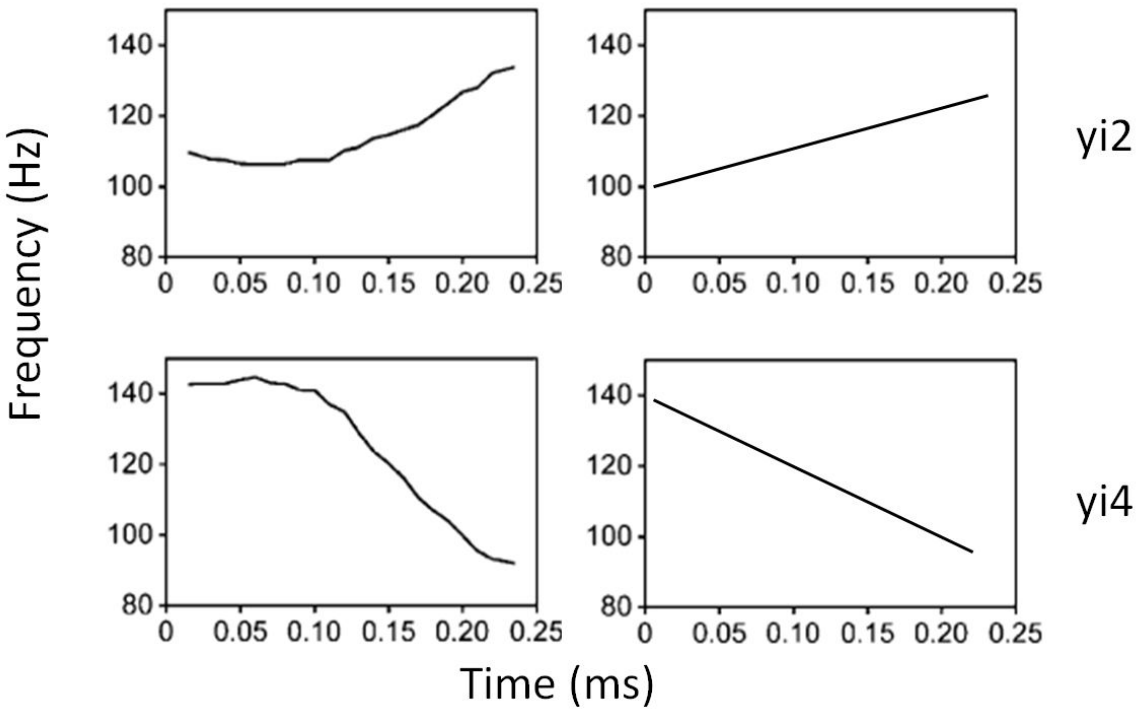


Figure 14. Illustration des différences de trajectoire de la fréquence fondamentale des stimuli Mandarins avec (à gauche) ou sans contexte linguistique (à droite). Images issues et adaptées de Xu et al. (2006) et Krishnan et al. (2005).

En conclusion de ces études, même si les différences cross-linguistiques observées entre Chinois et Anglais dans l'encodage sous-cortical du *pitch* émergent d'une expérience linguistique différente, les effets d'une telle expérience ne sont pas spécifiques à un contexte langagier. Cela rejoint le fait que les bénéfices induits par l'expertise musicale ne sont pas spécifiques au domaine musical, mais se transfèrent à d'autres domaines comme la parole (Musacchia et al., 2007; Wong et al., 2007).

Ces découvertes ont permis de mieux comprendre l'influence de l'expertise linguistique sur les réorganisations des neurones sous-corticaux. Néanmoins, si les langues tonales utilisent les variations de la F0 pour catégoriser les mots monosyllabiques, les langues non-tonales, comme le Français ou l'Anglais, utilisent des processus de catégorisation phonémique qualitativement différents comme le temps de voisement (VOT), les trajectoires formantiques ou encore le rapport entre le premier et le second formant. Nous pouvons donc nous demander si de tels effets d'expertise linguistique seraient observables en comparant des langues beaucoup plus proches, utilisant le même système de catégorisation phonémique. C'est à travers notre première étude que nous avons tenté de répondre à cette question, en comparant l'encodage sous-cortical d'une syllabe française ou anglaise, chez des Américains et Français. Etant donné que Krishnan et collaborateurs ont rapporté que les effets d'expertise se retrouvaient spécifiquement au niveau de certaines dimensions linguistiques importantes, nous avons postulé que, si de tels effets existaient en comparant nos deux populations, ils devraient se retrouver au niveau d'indices acoustiques tels que les formants jouant un rôle important dans la discrimination des voyelles.

Transition de l'expertise au contexte

Nous avons vu au cours de ce chapitre qu'une expérience auditive, comme la pratique d'un instrument de musique, ou plus simplement de notre langue maternelle, peut influencer la manière dont le cerveau traite les informations auditives. Ces effets liés à l'expérience sont le produit d'une réorganisation neuronale du système auditif, qui devient alors plus performant pour traiter les informations les plus pertinentes au niveau comportemental et/ou linguistique. Néanmoins, si de telles réorganisations sont observables après des années de pratique musicale ou linguistique, il devrait exister des effets évanescents à plus court terme qui, avec le temps et la pratique, se cumuleraient pour ensuite s'ancrer de manière permanente au sein des réseaux de neurones. De plus, les effets induits par l'expérience auditive ne sont pas globaux, c'est-à-dire que le système auditif se réorganise uniquement pour mieux traiter certaines dimensions acoustiques importantes (e.g. la F0 pour les musiciens et les chinois Mandarins). Or, un des processus cognitifs à court terme permettant la sélection d'informations pertinentes est l'attention. C'est pourquoi nous pensons que les effets de plasticité dépendants de l'expérience observés au niveau sous-cortical puissent provenir de modifications à plus court terme comme celles induites par l'attention sélective. Etant donnée la littérature abondante concernant les effets attentionnels au

niveau cortical, nous allons nous focaliser sur les découvertes qui ont été faites récemment au niveau sous-cortical, et qui sont plus directement en lien avec le sujet de cette thèse.

2.3. Plasticité dépendante du contexte

Dans la vie de tous les jours, nous sommes confrontés à des conditions d'écoute qui ne sont pas toujours optimales. Il est rare, par exemple, lors d'une conversation que seule la voix du locuteur parvienne jusqu'à nos oreilles. En général, le signal de parole est mélangé avec du bruit environnant (e.g. autres locuteurs, bruits de circulation, etc...). Il est donc nécessaire pour celui qui reçoit le signal acoustique d'extraire du bruit environnant les informations qu'il juge pertinentes. La vue *bottom-up* classique du système auditif, dans laquelle les informations sensorielles sont relayées de la périphérie vers les centres, ne peut pas expliquer complètement comment nous sommes capables d'extraire ces informations du bruit de fond. Les entrées sensorielles sont donc probablement modulées par des processus *top-down*, afin d'augmenter le rapport signal sur bruit et permettre un encodage des sons plus optimal.

2.3.1. Effets du contexte attentionnel

Les effets du contexte attentionnel ont principalement été étudiés au niveau cortical. Il en ressort que les informations auditives auxquelles le sujet porte attention sont mieux traitées et les réponse corticales (N1 et P3) amplifiées (Teder-Sälejärvi and Hillyard, 1998). Dans des conditions d'écoutes difficiles, lorsque le sujet doit extraire une syllabe mélangée à d'autres voix (i.e. l'effet *cocktail party*), les réponses corticales évoquées par la syllabe sont plus importantes en condition d'écoute active que passive (Zhang et al., 2016). Certaines études ont même rapporté des effets attentionnels dans les réponses précoces du cortex auditif, à environ 50 ms après la présentation du stimulus (van Atteveldt et al., 2015), ce qui suggère que les mécanismes d'attention sélective pourraient prendre effet bien avant l'étage cortical, en filtrant les informations sensorielles entrantes au niveau des structures sous-corticales.

En revanche, les études concernant les effets de l'attention sur l'activité des structures sous-corticales sont peu abondantes et leurs résultats ont été sujets à débat pendant longtemps.

Les premières études ont par exemple montré que les *click-ABRs* ne sont pas affectés par l'attention (Hackley, 1993; Hackley et al., 1990). A l'inverse, des études plus récentes, utilisant des stimuli plus complexes, ont montré que la latence et l'amplitude de la FFR sont modulées par l'attention, avec une latence plus précoce (Hoormann, Falkenstein, & Hohnsbein, 1994; Jörg Hoormann, Falkenstein, & Hohnsbein, 2000) et une amplitude plus importante (Galbraith et al., 1998, 2003; Galbraith and Doan, 1995; Hoormann et al., 2000, 1994) en réponse aux stimuli attendus. De même, en présentant deux syllabes, l'une ayant une fréquence d'apparition élevée (sons fréquents) et l'autre faible (sons déviants), Slabu et al. (2012) ont montré des réponses de détection de nouveauté au niveau sous-cortical, telles que celles qui sont observées au niveau cortical avec la MMN (Näätänen et al., 2011). De la même manière, Hairston et al. (2013) ont montré que la mobilisation des ressources attentionnelles pour une autre tâche diminue l'amplitude de la FFR. Il est probable que l'absence d'effet attentionnel dans les premières études sur les *clicks-ABRs* soit liée au fait que les stimuli utilisés avaient un contenu acoustique trop pauvre et étaient présentés à une cadence trop rapide pour que les processus attentionnels puissent avoir une influence.

Même si la littérature sur ce sujet reste encore rare, ce corpus d'études semble indiquer que, dans des conditions d'écoute active, l'encodage neuronal des sons complexes au niveau sous-cortical est affecté par des processus attentionnels *top-down*, qui modifient le seuil d'excitabilité et/ou les propriétés de *phase-locking* des neurones sous-corticaux qui sous-tendent la génération de la FFR. Il existe en effet au sein du système auditif de nombreuses projections descendantes qui peuvent supporter de telles modulations attentionnelles cortico-sous-corticales (Kral and Eggermont, 2007).

L'attention peut être orientée de plusieurs manières : dans l'espace, lorsqu'il s'agit par exemple de localiser la source d'un bruit, et dans le temps, ce qui permet par exemple de prédire l'occurrence d'un son afin d'optimiser l'encodage des informations acoustiques attendues. L'attention spatiale renvoie à la notion d'*active sensing* qui tient compte du fait que le système moteur oriente les organes sensoriels en direction des stimuli (e.g. saccades oculaires, mouvements de la tête, reniflements, etc...), ce qui module les informations sensorielles entrantes par des interactions entre les systèmes moteurs, sensoriels et attentionnels. L'attention temporelle renvoie quant à elle à la *Dynamic of Attending Theory* (*DAT*) qui propose que, puisque la plupart des stimuli auditifs sont organisés de manière rythmique (e.g. la musique, la parole,...), l'attention temporelle devrait osciller dans le temps en fonction des régularités temporelles des indices acoustiques les plus pertinents (Jones, 1976; Jones and Boltz, 1989; Large and Jones, 1999). Nous

allons aborder brièvement ces deux notions et les questions qu'elles soulèvent quant aux modulations attentionnelles au niveau sous-cortical, questions auxquelles nous avons essayé de répondre au cours de notre troisième étude.

2.3.2. *Active sensing* et *Dynamic Attending Theory*

Active sensing

Les informations (visuelles, auditives, tactiles,...) présentes dans notre environnement ne parviennent pas à nos organes sensoriels de manière passive. Au contraire, l'exploration de l'environnement est un processus actif par lequel les organes sensoriels sont orientés en direction des stimuli les plus pertinents. Par exemple, les saccades oculaires permettent de lire ou d'orienter notre regard en direction d'un locuteur afin de percevoir les gestes qui accompagnent son discours, et ainsi mieux comprendre le message qu'il est en train de nous transmettre. Afin d'apprécier le volume et la forme d'un objet, nous devons bouger les doigts, les faire glisser à la surface de l'objet. Lorsque nous nous apprêtons à traverser la rue, si une nous entendons une voiture arriver, nous orientons notre tête en direction de celle-ci afin de l'éviter. L'*active sensing* est une notion qui suggère que l'environnement est échantillonné et structuré via des interactions continues entre les organes sensoriels et moteurs (Schroeder et al., 2010; Wachowiak, 2011). Une copie d'efference des commandes motrices serait envoyée par le système moteur en direction des organes sensoriels dans le but de moduler les informations *bottom-up* entrantes (Crapse and Sommer, 2008). Ce mécanisme *top-down* permet de prédire "où" et "quand" l'information sensorielle a le plus de chance d'apparaître (Sommer and Wurtz, 2008), ceci en vue d'améliorer son encodage.

L'audition est un sens particulier puisque, mis à part en réalisant des mouvements de la tête, l'Homme est incapable d'orienter ses organes sensoriels, c'est-à-dire ses oreilles, en direction des sons. Néanmoins, nous pouvons supposer que si le système moteur et attentionnel sont en interaction permanente au niveau de tous les autres sens, il se pourrait que cela soit aussi le cas pour l'audition. Comment alors ces interactions peuvent se produire si aucun mouvement des organes sensoriels n'est possible et aucune copie d'efference motrice n'est envoyée à la périphérie ? Une hypothèse est que les fluctuations de l'attention permettraient de moduler le flux acoustique entrant afin de renforcer l'encodage des indices acoustiques les plus pertinents (Schroeder et al., 2010, 2008). Selon cette approche, la perception auditive peut être vue comme

un processus actif dans lequel les systèmes auditifs, moteurs et attentionnels agissent en concert afin de traiter l'information acoustique de manière efficiente (Mottonen et al., 2014).

Dynamic of Attending Theory (DAT)

La *Dynamic of Attending Theory* (*DAT*) initialement proposée par Jones en 1976 suggère que l'attention fluctuerait dans le temps au même rythme que les régularités temporelles présentes dans l'environnement acoustique. Selon cette théorie, les niveaux attentionnels seraient entraînés par la structure temporelle des évènements externes, de telle manière que les évènements acoustiques apparaissant en phase avec des niveaux attentionnels élevés seraient encodés de manière plus efficiente (Jones and Boltz, 1989; Large and Jones, 1999). Dans la parole et la musique, les régularités temporelles sont véhiculées à travers plusieurs échelles : la parole peut être segmentée en phrases, mots, syllabes et phonèmes, et la musique en mélodies, rythmes et notes. Cette organisation temporelle permet de prédire l'occurrence des évènements à venir afin d'allouer plus de ressources attentionnelles et améliorer leur encodage. Par exemple, la musique présente des temps forts et des temps faibles qui correspondent à sa structure métrique. Selon la *DAT*, les cycles attentionnels seraient entraînés par la métrique de la musique, de telle manière que les notes apparaissant à des moments forts seraient associées à des niveaux attentionnels plus élevés, alors que le phénomène inverse se produirait aux moments faibles de la métrique (voir l'illustration de cet exemple **Figure 15).**

Les études comportementales et électrophysiologiques ont apporté des preuves venant appuyer la *DAT*. Dans la modalité auditive, Jones et al. (2002) ont montré que lorsque des sons sont présentés avec des intervalles inter-stimuli réguliers, c'est-à-dire avec une occurrence prédictible, les sujets étaient plus performants pour évaluer la hauteur tonale des sons que lorsque ces mêmes stimuli sont présentés avec des intervalles irréguliers. Des résultats similaires ont été observés pour la modalité visuelle : les sujets étaient plus précis pour détecter des stimuli visuels lorsqu'ils sont présentés dans un flux rythmique comparés à des flux visuels non-rythmiques (Cravo et al., 2013; Rohenkohl et al., 2012). Dans la parole, lorsque l'alternation régulière de syllabes stressées et non-stressées est perturbée, on observe un effet P600 au niveau des réponses corticales qui pourrait correspondre à un réajustement de la période et de la phase des oscillations neuronales nécessaire pour maintenir la synchronisation avec la structure métrique de la parole (Schmidt-Kassow and Kotz, 2009). Ces résultats montrent que les attentes temporelles des sujets

optimisent la perception des sons en modulant de manière dynamique le gain de l'information sensorielle.
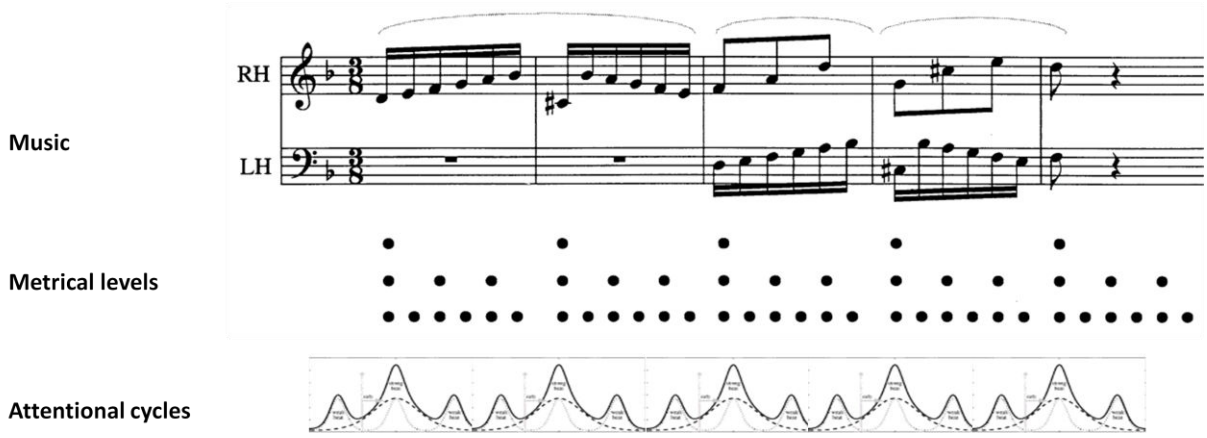


Figure 15. Illustration de la *Dynamic of Attending Theory*. Une pièce de musique complexe (en haut) est organisée dans le temps selon plusieurs niveaux de métriques (au milieu) caractérisés par des temps forts et des temps faibles. Les cycles attentionnels sont entraînés par la structure métrique avec des niveaux attentionnels plus élevés au niveau des temps forts et inversement au niveau des temps faibles. Adaptée de Large and Palmer (2002).

Les notions d'*active sensing* et de *Dynamic of Attending* sont étroitement liées si l'on considère que l'activité motrice aide à synchroniser les fluctuations temporelles de l'attention avec le déroulement temporel des évènements externes. Des études récentes montrent en effet que les activités oscillatoires beta, qui sont liées à l'activité motrice (Jenkinson and Brown, 2011), suivent la cadence de présentation isochrone des stimuli sonores (voir **Figure 16)**, et se resynchronisent juste avant la présentation du prochain stimulus. Ces résultats suggèrent que les oscillations neuronales beta modulent de manière dynamique l'interaction entre le système moteur, attentionnel et auditif, en fournissant un mécanisme neuronal de prédiction temporelle des évènements (Fujioka et al., 2012). Morillon et collaborateurs ont également montré qu'une activité motrice rythmique, comme taper en rythme, améliore la détection de cibles sonores qui apparaissent en phase avec le rythme moteur alors que celles tombant en opposition de phase sont moins bien détectées (Morillon et al., 2014).

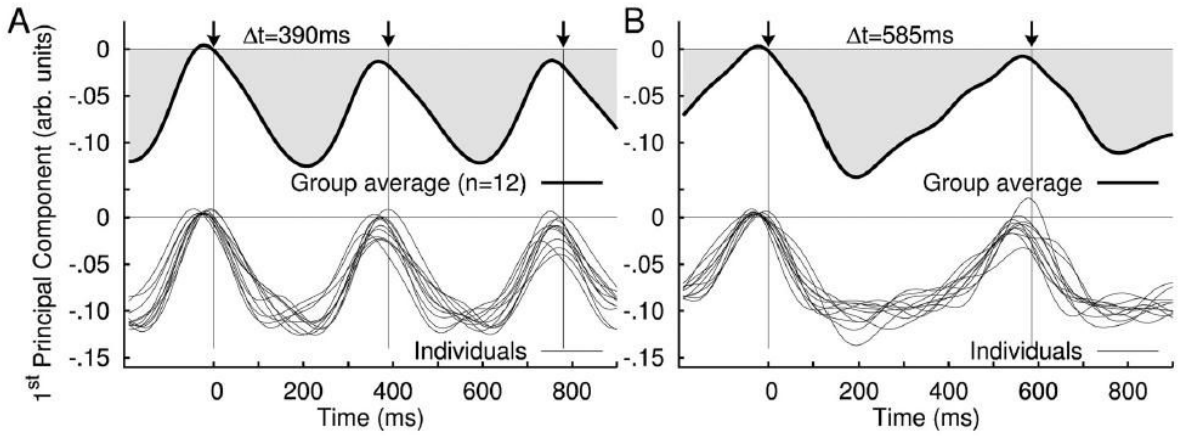


Figure 16. Patterns des activités oscillatoires beta en fonction de l'intervalle interstimulus. Les activités oscillatoires beta s'adaptent à la cadence de stimulation et se resynchronisent juste avant l'apparition du stimulus suivant, ce qui montre une mécanisme neuronal de prédiction temporel. Adapté de Fujioka et al. (2012).

Dans notre troisième étude, nous avons voulu déterminer si la notion d'*active sensing* et la *DAT* sont aussi valables au niveau sous-cortical. Le protocole que nous avons utilisé nous a permis de tester ces deux notions en même temps. Dans une première condition, nous avons enregistré les *cABRs* évoquées par une syllabe française, pendant que les sujets regardaient un film sous-titré, de manière à détourner leur attention des stimuli, et obtenir ainsi des réponses neuronales avec un niveau attentionnel le plus faible. Dans une deuxième condition, nous avons demandé aux sujets de taper au rythme d'un stimulus sur deux. En comparant les réponses neuronales recueillies durant la phase d'écoute passive et active, nous avons pu dans un premier temps tester la notion d'*active sensing*, à savoir si l'attention dirigée par l'activité motrice optimise le traitement des sons. Comme les mouvements phasique tels que le tapping sont supposés entraîner les ressources attentionnelles en phase avec le mouvement (Morillon et al., 2014; Mottonen et al., 2014), nous avons pu déterminer si les modulations rythmiques de l'attention avaient un effet sur l'encodage sous-cortical des sons verbaux (voir schéma du protocole **Figure 17**). Nous avons supposé que les stimuli tombant en phase avec la cadence de tapping devraient être associés à des niveaux attentionnels plus élevés et donc être traités de manière plus efficiente que ceux apparaissant en opposition de phase.

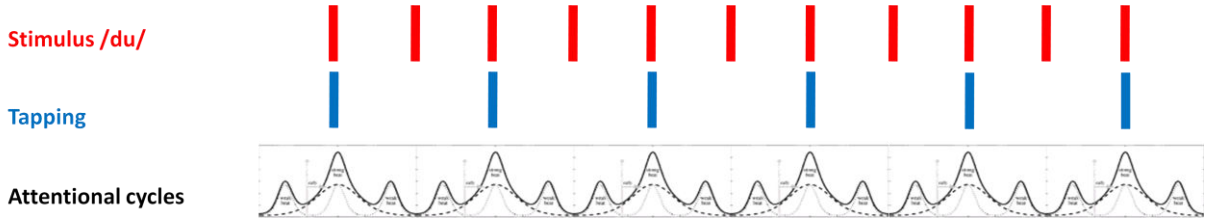


Figure 17. Schéma du protocole de tapping utilisé dans la condition d'écoute active durant la troisième étude. Les traits rouges représentent la présentation du stimulus. Les traits bleus représentent la tâche de tapping réalisée par le sujet (i.e. un stimulus sur deux). Les tracés noirs représentent les niveaux attentionnels prédits par la *DAT* en synchronisation avec les mouvements phasiques du tapping.

Objectifs et hypothèses

Dans les voies afférentes, le tronc cérébral se trouve à mi-chemin entre la périphérie et le centre. Son rôle est de relayer l'information sensorielle provenant de la cochlée vers le cortex. Dans les voies efférentes, il reçoit de nombreuses projections provenant du cortex. Cette structure cérébrale occupe ainsi une position clé lorsqu'il s'agit d'étudier l'interaction dynamique entre les voies primaires et non-primaires, entre les systèmes afférents et efférents. Selon le cadre théorique actuel développé ci-dessus, l'étude des *cABRs* peut nous aider à mieux comprendre les phénomènes de plasticité qui se produisent tout au long de la vie d'un individu, en fonction de ses expériences.

A travers une première étude, nous avons voulu déterminer si les expériences auditives répétées sur le long terme telles que la pratique de notre langue maternelle peuvent affecter de manière significative l'encodage sensoriel des sons au niveau sous-cortical. Les premières études sur l'influence de l'expertise linguistique sur les *cABRs* ont comparé les différences d'encodage de la fréquence fondamentale entre les locuteurs de langues tonales et non-tonales. Bien que ces études aient mis en évidence des différences d'expertise entre Chinois et Anglais dans l'encodage sous-cortical de la fréquence fondamentale, il semblerait que cette plasticité ne soit pas spécifique à un contexte langagier, probablement car l'encodage de la hauteur tonale fait appel à des processus généraux du traitement de la fréquence fondamentale, qui dépendent des propriétés de *phase-locking* des neurones sous-corticaux. De plus, la distinction de mots par variations de la hauteur tonale est un processus phonologique bien particulier, et spécifique aux langues tonales. A l'inverse, dans les langues non-tonales les mots se distinguent par leur combinaison de phonèmes, eux-mêmes se distinguant entre eux par les trajectoires formantiques pour les consonnes, et par le rapport entre formants pour les voyelles. Il serait donc intéressant de déterminer si des différences d'expertise linguistique sont aussi observables entre deux langues non-tonales utilisant des processus phonologiques plus proches comme le Français et l'Anglais. Dans notre première étude, nous avons ainsi comparé les *cABRs* de français et américains natifs en réponse à un son soit appartenant à leur inventaire phonémique (i.e. la syllabe française /ru/ pour les français et la syllabe anglaise /thae/ pour les américains), soit n'en faisant pas partie (i.e. la syllabe /thae/ pour les français et /ru/ pour les américains). En comparant les *cABRs* des deux populations pour chaque son, nous avons ainsi pu étudier l'effet de l'expertise linguistique sur l'encodage sous-cortical des sons.

Dans une deuxième étude, nous avons voulu savoir si ces différences d'expertise linguistique peuvent être réduites voire supprimées par la pratique musicale. En effet, un

ensemble important d'études menées principalement par l'équipe de Kraus a montré que l'expertise musicale induit des bénéfices dans l'encodage sous-cortical non seulement des sons musicaux mais aussi des sons de parole. Notre hypothèse était que la pratique musicale entraînerait des adaptations neuronales qui permettraient aux musiciens de traiter des sons non-natifs (i.e. la syllabe anglaise /thae/) avec autant d'acuité qu'un natif. Afin de tester cette hypothèse, nous avons comparé les *cABRs* évoqués la syllabe anglaise /thae/ pour un groupe de français natifs non-musiciens, de français natifs musiciens et d'américains natifs.

Si de telles adaptations neuronales sont observables sur le long terme, nous pouvons supposer qu'elles émergent à la suite d'une série d'adaptations rapides et évanescentes à plus court terme. Néanmoins, si ces processus d'adaptation rapide sont renforcés par répétition (e.g. par exposition répétées aux mêmes stimuli pendant des années), ces adaptations pourraient se cumuler et s'ancrer de manière permanente, même au niveau des régions les plus primaires. Les découvertes récentes ont déjà montré que des phénomènes d'adaptation rapide se produisent au niveau du cortex auditif primaire, et qu'ils sont dépendants de processus attentionnels (Fritz et al., 2007, 2005). Cependant, à ce jour très peu études ont exploré ces phénomènes au niveau sous-cortical. A travers notre troisième étude nous avons ainsi voulu explorer les réorganisations rapides du tronc cérébral en fonction de l'état attentionnel du sujet, c'est-à-dire si la notion d'*active sensing* se maintient au niveau sous-cortical. De plus, nous avons voulu étudier si les ces effets attentionnels repose sur une dynamique oscillatoires tels que ceux décrits par la *Dynamic of attending theory* (*DAT*).

Deuxième partie - MATERIELS et METHODES

Matériels et méthodes

Nous allons voir dans ce chapitre le matériel et la méthodologie nécessaires à l'enregistrement des réponses complexes du tronc cérébral (*cABRs*). Bien que les mesures électrophysiologiques (EEG, MEG) soient courantes au sein de notre équipe de recherche, l'enregistrement des *cABRs* était une technique nouvelle pour l'équipe. La méthodologie des *cABRs* est en effet assez particulière, autant du point de vue du design expérimental, que de l'enregistrement et de l'analyse des réponses neuronales. Une grande partie de ma thèse fut donc dédiée à la mise en place du matériel, à la vérification de son bon fonctionnement, et à l'implémentation de scripts en vue de réaliser toute une batterie d'analyses des réponses neuronales (FFT, corrélation stimulus-réponse, inter-trial phase-coherence, neural response consistency, etc…). Cette partie est organisée sous la forme d'un tutoriel, basé sur celui proposé par Skoe & Kraus (2009) et adapté en fonction de mon expérience avec l'enregistrement des *cABRs*, en rapportant les difficultés techniques que j'ai pu rencontrer et les solutions que nous y avons apportées. Ce tutoriel est destiné entre autres aux futurs étudiants qui prendront le relais sur ce projet *cABRs* afin de leur éviter les erreurs que j'ai pu faire, et de leur faciliter la maîtrise théorique et pratique des *cABRs*.

Dans un premier temps, nous verrons quels sont les facteurs qui vont conditionner le choix du stimulus. Puis, nous aborderons en détails les précautions à prendre dans le choix des paramètres de stimulation mais aussi d'enregistrement des *cABRs*. Nous nous pencherons sur les différents types d'analyses qu'il est possible de réaliser sur les *cABRs*. Enfin, nous verrons quelles sont les forces et les limites liées à la technique des *cABRs*.

3. Choix du stimulus

3.1. Sons artificiels ou naturels

Dans la littérature, l'utilisation des sons artificiels est privilégiée par rapport aux sons naturels. L'avantage des sons synthétiques ou artificiels est qu'il est possible de contrôler finement les différents paramètres du stimulus, ce qui est utile si l'on veut faire varier un seul paramètre

tout en gardant les autres paramètres identiques afin d'étudier l'effet de ce paramètre au niveau du traitement sous-cortical indépendamment de tout autre facteur confondant. Par exemple, la distinction des syllabes /ba/, /da/ et /ga/ repose principalement sur une différence de trajectoire du deuxième formant. En faisant varier ce paramètre tout en maintenant les autres constants (fréquence fondamentale, autres formants), Parbery-Clark et al. (2012) ont montré que les neurones sont capables d'encoder ces différences de trajectoire par des différences de latence des réponses neuronales. Ces déductions n'auraient pas été possibles si d'autres facteurs confondants étaient en jeu, comme la variation de la trajectoire d'autres fréquences.

Néanmoins, en utilisant des stimuli synthétiques nous nous éloignons nécessairement des conditions d'écoute écologiques. La richesse et la complexité de l'environnement sonore dans lequel nous évoluons ne peut se résumer à des variations linéaires de certains paramètres, indépendamment les uns des autres. Par exemple, les syllabes naturelles /ba/, /da/ et /ga/ ne se distinguent pas uniquement sur la trajectoire du deuxième formant mais sur plusieurs paramètres (Johnson et al., 2008). A l'inverse, les sons naturels sont inévitablement plus complexes, variables dans le temps et plus riches en informations acoustiques que les sons artificiels. Ils permettent donc de se rapprocher au mieux des conditions d'écoute naturelles, mais en contrepartie, il est impossible de contrôler finement tous les paramètres qui caractérisent ces sons.

Bien que la plupart des études à ce jour aient utilisé des stimuli synthétiques, l'utilisation des stimuli naturels est prometteuse puisque ces derniers sont représentés de manière optimale tout au long des différents relais du système auditif (Theunissen and Elie, 2014). Le choix entre un stimulus naturel ou artificiel déprendra avant tout de la question de recherche qui motive le chercheur. Si le but est d'étudier l'effet d'un ou plusieurs paramètres du stimulus ou de comparer différents stimuli qui ne diffèrent que sur une dimension acoustique, le choix va se porter sur les stimuli artificiels pour lesquels il sera possible de faire varier les paramètres indépendamment les uns des autres. Si le but est de mettre en évidence des effets de groupe (e.g. population saine vs. clinique ou Français vs. Anglais), ou des effets indépendants du stimulus (e.g. écoute passive vs. active), il sera alors possible d'utiliser des sons naturels puisque l'effet recherché dépendra du groupe ou de la condition expérimentale et non des caractéristiques acoustiques du stimulus.

Les questions qui ont motivé les études durant ma thèse concernaient avant tout des effets de groupe (effet de la langue maternelle entre Américains et Français, effet de la pratique musicale entre musiciens et non-musiciens), et des effets de condition d'écoute (passive vs. active), notre choix s'est donc porté vers des stimuli naturels.

3.2. Sons de parole et de musique

De par leur richesse et leur complexité, les stimuli verbaux et musicaux représentent un choix pertinent lorsque l'on veut mettre en évidence des différences au niveau du traitement fin des sons. Néanmoins, ces deux types de sons ne sont pas produits de la même façon ce qui leurs confère des caractéristiques spectro-temporales qui leurs sont propres.

Dans les sons verbaux, la fréquence fondamentale (F0) et ses harmoniques (H1, H2, H3, etc...) proviennent de la vibration des cordes vocales. Les formants (F1, F2, F3, etc...), ainsi que leurs trajectoires dans l'espace fréquentiel sont quant à eux déterminés par la configuration de l'appareil phonatoire. C'est la configuration spécifique des harmoniques et des formants qui donne le timbre unique de la voix et nous permet de reconnaître l'identité de notre interlocuteur. Dans les sons de parole, il existe des variations spectro-temporales très rapides, comme le temps de voisement d'une consonne ou la transition d'une consonne à la voyelle. Ces indices permettent de transmettre une quantité importante d'informations en peu de temps au locuteur. Par exemple, les trajectoires formantiques permettent entre autres de discriminer les consonnes, et la distinction des voyelles repose principalement sur le rapport entre le premier et le deuxième formant dans les langues occidentales. Dans les langues tonales comme le Chinois Mandarin, les variations dans le temps de la F0 servent d'indices lexicaux puisqu'elles permettent de faire la distinction entre des mots monosyllabiques (e.g. la syllabe /ma/ avec un contour haut-montant de la F0 signifie "chanvre" alors que /ma/ avec un contour bas-descendant de la F0 signifie "cheval").

La production des sons musicaux varie d'un instrument à l'autre : le frottement de l'archet contre les cordes du violon, le pincement des cordes de la guitare, ou le passage du flux d'air dans la clarinette permettent de jouer la même note de musique (i.e. la même hauteur tonale), mais nous serons capables de distinguer ces trois notes de par leurs timbres. C'est la distribution particulière des harmoniques qui leur confère un timbre différent. A la différence des sons de parole, les notes de musique contiennent majoritairement des caractéristiques acoustiques soutenues dans le temps.

Le choix entre un stimulus verbal ou musical dépendra aussi de la question de recherche. Dans nos études, nous souhaitions étudier le traitement fin des sons au niveau sous-cortical, et en particulier l'effet de la langue maternelle ou de la pratique musicale sur le traitement des sons de parole. C'est donc naturellement que notre choix s'est porté sur des stimuli verbaux. En revanche,

le choix du stimulus peut être restreint à quelques stimuli selon l'étude que l'on veut mener. A titre d'exemple, dans notre première étude nous voulions étudier l'effet de l'expertise linguistique en comparant deux langues non-tonales, l'Anglais et le Français, dont les répertoires phonémiques sont plutôt proches. Afin de potentialiser l'effet de la langue sur le traitement sous-cortical des sons, il fallait trouver des syllabes présentes dans une langue et absentes dans l'autre. Notre choix s'est donc restreint à la syllabe anglaise /thae/ et la syllabe française /ru/ pour lesquelles la consonne et la voyelle n'existent pas dans l'autre langue.

3.3. Composantes acoustiques transitoires et continues

Les composantes transitoires des stimuli vont générer des réponses neuronales également transitoires caractérisées par des variations spectro-temporelles rapides. Par exemple, le début et la fin d'un son vont générer des réponses de type *onset* et *offset*. De même, la transition entre une consonne et une voyelle va générer des réponses transitoires, souvent de courte durée et de grande amplitude. Les caractéristiques d'attaque et de descente des sons vont également jouer un rôle important sur la robustesse des réponses *onset* et *offset*. Les sons ayant un temps d'attaque court comme les consonnes occlusives voisées /d/, /b/, /g/ vont évoquer une réponse *onset* plus robuste que les consonnes fricative comme le /v/, ou approximantes telles que le /r/ (Skoe et Kraus, 2010). Dans le cas des sons de musique, les instruments à percussion, comme le piano, produisent des sons ayant des phases d'attaque rapides, ce qui génère des réponses *onset* plus amples que les instruments à cordes frottées tels que le violon (voir exemple **Figure 3 p8**). De même, les sons dont l'amplitude décroît rapidement seront plus à même de générer des réponses *offset* que ceux dont l'amplitude décroît lentement tels que les voyelles et les notes de musiques glissantes.

Les composantes continues telles que les voyelles, les notes de musique soutenues, ou les sons sinusoïdaux génèrent des réponses neuronales soutenues dans le temps telles que la *Frequency-Following Response* (*FFR*). Ces réponses, qui ont la particularité de suivre avec précision les variations de fréquence présentes dans le stimulus, sont générées par la décharge synchrone des neurones sous-corticaux en phase avec l'enveloppe temporelle présente dans le son. Par exemple, un son ayant une fréquence fondamentale de 100 Hz va générer l'activation synchrone des neurones sous-corticaux à la même fréquence. Cependant, les capacités de suivi de fréquence ou *phase-locking* des neurones sous-corticaux présentent une limite basse d'environ 80-100 Hz et haute d'environ 1.5-2 kHz (Aiken and Picton, 2008; Moushegian et al., 1973). Lors du choix du

stimulus, il faut donc prendre en compte ces limites de *phase-locking* selon les composantes spectro-temporelles du stimulus que l'on souhaite étudier.

Les composantes transitoires et continues des réponses neuronales ne sont pas affectées de la même manière selon les paramètres de stimulation (e.g. l'intensité sonore, Akhoun et al., 2008), les conditions d'écoutes (e.g. active vs. passive, Galbraith et al., 2003; Hackley, 1993; Hackley et al., 1990; Hoormann et al., 1994) ou les niveaux d'expertise (Parbery-Clark et al., 2012), ce qui suggère que ces deux types de réponses auraient des générateurs neuronaux différents. Ces différences sont à prendre en compte en fonction de la question de recherche.

De plus, la nature transitoire ou continue du stimulus va conditionner les types d'analyses qu'il sera possible de réaliser sur les réponses neuronales (voir **section 7** pour plus de détails), certaines étant plus adaptées à l'analyse des réponses *onset/offset*, alors que d'autres comme la corrélation stimulus-réponse sont plus adaptées à l'analyse de la *FFR*. D'ailleurs, dans le cas des stimuli présentant à la fois des composantes transitoires et continues, tels que les syllabes, il est possible réaliser des analyses séparées sur la composante transitoire et continue des réponses. Il sera alors utile de fenêtrer le signal en deux parties. La taille des fenêtres dépendra de la durée des composantes présentes dans le stimulus, à laquelle il faudra ajouter un délai lié au temps de propagation nerveuse. Par exemple, pour la syllabe /da/ d'une durée de 100 ms, avec une consonne de 20 ms et une voyelle de 80 ms, les analyses des réponses seront réalisées sur deux fenêtres temporelles, la première de 10 à 30 ms, et la deuxième de 30 à 90 ms, pour un temps de propagation neuronal de 10 ms.

4. Présentation du stimulus

4.1. Matériel de stimulation

En général, la présentation des stimuli et l'enregistrement des réponses neuronales se fait séparément à l'aide de deux systèmes indépendants. Or, les réponses du tronc cérébral reflètent les variations très rapides de l'activité synchrone d'une population de neurones ; aussi, des retards de quelques centaines de microsecondes peuvent être révélateurs de troubles fonctionnels des structures sous-corticales (Skoe and Kraus, 2010). De ce fait, il est nécessaire que les systèmes de stimulation et d'enregistrement soient précisément synchronisés. En effet, les réponses du tronc cérébral sont de très faibles amplitudes (de l'ordre de la centaine de nanovolts), des décalages

infimes d'essais en essais dus à une mauvaise synchronisation peuvent totalement annihiler les réponses neuronales lors du moyennage. Si les triggers correspondants à la présentation du stimulus sont générés avec une variabilité de plus d'une milliseconde par rapport à la présentation réelle du stimulus, la réponse sera au mieux fortement réduite et distordue, sinon absente. Lors de la mise en place d'un nouveau système d'enregistrement des *cABRs*, il est donc primordial de vérifier qu'il soit bien synchronisé avec le système de stimulation.

Pour les études que nous avons réalisé durant cette thèse, nous avons utilisé un système qui gère à la fois la stimulation et l'enregistrement des réponses neuronales (µV-*ABR*, Microvitae Technologies). La synchronisation est directement intégrée dans le système en hardware. Cependant, comme il s'agissait des premiers enregistrements *cABRs* pour notre équipe, nous avons tout de même voulu vérifier qu'il n'existait pas de jitters trop importants à travers les essais. Pour cela, nous avons placé un micro au niveau de la sortie audio du système de stimulation afin d'enregistrer les stimuli effectivement envoyés par le système, et nous avons relié la sortie du micro à un câble d'enregistrement que nous avons branché sur le système d'acquisition. De cette façon, le micro remplace l'électrode EEG. En générant une série de clics, nous avons pu vérifier la variabilité du signal au fur et à mesure des essais. Dans ce montage, le décalage inter-essais mesure le jitter entre stimulation acoustique et enregistrement EEG. Nous avons pu observer que ce décalage ne variait pas plus d'un dixième de milliseconde entre les essais (6000 essais enregistrés), ce qui a confirmé la fiabilité du système pour nos futurs enregistrements.

Un autre avantage de ce système est sa portabilité. Il est tient en effet dans une mallette portative (c.f. photo ci-après), ce qui nous a permis d'utiliser le même système pour l'enregistrement des sujets Français et Américains, et de s'affranchir ainsi de tout éventuel biais lié au matériel.



Figure 18. Photo du matériel portatif de stimulation et d'enregistrement des *cABR* (µV-*ABR*, Microvitae Technologies). Photo issue du site web www.microvitae.com.

4.2. Intensité

La latence et l'amplitude des réponses du tronc cérébral sont dépendantes de l'intensité sonore : plus l'intensité est élevée, plus les réponses neuronales sont amples et précoces et vice-versa (Akhoun et al., 2008; Krishnan, 2002). Cependant, les différents types de réponses ne sont pas affectés de la même manière. En utilisant la syllabe /ba/, Akhoun et al. (2008) ont par exemple montré qu'en augmentant l'intensité sonore (de 0 à 60 dB par pas de 10 dB), la latence des réponses *onset* diminuait plus lentement que celle des réponses *FFR*, ce qui pourrait suggérer que ces deux types de réponses reflètent des processus neuronaux distincts. Plus spécifiquement, Krishnan et al. (2002) ont étudié l'influence de l'intensité sonore (de 55 à 85 dB par pas de 10 dB) sur la représentation des différentes composantes spectrales contenues dans la *FFR* pour les voyelles anglaises /u/ et /)/. Bien que l'amplitude des harmoniques tende à augmenter en fonction du niveau d'intensité, la courbe amplitude-intensité n'était pas identique pour toutes les harmoniques.

Ces études mettent en évidence l'importance d'avoir un matériel de stimulation bien calibré avant toute nouvelle étude. Pour nos études, nous avons mesuré le niveau d'intensité sonore directement à la sortie des inserts en utilisant un sonomètre (Iso-Tech, SLM 52N) relié à un coupler 2cc. En effet, la sortie audio étant trop petite pour pouvoir réaliser la mesure directement avec le micro du sonomètre, il convient d'utiliser un coupler 2cc qui permet non seulement de relier le sonomètre à l'insert avec étanchéité, mais aussi de simuler la cavité de l'oreille, rendant ainsi la mesure la plus proche des conditions d'écoute réelles.

En réalisant ces mesures de contrôle, nous avons pu nous rendre compte que l'intensité sonore variait en fonction des caractéristiques acoustiques présentes dans le stimulus, mais aussi en fonction de la cadence de stimulation. Plus le temps d'attaque des sons est long, plus l'intensité sonore mesurée était faible, et vice-versa ; aussi, plus la cadence de stimulation était élevée, plus le niveau mesuré était constant dans le temps, probablement lié à un effet plateau du niveau sonore à haute cadence. Ce dernier effet est directement lié à la manière dont le sonomètre calcule l'intensité. La mesure peut être « instantanée », ce qui se traduit par des variations d'intensité plus importantes, ou se faire en calculant une moyenne pendant une durée donnée. Sachant que nous avons utilisé des cadences de stimulation relativement faible (~3-4 Hz), nous avons préféré réaliser une mesure moyenne de l'intensité sonore délivrée par le système sur une durée de 5 secondes. Cette mesure a été répétée plusieurs fois, à plusieurs jours d'intervalles, afin de s'assurer de la stabilité du système de stimulation.

4.3. Cadence de stimulation

La cadence de stimulation n'est ni plus ni moins que l'inverse du temps séparant la présentation de deux stimuli. Il y a trois points importants à retenir lorsque l'on veut définir la cadence de stimulation. Le premier est que les *cABRs* sont sensibles à la cadence de stimulation. Krizman et al. (2010) ont enregistré les réponses neuronales évoquées par des clics et par la syllabe /da/présentés à trois cadences de stimulation (6.9, 10.9 et 15.4 Hz). Les résultats ont révélé que la latence des *click-ABRs* n'était pas affectée par la cadence de stimulation, alors que celle des réponses *onset* (III, V et A) évoquées par la syllabe /da/ augmentait en fonction de la cadence de stimulation. En revanche, la latence des réponses *FFR* restait constante quelque soit la cadence utilisée. D'autres effets ont été observés au niveau de l'amplitude des représentations neuronales. Bien que pour les basses fréquences (<410 Hz) aucun effet de la cadence de stimulation n'ait été observé, pour les hautes fréquences (>410 Hz) il existait une relation inverse

entre l'amplitude et la cadence de stimulation : plus la cadence de stimulation était élevée, plus l'amplitude des composantes hautes fréquences diminuait.

Le deuxième point important à prendre en compte est que la cadence choisie doit permettre d'inclure une ligne de base d'au moins 30 ms avant la présentation du stimulus, et un retour de la réponse à la ligne de base en laissant au moins 30 ms de silence après la fin du stimulus. Il sera ainsi possible de calculer le rapport signal sur bruit.

Enfin, pour éviter la contamination du signal neuronal par les fréquences du secteur (50 Hz), la cadence de stimulation choisie ne doit pas être un multiple entier de la fréquence du secteur. Par exemple, une cadence de stimulation à 10 Hz n'est pas recommandée car la stimulation sera synchronisée avec la phase du 50 Hz, ce qui augmente la contamination liée au secteur, alors qu'à 10.3 ou 11 Hz, la présentation n'est plus en phase avec celle du 50 Hz, ce qui permet lors du moyennage des essais de réduire l'influence du 50 Hz sur le signal neuronal.

4.4. Polarité du stimulus

Une onde sonore est la propagation de proche en proche des variations de pression à travers les molécules d'air. Les ondes acoustiques correspondent donc à une alternance de compressions (i.e. de condensations) et de décompressions (i.e. de raréfactions) des molécules d'air. On peut représenter ces variations de pression dans le temps comme une alternance d'ondes positives et négatives. Les clics qui consistent en une onde positive ou négative, sont définis comme étant une condensation ou une raréfaction. En revanche, comme les stimuli périodiques oscillent entre des états de raréfaction et de condensation, nous ne pouvons pas utiliser la même terminologie. La polarité d'un stimulus sera d'abord définie de manière arbitraire. Par exemple, un stimulus A correspondant à la polarité positive et un stimulus B à la polarité négative (voir **Figure 19**) . Pour passer d'une polarité à une autre il suffit de retourner l'onde à 180°, ce qui revient en pratique à multiplier le stimulus par -1.

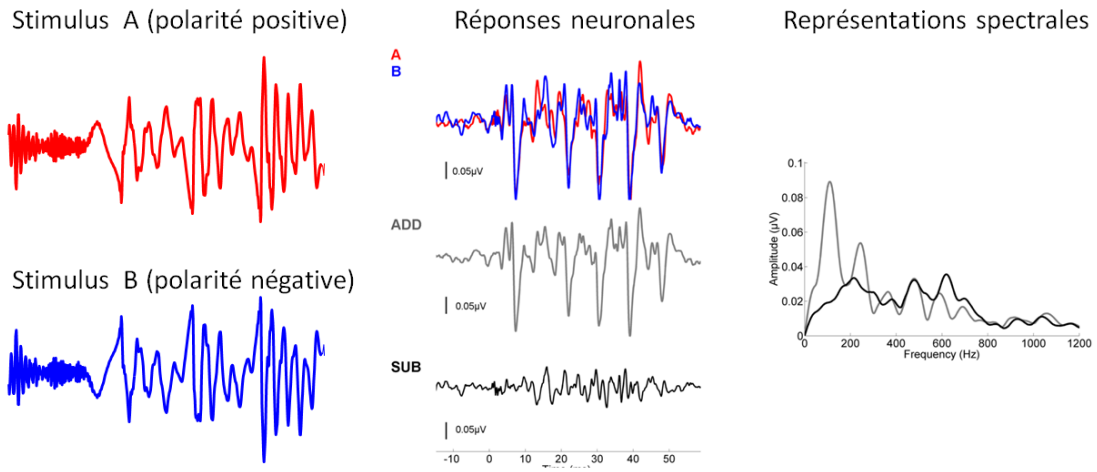


Figure 19. Polarité du stimulus et son effet sur les réponses neuronales. Le stimulus /da/ (panel gauche) peut être présenté avec une polarité positive (en rouge) ou négative (en bleu). Les réponses neuronales à ces deux polarités sont très similaires au niveau des basses fréquences (panel central), mais présentent des différences de phase au niveau des hautes fréquences. Le moyennage par addition (add) amplifie la représentation de l'enveloppe (basses fréquences) qui est indépendante de la phase du stimulus, alors que le moyennage par soustraction (sub) amplifie la représentation des hautes fréquences (harmoniques et formants) au détriment des basses fréquences. Issue et adaptée de Skoe et Kraus (2010).

Bien que les deux polarités d'un même stimulus ne soient pas perçues différemment, les réponses neuronales qu'elles engendrent présentent de légères différences, en particulier lorsque l'onde sonore n'est pas symétrique (Greenberg, 1980). Les différences entre les réponses neuronales sont dues principalement à des décalages de phase au niveau des hautes fréquences. En effet, l'encodage des hautes fréquences (harmoniques, formants) est étroitement dépendant de la phase. Du fait que les oscillations des neurones sous-corticaux se calent sur la phase de ces fréquences, nous observerons des décalages de phase d'une demi-période entre les réponses neuronales aux deux polarités. En superposant les deux réponses (**Figure 19**, panneau central), nous pouvons voir entre les larges déflections négatives que certains pics sont en opposition de phase. A l'inverse, l'encodage des basses fréquences comme l'enveloppe temporelle du stimulus n'est pas dépendant de la phase car l'enveloppe temporelle n'est pas réellement présente dans le stimulus, mais est introduite au sein des voies auditives via des processus de traitement cochléaire non-linéaires (Kumar et al., 2014). De ce fait, la représentation neuronale de l'enveloppe varie très peu en fonction de la polarité, ce que l'on voit bien au niveau des larges déflections négatives des réponses A et B.

L'addition des réponses aux deux polarités (i.e. (A+B)/2) maximise l'enveloppe et minimise les hautes fréquences ainsi que l'artéfact lié au stimulus, puisque ces derniers sont dépendants de la phase. A l'inverse, la soustraction des polarités (i.e. (A-B)/2) maximise la

représentation des hautes fréquences et l'artéfact lié au stimulus au détriment des basses fréquences (**Figure 19**, panneau central et droit). Néanmoins, même si en théorie l'effet de la polarité est présent pour tous les stimuli, il a surtout été étudié pour des stimuli simples tels que les clics, et il est moins évident pour des stimuli plus complexes tels que les tones-burst (Don et al., 1996; voir aussi Starr et al., 1996).

4.5. Stimulation monaurale et binaurale

Dans des conditions d'écoute naturelles, nous percevons généralement les sons avec nos deux oreilles. L'écoute binaurale permet d'une part de localiser les sons dans l'espace, mais aussi d'extraire la voix d'un locuteur du bruit de fond. Il existe aussi un effet de sommation de l'intensité sonore lié à l'écoute binaurale. C'est-à-dire qu'un son présenté avec la même intensité sonore aux deux oreilles sera perçu plus fort (environ +6dB) que s'il est présenté à une seule oreille. De ce fait, au niveau des réponses sous-corticales, cet effet de sommation se traduit par des réponses plus amples et plus robustes comparées aux réponses recueillies en condition d'écoute monaurale. Chez l'adulte normo-entendant, il est en général préférable d'utiliser une stimulation binaurale. En revanche, si le but est d'étudier l'influence de processus *top-down* ayant une latéralisation hémisphérique, comme les processus langagiers ou musicaux, il sera recommandé de stimuler l'oreille controlatérale à la latéralisation hémisphérique du processus en question. Par exemple, dans nos études nous avons étudié entre autres comment l'encodage des sons de parole est modulé au niveau sous-cortical par des processus *top-down*. Etant donné le gradient de latéralisation du langage au niveau de l'hémisphère gauche, nous avons choisi de présenter les stimuli au niveau de l'oreille droite des sujets, afin de maximiser l'influence potentielle des processus *top-down* liés au traitement du langage.

Cependant, chez les sujets présentant une asymétrie des seuils auditifs (e.g. patients sourds appareillés ou non) il conviendra mieux d'utiliser une stimulation monaurale.

La stimulation de l'oreille droite et gauche engendre des réponses très similaires mais pas strictement identiques (Akhoun et al., 2008). Il existe par exemple un avantage de l'oreille droite dans le traitement des sons de parole, caractérisé par des latences plus précoces et une représentation des harmoniques et des formants plus robuste par rapport à l'oreille gauche (Hornickel et al., 2008).

4.6. Transducteurs

Le temps entre la présentation du stimulus et le début de la réponse sous-corticale est très court, en moyenne entre 5 et 12 ms selon les paramètres de stimulation (stimulus, cadence de stimulation, intensité, ...). Or, la plupart des stimuli utilisés pour les *cABRs* durent au moins quelques dizaines, sinon quelques centaines, de millisecondes. Il existe donc un recouvrement temporel important entre l'artéfact électromagnétique généré par la présentation du stimulus à travers les écouteurs et les réponses neuronales. De plus, il existe une grande similarité entre l'artéfact du stimulus et les *cABRs*. Il sera donc difficile de distinguer l'artéfact de la réponse neuronale si l'enregistrement est contaminé. Afin de limiter l'artéfact lié au stimulus, en plus des précautions à prendre décrites précédemment (i.e. l'alternance des polarités et moyennage par addition), il est important de placer les stimulateurs acoustiques, sources d'artéfacts, le plus loin possible des électrodes d'enregistrement. Il est nécessaire d'utiliser dans ce cas des écouteurs de type inserts reliés au stimulateur acoustique via des tubes en plastique, tels que les inserts ER-3A (Etymotic Research) que nous avons utilisé pour toutes nos études.



Figure 20. Photo des écouteurs de type inserts ER-3A. Issue du site www.unicog.org.

L'utilisation de tels écouteurs ne garantit pas forcément un enregistrement totalement dépourvu d'artéfact, mais il réduit néanmoins grandement les chances de contamination du signal (Akhoun et al., 2008). Aussi, il existe plusieurs manières de vérifier la présence d'un artéfact a posteriori (voir **section 6** pour plus de détails sur le rejet des artéfacts).

Il faut noter que les tubes en plastique des inserts vont d'une part, générer une légère distorsion du signal acoustique (voir Akhoun et al., 2008), et d'autre part induire un délai supplémentaire dans les réponses neuronales lié au temps de propagation du son dans les tubes (estimé à environ 1 ms pour les inserts ER-3A). Ce délai sera à prendre en compte dans les analyses, s'il n'est pas déjà corrigé en amont au niveau du système de stimulation ou d'enregistrement.

5. Enregistrements électrophysiologiques

5.1. Montage

Les réponses complexes du tronc cérébral (*cABRs*) peuvent être enregistrées avec différents montages d'électrodes : vertical (vertex référencé aux mastoïdes, lobes d'oreilles, ou au niveau de la nuque) ou horizontal (mastoïde ou lobe d'oreille référencé à la mastoïde ou lobe d'oreille controlatéral). Aussi, les électrodes de références peuvent être placées à différents endroits (Fz, lobes d'oreilles, nuque, mastoïdes), ce qui peut affecter la forme des réponses enregistrées.

Picton et al. (1974) ont montré qu'un enregistrement vertical permet d'enregistrer une onde V plus distincte qu'un enregistrement horizontal, tandis que l'onde III apparaît aussi distinctement pour les deux types de montage. L'effet du montage a aussi été testé sur l'enregistrement de la *FFR*. En utilisant un montage vertical (vertex-lobes d'oreilles) et horizontal (oreille droite-gauche), Stillman et collaborateurs ont identifié deux types de réponses *FFRs* : FFP1 et FFP2 (P pour potentiel). Bien que la FFP1 soit bien représentée par les deux types de montages, la FFP2 apparaît plus nettement en utilisant un montage vertical (Stillman et al., 1978). Suite à ces résultats, les auteurs ont proposé que la *FFR* enregistrée au niveau du scalp est probablement générée par plusieurs sources. Dans une étude plus récente, en enregistrant les *FFRs* évoquées par un son dépourvu de sa fondamentale, Galbraith (1994) a montré que la fondamentale absente apparaît uniquement en utilisant un montage vertical, alors qu'elle reste absente dans les enregistrements réalisés avec un montage horizontal. Ces découvertes suggèrent que la fréquence fondamentale est recréée avant d'atteindre le cortex auditif, non pas au niveau des structures situées à l'extrémité caudale du tronc cérébral (noyaux cochléaires), mais plutôt au niveau des structures rostrales. L'ensemble de ces études montrent que les montages horizontaux reflèteraient l'activité des structures plus caudales du tronc cérébral, telles que le noyau cochléaire, alors que les montages verticaux reflèteraient plutôt l'activité des structures rostrales du tronc cérébral, probablement le lemnisque latéral ou le colliculus inférieur (Chandrasekaran and Kraus, 2010).

Le choix du référencement est aussi un facteur important à prendre en considération lors de l'enregistrement des réponses du tronc cérébral. L'effet du positionnement de l'électrode de référence sur les ondes précoces (ondes I à V) a été particulièrement bien détaillé dans l'étude de Terkildsen et Osterhammel (1981). Dans cette étude, l'activité électrophysiologique a été

enregistrée de manière simultanée au niveau du vertex, des lobes d'oreilles, et de la nuque (côté droit et gauche), puis les réponses ont été calculées selon plusieurs différentiels d'activité avec des montages verticaux (vertex - lobes d'oreilles reliés, vertex - lobes d'oreille contro- ou ipsilatéral, vertex - nuque contro- ou ipsilatéral) et horizontaux (oreille ipsi - oreille controlatérale, nuque ipsi - controlatérale, oreille contro ou ipsilatéral - nuque). Avec les montages verticaux, il apparaît que les réponses neuronales sont plus amples lorsque la référence est placée du côté ipsilatéral à la stimulation de l'oreille plutôt que du côté controlatéral, alors que le phénomène inverse est observé dans les montages horizontaux. De même, les montages verticaux avec références reliées (lobes d'oreille ou nuque droite et gauche reliés) donnent des réponses neuronales plus faibles en amplitude comparées aux montages verticaux utilisant une seule électrode de référence du côté ipsilatéral à la stimulation. En revanche, lorsque l'électrode est placée sur la nuque les auteurs ne rapportent aucun effet de la latéralité.

Enfin, lors des tests pilotes que nous avons menés au début de ma thèse, nous avons enregistré des *FFRs* avec un montage vertical et une électrode de référence placée au niveau de la mastoïde ou au niveau du lobe d'oreille du côté ipsilatéral. Les enregistrements ont révélé chez certains sujets un artéfact en début de réponse (voir **Figure 21**) lorsque l'électrode était placée au niveau de la mastoïde. En montrant ces enregistrements à l'équipe de Kraus, il s'est avéré que cet artéfact reflétait probablement l'activation réflexe du muscle stapédien. Le réflexe stapédien intervient pour réduire la tension s'exerçant sur la fenêtre ovale lorsque l'intensité sonore est trop élevée. Bien que le niveau sonore ait été calibré à 80 dB SPL, certains sujets ont néanmoins montré cette activité artéfactuelle. Cet artéfact apparaît avec une latence précoce et une amplitude supérieure à celle de la *FFR*. Selon l'équipe de Kraus, l'électrode placée sur le lobe de l'oreille permet non seulement d'éviter l'artéfact lié au réflexe stapédien, mais aussi de limiter les artéfacts liés à la vibration des os (Skoe and Kraus, 2010).

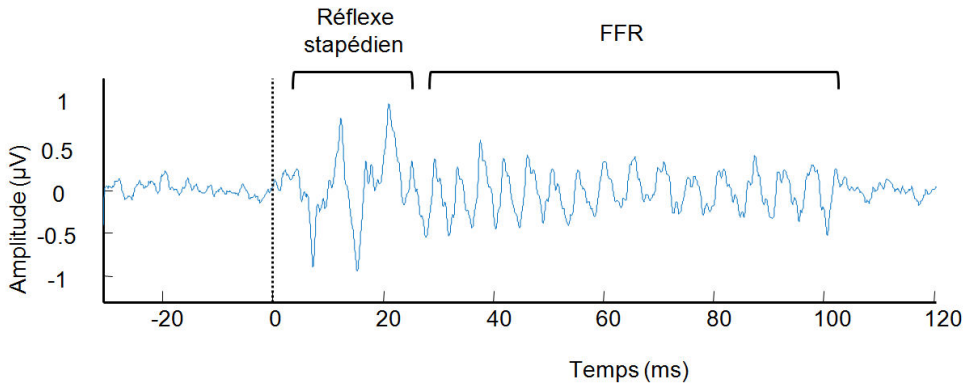


Figure 21. Exemple de *FFR* contaminée par l'activité musculaire liée au réflexe stapédien. Figure issue des données recueillies durant ma thèse.

En résumé, si le but est d'enregistrer les réponses neuronales précoces (ondes I à III) des structures caudales du tronc cérébral (nerf auditif, noyau cochléaire), un montage horizontal avec une référence placée du côté controlatéral à la stimulation sera le plus optimal. A l'inverse, pour enregistrer les réponses neuronales plus tardives (onde V, *FFR*, *cABRs*) des structures rostrales du tronc cérébral (lemnisque latéral, colliculus inférieur), il vaudra mieux utiliser un montage vertical avec une référence placée du côté ipsilatéral à la stimulation.

5.2. Fréquence d'échantillonnage

La fréquence d'échantillonnage (Fs) d'un signal est le nombre de mesures ou échantillons du signal recueillis par unité de temps. En théorie, il faut une Fs deux fois supérieure à la plus haute fréquence présente dans le signal (fréquence de Nyquist). En pratique, il vaut mieux suréchantillonner le signal à une Fs au moins 4 fois supérieure. Une Fs élevée permet de limiter le nombre d'erreur d'échantillonnage, et augmente la résolution temporelle du signal enregistré. Sachant que des différences de l'ordre du dixième de milliseconde dans les réponses *cABRs* peuvent être représentatives de troubles fonctionnels ou de phénomènes de plasticité, il est important d'avoir une haute précision temporelle. En contrepartie, plus la Fs est haute plus, par définition, le nombre d'échantillons est grand, donc plus les fichiers de données seront volumineux. La fréquence d'échantillonnage peut aussi être limitée par le système d'acquisition. Pour toutes nos expériences, une Fs de 30 kHz s'est avérée largement suffisante pour enregistrer toutes les composantes fréquentielles d'intérêt tout en limitant la taille des fichiers de données.

5.3. Nombre d'essais

A l'inverse des réponses corticales qui montrent un phénomène d'habituation important en fonction du nombre d'essais, les réponses sous-corticales sont très robustes. De plus, du fait de leur faible amplitude, il est nécessaire de recueillir un grand nombre d'essais pour pouvoir extraire les réponses sous-corticales du bruit de fond. Pour collecter les *click-ABRs* et les *FFRs* évoquées par tones-burst, 1000 à 2000 essais sont suffisants pour obtenir des réponses robustes (Skoe and Kraus, 2010). En revanche, pour les *cABRs* il est préférable de recueillir plus d'essais, environ 2000 à 3000 par polarité, soit 4000 à 6000 au total.

Le rapport signal sur bruit (SNR) est proportionnel à la racine carrée du nombre d'essais. De ce fait, dans un premier temps le SNR augmente rapidement au début puis atteint progressivement un plateau en fonction du nombre d'essais. Néanmoins, il semblerait que la courbe du SNR en fonction du nombre d'essais ne soit pas la même pour toutes les composantes du signal. Plus spécifiquement, si les analyses sont réalisées uniquement dans le plan spectral, des réponses significatives (i.e. au dessus du bruit de fond) peuvent être détectées avec un nombre d'essais plus faible (Aiken and Picton, 2008, 2006; Dajani et al., 2005).

La meilleure approche est d'enregistrer plus d'essais que nécessaire, puisque d'une part certains essais seront contaminés par des artéfacts musculaires (voir **section 6**) et donc rejetés, et d'autre part un nombre d'essais important augmente la fiabilité et la robustesse des analyses. Il sera par exemple possible d'étudier l'évolution du signal en fonction du temps d'expérimentation en comparant des sous-moyennages. Il est aussi plus fiable de tester la robustesse du signal (voir **section 7**), avec un nombre d'essais important. Pour nos étude, nous avons choisi de recueillir 6000 essais bruts (3000 par polarité) pour chaque stimulus utilisé.

Les principaux facteurs limitant le nombre d'essais sont le temps d'expérimentation et la taille des fichiers de données. Un temps d'expérimentation trop long peut pour certains sujets favoriser la tension musculaire, et ainsi dégrader la qualité du signal; et générer des fichiers trop volumineux pour pouvoir être stockés et/ou analysés.

5.4. Fenêtre d'enregistrement

La fenêtre temporelle d'enregistrement comprend généralement une ligne de base (période de silence pré-stimulus), la réponse neuronale et une période de silence post-stimulus. La ligne de base est nécessaire pour pouvoir calculer le rapport signal sur bruit (SNR). Le SNR sera calculé comme le rapport entre l'activité pré- et l'activité post-stimulus. Il faut au minimum une ligne de base de 30 ms, sinon plus pour pouvoir calculer avec fiabilité le SNR. La période de silence qui suit la réponse neuronale est nécessaire pour que l'activité neuronale redescende au niveau de la ligne de base, ce qui permet de s'assurer que la fenêtre temporelle suivante ne soit pas contaminée par la réponse au stimulus qui précède. La taille de la fenêtre d'enregistrement dépend de la cadence de stimulation, de la durée du stimulus, de la durée de la ligne de base et de la durée de silence post-réponse.

5.5. Conditions d'enregistrement passives et actives

Les réponses sous-corticales peuvent être enregistrées quelque soit l'état d'éveil du sujet : endormi ou éveillé, en conditions d'écoute passive ou active. L'effet du niveau attentionnel sur les *ABRs* a été un long débat (Rinne et al., 2008). Les premières études ont montré que les *clicks-ABRs* ne sont pas affectés par l'attention (Hackley, 1993; Hackley et al., 1990). Plus récemment, des études utilisant des stimuli plus complexes ont monté que l'amplitude et la latence de la *FFR* est modulée par l'attention, avec des latences plus courtes (Hoormann et al., 2000, 1994) et une amplitude plus importante (Galbraith et al., 2003, 1998, 1995, Hoormann et al., 2000, 1994) lorsque le sujet porte attention aux stimuli. A l'inverse, lorsque l'attention du sujet est détournée du stimulus, en engageant le sujet dans une autre tâche, l'amplitude de la *FFR* est réduite. Ces résultats concordent avec un étude d'imagerie ayant montré une influence de l'attention sélective sur le niveau d'activation du colliculus inférieur (Rinne et al., 2008). L'ensemble de ces travaux suggère que le niveau attentionnel aurait un effet sur les réponses neuronales aux sons complexes, il est donc important de réaliser les enregistrements dans des conditions d'écoute constantes, à moins que l'effet de l'attention ne soit l'objet de l'étude (Musacchia et al., 2007).

Dans la plupart des études, les réponses sont collectées dans des conditions d'écoute passives, ce qui permet au sujet de se détendre, et de limiter les artéfacts musculaires. Il est possible de maintenir le sujet éveillé, tout en détourant son attention des stimuli en diffusant un film sous-titré. Certains laboratoires diffusent même le son du film en champ libre à un niveau faible (~40 dB SPL), lorsque la stimulation est monaurale, de manière à ce que le son soit perçu par l'oreille non stimulée, tout en évitant de masquer la stimulation de l'autre oreille.

Contrairement à ce qu'on pourrait penser, l'enregistrement des *cABRs* dans des conditions d'écoute passives permet d'étudier la nature dynamique du traitement auditif. En effet, les *cABRs* reflètent la manière dont le système auditif traite l'information acoustique en fonction des expériences passées. Nous verrons dans la discussion de cette thèse que ces réponses peuvent illustrer la manière dont les expériences à plus court terme telles que les effets attentionnels, se cumulent dans le temps jusqu'à s'ancrer de manière profonde au sein du traitement automatique des sons au niveau des structures sous-corticales.

6. Détection et réduction des artéfacts

De par leur faible amplitude et leur origine sous-corticale, les *cABR* sont des réponses électrophysiologiques très sensibles à l'activité électromagnétique environnante, que cette dernière soit d'origine physiologique (activité musculaire, corticale et microphone cochléaire) ou non-physiologique (artéfact du stimulus et 50 Hz en particulier), ce qui nécessite certaines précautions à prendre lors de leurs enregistrements.

6.1. Courant de secteur (50 ou 60 Hz)

C'est l'artéfact le plus fréquemment retrouvé dans les enregistrements électrophysiologiques, que ce soit au niveau des potentiels évoqués corticaux ou sous-corticaux. Ils ont pour origine les prises de courant et les appareils électroniques environnants. Il est donc conseillé de limiter au maximum l'utilisation de tels appareils dans la chambre d'enregistrement et d'éloigner ces sources de contamination le plus possible des électrodes d'enregistrement. L'idéal, sinon l'indispensable, est de réaliser les enregistrements dans un chambre isolée au niveau acoustique et électromagnétique (chambre Faraday anéchoïque), ce qui limite grandement les artéfacts liés au secteur.

Il est possible de vérifier la présence de courant de secteur (50 Hz) en moyennant les FFT calculées essai par essai. Sur le spectrogramme moyen, si le signal est contaminé, il existera en général un pic important à 50 Hz souvent associé à d'autres pics au niveau des harmoniques du 50 Hz (voir un exemple de signal contaminé **Figure 22**).

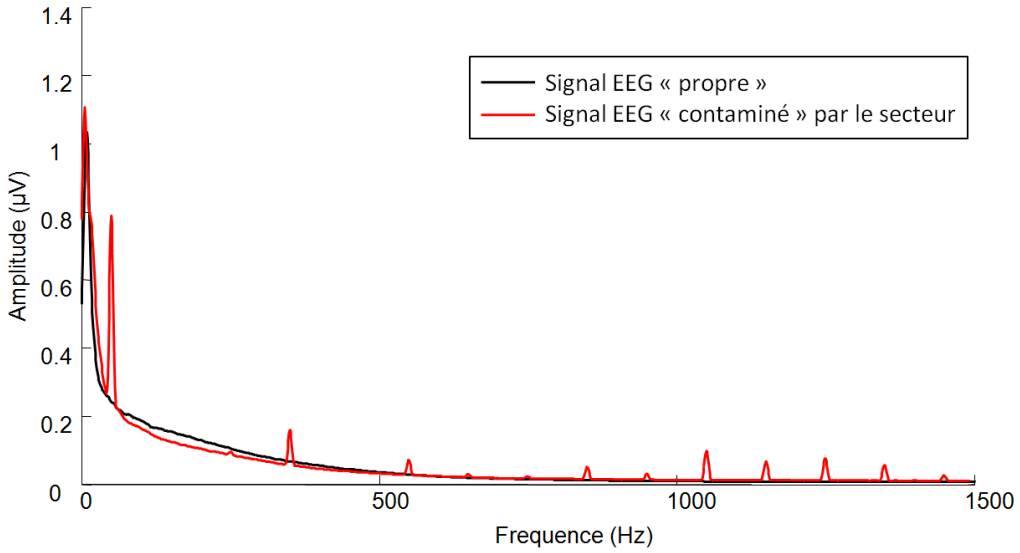


Figure 22. Exemple de signal EEG propre (tracé noir) et contaminé par le secteur 50 Hz (tracé rouge). Chacun des tracés a été obtenue en moyennant les FFT calculées essai par essai sur le signal brut d'un sujet. Un pic important apparaît à 50 Hz au niveau du signal bruité avec d'autres pics au niveau des harmoniques. Issue des données recueillies durant la thèse.

6.2. Artéfact du stimulus

L'avantage mais aussi l'inconvénient des *cABRs* est qu'ils reproduisent de manière fidèle les variations temporelles et fréquentielles du stimulus. Il est donc important de s'assurer que le stimulateur acoustique ne produise pas d'artéfact, également très similaire au stimulus, au niveau de l'enregistrement.

Il existe plusieurs manière de vérifier la présence d'un artéfact de stimulation. D'une part, il n'existe quasiment aucun délai entre la présentation du stimulus et l'artéfact, contrairement aux réponses neuronales qui apparaissent avec un délai d'environ 5-12 ms (Chandrasekaran and Kraus, 2010). En calculant la cross-corrélation stimulus-réponse (voir **section 8** pour plus de détails), si l'indice de corrélation est maximal pour un décalage temporel proche de 0 ms, alors il y a de grandes chances pour que la réponse enregistrée soit contaminée. D'autre part, l'artéfact est dépendant de la phase et de l'amplitude du stimulus, alors que la réponse, même si elle reflète avec fidélité les variations acoustiques du stimulus, présente des variations notables de la phase et de l'amplitude qui montrent qu'il ne s'agit pas d'une copie parfaite du stimulus (Worden and Marsh, 1968). En superposant les réponses enregistrées pour chaque polarité, si l'on observe une

parfaite opposition de phase entre les deux réponses, alors le signal est probablement contaminé (voir **Figure 23** pour un exemple de signal contaminé).

Afin de réduire cet artéfact, il faut vérifier que le stimulateur soit situé le plus loin possible des électrodes d'enregistrement. Aussi, le moyennage par addition des deux polarités permet de réduire l'artéfact. Néanmoins, si l'artéfact est trop important la réponse neuronale sera quand même contaminée, c'est pourquoi il est important d'avoir des conditions d'enregistrement les plus optimales.

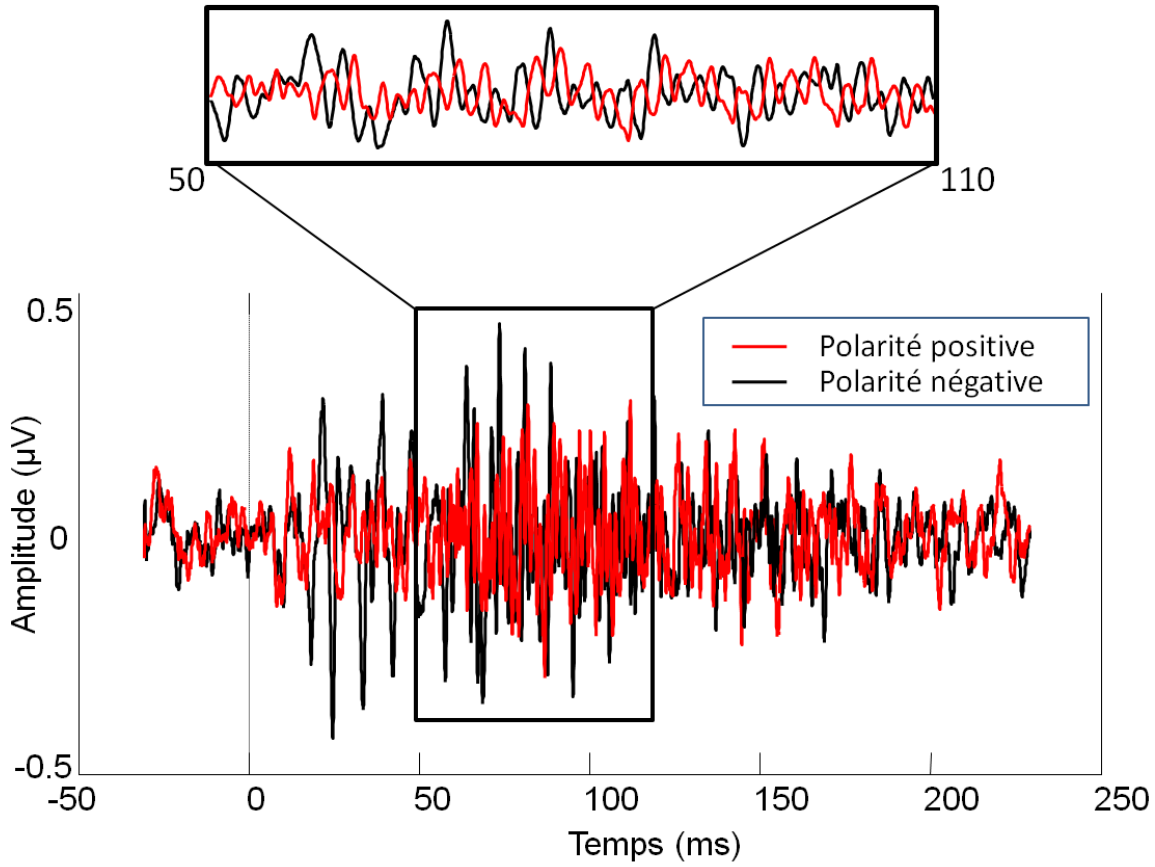


Figure 23. Exemple de réponse neuronale contaminée par l'artéfact du stimulus. Les tracés rouges représentent les réponses obtenues avec la polarité positive, et les tracés noirs représentent celles obtenues avec la polarité négative. L'encadré met en évidence l'opposition de phase des deux tracés qui suggère fortement la présence d'un artéfact de stimulation. Issue des données recueillies durant la thèse.

6.3. Microphone cochléaire

Les cellules ciliées de la cochlée génèrent une activité électrique appelée microphone cochléaire (MC) qui peut être enregistrée par les électrodes de scalp. Tout comme la *FFR*, le MC reproduit avec fidélité le signal acoustique entrant, de ce fait ces deux réponses peuvent être confondues. Néanmoins, il existe plusieurs manières de distinguer ces deux réponses. D'une part, le MC est quasiment synchronisé sur la présentation du stimulus, alors que la *FFR* apparaît avec une latence de 5-12 ms. En calculant l'indice de cross-corrélation entre stimulus et réponse, si le niveau de corrélation est maximal pour un décalage temporel entre les deux signaux très proche de zéro, alors le signal enregistré est soit contaminé par l'artéfact du stimulus (voir ci-dessus), soit par le MC. D'autre part, le MC et la *FFR* sont affectés différemment par la cadence de stimulation, l'intensité sonore et le bruit. En étudiant l'évolution de la réponse en fonction des variations de ces différents paramètres de stimulation, il est possible de déterminer si la réponse est contaminée ou non par le MC. Enfin, le MC est dépendant de la phase du stimulus, alors que la relation entre la phase du stimulus et celle de la *FFR* est beaucoup plus variable, en particulier dans les basses fréquences. Il est donc possible de minimiser l'influence du MC sur la réponse neuronale en alternant la polarité du stimulus, et calculant la moyenne par addition des polarités (voir **section 4**).

6.4. Artéfacts musculaires

Les potentiels évoqués sous-corticaux sont de très faibles amplitudes (de l'ordre du nanovolt), ils sont donc très sensibles à l'activité musculaire dont l'amplitude est beaucoup plus importante. Une manière de réduire les artéfacts musculaires est de rejeter les essais pour lesquels l'activité excède un certain seuil (voir **section 7** pour plus de détails).

A la différence du signal EEG classique, les réponses *cABRs* ne sont pas contaminées par les mouvements oculaires car ils génèrent une activité dans des gammes de fréquences bien plus basses que celles des réponses *cABRs*. L'extraction des réponses sous-corticales par filtrage (voir **section 7**) va automatiquement éliminer ces artéfacts musculaires.

Même si cette méthode permet de rejeter les artéfacts musculaires les plus importants, cela ne signifie pas pour autant que le signal sera complètement dépourvu de toute activité musculaire. De ce fait, il est important de maintenir le sujet dans un état de relaxation tout au long de la session d'enregistrement.

7. Traitement des données

Durant ma première année de thèse, une grande partie de mon temps fut dédiée à l'implémentation des scripts de traitement des données électrophysiologiques. En effet, le traitement du signal *cABRs* est notablement différent de celui du signal EEG classique. J'ai donc développé sous Matlab toute une pipeline de prétraitements, de traitements et de post-traitements qui vont être détaillés ci-dessous.

7.1. Données brutes

Avant tout prétraitement, il est important d'observer les données brutes afin d'apprécier la qualité du signal ou d'identifier la présence éventuelle d'artéfacts. Dans notre pipeline, les données brutes sont présentées sous deux formes : la première est un *rasterplot* qui est une représentation de l'amplitude du signal au cours du temps, essai par essai (**Figure 24**); la deuxième est un spectre moyen calculé à partir des FFTs essai par essai (voir un exemple **Figure 22**).

Le raster plot permet de se rendre compte de la qualité globale du signal au cours de la session d'enregistrement. Dans l'exemple ci-dessous, on peut observer que la qualité du signal s'est dégradée vers la fin de l'enregistrement (à partir de l'essai 1600), probablement en raison de mouvements du sujet.

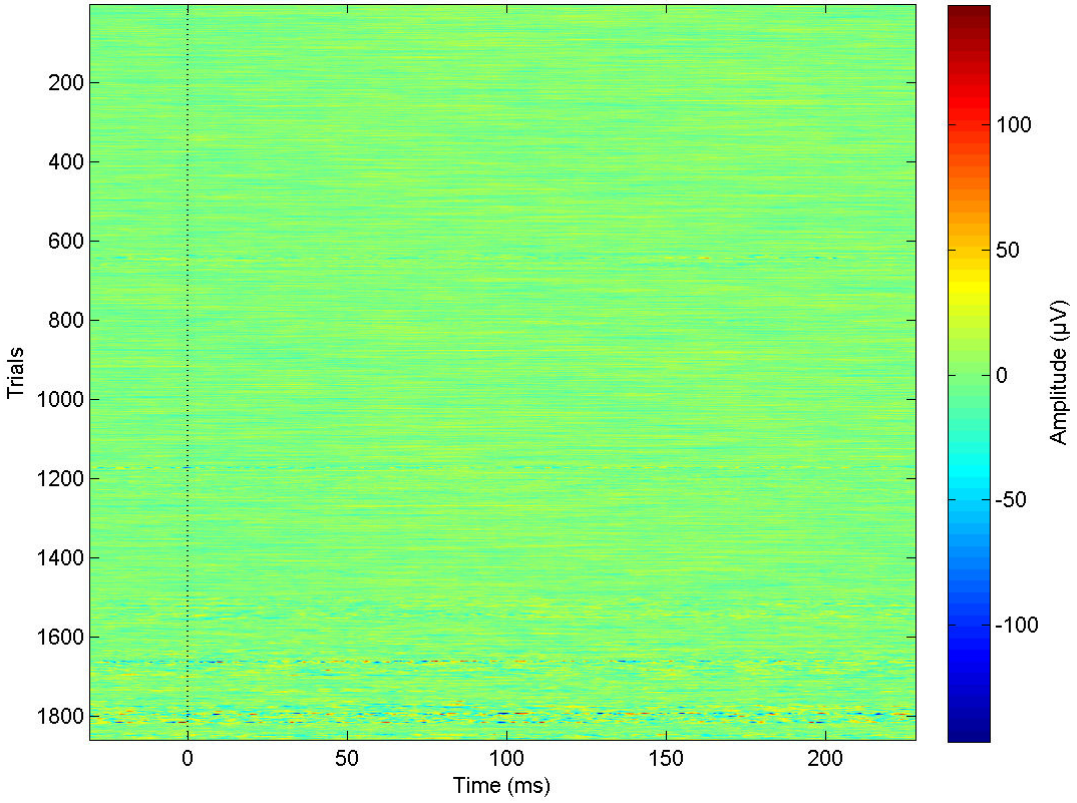


Figure 24. *Rasterplot* d'une session d'enregistrement de *cABRs* chez un sujet. L'axe horizontal est le temps, l'axe vertical sont les numéros des essais en fonction de leurs occurrences, et le code couleur correspond à l'amplitude du signal en microvolts. Le signal apparaît plus bruité vers la fin de la session d'enregistrement (à partir de l'essai 1600). Issue des données recueillies durant la thèse.

Il est souvent difficile de discerner la présence de 50 Hz sur le signal temporel moyenné, d'autant plus que la fenêtre d'analyse est souvent courte pour les *cABR*s. L'analyse spectrale essai par essai puis moyennée permet de faire ressortir la présence éventuelle de 50 Hz (voir **Figure 22**).

7.2. Rejet des essais bruités

La première étape de prétraitement consiste à rejeter les essais trop bruités, c'est-à-dire ceux dont l'amplitude absolue maximum dépasse un certain seuil. Dans la littérature, ce seuil varie de +/- 20 à +/- 75 µV (Akhoun et al., 2008; Galbraith and Doan, 1995). Pour nos études, nous avons utilisé un seuil de +/- 30 µV pour lequel moins de 10% d'essais étaient rejetés tout en conservant un bon rapport signal sur bruit pour un maximum de sujets.

Selon le système de stimulation et d'enregistrement, il est possible de rejeter les essais en ligne. Cette option est pratique puisqu'elle permet d'observer la qualité de la réponse neuronale dépourvue des artéfacts musculaires grâce au moyennage en temps réel, mais aussi d'observer le pourcentage d'essais rejetés, ce qui donne une autre appréciation de la qualité du signal. En revanche, même si cette option est activée, il est préférable d'enregistrer tous les essais afin de pouvoir utiliser d'autres valeurs de seuil de rejet *a posteriori*. L'idéal est d'obtenir un taux de rejet inférieur à 10%, ce qui garanti un signal peu bruité.

7.3. Filtrage

Le filtrage du signal EEG permet d'extraire les réponses du tronc cérébral de l'activité corticale. Le tronc cérébral répond dans des gammes de fréquences allant d'environ 100 Hz à 1500 Hz (Aiken and Picton, 2008; Moushegian et al., 1973), alors que le cortex répond dans des gammes de fréquences <100 Hz, bien que des travaux récents aient montré des activités de type *FFR* au niveau cortical pour des fréquences fondamentales allant jusqu'à 200 Hz (Bellier et al., en préparation). En général, le signal sera filtré entre 100 et 1500 à 3000 Hz. Dans certains cas la limite basse du filtre peut être étendue, par exemple lorsque la F0 ou les produits de distorsion sont inférieurs à 100 Hz. Dans nos études, nous avons utilisé un filtre passe bande de 70-2000 Hz d'ordre 2 car certains stimuli avaient une F0 légèrement inférieure à 100 Hz.

Il est important de noter que certains filtres peuvent induire une distorsion au niveau de la phase des fréquences, en particulier lorsque la fréquence de coupure est proche des gammes de fréquences de la réponse neuronale. Or, pour certaines analyses telles que la cohérence de phase inter-essais, une distorsion du signal dans l'espace des phases peut générer des résultats incohérents, voire difficilement interprétables. Pour éviter cela, il faut filtrer le signal dans les deux sens (du début vers la fin et vice-versa), de manière à ce que les effets de distorsions se compensent. Le filtre Butterworth implémenté dans Matlab inclut cette fonction de manière automatique (fonction "butter.m").

7.4. Moyennage

Une fois le signal "nettoyé", il est possible de moyenner les essais selon plusieurs méthodes : par polarité, et par addition ou soustraction des polarités.

Les avantages et les inconvénients du moyennage par addition et soustraction sont décrits plus haut (**section 4**). En résumé, le moyennage par addition amplifie les basses fréquences (F0, enveloppe) au détriment des hautes fréquences (harmonique et formants), et inversement pour le moyennage par soustraction. De plus, le moyennage par addition limite l'influence des artéfacts liés au microphone cochléaire et au stimulus, alors que le moyennage par soustraction va au contraire les amplifier.

Il faut noter que les analyses plus avancées sont plus ou moins appropriées selon le type de moyennage. Par exemple, l'indice corrélation stimulus-réponse dépend étroitement des basses fréquences (enveloppe temporelle et F0), le moyennage par addition sera donc plus adapté, alors que pour analyser les représentations spectrales des formants, le moyennage par soustraction sera privilégié (Intartaglia and Schön, en révision; Krishnan, 1999).

Il est intéressant aussi d'étudier le signal moyenné par polarité car cette méthode de moyennage conserve à la fois les hautes et les basses fréquences. En outre, en superposant les réponses aux deux polarités, il est possible d'identifier la présence d'artéfacts liés au stimulus, ce qui se traduira par une opposition de phase quasi-parfaite entre les deux signaux (voir exemple de signal artéfacté **Figure 23**).

7.5. Rapport signal/bruit (*SNR*)

Le calcul du rapport signal sur bruit (*Signal-to-Noise Ratio, SNR*) est une mesure permettant d'apprécier la qualité du signal "nettoyé" et moyenné. Le *SNR* s'obtient sur le signal moyenné par addition des polarités en calculant le rapport entre la racine carrée de la moyenne de l'amplitude au carré du signal (*Root-Mean-Square, RMS*) et celle de l'amplitude au carré de la ligne de base (*SNR* = *RMS* signal/*RMS* ligne de base). Si le *SNR* est inférieur à 1, cela signifie que l'activité pré-stimulus est plus importante que l'activité de la réponse. D'après Skoe et Kraus (2010), si le *SNR* est inférieur à 1.5, il vaut mieux rejeter les données du sujet.

8. Analyses des données

Il existe un multitude d'analyses possibles sur le signal *cABRs*. Le choix des analyses dépendra à la fois du stimulus utilisé et de la question de recherche. Cette section décrit les différents types d'analyses des *cABRs* sur le plan temporel et spectral. La liste n'est pas exhaustive, mais elle regroupe les analyses les plus couramment utilisées dans la littérature sur les *cABRs*. Certaines analyses sont plus adaptées aux réponses de type transitoire et d'autres aux réponses continues, c'est pourquoi nous avons distingué deux groupes d'analyses en fonction de la nature transitoire ou continue de la réponse étudiée.

8.1. Réponses transitoires (*onset, offset,...*)

8.1.1. Latence et amplitude des pics

La latence d'un pic correspond au délai entre la présentation du stimulus et l'amplitude absolue maximum du pic. La mesure de la latence peut être appliquée aux *click-ABRs*, aux réponses *onset*, *offset* et transitoires des *cABRs*, et dans certains cas aux réponses continues (*FFR*) comme le montre l'exemple qui suit. La présence des réponses de type transitoire n'est pas certaine, elle dépend avant tout des phases d'attaque et de descente du stimulus. Dans le cas des réponses transitoires, il est possible de réaliser des mesures de la latence, de l'amplitude, de l'aire sous la courbe et de pente.

Lorsqu'un grand nombre de sujets ont été enregistrés avec le même stimulus, il est possible de définir des intervalles de confiance qui serviront à estimer la distribution des mesures. L'équipe de Kraus a par ailleurs estimé des valeurs normatives pour la syllabe /da/ de 40 ms pour laquelle les réponses neuronales sont robustes à travers les sujets (**Figure 25**). Les pics I, III, V et A correspondent aux réponses *onset*, le pic C à la transition entre consonne et voyelle, les pics D, E, F à la réponse soutenue type *FFR*, et le pic O à la réponse de type *offset*. Les latences de ces pics ainsi que la pente entre le pic V et A sont identifiées pour chaque sujet et comparées aux valeurs normatives. L'intérêt de développer des valeurs normatives pour les *cABRs* est de pouvoir identifier à un niveau plus fin les troubles ou les avantages de certains sujets au niveau de

l'encodage neuronal des différentes composantes transitoires et continues présentes dans le stimulus.

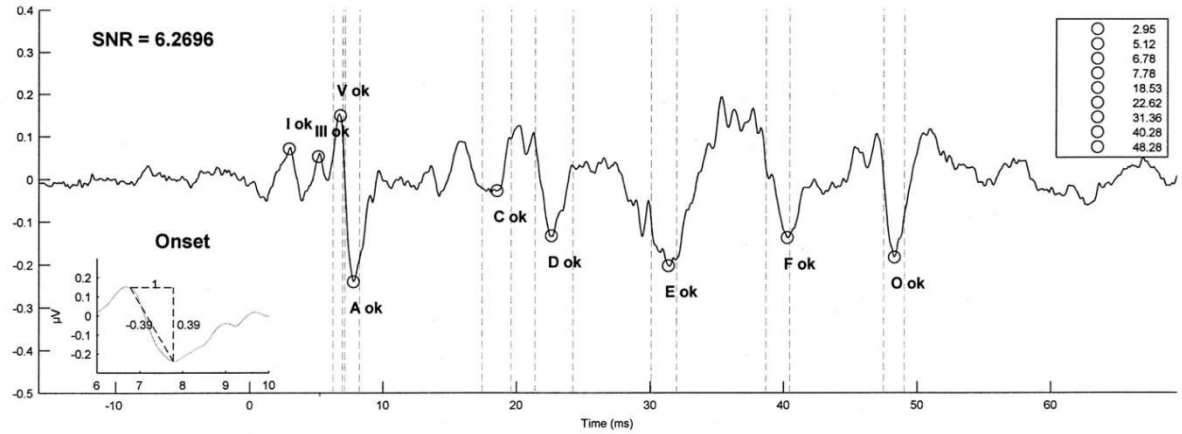


Figure 25. Exemple de *cABR* évoquée par la syllabe /da/ présentée au niveau de l'oreille droite et les marqueurs biologiques associés. Les pics I, III, V et A correspondent aux réponses *onset*, le pic C à la transition entre consonne et voyelle, les pics D, E, F à la réponse soutenue type *FFR*, et le pic O à la réponse de type *offset*. Les lignes verticales en pointillées représentent les intervalles de confiance calculés sur un grand nombre de sujet. L'encadré correspond aux latences des différents pics (en ms). La pente entre les pics V et A est représentée en bas à gauche du tracé. *SNR = Signal to Noise Ratio.* Issue des mesures et analyses réalisées pendant une visite à l'Auditory Neuroscience Lab de l'équipe de Kraus.

L'amplitude des pics peut aussi être mesurée. Pour qu'un pic soit considéré comme significatif, il faut que son amplitude absolue soit supérieure à l'amplitude moyenne de la ligne de base. Bien que l'amplitude des pics puisse présenter un intérêt, nous avons constaté une grande variabilité inter-sujets. Pour cette raison, nous avons préféré l'exclure de nos analyses.

8.1.2. Cross-phaseogramme

Une technique d'analyse plus avancée a été mise au point par Skoe et al. (2011) afin d'étudier les différences de latences entre plusieurs conditions/stimuli en fonction du temps et de la fréquence. Cette analyse nommée cross-phaseogramme mesure la différence de phase entre deux conditions/stimuli au cours du temps et pour chaque fréquence. Par exemple, dans une étude, ils ont montré que les différences de trajectoires formantiques entre les syllabes /ba/, /da/ et /ga/ sont représentées au niveau des réponses sous-corticales par des différences de phase avec une latence plus précoce pour le /ga/ suivit du /da/ et du /ba/ (**Figure 26**).

En pratique, il s'agit de calculer sur le signal moyenné une transformée de Fourier avec fenêtre glissante. A chaque pas de temps et pour chaque fréquence, les valeurs de phase sont extraites pour chaque stimulus/condition. Les différences de phase entre stimuli/conditions sont ensuite calculées pour chaque point temps-fréquence. Le résultat est représenté sous la forme d'une image avec la fréquence en ordonnée, le temps en abscisse et un code couleur pour le décalage de phase (en radian).

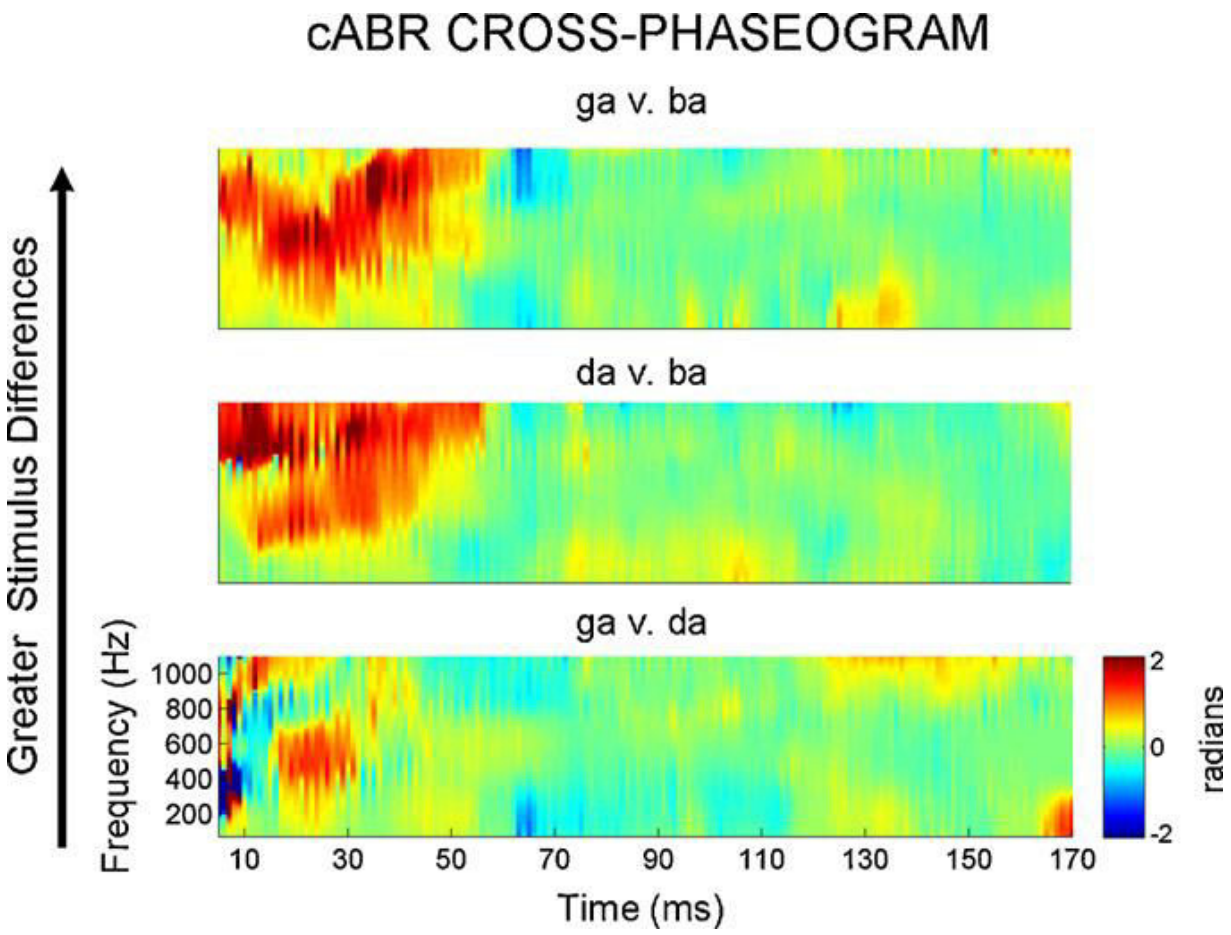


Figure 26. Exemple de cross-phaseogramme réalisé entre les *cABRs* évoquées par les syllabes /ba/, /da/ et /ga/. Les fréquences sont en ordonnée (en Hz), le temps en abscisse (en ms) et la couleur code pour le décalage de phase (en radian). Les couleurs chaudes signifient que le premier stimulus devance le deuxième et inversement pour les couleurs froides (e.g. sur le panel du haut, /ga/ devance /ba/ durant les 50 premières millisecondes). Issues de Skoe et al. (2011).

8.2. Réponses continues (*FFR*)

8.2.1. Corrélation stimulus-réponse

La similarité entre stimulus et réponse neuronale permet notamment de calculer le coefficient de corrélation entre ces deux signaux. La corrélation stimulus-réponse va fournir un indice objectif de la précision temporelle avec laquelle est encodé le stimulus au niveau du tronc cérébral. Plus le coefficient est élevé, plus la représentation sous-corticale est fidèle au son. Le coefficient de corrélation est calculé en fonction du décalage temporel entre les deux signaux (**Figure 27**). Le niveau de corrélation stimulus-réponse atteint un maximum au niveau du décalage temporel correspondant au délai entre l'*onset* du stimulus et celui de la réponse, ce qui donne une indication du délai de conduction neuronal.

En théorie, le coefficient de corrélation atteint un maximum au décalage temporel correspondant au délai de conduction neuronal entre la cochlée et les noyaux du tronc cérébral (lemnisque latéral, colliculus inférieur), soit environ 5-12 ms. En pratique, il existe souvent plusieurs pics, correspondant à des multiples de la période de la fréquence fondamentale. En effet, puisque l'enveloppe temporelle de ces deux signaux est dominée par la même fréquence fondamentale (F0), si l'on décale progressivement les deux signaux l'un par rapport à l'autre, les oscillations de la F0 des deux signaux vont successivement se retrouver en phase et en opposition de phase en fonction du décalage temporel. De plus, l'indice de corrélation maximum n'est pas forcément positif puisque l'encodage de l'enveloppe temporelle est indépendant de la phase du stimulus. Pour déterminer le bon délai, il faut donc dans un premier temps inspecter visuellement la fonction de corrélation de plusieurs sujets, ce qui donne une idée du délai de réponse moyen pour un stimulus donné. Dans un deuxième temps, il faut définir une fenêtre de décalage temporelle dans laquelle on ira rechercher le coefficient de corrélation absolu maximum pour chaque sujet.

La corrélation stimulus-réponse peut également être calculée sur une fenêtre glissante (d'une largeur de 20 à 40 ms), ce qui permet d'apprécier la similarité entre les deux signaux en fonction du temps. En général, le coefficient de corrélation sera plus important sur la partie stationnaire du signal, telle que la voyelle, pour laquelle le suivi de fréquence des réponses neuronales est plus robuste.

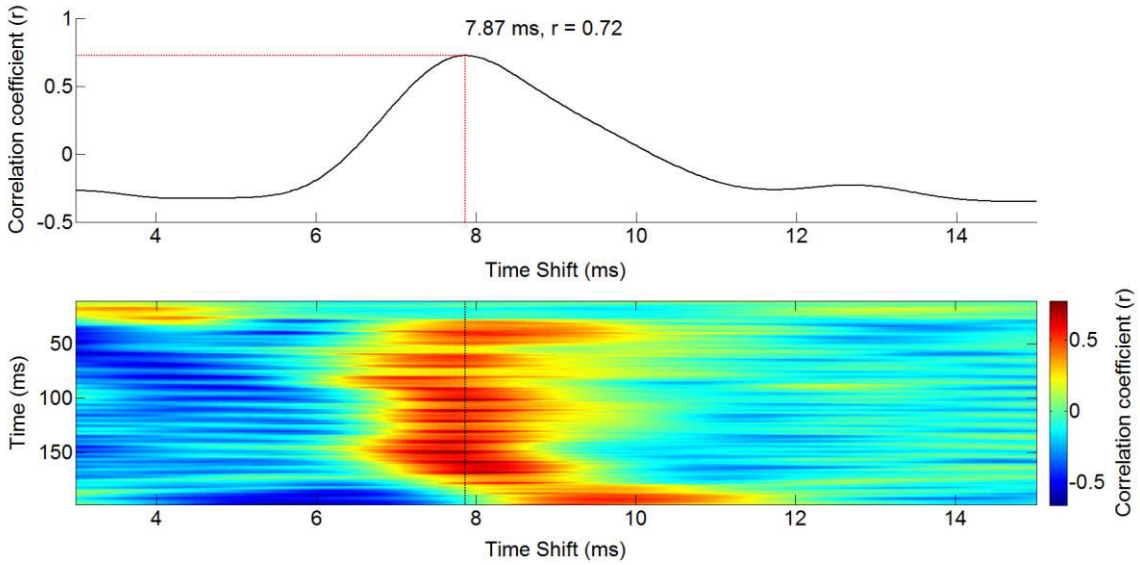


Figure 27. Corrélation stimulus-réponse. **Haut** : Cette figure représente le niveau de corrélation entre le stimulus et la réponse en fonction du décalage temporel entre les deux signaux. La corrélation maximale est atteinte à 7.87 ms de décalage temporel, qui est une indication du délai de conduction neuronal. **Bas** : Analyse en fenêtre glissante de 20 ms de la corrélation stimulus-réponse. Cette approche permet d'apprécier la similarité entre les deux signaux en fonction du temps. Dans cet exemple, les deux signaux sont plus similaires durant la partie stationnaire du signal (i.e. la voyelle /u/). Ces figures ont été obtenues d'après les données issues de Intartaglia et al. (2016).

8.2.2. Autocorrélation

L'autocorrélation permet de trouver de manière objective des patterns temporels répétitifs au sein du signal. L'autocorrélation consiste à calculer la corrélation du signal avec lui-même. En pratique, l'indice de corrélation est calculé en faisant une copie de la réponse neuronale et en la décalant progressivement dans le temps (**Figure 28**), ce qui donne une fonction d'autocorrélation ou autocorrélogramme qui est une représentation du coefficient de corrélation en fonction du décalage temporel.

Dans la *FFR*, le pattern temporel dominant est celui de l'enveloppe temporelle ou de la fréquence fondamentale (F0). De ce fait, l'indice de corrélation sera maximal pour un délai correspondant à la période de la fréquence fondamentale (période = 1/fréquence). La valeur du coefficient de corrélation maximal sera représentatif de la robustesse de l'encodage de la F0. De

même, une fonction d'autocorrélation étalée reflètera un encodage de la F0 moins robuste qu'une fonction plus pointue (Krishnan et al., 2005; Lee et al., 2009).

Il est aussi possible d'appliquer la technique d'autocorrélation sur une fenêtre glissante, ce qui permet d'apprécier la robustesse du suivi de fréquence de la réponse neuronale au cours du temps. Le résultat est représenté sous la forme d'une image à trois dimensions avec le décalage temporel en ordonnée, le temps en abscisse et un code couleur pour la valeur du coefficient de corrélation.

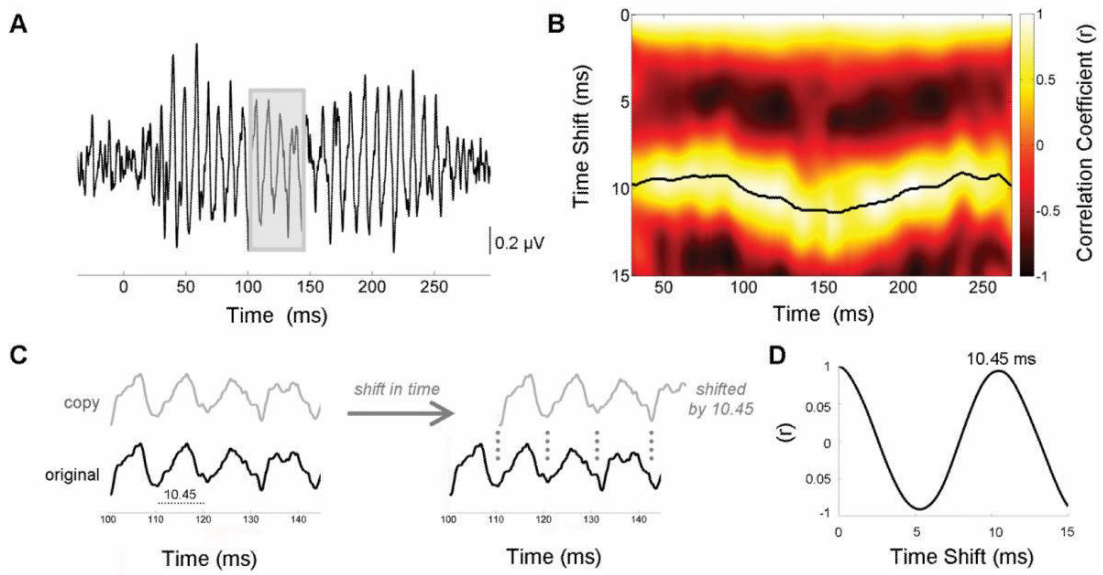


Figure 28. Illustration de la technique d'autocorrélation. **A**. Cette figure illustre la *cABR* évoquée par la syllabe tonale /mi/ ayant un contour de F0 variable au cours du temps. L'encadré gris transparent représente la fenêtre glissante de 40 ms utilisée pour calculer l'indice d'autocorrélation en fonction du temps. **B.** Représentation de la fonction d'autocorrélation appliquée sur une fenêtre glissante. La couleur représente la valeur du coefficient de corrélation. Le trait noir représente la coefficient de corrélation maximal au cours du temps qui correspond à la période fondamentale . **C.** Illustration du calcul de la fonction d'autocorrélation. La réponse neuronale et sa copie sont décalées progressivement, et à chaque pas de temps l'indice de corrélation entre les deux signaux est calculé. **D.** Fonction d'autocorrélation. Dans cet exemple, le coefficient de corrélation atteint une valeur maximale pour un délai de 10.45 ms correspondant à la période de la fréquence fondamentale (1/10.4 ms = 96 Hz). Issue de Skoe et Kraus (2010).

La fonction d'autocorrélation avec fenêtre glissante permet aussi d'estimer la qualité et la robustesse du suivi de la F0 et de l'enveloppe temporelle ou *frequency tracking*, ce qui peut être intéressant pour des stimuli dont la fréquence fondamentale varie au cours du temps tels que les syllabes tonales chinoises (voir exemple **Figure 29**). Dans un premier temps, une fonction d'autocorrélation avec fenêtre glissante est obtenue sur le signal neuronal et sur le stimulus. La trajectoire de la F0 de chaque signal est estimée à partir de la valeur de coefficient de corrélation

maximale pour chaque point-temps. Enfin, la robustesse du suivi de fréquence est estimée en calculant un indice de déviation des deux trajectoires au cours du temps, et la puissance du suivi de fréquence en calculant la moyenne de l'indice d'autocorrélation maximum de la réponse neuronale au cours du temps (i.e. en moyennant les valeurs de corrélation de la trajectoire de F0 du signal neuronal).

L'équipe de Kraus a par ailleurs développé une toolbox Matlab qui permet de réaliser facilement ces analyses d'autocorrélation et qui contient des explications plus avancées sur cette méthode d'analyse. Cette toolbox est téléchargeable gratuitement en remplissant un formulaire de demande sur leur site web : http://www.brainvolts.northwestern.edu/form/freeware.php.

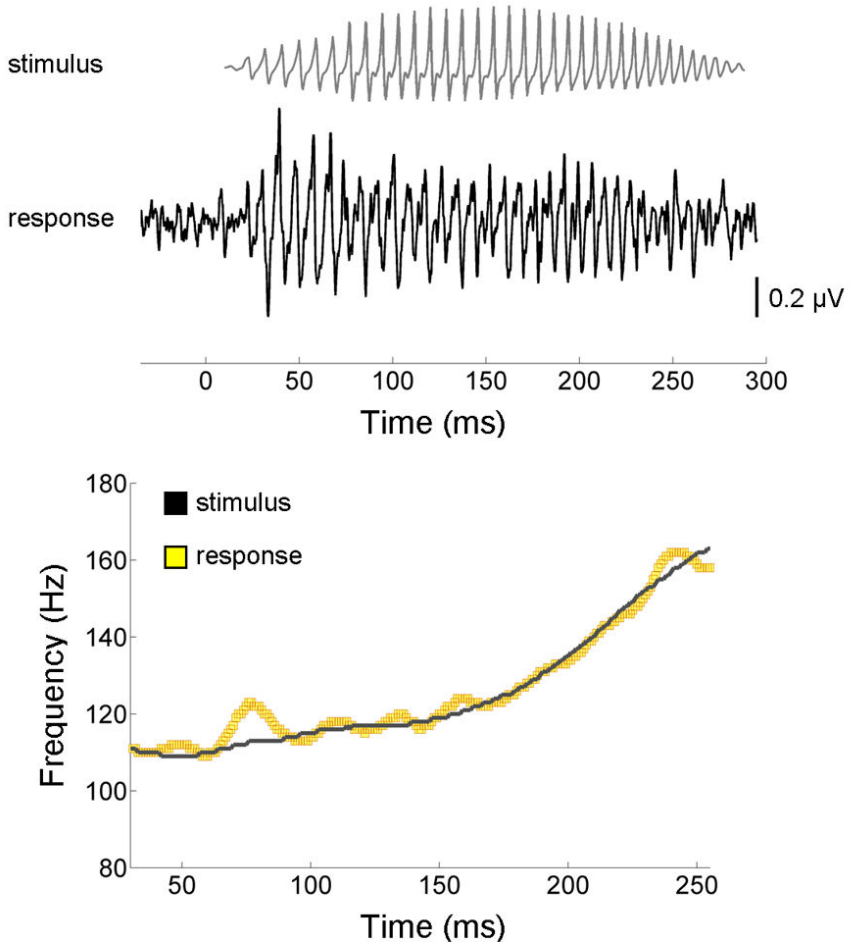


Figure 29. Illustration du *frequency-tracking* par la méthode d'autocorrélation par fenêtre glissante. **Haut :** Le stimulus utilisé est une syllabe tonale chinoise ayant une hauteur tonale montante. La réponse neuronale évoquée par ce stimulus suit les variations de la fréquence fondamentale (F0) avec fidélité. **Bas :** La précision du suivi de fréquence des *cABRs* est illustrée sur ce graphique. Chaque temps-point correspond à la fréquence estimée en calculant l'inverse du décalage temporel pour lequel la fonction d'autocorrélation atteint son maximum. La trajectoire de la F0 résultante présente dans le stimulus (en noir) et dans la réponse neuronale (en jaune) sont ensuite superposées. Issue de Skoe et Kraus (2010).

8.2.3. Transformée de Fourier (FFT)

L'analyse la plus simple dans le plan spectral est la transformée de Fourier. Cette technique est utilisée pour estimer la précision et l'amplitude du verrouillage de phase des réponses neuronales au niveau de certaines fréquences ou bandes de fréquences.

Tout signal temporel peut être décomposé en une somme de sinusoïdes. La transformée de Fourier permet de décomposer le signal en un ensemble d'ondes sinusoïdales. L'amplitude de chaque onde représente la quantité d'énergie contenue dans le signal d'origine à la fréquence de cette sinusoïde. La composition spectrale d'un signal peut être représentée en traçant l'amplitude en fonction de la fréquence des ondes sinusoïdales (**Figure 30**).

La résolution fréquentielle de la FFT dépend de la taille de la fenêtre d'analyse temporelle. Plus la fenêtre est grande, plus la résolution fréquentielle est fine (résolution = 1/T, où T = durée en secondes de la fenêtre). Il existe des phénomène de "fuites" ou *leakage* liés à la résolution fréquentielle. Par exemple, pour un signal de 50 ms, la résolution fréquentielle est de 1/0.05 soit 20 Hz. De ce fait, la transformée de Fourier contiendra des informations seulement au niveau des multiples entiers de 20 Hz (i.e. 0, 20, 40, 60 Hz...). Si le signal contient une fréquence à 130 Hz, qui n'est pas un multiple entier de 20 Hz, cette fréquence va "fuir" vers les fréquences voisines (i.e. 120 et 140 Hz, voir exemple **Figure 30**). Pour éviter ce phénomène de *leakage*, la meilleure option, mais pas toujours disponible, est d'allonger la durée de la fenêtre d'analyse. En général, il faut que la durée T de la fenêtre soit égale ou supérieure à 2 ou 4 cycles (ou périodes) de la fréquence d'intérêt la plus basse. Par exemple, si l'on veut estimer l'amplitude d'une fréquence de 100 Hz dans le signal, la durée T doit être au moins de 20 ms (i.e. (1/100)*2). S'il n'est pas possible d'allonger la fenêtre d'analyse, une manière d'augmenter virtuellement la longueur de la fenêtre d'analyse est d'ajouter une série de zéros au début ou à la fin du signal. Cette technique appelée *zero padding* n'améliore pas la résolution fréquentielle, puisqu'il n'y a pas plus d'information dans le signal, mais réduit les phénomènes de *leakage*. La FFT sera en quelque sorte lissée, de manière à ce que les pics d'amplitude se retrouvent au niveau des bonnes fréquences. En général, on rajoute autant de zéros que nécessaire pour atteindre un nombre d'échantillons équivalent à la fréquence d'échantillonnage du signal, ce qui augmente virtuellement la résolution fréquentielle à 1 Hz. Par exemple, si notre signal de 50 ms est échantillonné à 20 kHz (1000 point-échantillons), pour faire passer la résolution de 20 Hz à 1 Hz, il faut rajouter 19000 zéros au début ou à la fin du signal avant de calculer la FFT.

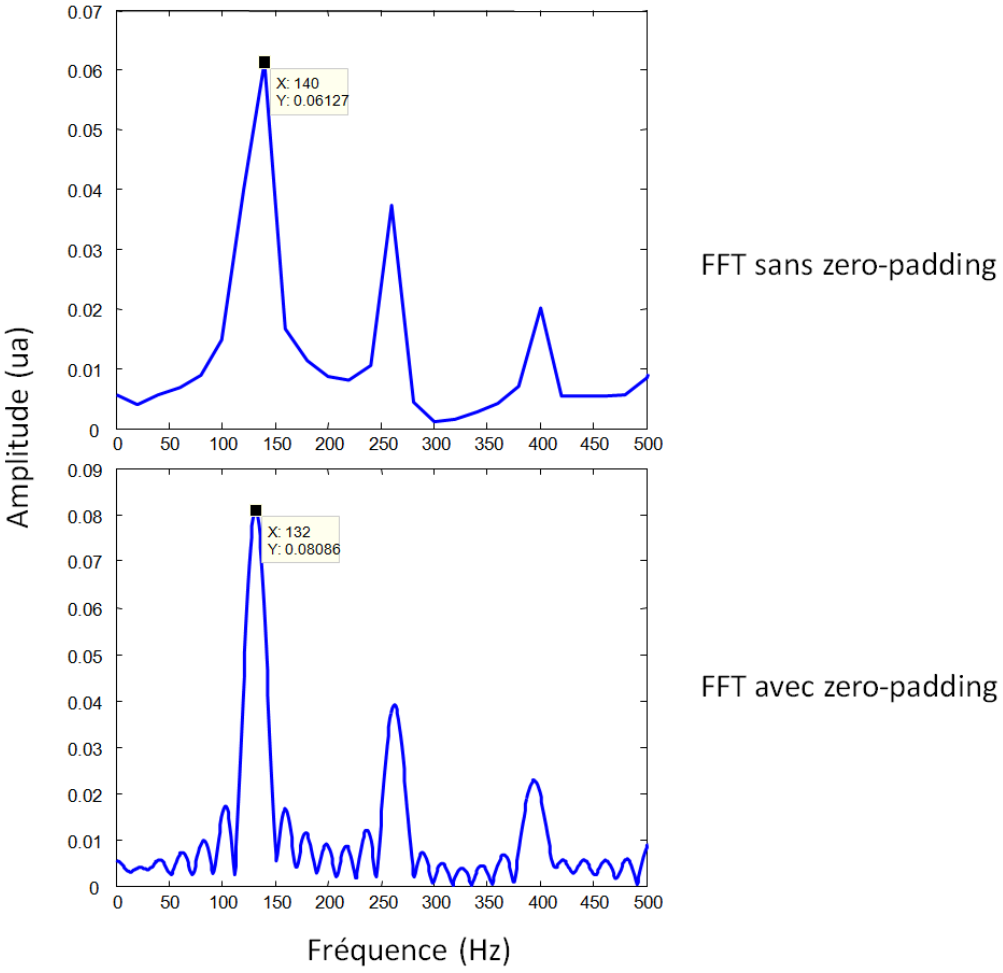


Figure 30. Illustration du phénomène de *leakage* et des effets de la technique du *zero-padding*. Dans cet exemple, une FFT a été calculée sur un signal d'une durée de 50 ms contenant une fréquence fondamentale (F0) à environ 130 Hz et ses harmoniques (multiples entiers de la F0). La résolution fréquentielle est de 1/0.05 sec, soit 20 Hz. Or, le signal contient une fréquence à 130 Hz qui n'est pas un multiple entier du pas fréquentiel, elle va donc "fuir" vers les fréquences voisines, ici à 140 Hz (FFT sans zero-padding). La technique du zero-padding n'améliore pas la résolution fréquentielle, mais permet de réduire le phénomène de *leakage*, le pic se retrouve alors à 130 Hz. Issue des données recueillies durant la thèse.

Un autre point à prendre en compte est que lors du calcul de la FFT le signal est considéré comme une boucle continue, dans laquelle le premier et le dernier échantillon sont contigües. De ce fait, s'il existe une différence d'amplitude trop importante entre le début et la fin du signal, cette discontinuité va générer un artéfact semblable à un clic présent dans le signal, ce qui va se traduire par une activité à large bande sur le spectre. Afin d'éviter cet artéfact, il est possible de multiplier le signal par une fonction de fenêtrage, qui permet de réduire l'amplitude en début et en fin de signal, de telle manière que le signal commence et finisse par zéro. Une des

fonctions de fenêtrage les plus utilisée est la fenêtre de Hanning en forme de cloche (**Figure 31**). Néanmoins, il est important de noter que les fonctions de fenêtrage déforment le signal, ce qui peut réduire grandement l'amplitude de certaines composantes fréquentielles présentes en début ou en fin de signal telles que les réponses transitoires.

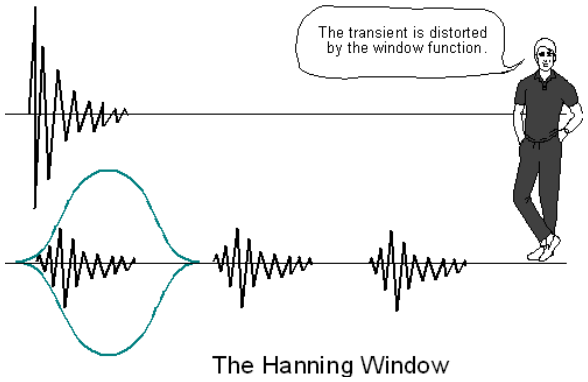


Figure 31. Illustration de la fenêtre de Hanning. Le début et la fin du signal sont ramenés à zéro, ce qui évite l'effet de discontinuité et prévient la génération d'un signal artéfactuel lors du calcul de la FFT. En revanche, comme le signal est déformé, une partie de l'information est perdue par le fenêtrage, en particulier au début et à la fin du signal qui sont les parties les plus affectées. Image issue de http://azimadli.com/vibman/thehanningwindow.htm

Comme les *cABRs* sont des réponses neuronales très similaires au stimulus, les pics d'amplitude du spectre de la réponse neuronale sont souvent proches de ceux des fréquences présentes dans le stimulus (**Figure 5**). De ce fait, il peut être intéressant d'extraire les pics d'amplitude maximale dans une fenêtre fréquentielle donnée. Par exemple, si le stimulus contient une fréquence à 100 Hz, on cherchera un pic d'amplitude à 100 +/- 20 Hz au niveau du spectre de la réponse neuronal. La largeur de la fenêtre fréquentielle varie selon les études de 20 à 60 Hz (Parbery-Clark et al., 2012; Parbery-Clark et al., 2009) ou est même ajustée après inspection visuelle du spectre de la réponse neuronale. Une inspection visuelle du grand moyennage des FFTs permettra de définir la taille de fenêtre la plus adaptée. Il est aussi possible de calculer l'amplitude spectrale moyenne sur une large bande, ce qui peut être utile lorsqu'il s'agit d'analyser des composantes qui varient dans le temps telles que les formants (Banai et al., 2009).

La transformée de Fourier peut aussi être calculée sur une fenêtre glissante, ce qui permet d'étudier les variations des composantes spectrales en fonction du temps. Le résultat est représenté sous la forme d'un graphique à trois dimensions nommé spectrogramme ou *short-term Fourier transform* : les fréquences en ordonnée, le temps en abscisse et un code couleur pour l'amplitude spectrale (**Figure 32**). Dans ce cas, il est important d'appliquer une fonction de

fenêtrage à chaque pas de temps, sans quoi l'analyse résultante sera contaminée par de nombreux artéfacts à large bande.

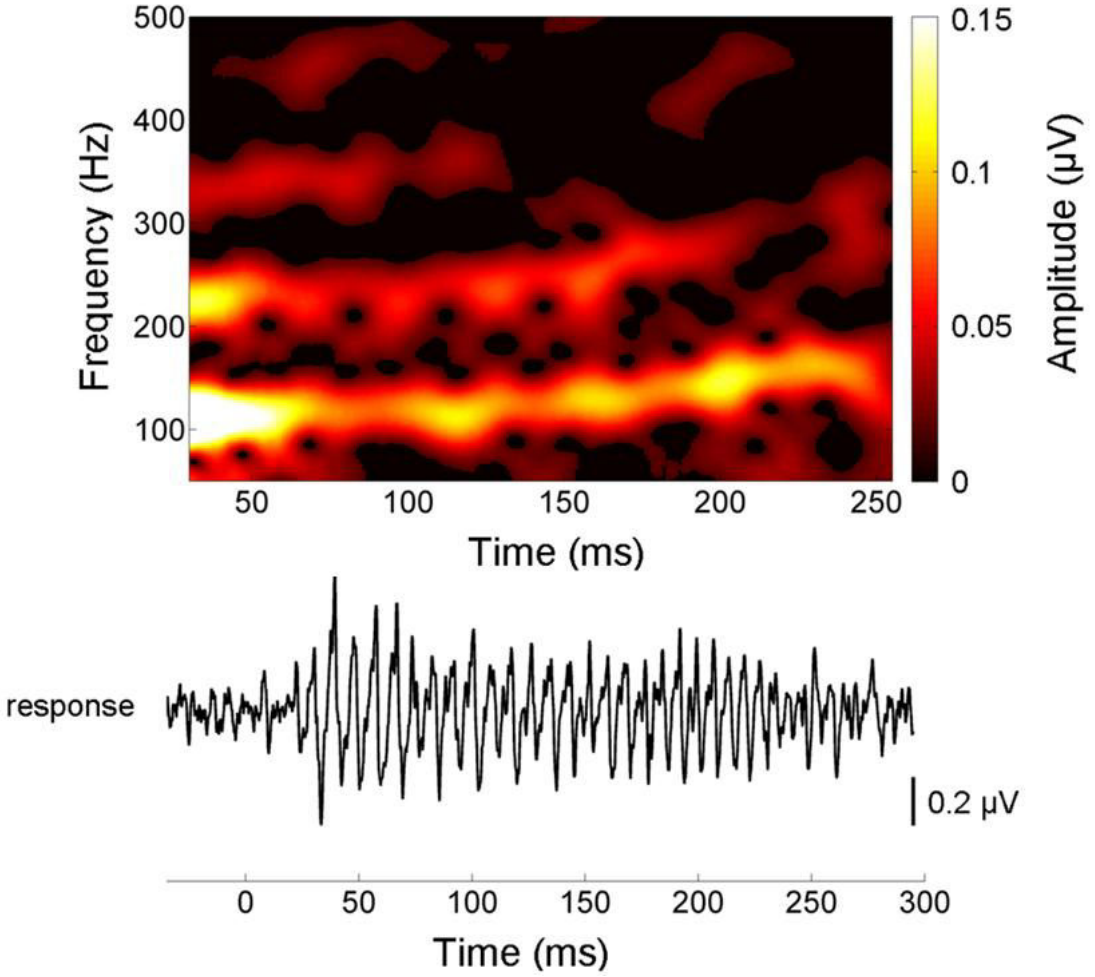


Figure 32. Illustration d'une analyse en temps-fréquence. Spectrogramme obtenu par fenêtre glissante sur la réponse sous-corticale évoquée par la syllabe chinoise /mi/ ayant un contour tonal montant, avec la fréquence en ordonnée (en Hz), le temps en abscisse (en ms) et un code couleur pour l'amplitude spectrale (en µV). Les couleurs claires illustrent une amplitude plus élevée. Cette méthode permet de visualiser le suivi de fréquence des *cABRs* en fonction de la fréquence et du temps. Issue et adaptée de Skoe et Kraus (2010).

8.2.4. Consistance temporelle des *cABRs* (*neural response consistency*)

La robustesse du signal à travers les essais peut être estimée de deux façons, dans le plan temporel avec la technique de consistance temporelle (*neural response consistency*) et dans le plan spectral avec la technique de la cohérence de phase inter-essais (*inter-trial phase-locking*, voir ci-dessous).

La technique de consistance temporelle, précédemment utilisée dans des études (Hornickel et al., 2012) et décrite par Tierney et Kraus (2013), consiste en trois étapes répétées un grand nombre de fois (i.e. *bootsrapping*) : (1) un sous-échantillon du nombre total d'essais enregistrés est tiré au hasard et moyenné. Par exemple, pour 6000 essais enregistrés, 3000 seront tirés au hasard et moyennés. (2) les 3000 essais restants sont aussi moyennés. Ces deux étapes donnent deux signaux moyennés. Au plus un signal est robuste dans le temps, au plus les deux signaux seront similaires. (3) afin de quantifier cette similarité, le coefficient de corrélation entre ces deux signaux est calculé. Ces trois étapes sont répétées un grand nombre de fois, en général 100 fois, correspondant à 100 tirages aléatoires des 6000 essais, résultant ainsi en 100 valeurs de corrélations. Enfin, ces 100 valeurs de corrélation sont moyennées pour former un indice global de la consistance temporelle du signal à travers les essais.

8.2.5. Cohérence de phase inter-essais (*inter-trial phase-coherence*)

La robustesse du signal à travers les essais peut être estimée dans le plan spectral grâce à la technique de la cohérence de phase inter-essais, utilisée précédemment et décrite par Tierney et Kraus (2016, 2013). Cette technique permet d'apprécier la robustesse du signal à travers les essais et en fonction de la fréquence. En pratique, une transformée de Fourier est calculée pour chaque essai, ce qui donne pour chaque fréquence deux valeurs, l'une correspondant à l'amplitude et l'autre à la phase. Seule la phase, qui véhicule les informations temporelles du signal, est gardée pour la suite des analyses. Plus le signal est robuste à travers les essais, moins la phase des fréquences sera variable. La variabilité des phases peut être représentée dans un espace trigonométrique (**Figure 33**). Chaque phase est représentée par un vecteur orienté dans une direction. Le vecteur moyen de toutes les phases ou vecteur résultant permet d'apprécier la variabilité des phases. L'angle du vecteur résultant illustre la direction moyenne de l'ensemble des phase, alors que la longueur du vecteur illustre la variabilité de la phase au cours des essais. Un signal peu robuste va générer des phases ayant des directions aléatoires, ce qui donnera un vecteur résultant proche de zéro. Inversement, un signal robuste aura des phases peu variables à travers les essais, ce qui donnera un vecteur résultant élevé. Le vecteur résultant peut être facilement calculé à l'aide de la fonction circ_r.m de la toolbox Matlab CircStat (Berens, 2009). La valeur de cohérence de phase inter-essais varie de 0 (aucune cohérence de phase) à 1 (cohérence de phase parfaite).

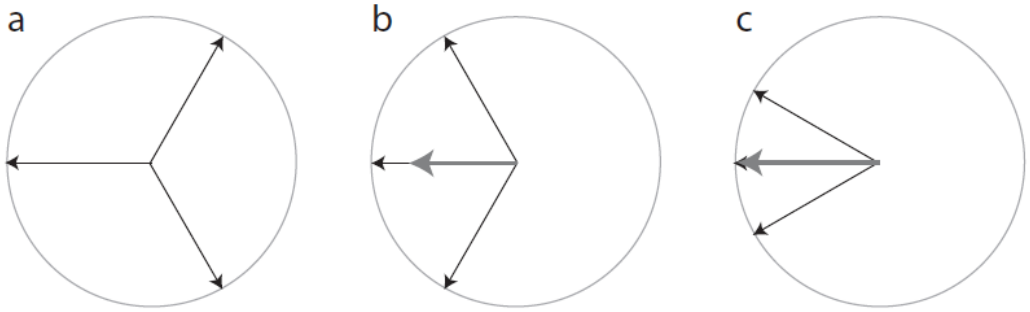


Figure 33. Illustration de la cohérence de phase inter-essais dans le plan trigonométrique. Dans cet exemple, la phase d'une fréquence est extraite pour trois essais et représentée sous la forme d'un vecteur. **a**. Les phases sont aléatoires donc la direction de leurs vecteurs s'oppose globalement, ce qui génère un vecteur résultant nul. **b.** Les phases sont moins variables, le vecteur résultant est positif. **c.** Les phases sont peu variables, ce qui génère un vecteur résultant élevé. Issus de Berens (2009).

Il est possible d'extraire les valeurs de cohérence maximale dans une fenêtre fréquentielle donnée. En effet, comme les *cABRs* sont des réponses neuronales très similaires au stimulus, les valeurs de cohérence maximales vont se retrouver au niveau des pics de fréquences présentes dans le stimulus (**Figure 34**). Comme pour l'analyse des représentations spectrales de la FFT, les pics de cohérence seront recherchés dans des fenêtres spectrales dont la largeur varie de +/- 20 Hz à +/- 60 Hz. Une inspection visuelle du grand moyennage des tracés de cohérence de phase permettra de définir la taille de fenêtre la plus adaptée.

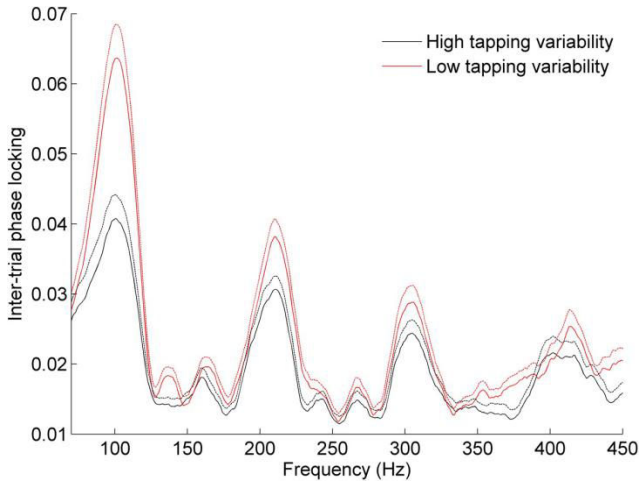


Figure 34. Illustration d'une analyse en cohérence de phase inter-essais calculée sur des réponses neuronales évoquées par la syllabe /da/ ayant une fréquence fondamentale de 100 Hz. Dans cette étude les sujets ont été séparés en deux groupes en fonction de leurs performances de tapping : les tracés rouges représentent la moyenne de cohérence de phase en fonction de la fréquence pour les bons tapeurs, et les tracés noirs pour les mauvais tapeurs. Les tracés pointillés représentent + 1 erreur standard. Issue de Tierney et Kraus (2016).

Cette technique peut aussi être réalisée sur une fenêtre glissante dans le temps, ce qui permet d'étudier la cohérence de phase inter-essais en fonction du temps et de la fréquence (**Figure 35**).

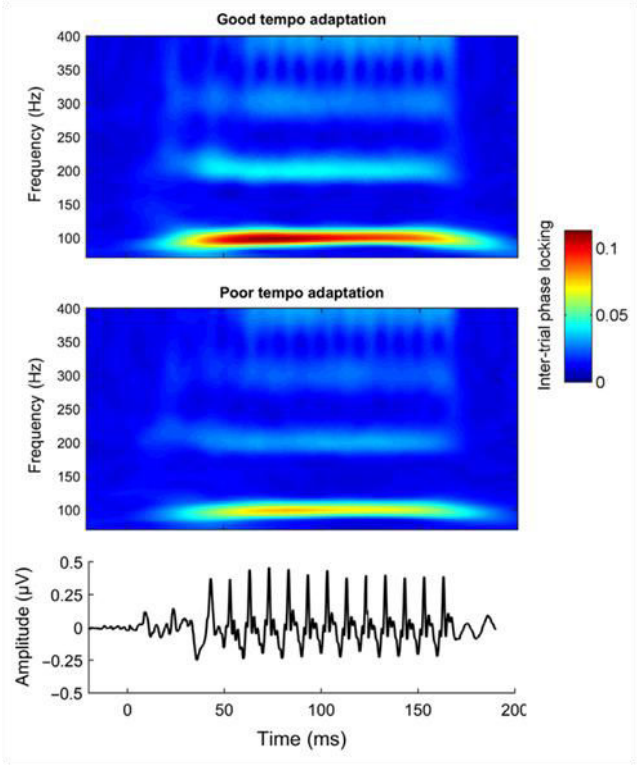


Figure 35. Exemple de cohérence de phase inter-essais en fonction de la fréquence et du temps. Dans cette étude les sujets ont été séparés en deux groupes en fonction de leurs performances de tapping : en haut ceux qui ont de bonnes performances, en dessous ceux qui ont de moins bonnes performances. En bas, le tracé représente le grand moyennage des réponses neuronales de tous les sujets. Issue et adaptée de Tierney et Kraus (2016).

9. Forces et limites de la technique des *cABRs*

La technique des *cABRs* présente des avantages, mais aussi des limites qu'il faut prendre en considération avant tout enregistrement.

Tout d'abord sur le plan pratique, la mise en place des électrodes est très rapide puisque seules trois électrodes sont réellement nécessaires pour recueillir les réponses du tronc cérébral à la différence des potentiels corticaux qui nécessitent généralement la pose d'un casque EEG complet, pour lequel le nombre d'électrodes peut aller jusqu'à 256. De plus, l'enregistrement des réponses neuronales ne nécessite pas la participation active du sujet, sauf dans certains cas (e.g. études des effets attentionnels), il n'y a donc pas les mêmes risques associés aux taches comportementales comme les artéfacts musculaires, la mauvaise réalisation de la tâche, etc... En général, le sujet est assis sur un fauteuil incliné avec la possibilité de regarder un film en sous-titré afin de maintenir son état d'éveil tout en restant détendu, ce qui rend l'expérience plutôt confortable pour le sujet. Et même si la durée de l'expérience peut parfois être longue en raison

du grand nombre d'essais nécessaires à l'émergence des *cABRs,* les sujets demandent rarement des pauses et sont en général satisfaits de la durée de l'expérience.

Sur le plan fondamental, la technique des *cABRs* permet d'étudier l'encodage automatique des sons de manière beaucoup plus fine que les autres mesures électrophysiologiques telles que les *click-ABRs* ou les potentiels évoqués auditifs corticaux. Comme les *cABRs* sont riches en informations temporelles et spectrales, il est possible d'utiliser une multitude de mesures pour disséquer les différentes composantes présentes dans la réponse afin d'étudier différents aspects du traitement auditif. De plus, même si ces réponses sont généralement recueillies dans des conditions d'écoute passive, elles reflètent néanmoins l'interaction dynamique entre les processus cognitifs *top-down* et sensoriels *bottom-up*. Les *cABRS* sont des mesures objectives de l'encodage fin des sons de parole, sensibles aux troubles de l'audition et aux effets d'expertise (musicale, linguistique, bilinguisme...), ce qui permet d'enrichir de manière importante nos connaissances sur le fonctionnement du système auditif et de proposer de nouvelles stratégies thérapeutiques.

En revanche, comme nous l'avons vu au cours de ce chapitre, les réponses peuvent être contaminées par de nombreuses sources (stimulus, 50 Hz, microphone cochléaire,...), et bien qu'il existe de nombreuses techniques afin de prévenir ou de limiter l'influence de ces artéfacts, il peut parfois être difficile de distinguer la réponse neuronale de l'artéfact. Il faut donc prendre de nombreuses précautions avant, pendant et après tout enregistrement.

En outre, même si la plupart des études considère que ces réponses neuronales ont une origine sous-corticale, la question de l'origine des *cABRs* n'a pas encore fait l'objet de nombreux travaux. Les quelques études disponibles sur le sujet suggèrent que ces réponses neuronales seraient générées par un ou plusieurs noyaux sous-corticaux tels que le noyau cochléaire, le lemnisque latéral et le colliculus inférieur (Akhoun et al., 2010; Bidelman, 2015; Chandrasekaran and Kraus, 2010). Néanmoins, à l'aide d'enregistrement intracrâniens chez des patients épileptiques pharmaco-résistants, une étude a récemment montré qu'il existe des réponses de type *FFR* au niveau du cortex auditif primaire très similaires aux réponses enregistrées par les électrodes de surface (Bellier et al., en préparation). De plus, la *FFR* de surface présenterait une distribution temporale bilatérale, ce qui suggère une contribution importante du cortex auditif primaire dans la génération de la *FFR*. L'ensemble de ces études montrent qu'il n'existe pas encore de consensus sur l'origine des *cABRs*, il faut donc rester prudent quant aux conclusions des études utilisant cette technique.

Troisième partie - TRAVAUX EXPERIMENTAUX

Axe « Effet de l'expertise linguistique et musicale »

Les études précédentes ayant exploré les effets de l'expertise linguistique et les effets de transfert entre musique et langage au niveau des réponses neuronales sous-corticales se sont uniquement intéressées à la comparaison des langues tonales et non-tonales. Ces comparaisons ont permis de mettre en évidence un meilleur encodage de la fréquence fondamentale chez les chinois Mandarins et les musiciens pour lesquels il existe une exposition importante à une grande variété de patterns de hauteur tonale. Or, comme nous l'avons mentionné dans la partie introductive, ces avantages sont indépendants du domaine d'expertise linguistique ou musical, puisque des effets de transfert ont été observés d'un domaine à l'autre, de la musique vers les langues tonales, et inversement.

Afin de déterminer s'il existe des effets d'expertise spécifiques à la langue maternelle, et indépendants de processus de traitement plus généraux tels que ceux liés à l'encodage de la fréquence fondamentale, nous avons dans notre première étude comparé l'encodage sous-cortical de sons natifs et non-natifs entre deux langues non-tonales, le Français et l'Anglais, pour lesquelles les différences reposent principalement sur l'organisation de leur répertoire phonémique, lui-même dépendant des rapports de fréquences formantiques spécifiques à chaque langue.

Bien que les effets de transfert entre musique et langage et plus spécifiquement entre musique et langues tonales soient bien documentés, à la fois au niveau cortical et sous-cortical, nous en savons encore peu sur la manière dont la pratique musicale améliore les capacités de perception et de production dans les langues étrangères non-tonales. Le but de notre deuxième étude était de déterminer si, au niveau sous-cortical, les musiciens présentaient des avantages dans l'encodage des syllabes non-tonales étrangères. Notre hypothèse, basée sur les résultats de notre première étude et sur la littérature déjà existante, était que la pratique musicale permettrait de renforcer le traitement des indices linguistiques les plus importants, tels que les formants qui participent à la discrimination des phonèmes dans les langues non-tonales.

Ces deux études ont donné lieu à deux publications, la première étant publiée, et la seconde en voie de soumission.

10. Effet de l'expertise linguistique

L'étude intitulée : « *Native language shapes automatic neural processing of speech* » a été publiée en 2016 dans le journal à comité de lecture *Neuropsychologia*, numéro 89, pages 57-65. Cette étude est détaillée ci-après.

Native language shapes automatic neural processing of speech

Bastien Intartaglia[a,b,c,†], Travis White-Schwoch[d], Christine Meunier[c,e], Stéphane Roman[f], Nina Kraus[d,g,h], Daniele Schön[a,b,c]

[a] *Aix-Marseille Université, INS, Marseille, France*

[b] *INSERM, U1106, Marseille, France*

[c] *Brain and Language Research Institute, Labex BLRI*

[d] *Auditory Neuroscience Laboratory and Department of Communication Sciences, Northwestern University, Evanston, Illinois, United States of America*

[e] *Aix-Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France*

[f] *La Timone Children's Hospital, ENT Unit, Marseille, France*

[g] *Department of Neurobiology & Physiology, Northwestern University, Evanston, Illinois, United States of America*

[h] *Department of Otolaryngology, Northwestern University, Chicago, Illinois, United States of America*

[†]Corresponding author
Bastien Intartaglia
Institut de Neurosciences des Systèmes
Inserm UMR1106
Aix-Marseille Université
Faculté de Médecine, 27, Boulevard Jean Moulin
13005 Marseille, France
Email: bastien.intartaglia@etu.univ-amu.fr
Phone: 334 91 38 46 33

Abstract

The development of the phoneme inventory is driven by the acoustic-phonetic properties of one's native language. Neural representation of speech is known to be shaped by language experience, as indexed by cortical responses, and recent studies suggest that subcortical processing also exhibits this attunement to native language. However, most work to date has focused on the differences between tonal and non-tonal languages that use pitch variations to convey phonemic categories. The aim of this cross-language study is to determine whether subcortical encoding of speech sounds is sensitive to language experience by comparing native speakers of two non-tonal languages (French and English). We hypothesized that neural representations would be more robust and fine-grained for speech sounds that belong to the native phonemic inventory of the listener, and especially for the dimensions that are phonetically relevant to the listener such as high frequency components. We recorded neural responses of American English and French native speakers, listening to natural syllables of both languages. Results showed that, independently of the stimulus, American participants exhibited greater neural representation of the fundamental frequency compared to French participants, consistent with the importance of the fundamental frequency to convey stress patterns in English. Furthermore, participants showed more robust encoding and more precise spectral representations of the first formant when listening to the syllable of their native language as compared to non-native language. These results align with the hypothesis that language experience shapes sensory processing of speech and that this plasticity occurs as a function of what is meaningful to a listener.

Keywords

Experience-dependent plasticity; Auditory brainstem responses; Speech perception

1. Introduction

While the number of consonants and vowels across world's languages is large, each language only uses a few dozen basic units. The development of this specific phoneme inventory during childhood is language dependent, meaning that it is driven by the acoustic-phonetic properties of a listener's native language. During the first months of life, infants are able to discriminate speech sounds that are not used in their native language but with growing exposure to their mother tongue, this ability declines, to finally disappear in adulthood (Werker and Tees, 2002). For example, in their cross-linguistic and longitudinal study, Werker and Tees (2002) showed that at 6-8 months of age, English infants' ability to discriminate Hindi or Salish speech contrasts is as good as native infants of the same age. Yet by 10-12 months of age, their performance drops drastically and remains as poor English-speaking adults'. This decline is not restricted to Western languages: it has been also observed in Eastern languages, such as Japanese. For Japanese adults, the perceptual distinction of two acoustically close - but distinct - phonemes /r/ and /l/, which are not distinct in Japanese, is impossible (Iverson et al., 2003; Zhang et al., 2005). It is worth noting that this language-dependent reorganization of the phonemic inventory relies on two concomitant and opposite developmental patterns. Indeed, the infant's ability to discriminate foreign speech sounds decreases, while at the same time the ability to discriminate native speech sounds improves (Cheour et al., 1998; Kuhl et al., 2006, 1992; Rivera-Gaxiola et al., 2005).

Electrophysiological studies confirm the hypothesis that children develop neural representations that become attuned to the processing of their native language to the detriment of foreign languages (Kuhl, 2004; Mehler et al., 1994; Ortiz-Mantilla et al., 2013). For instance, Cheour et al. (1998) found that from 6 to 12 months of age, mismatch negativity (MMN, an index of slower cortical activity occurring < 10 ms) amplitude drops significantly for non-native phonemes. Likewise, in a longitudinal study, Rivera-Gaxiola and colleagues (2005) found that discriminatory event-related potentials (ERP) to non-native contrasts were present at 7 months of age, but were largely reduced by 11 months of age, while at the same time, the responsiveness to native language contrasts increased over time. In adults, Dehaene-Lambertz et al. (2000) showed that acoustically-close French phonemes elicit an MMN peaking around 130 ms in native speakers (French), whereas the MMN is reduced or absent in non-native speakers (Japanese) who are unable to discriminate these phonemes. Recently, Raizada et al. (2010) compared English and Japanese participants listening to two acoustically-close English syllables, /ra/ and /la/. The separability of brain metabolic patterns predicted subject's behavioral ability to discriminate the

two syllables. Altogether these studies show that the neural attunement to native language takes place early during development and continues throughout the life-span, shaping the auditory system to become more efficient in processing phonemes that belong to the native language. Interestingly, Jeng et al. (2011) have shown that this pattern of developmental change is also present at the subcortical level. Compared to Mandarin-speaking infants, adults have stronger subcortical pitch encoding, a lexically-relevant feature to discriminate Mandarin syllables.

Two hypotheses have been proposed to explain this early attunement to native language. The bottom-up hypothesis assumes that infants extract discrete units from continuous speech through statistical learning. For instance, infants' ability to discriminate phonemes seems to rely heavily on the statistical distribution of speech sounds in the native language (Maye et al., 2002). In contrast, the top-down hypothesis suggests that learning low-level linguistic units involves higher-level units (i.e. words) (Fourtassi and Dupoux, 2014). According to this view, the English infant would learn to discriminate two similar phonemes (/æ/ and /e/), because they are relevant to discriminate two different words (bad vs. bed).

For many decades, the bottom-up hypothesis of speech processing was predominant, conveying the idea that, as speech sound is processed along the auditory pathway, neural structures' sensitivity to the acoustic content decreases while the sensitivity to abstract features (syllables, words, intelligibility) increases (Okada et al., 2010). In a commentary on Okada's article, Peelle et al. (2010) proposed a hierarchical model of speech processing that starts from Heschl's gyrus exhibiting high acoustic sensitivity and gradually shows higher acoustic invariance in anterior and posterior temporal regions.

An alternative view to the bottom-up and top-down hypothesis is a more interactive and dynamic model based on interplay between high and low levels of speech representation. Due to the high acoustical variability of real-life speech tokens, phonemic categories exhibit a certain degree of overlap (Hillenbrand et al., 1995), therefore, the bottom-up hypothesis is not sufficient to explain the whole development of a phonemic inventory. Indeed, computational studies show that top-down influences are needed to refine phonemes categories with a high degree of accuracy (Fourtassi and Dupoux, 2014). Moreover, Lew-Williams and Saffran (2012) showed that previous exposure to specific word lengths (bi- or tri-syllabic words) influences infants' ability to segment fluent speech. In other words, prior linguistic knowledge builds expectations that influences speech processing in a top-down manner. In the Reverse Hierarchy Theory (RHT), Ahissar and Hochstein (2004) postulate that perceptual learning starts at high-level cortical areas. Then, through long-term exposure to a given context, plasticity would gradually reach lower-level areas, via top-down dynamics. The RHT was originally proposed for visual perception, but has

been recently extended to auditory perception (Gutschalk et al., 2008; Suga, 2008). For example, electrical stimulation of the primary auditory cortex modulates activity in subcortical auditory structures such as the inferior colliculus (Gao and Suga, 2000) and the cochlea (Perrot et al., 2006). Together, these studies support the hypothesis that refinement of neuronal representations to native speech sounds is a result of continuous interactions between primary and associative auditory structures and subcortical auditory structures (Kraus and Chandrasekaran, 2010; Tzounopoulos and Kraus, 2009) and are consistent with an emerging view of the auditory system as a distributed, but integrated, circuit (Kraus and White-Schwoch, 2015).

The anatomical organization of the auditory system supports these top-down and bottom-up interactions. Peripheral auditory structures such as the cochlea send neural firings from the auditory nerve to the auditory cortex via a series of brainstem nuclei. In addition, central auditory structures such as the primary auditory cortex and associative cortices send back top-down projections to periphery (Kral and Eggermont, 2007). Thus, the neural representation of speech sounds is the result of bottom-up mechanisms that can be modulated via the descending cortico-fugal system acting on subcortical structures. According to this interactive model, the auditory midbrain, where afferent and efferent projections converge, presents an excellent model to study the effects of language experience on speech processing.

Research on language-dependent brain plasticity in the subcortical auditory system, focusing on faster neural activity occurring < 1 ms, is an emerging area of study. Krishnan et al. (2005) compared auditory brainstem responses evoked by Mandarin tones in native speakers of Chinese Mandarin and native speakers of American English. They found that Chinese participants have a more robust and faithful representation of the fine pitch variations of Mandarin tones as compared to American participants. Indeed, in Mandarin Chinese, dynamic variations in voice pitch (i.e. the fundamental frequency) provide a major acoustic cue to discriminate two monosyllabic words. For instance, the syllable /yi/ with high-rising pitch contour means "aunt", whereas /yi/ with a high-falling pitch contour means "easy". In contrast, pitch variations in non-tonal languages (e.g. English) are not lexically relevant to discriminate words or syllables; rather they convey supra-lexical information such as stress and intonation patterns (Krishnan and Gandour, 2014). However, in a subsequent study, Krishnan and colleagues used iterated rippled noise (IRN) to simulate Mandarin tones without any speech context, and found that Mandarin speakers exhibited better pitch representation at the subcortical level as compared to American speakers. Thus, these effects may not be necessarily language-specific (Krishnan et al., 2009). Similar to musicians, who, via intensive training, develop outstanding abilities to track the fundamental frequency (i.e. the pitch) of music sounds,

Mandarin speakers develop, through long-term exposure to tonal speech sounds, excellent skills to process fine variations of the pitch in subcortical systems (Bidelman et al., 2011). Overall, since tonal languages use qualitatively different phonemic contrasts as compared to non-tonal languages (i.e. pitch contour), it remains unclear whether the differences described above are due to top-down influences of long-term phonemic representations on subcortical functioning or to a more precise pitch tracking computation, independently of whether the stimulus is part of phonemic inventory of the language system.

The aim of this cross-language study is to determine how far subcortical encoding of speech sounds is sensitive to language experience. Comparing neural responses in native speakers of two non-tonal languages (American English and French), listening to syllables of both languages, gives us the opportunity to study language-dependent plasticity at the subcortical level without confounding factors such as those described above when comparing tonal and non-tonal languages.

We hypothesized that subcortical processing would be more robust and fine-grained for speech sounds that belong to the native phonemic inventory of the listener. In other words, American native listeners should exhibit a more robust and faithful subcortical representation of an American English syllable as compared to French native listeners. Conversely, French listeners should have better representation of a French syllable as compared to non-native listeners. This hypothesis is precisely tested by the interaction term of a factorial design: we presented to French and American participants a French and an American English syllable, [ru] and [thae], respectively. Both phonemes are "illegal" in the non-native language, that is, both consonant and vowel of each syllable do not exist in the other language. This should maximize differences in long-term memory representations of these two syllables, thus increasing any potential top-down effect of language experience on auditory processing. Because the distinction of consonants and vowels mostly relies on the formants' properties, we hypothesized that differences on neural responses would be maximal over high frequency components of the EEG spectrum (200-800 Hz) such as harmonics at formant frequencies but not at the fundamental frequency.

2. Materials and methods

2.1. Participants

Twenty-six (18 females and 8 males) adult native speakers of American English and thirty-five (21 females and 14 males) native speakers of French, ranging in age from 18 to 36

years, participated in the study. American participants were recruited at Northwestern University (Chicago, USA) and French participants were recruited at Aix-Marseille University (Marseille, France). Inclusion criteria were a high-school level of education and click-evoked brainstem response latencies within lab-internal normal limits (Wave V latency 5.41-5.97 ms in response to a click presented at 80 dB sound pressure level, rate of 31/s). To make sure that there were no different hearing thresholds, we compared wave V latencies and amplitudes between groups. Analyses showed that groups were matched for both measures (wave V latency : $U = 250$, $p = 0.633$; wave V amplitude : $U = 252$, $p = 0.664$). The two language groups were also matched in term of age (Americans : 22 ± 3 years ; French : 23 ± 3 years ; $F(1,46) = 0.19$, $p = 0.662$) and of musicianship (Americans : 8 ± 5 years; French : 5 ± 6 years of musical practice ; $F(1,46) = 0.87$, $p = 0.355$). The Northwestern University Institutional Review Board and INSERM approved all procedures. Participants gave their informed consent and were paid for their participation.

An ad-hoc questionnaire was used to measure language proficiency both in American English and French. On a scale from 0 (novice) to 10 (expert), participants self-rated their proficiency for oral and written expression, and oral and reading comprehension for both languages (English and French). All subjects reported high proficiency for their native language without any significant difference between groups. However, French participants reported significantly higher proficiency for English than American participants for French (see **Table 1**). This latter point will be further discussed in the discussion. One American participant was bilingual English-Spanish, and one French participant was bilingual French-Vietnamese.

	Skill	French	American	*p*-value
Native language	Understanding	10 (0)	9.95 (0.22)	0.35
	Reading	10 (0)	9.95 (0.22)	0.35
	Speaking	10 (0)	9.95 (0.22)	0.35
	Writing	10 (0)	9.90 (0.30)	0.06
Non-native language	Understanding	5.74 (2.71)	1.43 (3.09)	<0.001
	Reading	6.65 (2.44)	1.33 (2.88)	<0.001
	Speaking	5.13 (2.44)	1.28 (2.70)	<0.001
	Writing	5.65 (2.53)	0.86 (1.90)	<0.001

Table 1. Native and non-native language proficiency. Mean, standard deviations, and significance values for the French and American groups' self-rated proficiency of their oral and reading comprehension, and oral and written expression.

2.2. Stimuli

The two stimuli used were natural speech syllables, recorded in an anechoic chamber by an American English male speaker and a French male speaker. The French syllable [ʁy] (henceforth ru) and the English syllable [ðæ] (henceforth thae) were chosen because they are both « illegal » speech sounds in the other language, which means that both the consonant and the vowel do not exist in the other language (i.e., [ʁ] and [y] do not exist in English and [ð] and [æ] do not exist in French). Note that while the phoneme [r] exists in both English and French, its realization is very different across languages, with an uvular realization in French and a retroflex realization in English. This choice should maximize the differences between the two languages and should consequently maximize the expected effect of language experience on neural responses (see waveforms in **Figure 1**).

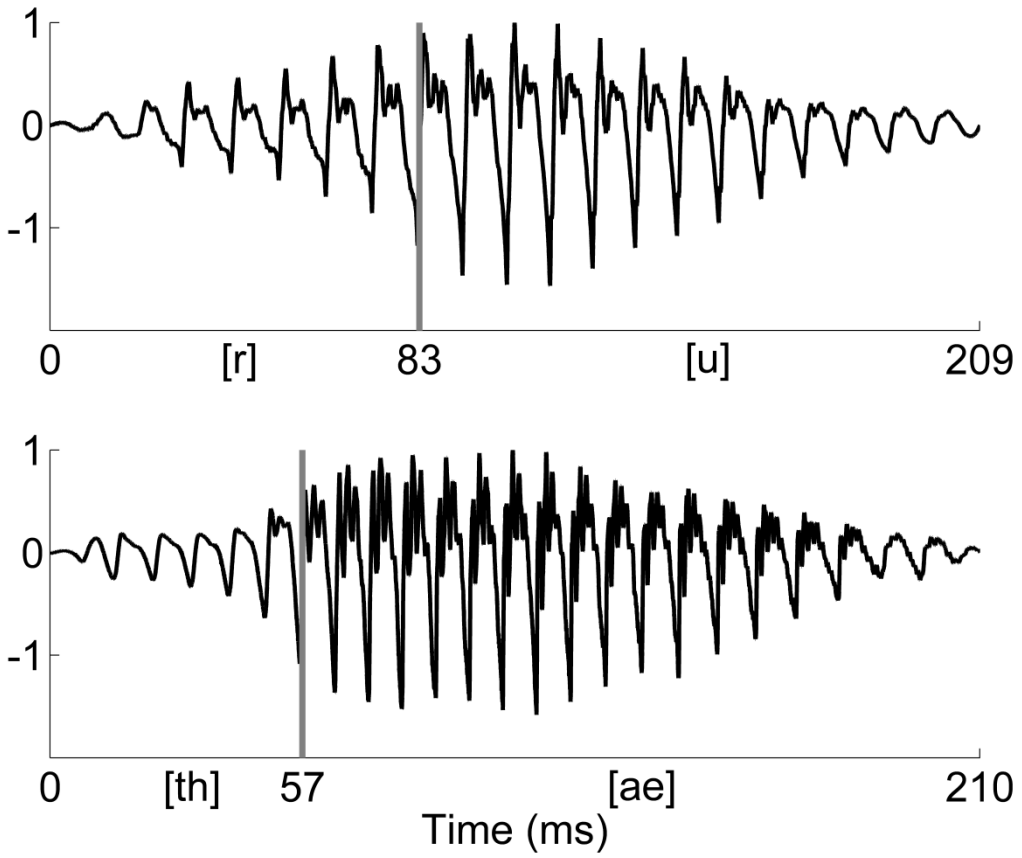


Figure 1. Waveforms of the two stimuli (normalized amplitudes). Top, waveform of the French stimulus [ru]. Bottom, waveform of the English stimulus [thae]. Vertical gray lines indicate the boundaries between the consonants and vowels, as established according to the spectral changes by an experienced phonetician.

The syllables were matched in duration (209 ms for [ru] stimulus and 210 ms for [thae] stimulus). Frequencies of interest included the fundamental frequency (*F*0) and the second to sixth harmonics (*H*2-*H*6). These latter were chosen because they encompass the first formant range (*F*1) which is an acoustically relevant cue in discriminating vowels (see **Table 2** for more details). Since the stimuli were natural speech sounds, the *F*0 and its subsequent harmonics *H*2-*H*6 vary slightly over time and were thus estimated by computing a fast Fourier transform (FFT) on three time regions (whole stimuli, consonant and vowel). These time regions were established according to the spectral changes by an experienced phonetician. Importantly the ranges of the first two vowel formants ([u]: *F*1 = 242-298 Hz, *F*2 = 1449-1732 Hz; [ae]: *F*1 = 420-705 Hz, *F*2 = 1408-1556 Hz, from Praat: Boersma and Weenink, 2001) showed that for both stimuli *F*2 range exceeded the phase-locking limit of the auditory brainstem which is around 1.5 kHz and was thus not taken into consideration in further analyses (Aiken and Picton, 2008; Moushegian et al., 1973).

It is important to make clear that, while the two stimuli have different spectral features (see FFT in **Figure 2** and **Table 2**), the aim of the study was to determine whether there would be an effect of language expertise on stimulus processing. This is reflected in the interaction term of the language expertise and stimulus factors. As such, a main effect of stimulus would not be particularly meaningful here because it would simply reflect the sensitivity of the neural system to acoustic differences in the stimuli.

Stimulus	FOI	Whole	Consonant	Vowel
[ru]	*F*0	101	93	105
	*H*2	205	188	210
	*H*3	311	279	316
	*H*4	401	373	421
	*H*5	487	461	528
	*H*6	590	549	632
	*F*1 range	N/A	N/A	242-298
[thae]	*F*0	129	118	125
	*H*2	261	235	258
	*H*3	393	355	386
	*H*4	531	472	528
	*H*5	665	590	665
	*H*6	787	711	788
	*F*1 range	N/A	N/A	420-705

Table 2. Frequencies of interest (FOI) for each time region and stimulus. Values represent the frequency peaks of the fundamental frequency (*F*0), its second through sixth subsequent harmonics (*H*2-*H*6) and frequency range of the first formant (*F*1 range) for each time region and stimulus (in Hz). Note that because stimuli are natural, FOI vary slightly over time and harmonics are not exactly whole integers of the *F*0.

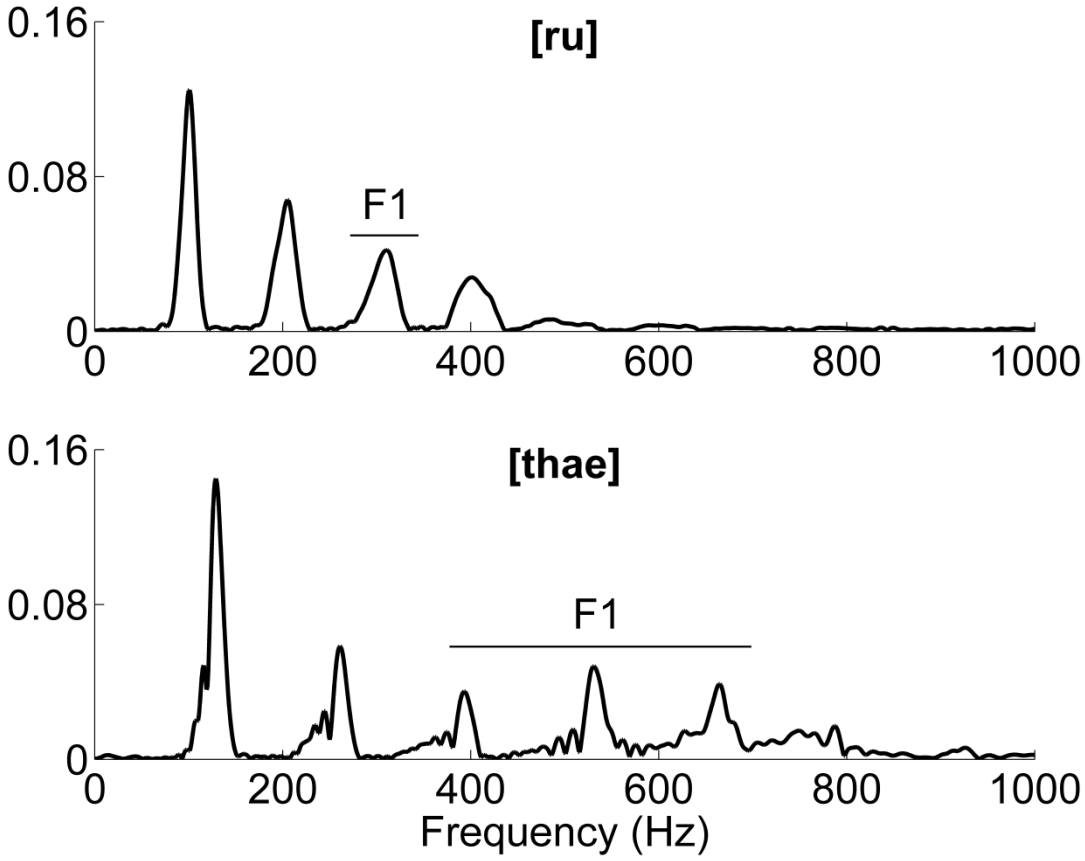


Figure 2. Fast Fourier transforms computed on the whole stimuli (normalized amplitudes). Top, French stimulus [ru]. Bottom, English stimulus [thae]. Horizontal lines indicate the range of the first formant.

Each stimulus was presented monaurally to the right ear at 80 dB SPL at a rate of 3.8 Hz with alternating polarities through magnetically shielded insert earphones (ER-3A, Etymotic Research) using the stimulus presentation software Microvitae (μV-ABR, Microvitae Technologies).

2.3. Electrophysiological recording

During electrophysiological recordings, participants sat in a comfortable reclining chair in an electrically-shielded, sound-attenuated room and were instructed to watch a subtitled movie of their choice to maintain relaxation and prevent drowsiness. Brain responses were collected at 30 kHz sampling rate using Microvitae recording system (μV-ABR, Microvitae Technologies) with three Ag-AgCl scalp electrodes in a vertical montage (Cz active, forehead ground, and right earlobe reference). Electrode impedances were kept <5 KΩ. Six-thousand sweeps were collected for each stimulus (two blocks of three-thousand sweeps). The order of presentation of the two

stimuli was counterbalanced across participants. One of the authors (BI) was in charge of data acquisition in both countries using the same portable EEG system. This prevents the possibility of having a bias due to different experimental setups, participant preparation, and instruction.

2.4. Data analysis

All analyses were performed using custom MATLAB scripts (MathWorks). First, electrophysiological recordings were bandpass filtered from 70 to 2000 Hz (12 dB/octave roll-off) using a Butterworth filter. Then, sweeps with activity exceeding ± 30 μV were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period (-30 to 0 ms). Neural responses were then averaged over a -30 to 229 ms window for [ru] stimulus and -30 to 230 ms for [thae] stimulus. The signal-to-noise ratio (SNR) was computed using the quotient of response root mean square (RMS) amplitude and pre-stimulus baseline RMS amplitude (see Skoe and Kraus, 2010). If the SNR was less or equal to 1.4 for one or both stimuli, the participant was excluded. This resulted in excluding 5 American and 8 French participants.

2.4.1. Spectral amplitude

The spectral energy of the neural responses was analyzed by computing Fast Fourier transform on three time regions of the response (whole response, consonant and vowel). These time regions were defined on the basis of the stimuli by a phonetician also taking into account a 10 ms neural delay in the response: whole response (10-220 ms for both stimuli), consonant (10-93 and 10-67 ms for the French and English stimuli respectively) and vowel (93-220 and 67-220 ms for the French and English stimuli respectively). For each time region, the maximum spectral amplitudes of the fundamental frequency (*F*0) and its second to sixth harmonics (*H*2-*H*6) were extracted in a bandwidth of 40 Hz surrounding the peak in the stimulus fast Fourier transform (e.g. for a peak at 105 Hz, values were extracted between 85 and 125 Hz). For the whole response and the consonant, the five values extracted from *H*2 to *H*6 were then averaged to form a global measure of harmonics' representation. For the vowel, since our hypothesis was that differences on neural responses would be maximal over frequencies that are relevant to discriminate vowels, harmonics were separated in two groups. Harmonics falling within *F*1 range ± 20 Hz were averaged to form a global measure of *F*1 spectral representation, while harmonics falling outside *F*1 range were averaged to form a global measure of non-formant spectral representation. For the vowel [u] the *F*1 range ± 20 Hz included only the third harmonic (*H*3), and the remaining

harmonics were considered as non-formant harmonics (*H*2, *H*4-*H*6). For the vowel [ae] the *F*1 range ± 20 Hz included the third to fifth harmonics (*H*3-*H*5), and the remaining harmonics were considered as non-formant frequencies (*H*2, *H*6).

2.4.2. Inter-trial phase-coherence

We used the same procedure described in Tierney and Kraus (2013). This technique measures the phase consistency across trials of each frequency component in the neural responses. To summarize, for the three time regions (whole response, consonant and vowel) a fast Fourier transform was performed on each trial that resulted in two values, the amplitude and the phase for each frequency component. Since we were interested in phase variability across trials, only the phase values were kept. The vector's length of each frequency was computed using the Matlab toolbox CircStat Version 2012a (Berens, 2009). The length of the resultant vector represents the phase-coherence across trials for each frequency component. This measure ranges from 0 (no phase coherence) to 1 (perfect phase coherence). For each time region, the maximum phase-coherence values of the fundamental frequency (*F*0) and its second to sixth harmonics (*H*2-*H*6) were extracted in a bandwidth of 40 Hz surrounding the frequencies of interest. Then, the same procedure described above for spectral analyses was used for the whole, the consonant and the vowel time regions.

2.5. Statistical analyses

All statistical analyses were performed using Statistica Version 7.1 (StatsSoft, Tulsa, OK). Repeated measure analyses of variance (RMANOVA) were used for group (American vs. French) x stimulus ([ru] vs. [thae]) comparisons for spectral representation and inter-trial phase-coherence. *Post-hoc* tests were used when appropriate (Fisher LSD).

3. Results

3.1. Spectral representation

3.1.1. Whole stimuli

In the neural responses to the whole stimuli, American participants had larger spectral amplitudes in response to the fundamental frequency (*F*0) for both [ru] and [thae] (main effect of group: $F(1,45) = 3.580$, $p = 0.045$). This did not differ across stimuli (group x stimulus interaction: $F(1,45) = 0.072$, $p = 0.789$).

Across both stimuli (thae/ru), the magnitude of the responses to the harmonics was equivalent for both groups ($F(1,42) = 0.057$, $p = 0.812$) suggesting that the main effect of group found for the *F*0 cannot be attributed to a global enhancement of stimuli representations in one group. Most importantly, stimuli neural representations of the harmonics (*H*2-*H*6) differed as a function of the native language (group x stimulus interaction: $F(1,42) = 5.704$, $p = 0.021$).

3.1.2. Consonants

Spectral representation of the *F*0 did not show group differences for both stimuli (main effect of group: $F(1,44) = 0.262$, $p = 0.611$; group x stimulus interaction: $F(1,44) = 0.022$, $p = 0.884$).

When turning to the harmonics (*H*2-*H*6), while the two groups did not differ overall (main effect of group: $F(1,44) = 0.440$, $p = 0.511$), stimuli neural representation did differed as function of the native language (group x stimulus interaction: $F(1,44) = 7.128$, $p = 0.011$). Indeed, *post-hoc* tests revealed that American participants had a stronger spectral representation of the harmonics of the American consonant ([th]: p=0.056), while this was not the case for the French consonant ([r]: $p = 0.421$, **Figure 3 and 4 A**).

3.1.3. Vowels

Analysis of the *F*0 did not reveal any significant effect (main effect of group: $F(1,44) = 1.165$, $p = 0.286$; group x stimulus interaction: $F(1,44) = 0.036$, $p = 0.850$).

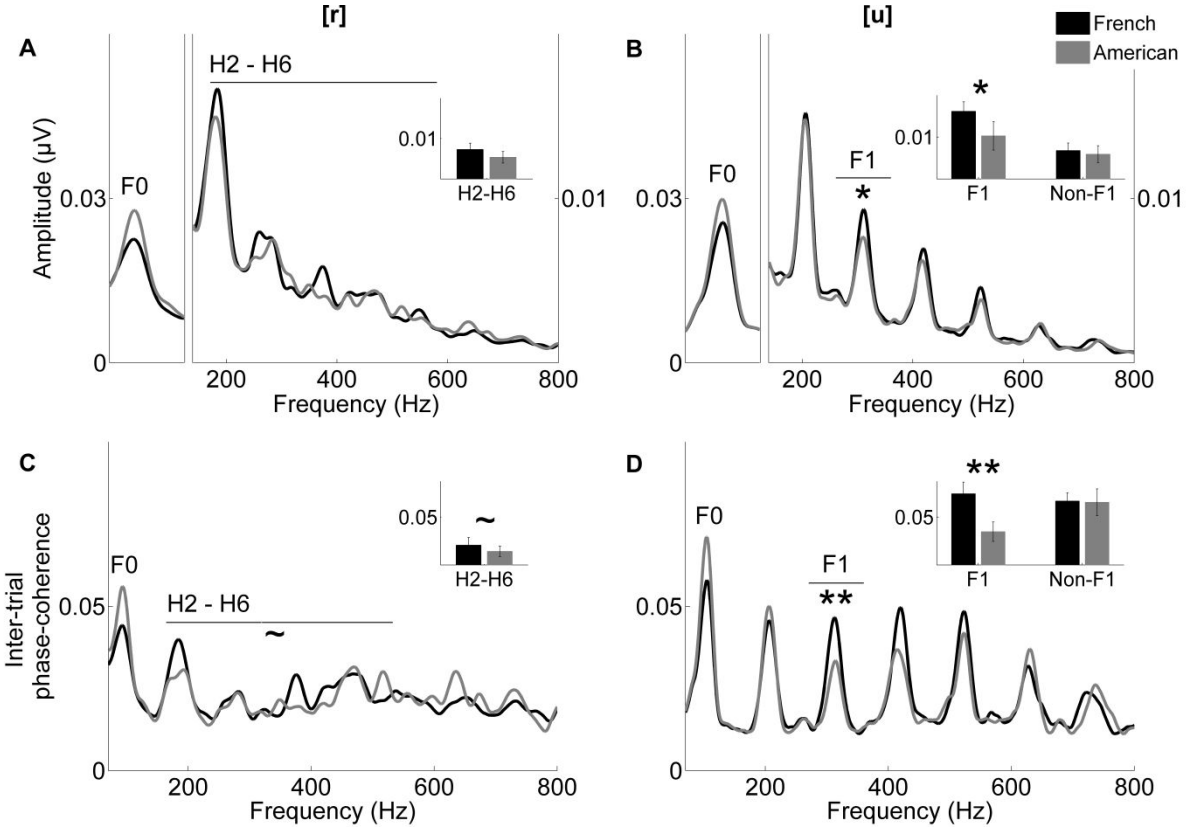


Figure 3. Spectral representation (top) and inter-trial phase-coherence (bottom) of the French consonant [r] (left panels) and vowel [u] (right panels). ***A***,***C***, The results show that for the French consonant [r], French (black) and American (gray) participants have equivalent representation of the *F*0 and its higher harmonics (*H*2-*H*6), while phase-coherence analyses show that French participants tend to have more robust representation of the higher harmonics only. ***B***,***D***, For the French vowel [u], French participants exhibit an advantage over American participant with a greater and more robust representation of the harmonics at formant frequencies (*F*1). * $p < 0.01$, ** $p < 0.005$. Error bars represent ± 1SE.

Analysis of the harmonics at formant frequencies (*F*1) showed that vowel neural representations differed as function of the native language (group x stimulus interaction: $F(1,43) = 4.295$, $p = 0.044$) with French showing an advantage over American participants in encoding *F*1-like harmonics of the French vowel [u]. *Post-hoc* tests revealed that for the French vowel [u] only, French participants had stronger spectral representations of *F*1 range harmonics frequencies ([u]: $p = 0.005$, effect size = 0.002 μV; [ae]: $p = 0.998$, **Figure 3 and 4 B**).

Interestingly, analyses at non-formant frequencies did not reveal any significant difference (main effect of group: $F(1,41) = 0.659$, $p = 0.421$; group x stimulus interaction: $F(1,41) = 0.008$, $p = 0.931$), showing that for the French vowel [u], French participants had a specific neural enhancement at harmonics within *F*1 range but not at harmonics outside *F*1 range.

3.2. Inter-trial phase-coherence

3.2.1. Whole stimuli

Analyses of inter-trial phase-coherence at the fundamental frequency (*F*0) did not reveal any significant effect (main effect of group: $F(1,43) = 0.550$, $p = 0.462$; group x stimulus interaction: $F(1,43) = 0.187$, $p = 0.667$).

Robustness of neural responses at harmonics (*H*2-*H*6) was equivalent between group (main effect of group: $F(1,40) = 0.098$, $p = 0.755$). However, robustness of stimuli neural representations of the harmonics (*H*2-*H*6) differed as a function of the native language (group x stimulus interaction: $F(1,40) = 4.414$, $p = 0.042$).

3.2.2. Consonants

Analyses of inter-trial phase-coherence at the *F*0 did not reveal any significant effect (main effect of group: $F(1,41) = 0.021$, $p = 0.885$; group x stimulus interaction, $F(1,41) = 0.002$, $p = 0.961$).

Inter-trial phase-coherence at harmonics (*H*2-*H*6) was equivalent between group (main effect of group: $F(1,43) = 0.256$, $p = 0.615$). However, robustness of stimuli neural representations differed as function of the native language (group x stimulus interaction: $F(1,43) = 5.179$, $p = 0.030$). This was mostly due to French participants having more robust encoding of the French consonant's harmonics compared to American participants, although this difference did not reach significant in *post-hoc* tests ([r]: $p = 0.102$; [th]: $p = 0.426$, **Figure 3 and 4 C**).

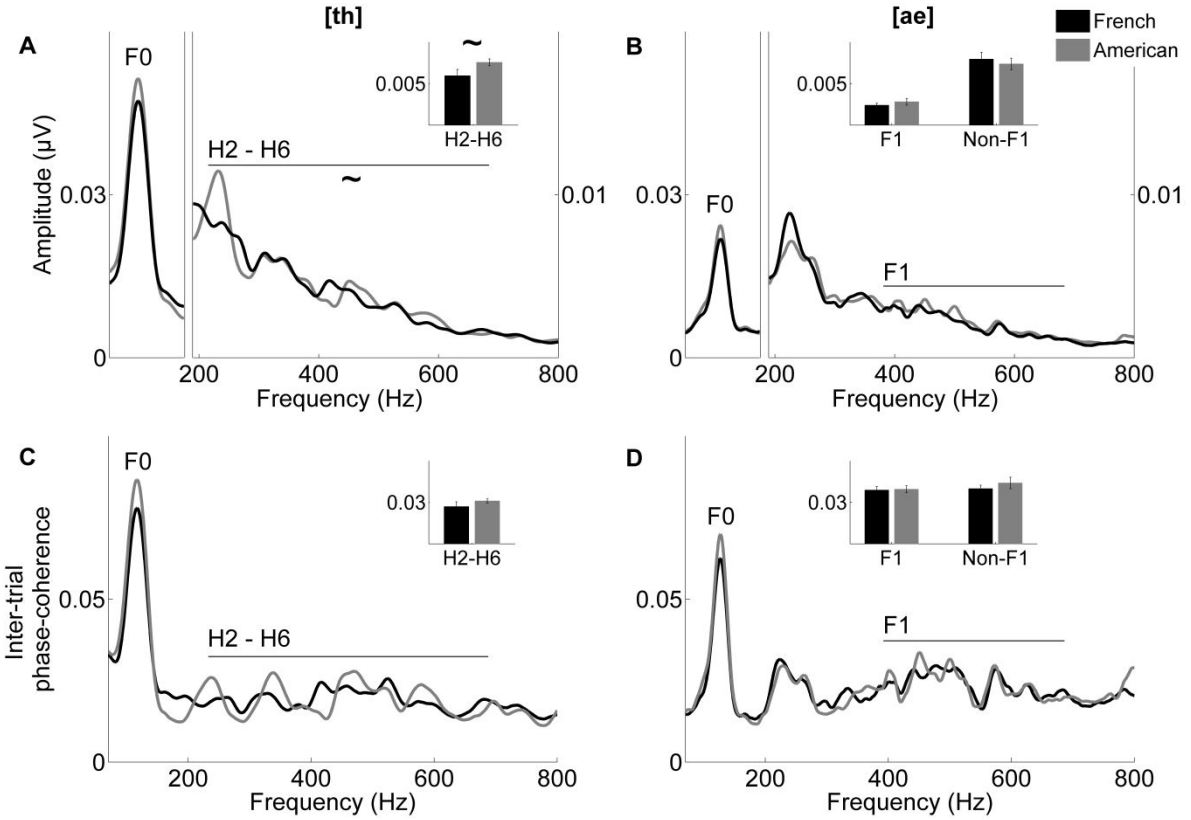


Figure 4. Spectral representation (top) and inter-trial phase-coherence (bottom) of the American consonant [th] (left panels) and vowel [ae] (right panels). ***A***,***C***, The results show that, while both groups have equivalent representations of the *F*0, American participants (gray) have a stronger spectral representations of the harmonics (*H*2-*H*6) for the American consonant [th]. ***B***,***D***, For the American vowel [ae], both groups show equivalent representations of the *F*0, and harmonics at formant and non-formant frequencies. Error bars represent ± 1SE.

3.2.3. Vowels

Analyses of inter-trial phase-coherence at the *F*0 did not reveal any significant effect (main effect of group: $F(1,42) = 0.587$, $p = 0.448$; group x stimulus interaction, $F(1,42) = 0.388$, $p = 0.537$).

Although inter-trial phase-coherence of the harmonics at formant frequencies (*F*1) was equivalent between groups (main effect of group: ($F(1,42) = 2.763$, $p = 0.104$), the robustness of stimuli neural representations at *F*1 frequencies strongly differed as function of the native language ($F(1,42) = 10.567$, $p = 0.002$). *Post-hoc* tests revealed that for the French vowel, French participants had more robust neural representations of the harmonics at *F*1 frequencies while for the American vowel there was no significant difference ([u]: $p = 0.002$, effect size = 0.016 μV ; [ae]: $p = 0.592$, **Figure 3 and 4 D**).

Consistent with the results found for spectral representation, analyses of inter-trial phase-coherence at non-formant frequencies did not reveal any significant effect (main effect of group: ($F(1,41) = 0.614$, $p = 0.438$; group x stimulus interaction: $F(1,41) = 1.924$, $p = 0.173$), showing that the advantage of French participants in encoding the French vowel [u] is specific to harmonics within *F*1 range.

4. Discussion

The goal of this study was to test the hypothesis that language experience shapes neural representations of speech sounds. More precisely, fast subcortical processing should be more accurate for speech sounds that belong to the native phonemic inventory of the listener. To test this hypothesis, we recorded speech-evoked subcortical responses in American and French native speakers using an American English and a French syllable. Importantly, both consonants and vowels of these two syllables - [thae] and [ru] - do not exist in the other language, which means that for each participant language experience should be maximal with one sound and minimal with the other sound. Taken together, results are consistent with the hypothesis that language experience shapes the neural processing of speech, and that this plasticity occurs as a function of what is behaviorally-relevant to a listener. Importantly, the legacy of this linguistic experience was apparent during a passive listening task, suggesting that language experience sculpts automatic response properties of auditory nuclei. We will focus our discussion on two main findings. Firstly, independently of the stimulus type, American participants showed a greater subcortical representation of the fundamental frequency (*F*0) compared to French participants. Secondly, participants exhibited more robust encoding and more precise spectral representations when listening to phonemes of their native language as compared to non-native language, especially at frequencies that are linguistically-relevant (formants).

4.1. Effect of language expertise on F0 representation

The subcortical representation of the fundamental frequency (*F*0) was enhanced in American, compared to French, participants across both stimuli. Since previous research has shown that musical expertise can have an effect on representation of the *F*0 (Bidelman et al., 2011, 2009; Wong et al., 2007), we carefully verified that American and French participants were matched in term of musicianship (see Materials and Methods). Then, the global enhancement in subcortical encoding of the *F*0 in the American group may be related to differences in language

experience, particularly the role that *F*0 cues play in American English as compared to French. Indeed, although English is not considered as a tonal language, it is nonetheless characterized by a large range of pitch dynamics akin to tonal languages (Duanmu, 2004). By contrast, French exhibits less variability in pitch (*F*0) at the utterance level and is classified as a non-tonal and non-stress language at the word level (Braun et al., 2014; Vaissière, 1991). Moreover, pitch carries segmental information in American English while it does not in French, which could explain the subcortical strengthening of this stimulus feature in American listeners independently of whether the stimulus belongs to their native language or not.

Interestingly, in a cross-language experiment, Braun et al. (2014) tested whether the complexity of the pitch system in the native language modulates encoding of non-native tonal speech sounds. In this experiment, native speakers of a non-stress language (French), after learning associations between pictures and non-words distinguished only by their tonal contrasts, exhibited more difficulties remembering these associations than native speakers of a stress language (German). These results suggest that languages without stress at the word level (e.g. French) are less sensitive to tonal contrasts (i.e. pitch variations) as compared to stress languages. Dupoux et al. (1997) have shown that French listeners exhibited significantly more difficulties than Spanish listeners to discriminate words that differ only by their accent. This reduced sensitivity to stress patterns, referred to as "stress deafness", could result from the difficulty for French listeners to represent stress at the phonological level (Dupoux et al., 2008). Since *F*0 variations are a major marker of stress, the findings of a reduced sensitivity to stress patterns go well along with our findings of a poorer representation of *F*0 in French participants. Nonetheless, these group differences on the *F*0 have to be interpreted with care because they were only visible on the spectral analyses of the whole stimuli and were absent in the phase coherence analyses of the whole stimuli and, most importantly, in the analyses separating consonants and vowels (the latter being the typical stress carrier in a word).

4.2. Specific neural enhancement of native-language sounds

Learning a language requires the ability to discriminate subtle differences in the phonemic inventory. This specialization for sounds of the native language may take place at the detriment of phonemes of other languages (Werker and Tees, 2002). Our results showed a significant interaction of language expertise and stimulus type on the responses to the whole stimuli when considering harmonics higher than the fundamental frequency (*H*2-*H*6). Analyses on the consonants and the vowels separately revealed two main findings. First, in both analyses of

spectral density (the amplitude of the signal in a spectral representation) and inter-trial phase-coherence (the stability of phase over trials), spectral representations of the consonants were sensitive to the native language of the listeners, but only at harmonics higher than the fundamental frequency (*H*2-*H*6). While differences are not striking when considering only one stimulus at a time, the analyses of the interaction between groups and stimuli is well designed to reveal subtle differences of language expertise on subcortical representations, independently of any possible effect driven by differences in the acoustic content of the stimuli. The results demonstrate a fine-grained subcortical representation of the consonant that belong to the phonemic inventory of the listener. Second, this sensitivity to the native language of the listeners was also present in the analyses of the vowels, but only when considering the harmonics at *F*1 frequency range. The fact that the interaction was visible at *F*1 frequencies, rather than on the fundamental frequency or other harmonics, is consistent with our prediction that language expertise shapes linguistically relevant neural responses.

Although the effect of the native language on stimulus processing was present for both stimuli, it was more robust with the stimulus [ru], with French participants showing a clear advantage processing the French vowel, while group differences for the stimulus [thae] were marginal. This could be due to the fact that French subjects are necessarily more familiar with English speech sounds than are American subjects with French speech sounds, since English is mandatory in France at school from the age of 12. Indeed, French subjects reported a better proficiency for English than American for French (See Material and Methods, **Table 1**). According to Song et al. (2008), subcortical plasticity can occur even in adults after short-term auditory training (see also Carcagno and Plack, 2011; Chandrasekaran et al., 2012). These results suggest that exposure to English language in French schools could induce subcortical plasticity effects that may reduce group differences for the English syllable. Put differently, the need for French listeners to distinguish French and English could accentuate the contrast in language-dependent processing.

Another explanation for the lack of difference between the French and American participants for the stimulus [thae] could be related to the fact this speech sound did not evoke distinct responses to the harmonics, thus potentially blurring the difference between groups. Working with natural tokens is in our opinion the best way to study language dependant effects on subcortical representations, but it comes with the price of having more complex and time-varying harmonic structure as compared to synthetic sounds. In the present experiment, the [thae] stimulus has more varying harmonics over time as compared to the [ru] stimulus, which results in more spread out peaks on the Fourier transform. Because of the phase-locking property

of subcortical structures, representations of these frequency components are therefore not stable over time resulting in a lower and more spread out peaks on the response FFT. The difference in pitch between the two stimuli could also explain a poorer harmonics representation for the [thae] stimulus. Compared to the [ru], the [thae] syllable has a higher fundamental frequency and harmonics (see **Table 2**). Given that neural phase-locking of subcortical structures becomes weaker with increasing frequency (Greenberg, 1980; Greenberg et al., 1987), higher and more varying harmonics for the syllable [thae] may induce less robust neural representations.

Overall, this study reveals that language-dependent effects do not result in a global enhancement of subcortical encoding of native speech sounds, but rather in a strengthening of specific stimulus features that are linguistically-relevant (in this case, the high frequency components that are critical to discriminate vowels). This is in line with previous studies on experience-dependent plasticity that emphasize that it occurs along dimensions that are behaviorally-relevant to an individual (Kraus and White-Schwoch, 2015). For instance, long-term musical practice strengthens subcortical encoding of specific stimulus features of music sounds (Lee et al., 2009) and speech sounds (Parbery-Clark et al., 2012) that may also have behavioral relevance. Even more relevant to the goal of this study, Krishnan et al., (2009, 2005) have shown that long-term language experience does not induce an overall enhancement of stimulus processing, but rather a specific strengthening of stimulus features that are linguistically-relevant (see also Strait et al., 2009). While in tonal languages, representation of pitch variations (i.e. *F*0) carrying lexical information is enhanced in native speakers, our results show that in non-tonal languages, spectral representations of formant frequencies (*F*1) that are relevant to the phonetic system are specifically strengthened in native speakers whereas other non- relevant spectral features are not affected by language experience.

The mechanisms behind these language-dependent effects likely relate both to bottom-up and top-down hypotheses. From birth and even before, the infant is exposed to sounds, some of which are more frequent and relevant than others. More precisely, native speech sounds are more prevalent in the infant environment than non-native speech sounds. Thus, development of speech perception is, at first, essentially a bottom-up mechanism whereby neural representations along the auditory pathway are shaped in response to the statistical distribution of stimuli in the external world. With growing exposure to the mother tongue, the infant develops high-level representations of linguistic units (phonemes) that become resistant to the inherent variability present in regular speech—that is, different utterances of the same syllable [ru] are categorized as

the same syllable. Once lexical representations become stable, one can make predictions about the upcoming words and syllables. These effects are likely to stem from a cascade of top-down processes that enhance linguistically-relevant, and prune non-relevant, information (Ahissar and Hochstein, 2004; Gilbert and Sigman, 2007; Kraus and White-Schwoch, 2015; Krishnan and Gandour, 2014; Tzounopoulos and Kraus, 2009). This means that the relevant auditory stimulus features (e.g. first and second formants) may be also anticipated.

One direct consequence of anticipating upcoming events is that attention can be more efficiently directed towards these specific stimulus features and this will in turn render more robust neural representations (Fritz et al., 2010). Interestingly, the impact of attention on neural activity is evident throughout the auditory system (Perrot et al., 2006). Our results are in line with the recent views of the auditory system, describing auditory perception as an active process that involves interaction between cognitive and sensory levels, where top-down processes take place (Kraus and White-Schwoch, 2015).

Overall, our results can be explained both in terms of bottom-up and top-down processes. On the one hand, long-term exposure to the mother tongue implies that a phoneme of the native language will be heard a huge amount of times. On the other hand, the existence in each language of a limited phonemic inventory and lexicon allows one to predict auditory events and focus attention to relevant stimulus features. These two explanations may account for more accurate subcortical representations of native language stimulus features.

Future studies are needed to explore language-dependent effects on subcortical processing by comparing possibly more distant languages in terms of phonetic repertoire. This may maximize the effects that we showed here for two Indo-European languages. It will also be important to recruit participants that have a minimal or absolutely no knowledge of the other language, which was not the case here for French participants. Finally, while natural speech sounds possibly maximize the language-dependent effects, it is not possible to carefully control them in terms of their acoustic features. Thus, a possible development of our work could also be the use of resynthesized natural tokens, carefully controlling the spectro-temporal variability.

Acknowledgements

This work was supported by the Brain and Language Research Institute (BLRI, ANR-11-LABX-0036), CNRS & Université d'Aix-Marseille.

References

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464. doi:10.1016/j.tics.2004.08.011

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47. doi:10.1016/j.heares.2008.08.004

Berens, P., 2009. CircStat: A MATLAB Toolbox for Circular Statistics. J. Stat. Softw. 31, 1–21.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77, 1–10. doi:10.1016/j.bandc.2011.07.006

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2009. Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. J. Cogn. Neurosci. 23, 425–434. doi:10.1162/jocn.2009.21362

Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer.

Braun, B., Galts, T., Kabak, B., 2014. Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. Second Lang. Res. 30, 323–350.

Carcagno, S., Plack, C.J., 2011. Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. J. Assoc. Res. Otolaryngol. 12, 89–100. doi:10.1007/s10162-010-0236-1

Chandrasekaran, B., Kraus, N., Wong, P.C.M., 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107, 1325–1336. doi:10.1152/jn.00923.2011

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1, 351–353.

Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. J. Cogn. Neurosci. 12, 635–647. doi:10.1162/089892900562390

Duanmu, S., 2004. Tone and non-tone languages: An alternative to language typology and parameters. Lang. Linguist. 5, 891–923.

Dupoux, E., Pallier, C., Sebastian, N., Mehler, J., 1997. A destressing "deafness" in French? J. Mem. Lang. 36, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., Peperkamp, S., 2008. Persistent stress "deafness": The case of French learners of Spanish. Cognition 106, 682–706. doi:10.1016/j.cognition.2007.04.001

Fourtassi and Dupoux, 2014. A Rudimentary Lexicon and Semantics help Bootstrap Phoneme Acquisition. Proc. Eighteenth Conf. Comput. Lang. Learn. 191–200.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., Shamma, S.A., 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. 13, 1011–1019. doi:10.1038/nn.2598

Gao, E., Suga, N., 2000. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. Proc. Natl. Acad. Sci. 97, 8081–8086. doi:10.1073/pnas.97.14.8081

Gilbert, C.D., Sigman, M., 2007. Brain States: Top-Down Influences in Sensory Processing. Neuron 54, 677–696. doi:10.1016/j.neuron.2007.05.019

Greenberg, S., 1980. WPP, No. 52: Temporal Neural Coding of Pitch and Vowel Quality. Dep. Linguist. UCLA.

Greenberg, S., Marsh, J.T., Brown, W.S., Smith, J.C., 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear. Res. 25, 91–114. doi:10.1016/0378-5955(87)90083-9

Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol 6, e138. doi:10.1371/journal.pbio.0060138
Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97, 3099–3111. doi:10.1121/1.411872
Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. 'ich, Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, B47–B57. doi:10.1016/S0010-0277(02)00198-1
Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., Lin, C.-D., 2011. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear. 32, 699–707.
Kral, A., Eggermont, J.J., 2007. What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Res. Rev. 56, 259–269. doi:10.1016/j.brainresrev.2007.07.021
Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599–605. doi:10.1038/nrn2882
Kraus, N., White-Schwoch, T., 2015. Unraveling the Biology of Auditory Learning: A Cognitive–Sensorimotor–Reward Framework. Trends Cogn. Sci. 19, 642–654. doi:10.1016/j.tics.2015.08.017
Krishnan, A., Gandour, J.T., 2014. Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: Electrophysiological evidence. Acoust. Aust. Aust. Acoust. Soc. 42, 166–178.
Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009. Experience-dependent neural representation of dynamic pitch in the brainstem: NeuroReport 20, 408–413. doi:10.1097/WNR.0b013e3283263000
Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Cogn. Brain Res. 25, 161–168. doi:10.1016/j.cogbrainres.2005.05.004
Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. Nat. Rev. Neurosci. 5, 831–843. doi:10.1038/nrn1533
Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. 9, F13–F21. doi:10.1111/j.1467-7687.2006.00468.x
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.
Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective Subcortical Enhancement of Musical Intervals in Musicians. J. Neurosci. 29, 5832–5840. doi:10.1523/JNEUROSCI.6133-08.2009
Lew-Williams, C., Saffran, J.R., 2012. All words are not created equal: Expectations about word length guide infant statistical learning. Cognition 122, 241–246. doi:10.1016/j.cognition.2011.10.007
Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111. doi:10.1016/S0010-0277(01)00157-3
Mehler, J., Dupoux, E., Pallier, C., Dehaene-Lambertz, G., 1994. Cross-linguistic approaches to speech processing. Curr. Opin. Neurobiol. 4, 171–176. doi:10.1016/0959-4388(94)90068-X
Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr. Clin. Neurophysiol. 35, 665–667. doi:10.1016/0013-4694(73)90223-X
Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic

Invariance in the Response to Intelligible Speech. Cereb. Cortex 20, 2486–2495. doi:10.1093/cercor/bhp318
Ortiz-Mantilla, S., Hämäläinen, J.A., Musacchia, G., Benasich, A.A., 2013. Enhancement of Gamma Oscillations Indicates Preferential Processing of Native over Foreign Phonemic Contrasts in Infants. J. Neurosci. 33, 18746–18754. doi:10.1523/JNEUROSCI.3260-13.2013
Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4. doi:10.3389/fnagi.2012.00030
Peelle, J.E., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front. Hum. Neurosci. doi:10.3389/fnhum.2010.00051
Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., Mauguière, F., Collet, L., 2006. Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cereb. Cortex 16, 941–948. doi:10.1093/cercor/bhj035
Raizada, R.D.S., Tsao, F.-M., Liu, H.-M., Kuhl, P.K., 2010. Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences. Cereb. Cortex 20, 1–12. doi:10.1093/cercor/bhp076
Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005. Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. Dev. Sci. 8, 162–172. doi:10.1111/j.1467-7687.2005.00403.x
Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302–324. doi:10.1097/AUD.0b013e3181cdb272
Song, J.H., Skoe, E., Wong, P.C.M., Kraus, N., 2008. Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training. J. Cogn. Neurosci. 20, 1892–1902. doi:10.1162/jocn.2008.20131
Strait, D.L., Kraus, N., Skoe, E., Ashley, R., 2009. Musical experience and neural efficiency - effects of training on subcortical processing of vocal expressions of emotion. Eur. J. Neurosci. 29, 661–668. doi:10.1111/j.1460-9568.2009.06617.x
Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A 194, 169–183. doi:10.1007/s00359-007-0274-2
Tierney, A., Kraus, N., 2013. The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. J. Neurosci. 33, 14981–14988. doi:10.1523/JNEUROSCI.0612-13.2013
Tzounopoulos, T., Kraus, N., 2009. Learning to Encode Timing: Mechanisms of Plasticity in the Auditory Brainstem. Neuron 62, 463–469. doi:10.1016/j.neuron.2009.05.002
Vaissière, J., 1991. Rhythm, accentuation and final lengthening in French 108–120.
Werker, J.F., Tees, R.C., 2002. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Dev., 25th Anniversary Special Issue 25, 121–133. doi:10.1016/S0163-6383(02)00093-0
Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422. doi:10.1038/nn1872
Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y. 'ichi, 2005. Effects of language experience: Neural commitment to language-specific auditory patterns. NeuroImage 26, 703–720. doi:10.1016/j.neuroimage.2005.02.040

11. Effet de transfert entre musique et langue étrangère

L'étude intitulée : « *Musical training enhances the automatic neural processing of foreign language* » est en voie de soumission au journal à comité de lecture *The Journal of Neuroscience*. Cette étude est détaillée ci-après.

Musical training enhances the automatic neural processing of foreign language

Bastien Intartaglia[a,†], Travis White-Schwoch[b], Nina Kraus[b,c,d], Daniele Schön[a]

[a] *Aix Marseille Univ, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France*

[b] *Auditory Neuroscience Laboratory and Department of Communication Sciences, Northwestern University, Evanston, Illinois, United States of America*

[c] *Department of Neurobiology & Physiology, Northwestern University, Evanston, Illinois, United States of America*

[d] *Department of Otolaryngology, Northwestern University, Chicago, Illinois, United States of America*

[†]Corresponding author

Bastien Intartaglia

Institut de Neurosciences des Systèmes

Inserm UMR1106

Aix-Marseille Université

Faculté de Médecine, 27, Boulevard Jean Moulin

13005 Marseille, France

Email: bastien.intartaglia@etu.univ-amu.fr

Phone: 334 91 38 46 33

Abstract

Music and speech are both highly complex acoustic signals that share several acoustic properties, as well as common neural resources. The abundant literature about the transfer effects between music and language mainly focused on investigating the advantages of musical practice on native language or non-native tonal language processing. The aim of the present study is to assess whether musical training would also benefit the processing of non-tonal foreign languages. We hypothesized that musical practice would enhance subcortical encoding of non-tonal syllables, and that this plasticity should occur along linguistically-relevant acoustic features such as the formant frequencies. We recorded neural responses to an English syllable in French musicians and nonmusicians, and American participants. Results revealed that, while French nonmusicians exhibited significantly lower amplitudes and more variable neural encoding compared to American participants, these differences disappeared when comparing American participants to French musicians. Importantly, these differences were visible for both the encoding of the fundamental frequency and the formant frequencies. Thus, the present results go beyond the previous literature on pitch processing in non-native tonal languages by showing that musical experience strengthens the neural processing of non-tonal foreign languages along phonetic dimensions that are specifically linguistically-relevant (formants), thus pointing to a selective enhancement rather than an overall gain effect.

Keywords

Speech ABR, plasticity, musical expertise, linguistic expertise, cross-domain effects

1. Introduction

Both music and speech are universals of the human societies and require the perception, manipulation and production of complex sound sequences. These sequences are hierarchically organized (i.e. syllables, word, sentences in speech and notes, beats and phrases in music) and their decoding requires, for instance, an efficient representation of rapidly evolving sound spectro-temporal dynamics, selecting the relevant information, building temporary structures taking into account syntactic rules and pre-existing lexical traces and many other cognitive functions. It is thus not surprising that music and speech processing share common neural resources (Maess et al., 2001; Patel, 2011; Schön et al., 2010). The acoustic and structural similarities as well as the shared neural networks between speech and music suggest the idea that cognitive and perceptual abilities can transfer from one domain to the other via the reorganization of common neural circuits (Patel, 2011). This hypothesis has been verified by showing that musical practice not only improves music sound processing (Fujioka et al., 2004; Koelsch et al., 1999; Pantev et al., 1998; Trainor et al., 1999), but also enhances several levels of speech processing, such as the perception of prosody (Schön et al., 2004), consonant contrasts (Chobert et al., 2011), speech segmentation (Francois and Schön, 2011) and syntactic processing (Jentschke and Koelsch, 2009). Interestingly, these findings extend to the subcortical level, showing an enhancement of the neural representations of the pitch, timbre and timing of speech sounds by musical practice (Kraus et al., 2009, Kraus & Chandrasekaran, 2010). Subcortical responses to speech are more robust to noise in musicians, and this neural advantage has been correlated with better abilities to perceive speech in noisy background (Parbery-Clark et al., 2009). Overall, these studies suggest that the perceptual advantages induced by intensive music training rely on an enhancement of the neural coding of sounds, both cortical and subcortical structures and extending to speech sounds.

Interestingly, musical experience has also been associated with better perception and production of sounds in foreign languages (Chobert and Besson, 2013; Marie et al., 2010; Slevc and Miyake, 2006). At the cortical level, the slight pitch variations of both musical (i.e. harmonic sounds) and non-native speech syllables (i.e. Mandarin tones) evokes significantly larger MMN responses in non-native musicians as compared to non-native nonmusicians (Marie et al., 2010; Martínez-Montes et al., 2013). At the subcortical level, Wong et al. (2007) have shown that American musicians have more faithful neural representation of the rapid variations of the pitch of Mandarin tones as compared to American non-musicians. Moreover, this advantage was correlated with the amount of musical

expertise. The works of Krishnan and colleagues on the cross-domain effects between music and language experience revealed that long-term experience in pitch processing (e.g. in musicians and tonal language speakers) enhances the encoding of pitch regardless of the domain of expertise (Bidelman et al., 2009; Krishnan et al., 2012). These findings suggest that musical practice can compensate for the lack of language expertise when it comes to perceive foreign tonal speech sounds by sculpting through years of training the automatic neural coding of pitch. However, pitch is only one relevant acoustic dimension of speech, and musicians' advantage in foreign language perception are not restricted to tonal languages (Musacchia et al., 2007; A. Parbery-Clark et al., 2012; Slevc and Miyake, 2006). This raises the question whether musical expertise could compensate for the lack of language expertise in non-tonal languages by improving the processing of other relevant speech features, such as the formant frequencies that allow the listener to discriminate the vowels.

To this aim, we decided to compare subcortical responses to the English syllable [thae] between three groups of non-tonal language speakers: Americans non-musicians, French non-musicians and French musicians. Since French are necessarily familiar with English sounds, we chose a stimulus that is distant from the phonemic inventory of the French listeners, the English syllable [thae] for which both consonant and vowel do not exist in French. In a previous study we compared envelop-following responses to the syllable [thae] in American and French non-musicians (Intartaglia et al., 2016). Results showed that, compared to French, American non-musicians have more robust subcortical representations of the English phoneme [thae]. Thus, we hypothesized that American participants would confirm an advantage in neural encoding of linguistically-relevant features as compared to non-native speakers without musical practice, but that this language difference would be compensated by musical practice in non-native musicians.

2. Material and method

2.1. Participants

Forty-three (24 females and 19 males) young adults ranging in age from 18 to 30 years, participated in the study. In order to investigate the effect of musical practice on foreign language processing, the participants were separated in three groups depending on their native language and their musicianship. 14 were native speakers of American English and 29 were native speakers of French. American participants had received less than 3 years of musical practice and were categorized as nonmusicians while French participants were divided in two groups according to their musicianship. Eleven were musicians having started musical training before the age of 14 (mean = 7.6, SD = 0.65) and had consistently practiced for at least 8 years (mean = 14.8, SD = 1.84). The remaining 18 participants had received less than 3 years of musical practice and were categorized as nonmusicians. 9 of these nonmusicians had received no musical training at all. The French native nonmusicians are henceforth named non-native nonmusicians; French native musicians, henceforth named non-native musicians, and American native nonmusicians, henceforth named native nonmusicians. American participants were recruited at Northwestern University (Chicago, USA) and French participants were recruited at Aix-Marseille University (Marseille, France). Inclusion criteria were a high-school level of education, click-evoked brainstem response latencies within lab-internal normal limits (5.41-5.97 ms; 100-µs click stimulus presented at 80 dB SPL). The three groups did not differ in term of age (p = 0.863), sex (p = 0.722), education (p = 0.327) or hearing thresholds (p = 0.767). Participants gave their informed consent and were paid for their participation. The study was approved by the local ethics committee.

2.2. Stimulus

The stimulus used was the natural English syllable [ðæ], recorded in an anechoic chamber by an American English male speaker (see waveform **Figure 1**). This syllable was chosen because it is an « illegal » speech sound in French, which means that both the consonant and the vowel do not exist in French (i.e., [ð] and [æ] do not exist in French). This choice should maximize the differences between the two populations (French and American English) and should consequently maximize the

expected effect of language experience on neural responses. A fade-in and fade-out were applied to the initial and final 5 ms windows of the stimulus waveform (see waveform in **Figure 1**).

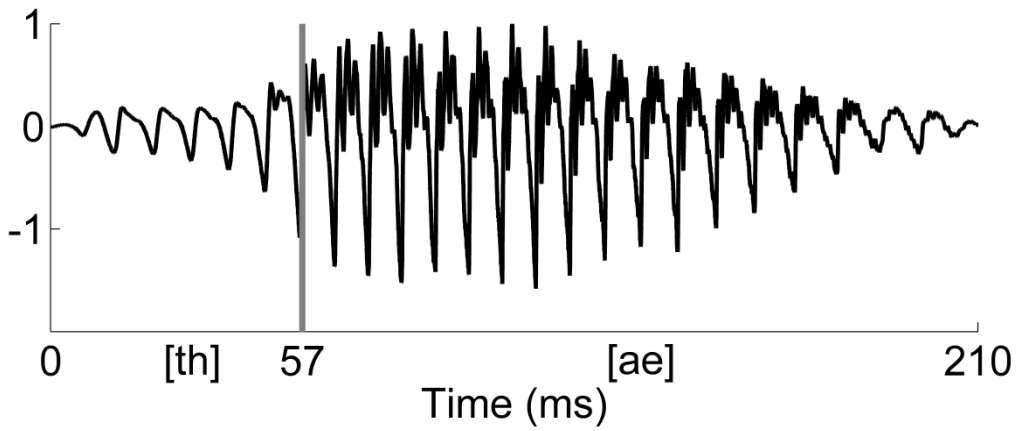


Figure 1. Waveform of the stimulus (normalized amplitude). The vertical gray line indicates the boundary between the consonant and vowel, as established according to the spectral changes by an experienced phonetician.

Frequencies of interest included the fundamental frequency (*F*0) and the second to sixth harmonics (*H*2-*H*6). These latter were chosen because they encompass the first formant range (*F*1) which is an acoustically relevant cue in discriminating vowels (see **Table 1** for more details). Since the stimulus was natural speech sounds, the *F*0 and its subsequent harmonics *H*2-*H*6 vary slightly over time and were thus estimated by computing a fast Fourier transform (FFT) on two time regions (consonant and vowel, see FFT **Figure 2**). These time regions were established according to the spectral changes by an experienced phonetician. Importantly the ranges of the first two vowel formants ([ae]: *F*1 = 420-705 Hz, *F*2 = 1408-1556 Hz, from Praat: (Boersma and Weenink, 2001)) showed that *F*2 range exceeded the phase-locking limit of the auditory brainstem which is around 1.5 kHz and was thus not taken into consideration in further analyses (Aiken and Picton, 2008; Moushegian et al., 1973).

FOI	Consonant	Vowel
*F*0	118	125
*H*2	235	258
*H*3	355	386
*H*4	472	528
*H*5	590	665
*H*6	711	788
*F*1 range	N/A	420-705

Table 1. Frequencies of interest (FOI) for each time region of the stimulus. Values represent the frequency peaks of the fundamental frequency (*F*0), its second through sixth subsequent harmonics (*H*2-*H*6) and frequency range of the first formant (*F*1 range) for each time region (in Hz). Note that because this is a natural stimulus, FOI vary slightly over time and harmonics are not exactly whole integers of the *F*0.

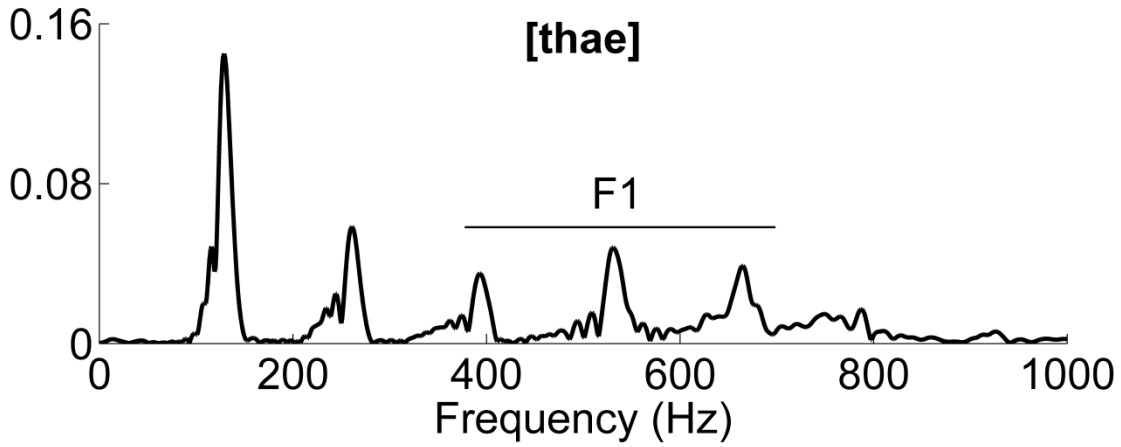


Figure 2. Fast Fourier transforms computed on the whole stimulus (normalized amplitude). The horizontal line indicates the range of the first formant.

2.3. Electrophysiological recordings

The stimulus was presented monaurally to the right ear at 80 dB SPL at a rate of 3.3 Hz through magnetically shielded insert earphones (ER-3A, Etymotic Research) using the stimulus presentation software Microvitae (μV-ABR, Microvitae Technologies) while participants sat in a comfortable reclining chair in an electrically-shielded, sound-attenuated room. Brain responses were collected at

30 kHz sampling rate using Microvitae recording system (μV-ABR, Microvitae Technologies) with three Ag-AgCl scalp electrodes in a vertical montage (Cz active, forehead ground, and right earlobe reference). Electrode impedances were kept <5 kΩ. A total of six-thousand sweeps were collected (three thousands for each polarity). The stimulus polarities were alternated across trials in order to minimize the contribution of stimulus artifact and cochlear microphonic (Aiken and Picton, 2008; Skoe and Kraus, 2010). One of the authors (BI) was in charge of data acquisition in both countries using the same portable EEG system. This prevents the possibility of having a bias due to different experimental setups, participant preparation, and instruction.

2.4. Data analysis

All analyses were performed using custom MATLAB scripts (MathWorks). First, electrophysiological recordings were bandpass filtered from 70 to 2000 Hz (12 dB/octave roll-off) using a Butterworth filter. Then, sweeps with activity exceeding ± 30 μV were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period (-30 to 0 ms). Neural responses were then averaged over a -30 to 230 ms window. The signal-to-noise ratio (SNR) was computed using the quotient of response root mean square (RMS) amplitude and pre-stimulus baseline RMS amplitude (Skoe and Kraus, 2010). If the SNR was less or equal to 1.4, the participant was excluded. This resulted in excluding two American and one French participants

2.5. Spectral amplitude

The fundamental frequency (F0) is the primary acoustic correlate of the pitch which is used, for instance, to identify the gender of the speaker, or to convey surpasegmental or segmental information in stressed and tonal languages respectively, while formant frequencies are used to discriminate phonemes, such as consonant and vowels, in non-tonal languages. As such, fundamental frequency and formant structure can be considered independent features in speech acoustics. Therefore, F0 and F1-related harmonics were analyzed separately, using two different procedures. Spectral representation of the F0 were quantified by computing Fast Fourier transform on the individual averages computed by adding the responses to the two polarities of the stimulus which accentuates the lower-frequency components (Aiken and Picton, 2008). In order to explore the neural coding at formant frequencies, we used the subtracting method, that is FFT were performed

on individual averages computed by subtracting the responses to the two polarities which maximizes the spectral response to the harmonics and formant frequencies (Krishnan, 2002; Skoe and Kraus, 2010).

For both procedures, FFT were performed on two time regions of the response (consonant and vowel). These time regions were defined on the basis of the stimulus by a phonetician also taking into account a 10 ms neural delay in the response: consonant (20-67 ms) and vowel (67-220 ms). For each time region, the maximum spectral amplitudes of the fundamental frequency (*F*0) and its second to sixth harmonics (*H*2-*H*6) were extracted in a bandwidth of 20 Hz surrounding the peak in the stimulus fast Fourier transform (e.g. for a peak at 118 Hz, values were extracted between 98 and 128 Hz, (Parbery-Clark et al., 2012). For the consonant, the five values extracted from *H*2 to *H*6 were then averaged to form a global measure of harmonics' representation. For the vowel, since our hypothesis was that differences on neural responses would be maximal over frequencies that are relevant to discriminate vowels, harmonics were separated in two groups. Harmonics falling within *F*1 range ± 20 Hz were averaged to form a global measure of *F*1 spectral representation, while harmonics falling outside *F*1 range were averaged to form a global measure of non-formant spectral representation. This included the third to fifth harmonics (*H*3-*H*5), and the remaining harmonics were considered as non-formant frequencies (*H*2, *H*6).

2.6. Statistical analyses

All statistical analyses were performed using Statistica Version 7.1 (StatsSoft, Tulsa, OK). To investigate the effect of language and musical experience, an analysis of variance (ANOVA) was performed using group as a between subjects factor (non-native nonmusicians vs. non-native musicians vs. native nonmusicians). In both analyses the dependent variables were power spectral density as a measure of spectral representation and inter-trial phase-coherence as a measure of neural consistency across trials. *Post-hoc* tests were used when appropriate (Fisher LSD).

3. Results

3.1. Spectral representation by adding polarities

Fundamental frequency (F0). Neural responses were divided into two time regions, corresponding to the response to the consonant and the response to the vowel. Spectral analysis of the time region corresponding to the consonant revealed a poorer F0 representation for non-native nonmusicians compared to the two other groups (**Figure 3A,C)**. However, the main effect of group was marginal ($F(2,37) = 2.46$, $p = 0.10$)**.** By contrast, analysis of time region corresponding to the vowel showed significant differences between the groups ($F(2,37) = 3.84$, $p = 0.030$, **Figure 3B,D**). Indeed, while non-native nonmusicians have a poorer representation of the F0 compared to both native nonmusicians and non-native musicians ($p = 0.042$ and $p = 0.015$, respectively), non-native musicians have an equivalent representation to that of native speakers ($p = 0.570$).

Harmonics (H2-H6). Analysis of time region corresponding to the consonant did not show any group difference on the spectral encoding of the harmonics ($F(2,37) = 0.34$, $p = 0.713$, **Figure 3A,C**).

Formant (F1) and non-formant (Non-F1) frequencies. Analysis of time region corresponding to the vowel did not show any group difference on the spectral encoding of the formant ($F(2,37) = 1.10$, $p = 0.344$) and non-formant frequencies ($F(2,37) = 1.30$, $p = 0.284$, **Figure 3B,D**).

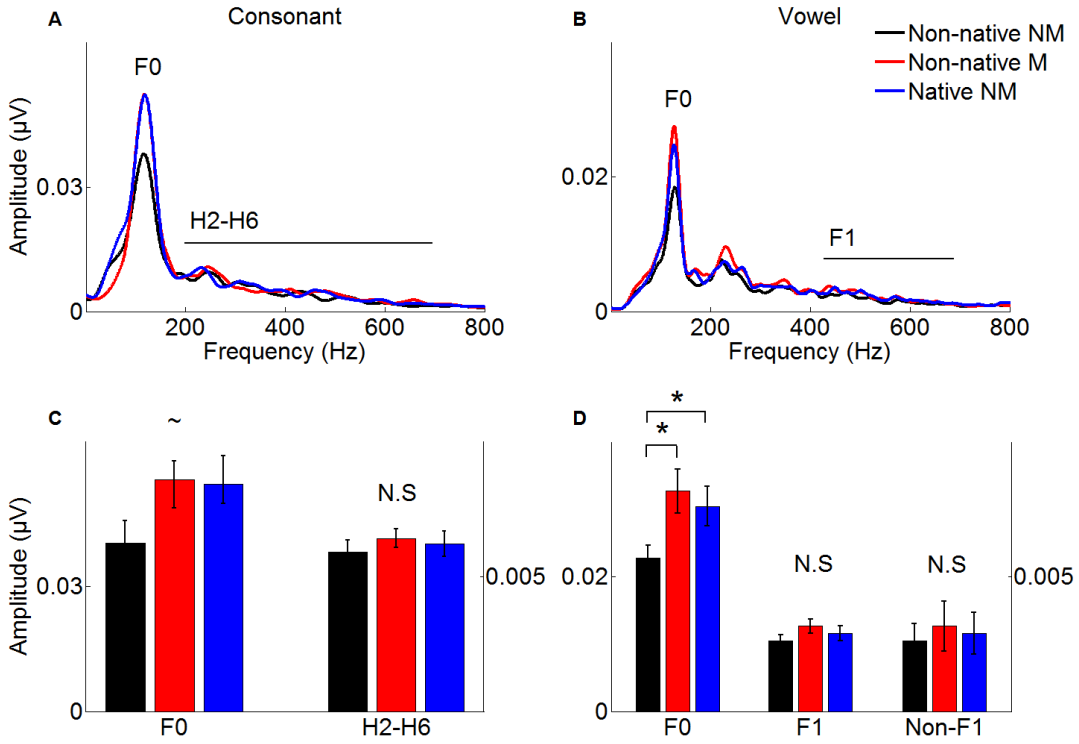


Figure 3. Spectral representation by adding polarities. **Top** : Fast Fourier transform of the neural response to the consonant (left panel) and the vowel (right panel) for non-native nonmusicians (black), non-native musicians (red) and native nonmusicians (blue). **Bottom** : Bar graphs corresponding to the fundamental frequency (F0) and its subsequent harmonics (H2-H6) for the consonant (left), and to the F0, the first formant (F1) and the non-formant frequencies (Non-F1) for the vowel (right). Left y axes correspond to the F0 and right y axes correspond to the harmonics (H2-H6), F1 and Non-F1 frequencies. Overall, language dependent differences are compensated by musical expertise * $p < 0.05$; error bars represent ± 1 standard error.

3.2. Spectral representation by subtracting polarities

Fundamental frequency (F0). Spectral analysis of the time region corresponding to the consonant and the vowel did not reveal any group difference in the encoding of the F0 (consonant : $F(2,37) = 2.10$, $p = 0.137$, **Figure 4A,C ;** vowel **:** $F(2,37) = 1.19$, $p = 0.315$, **Figure 4B,D**).

Harmonics (H2-H6). Analysis of time region corresponding to the consonant did not show any group difference on the spectral encoding of the harmonics ($F(2,37) = 1.98$, $p = 0.152$, **Figure 4A,C**).

Formant (F1) and non-formant (Non-F1) frequencies. Analysis of time region corresponding to the vowel showed significant group differences in the spectral encoding of the formant frequencies ($F(1,37) = 4.63$, $p = 0.016$, **Figure 4 B,D**). Indeed, while non-native nonmusicians exhibit poorer representation of the F1 compared to the native speakers and non-native musicians ($p = 0.006$ and $p = 0.042$, respectively), non-native musicians have an equivalent representation to that of native speakers ($p = 0.546$). A similar scenario is also visible for the non-formant frequencies although the main effect of group does not reach significance ($F(2,37) = 2.18$, $p = 0.127$).

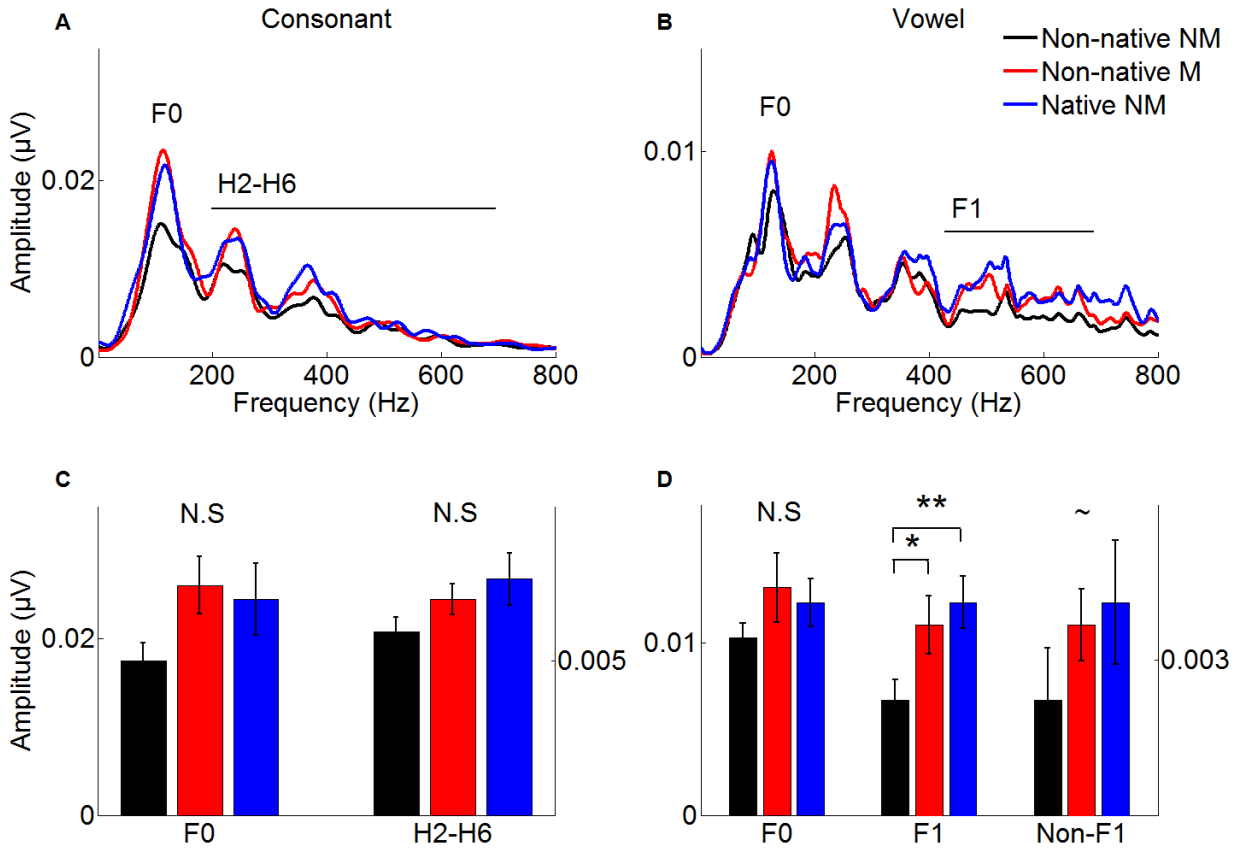


Figure 4. Spectral representation by subtracting polarities. **Top** : Fast Fourier transform of the neural response to the consonant (left panel) and the vowel (right panel) for non-native nonmusicians (black), non-native musicians (red) and native nonmusicians (blue). **Bottom** : Bar graphs corresponding to the fundamental frequency (F0) and its subsequent harmonics (H2-H6) for the consonant (left), and to the F0, the first formant (F1) and the non-formant frequencies (Non-F1) for the vowel (right). Left y axes correspond to the F0 and right y axes correspond to the harmonics (H2-H6), F1 and Non-F1 frequencies. Overall, language dependent differences are compensated by musical expertise * $p < 0.05$; ** $p < 0.01$; error bars represent ± 1 standard error.

4. Discussion

This study investigated the effect of both language and musical expertise on the neural encoding of non-tonal speech sounds. The results demonstrate that (1) subcortical encoding of pitch is sensitive to both musical and language experience; (2) neural processing of formant frequencies are dependent on language experience; (3) the lack of language experience is compensated by musical expertise.

Pitch processing and its malleability with sensory experience

Several experiments investigated how the neural processing of speech sounds are influenced either by linguistic or musical experience. These studies revealed that pitch processing is enhanced in speakers of tonal languages for whom pitch variations convey lexically-relevant information such as the phonemic categories (Krishnan et al., 2012, 2009, 2005) and in musicians who are trained to perceive subtle variations of the pitch (Besson et al., 2007; Chobert and Besson, 2013; Kraus et al., 2009). Moreover, cross-domain effects have been observed between speech and music experiences showing that musicians have more faithful neural encoding of the fine pitch variations of tonal speech sounds as compared to nonmusicians (Wong et al., 2007). Conversely, Chinese speakers have stronger representation of the pitch of musical sequences as compared to nonmusicians (Bidelman et al., 2011). These findings clearly suggest that pitch processing is dependent on the sensory experience with pitch that the listener acquired across his lifespan, and that this experience-dependent plasticity can benefit to other domains.

However, the difference in experience with pitch is huge when comparing for instance, Chinese and English nonmusicians, or experts musicians and nonmusicians. Thus, it was not clear whether these experience-dependent effects could hold comparing more similar languages, that is two non-tonal languages, in which the differences of experience with pitch are more subtle. In the present study, we investigated this question by comparing American and French speakers listening to an English syllable. Since pitch is an important acoustic cue in English, and that formant frequencies are relevant to discriminate vowels in non-tonal languages, we predicted that the language differences would occur specifically at the fundamental frequency and at the harmonics near, as compared to distant, to the first formant.

The results confirmed our previous findings that French non-musicians have less robust subcortical encoding of the fundamental frequency as compared to American speakers (Intartaglia et al., 2016), and most importantly showed that this language difference is compensated by musical expertise in non-native French musicians. French is a language with very little variations of the pitch, often classified as a non-stress language (Vaissière, 1991), and French are not sensitive to stress contrasts (Braun et al., 2014). In contrast, in English pitch carries segmental information and this language is characterized by a larger range of pitch dynamics (Duanmu, 2004). This difference in experience with pitch between the two populations, and more specifically, the fact that pitch does not have the same relevance in French and English, could explain the poorer neural representation of the fundamental frequency (i.e. the primary acoustic correlate of the pitch or stress patterns) found in French speakers compared to American speakers.

Interestingly, Bidelman and colleagues (2011) have shown that in Chinese speakers, who use intensively pitch variations to convey lexical meaning, subcortical representation of the F0 is significantly more robust than in American speakers (Bidelman et al., 2011). In addition, they have shown that the difference in language experience was compensated by musical practice, with American musicians showing equivalent representation of the F0 compared to Chinese speakers. Our results show similar effects of musical experience on pitch processing. Neural representation of the F0 was stronger in French musicians compared to nonmusicians, and equivalent to American speakers. These results strongly support the idea that musical expertise compensate for the lack of language experience, and that encoding of pitch is dependent on the sensory experience of the listener.

Specific enhancement of linguistically-relevant acoustic cues by musical expertise

In our previous study (Intartaglia et al., 2016), we have shown that language-dependent effects in subcortical structures are not restricted to pitch processing. French participants had more robust neural encoding of formant frequencies of their native language compared to American participants. However, American participants did not exhibit advantages in encoding formant frequencies of their native language. We interpreted this interaction of the language-dependent effect in terms of an asymmetry of second language experience. Indeed, while none of the English participants knew French, all French participants had some knowledge of English. However, another

possibility is that the analyses of the neural responses by adding the responses to the two polarities of the stimulus accentuated the lower-frequency components at the expenses of high frequency components (Aiken and Picton, 2008). Indeed, although using adding polarities technique was successful in showing distinct representations of the harmonics of the French stimulus, this was not the case for the English stimulus, which did not elicit clear neural responses to the harmonics. Therefore, in the present study we decided to analyze the responses using the two techniques of averaging.

Subtracting the responses to the two stimulus polarities, in addition to maximize neural responses to the harmonics, revealed stronger neural representations of the first formant in American participants compared to French nonmusicians. This extends the results of our previous study by showing that language-dependent effects are visible also when using stimuli for which the difference in familiarity between the groups is the smallest. Indeed also in the present study, French participants were more familiar with English than Americans were with French. Interestingly, spectral representations of formant frequencies (F1) that are relevant to the phonemic system are more strongly enhanced in native speakers whereas other non-relevant spectral features (non-F1 harmonics) are less affected by language experience. This is consistent with the hypothesis that language-dependent plasticity occurs as a function of what is functionally relevant to a listener. For instance, compared to English, Chinese speakers show stronger spectral encoding of the pitch on those sections that are characterized by rapid changes that are lexically-relevant in Chinese Mandarin (Bidelman et al., 2009). This specific neural enhancements has also been shown in musicians who are, for instance, particularly sensitive to the leading melodic voice in a chord (Lee et al., 2009) or the pitch frequencies that correspond to a note along the diatonic musical scale (Bidelman et al., 2009).

The present results reveal another major finding: the advantage of native language disappears when comparing natives to non-native musicians. In other words, language-dependent effects are compensated by musical practice. This is, to our knowledge, the first evidence that musical experience heightens the subcortical processing of formant frequencies in non-tonal foreign languages. Previous work has mainly focused on transfer effects between musical expertise and native language or tonal foreign language perception. For instance, musicians have more precise and distinct neural representations of native syllables that differ only on their second formant trajectories (Parbery-Clark et al., 2012; Strait et al., 2013), and this possibly allows for enhanced automatic processing of different consonant features in their native language (Chobert et al., 2011). Musical

practice has also been associated with better perceptual abilities to categorize native vowels along a speech continuum, and these behavioral advantages coincided with more robust subcortical encoding of salient speech cues, such as voice pitch and formant frequencies (Bidelman et al., 2014). Turning to foreign languages, several experiments examined the influence of musical expertise on the processing of tonal languages (Chobert and Besson, 2013). These studies revealed that musicians have better perceptual abilities (Alexander et al., 2005; Delogu et al., 2010; Lee and Hung, 2008), as well as more robust neural processing of non-native lexical tones (Marie et al., 2010; Wong et al., 2007). Now, while these results are not trivial, one may argue that the musicians advantage in processing lexical tones may be related to a more accurate pitch perception and/or memory for pitch, because the F0 trajectories are both relevant in tonal languages and in music. However, to our knowledge, the effects of musical practice on subcortical processing of non-tonal foreign languages have not been investigated yet. The present results go beyond the previous literature on pitch processing by showing that musical experience strengthens the neural processing of non-tonal foreign languages along phonetic dimensions that are specifically linguistically-relevant (formants), thus pointing to a selective enhancement rather than an overall gain effect..

These experience-dependent plasticity effects shed new light onto the literature pointing to a possible effect of music training onto perception and production abilities in a second language. For instance, it has been shown that musicians outperform nonmusicians in discriminating foreign speech sounds and are also better in learning novel phonetic categories of that same foreign language (Sadakata and Sekiyama, 2011). A similar advantage has been shown with speech segmentation, that is the ability to extract words from a continuous speech , in both adults and children learning an artificial language (François & Schön, 2011; François et al., 2013). Turning to the production abilities, better neural sensitivity to acoustic differences in music materials and higher musical aptitude are associated with superior production in a second language (Milovanov et al., 2010, 2008). Similarly, musically trained Indian children performed better on an English comprehension and vocabulary test than did nonmusicians children (Swaminathan and Gopinath, 2013). These studies demonstrate that musical expertise positively affects different aspects, such as phonological perception and production, of L2 proficiency. Our data extend these findings by showing that musical expertise strengthens the neural representation of linguistically-relevant acoustic cues, such as pitch and formant frequencies, which can account for the positive effects of musical expertise onto perception and production abilities in foreign languages.

According to the OPERA hypothesis, these transfer effects between music and speech can take place because they involve overlapping neural circuits to process similar acoustic information. By soliciting more intensively and repeatedly these neural circuits, musicians develop outstanding abilities to process fine spectral information, thus compensating for their lack of language-related experience. Importantly, such enhanced processing is not a global enhancement of one or several acoustic features of a stimulus, but rather seems to be a selective enhancement of those features that are functionally relevant, here linguistically relevant.

Our result may also reflect top-down processes that heighten relevant and filter non-relevant incoming sensory information in subcortical structures. Anatomically, there are several projections from the cortex to subcortical structures that can support they top-down dynamics (Kral and Eggermont, 2007). Krishnan et al. (2012) proposed a theoretical framework in which experience-dependent neural plasticity rely on both local mechanisms in the inferior colliculus (the presumed main neural generator of FFR, Chandrasekaran and Kraus, 2010), and dynamic feedback and feedforward interactions between subcortical and cortical structures. Stimulus features that are behaviorally-relevant for the listener would activate local, feedforward and feedback loops in a coordinated manner at each stage of processing. With growing experience, neural plasticity occurs in subcortical structures thus improving the neural processing of relevant acoustic cues.

In summary, we find that language-dependent effects occur along dimensions that are relevant to the listener, such as pitch and formant frequencies, and that musical training can fully compensate for the lack of language experience by strengthening the neural processing of these linguistically-relevant cues, bringing them to the level of the native speaker. According to the studies discussed above, these neural advantages in musicians can account for their better abilities in perception and production in a second language. These findings reinforce the link between music and speech, and support the hypothesis that music training may benefit second-language acquisition and may thus play an important role in the educational system.

References

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47. doi:10.1016/j.heares.2008.08.004

Alexander, J.A., Wong, P.C., Bradlow, A.R., 2005. Lexical tone perception in musicians and non-musicians., in: Interspeech. pp. 397–400.

Besson, M., Schön, D., Moreno, S., Santos, A., Magne, C., 2007. Influence of musical expertise and musical training on pitch processing in music and. Restor. Neurol. Neurosci. 25, 399–410.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77, 1–10. doi:10.1016/j.bandc.2011.07.006

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2009. Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. J. Cogn. Neurosci. 23, 425–434. doi:10.1162/jocn.2009.21362

Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur. J. Neurosci. 40, 2662–2673. doi:10.1111/ejn.12627

Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer.

Braun, B., Galts, T., Kabak, B., 2014. Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. Second Lang. Res. 30, 323–350.

Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47, 236–246. doi:10.1111/j.1469-8986.2009.00928.x

Chobert, J., Besson, M., 2013. Musical Expertise and Second Language Learning. Brain Sci. 3, 923–940. doi:10.3390/brainsci3020923

Chobert, J., Marie, C., François, C., Schön, D., Besson, M., 2011. Enhanced Passive and Active Processing of Syllables in Musician Children. J. Cogn. Neurosci. 23, 3874–3887. doi:10.1162/jocn_a_00088

Delogu, F., Lampis, G., Belardinelli, M.O., 2010. From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. Eur. J. Cogn. Psychol. 22, 46–61. doi:10.1080/09541440802708136

Duanmu, S., 2004. Tone and non-tone languages: An alternative to language typology and parameters. Lang. Linguist. 5, 891–923.

François, C., Chobert, J., Besson, M., Schön, D., 2013. Music Training for the Development of Speech Segmentation. Cereb. Cortex 23, 2038–2043. doi:10.1093/cercor/bhs180

Francois, C., Schön, D., 2011. Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures. Cereb. Cortex 21, 2357–2365. doi:10.1093/cercor/bhr022

Fujioka, T., Trainor, L., Ross, B., Kakigi, R., Pantev, C., 2004. Musical Training Enhances Automatic Encoding of Melodic Contour and Interval Structure. J. Cogn. Neurosci. 16, 1010–1021. doi:10.1162/0898929041502706

Intartaglia, B., White-Schwoch, T., Meunier, C., Roman, S., Kraus, N., Schön, D., 2016. Native language shapes automatic neural processing of speech. Neuropsychologia 89, 57–65. doi:10.1016/j.neuropsychologia.2016.05.033

Jentschke, S., Koelsch, S., 2009. Musical training modulates the development of syntax processing in children. NeuroImage 47, 735–744. doi:10.1016/j.neuroimage.2009.04.090

Koelsch, S., Schroger, E., Tervaniemi, M., 1999. Superior pre-attentive auditory processing in musicians. Neuroreport 10, 1309–1313.

Kral, A., Eggermont, J.J., 2007. What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Res. Rev. 56, 259–269. doi:10.1016/j.brainresrev.2007.07.021

Kraus, N., Skoe, E., Parbery-Clark, A., Ashley, R., 2009. Experience-induced Malleability in Neural Encoding of Pitch, Timbre, and Timing. Ann. N. Y. Acad. Sci. 1169, 543–557. doi:10.1111/j.1749-6632.2009.04549.x

Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192–201. doi:10.1016/S0378-5955(02)00327-1

Krishnan, A., Gandour, J.T., Bidelman, G.M., 2012. Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. NeuroReport 23, 498–502. doi:10.1097/WNR.0b013e328353764d

Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009. Experience-dependent neural representation of dynamic pitch in the brainstem: NeuroReport 20, 408–413. doi:10.1097/WNR.0b013e3283263000

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Cogn. Brain Res. 25, 161–168. doi:10.1016/j.cogbrainres.2005.05.004

Lee, C.-Y., Hung, T.-H., 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. J. Acoust. Soc. Am. 124, 3235. doi:10.1121/1.2990713

Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective Subcortical Enhancement of Musical Intervals in Musicians. J. Neurosci. 29, 5832–5840. doi:10.1523/JNEUROSCI.6133-08.2009

Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4, 540–545. doi:10.1038/87502

Marie, C., Delogu, F., Lampis, G., Belardinelli, M.O., Besson, M., 2010. Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. J. Cogn. Neurosci. 23, 2701–2715. doi:10.1162/jocn.2010.21585

Martínez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodríguez, L., Suárez-Murias, C., Valdés-Sosa, P.A., Besson, M., 2013. Musical expertise and foreign speech perception. Front. Syst. Neurosci. 7, 84. doi:10.3389/fnsys.2013.00084

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P.A.A., Tervaniemi, M., 2008. Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. Brain Res. 1194, 81–89. doi:10.1016/j.brainres.2007.11.042

Milovanov, R., Pietilä, P., Tervaniemi, M., Esquef, P.A.A., 2010. Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. Learn. Individ. Differ. 20, 56–60. doi:10.1016/j.lindif.2009.11.003

Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr. Clin. Neurophysiol. 35, 665–667. doi:10.1016/0013-4694(73)90223-X

Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl. Acad. Sci. 104, 15894–15898. doi:10.1073/pnas.0701498104

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., Hoke, M., 1998. Increased auditory cortical representation in musicians. Nature 392, 811–814. doi:10.1038/33918

Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4. doi:10.3389/fnagi.2012.00030

Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound. J. Neurosci. 29, 14100–14107. doi:10.1523/JNEUROSCI.3256-09.2009

Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 219, 111–119. doi:10.1016/j.neuroscience.2012.05.042

Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Audit. Cogn. Neurosci. 2, 142. doi:10.3389/fpsyg.2011.00142

Sadakata, M., Sekiyama, K., 2011. Enhanced perception of various linguistic features by musicians: A cross-linguistic study. Acta Psychol. (Amst.) 138, 1–10. doi:10.1016/j.actpsy.2011.03.007

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., Besson, M., 2010. Similar cerebral networks in language, music and song perception. NeuroImage 51, 450–461. doi:10.1016/j.neuroimage.2010.02.023

Schön, D., Magne, C., Besson, M., 2004. The music of speech: Music training facilitates pitch processing in both music and language. Psychophysiology 41, 341–349.

Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302–324. doi:10.1097/AUD.0b013e3181cdb272

Slevc, L.R., Miyake, A., 2006. Individual Differences in Second-Language Proficiency Does Musical Ability Matter? Psychol. Sci. 17, 675–681. doi:10.1111/j.1467-9280.2006.01765.x

Strait, D.L., O'Connell, S., Parbery-Clark, A., Kraus, N., 2013. Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30. Cereb. Cortex. doi:10.1093/cercor/bht103

Swaminathan, S., Gopinath, J.K., 2013. Music Training and Second-Language English Comprehension and Vocabulary Skills in Indian Children. Psychol. Stud. 58, 164–170. doi:10.1007/s12646-013-0180-3

Trainor, L.J., Desjardins, R.N., Rockel, C., 1999. A Comparison of Contour and Interval Processing in Musicians and Nonmusicians Using Event-Related Potentials. Aust. J. Psychol. 51, 147–153. doi:10.1080/00049539908255352

Vaissière, J., 1991. Rhythm, accentuation and final lengthening in French 108–120.

Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422. doi:10.1038/nn1872

Axe « Effet de l'attention »

La littérature sur les effets de l'attention sur l'encodage sous-cortical des sons de parole est encore peu abondante. Bien que les potentiels évoqués aux clics apparaissent insensibles aux effets de l'attention, ce n'est pas le cas des réponses neuronales aux sons complexes qui sont modulées en fonction du niveau attentionnel du sujet. En revanche, ces études ont été principalement menées avec des stimuli synthétiques, et aucune étude à ce jour n'a tenté de déterminer si ces effets attentionnels suivent des variations dynamiques et rapides au cours du temps, comme le suggère la *Dynamic of Attending Theory*. Nous avons tenté de répondre à cette question en entraînant l'attention du sujet de manière oscillatoire au rythme d'un stimulus sur deux via une tâche de tapping. Ce design expérimental est directement inspiré de la notion d'*active sensing* qui propose que la perception auditive est un phénomène actif dans lequel le système auditif, moteur et attentionnel agissent en synergie afin d'optimiser l'encodage sensoriel des informations dont l'occurrence temporelle est prédictible. Notre hypothèse était que la synchronisation du tapping avec la présentation d'un stimulus sur deux devrait entraîner une oscillation des niveaux attentionnels au même rythme, ce qui se traduirait par une modulation de l'activité sous-corticale en fonction de la phase de tapping.

Cette étude intitulée : « *Active sensing and predictive coding in subcortical structures* » a été soumise en 2016 au journal à comité de lecture *Journal of Cognitive Neuroscience*, et est actuellement en cours de révision. Cette étude est détaillée ci-après.

Active sensing and predictive coding in subcortical structures

Bastien Intartaglia[a,b,†], Daniele Schön[a,b]

[a] *Aix-Marseille Université, INS, Marseille, France*
[b] *INSERM, U1106, Marseille, France*

[†]Corresponding author
Bastien Intartaglia
Institut de Neurosciences des Systèmes
Inserm UMR1106
Aix-Marseille Université
Faculté de Médecine, 27, Boulevard Jean Moulin
13005 Marseille, France
Email: bastien.intartaglia@etu.univ-amu.fr
Phone: 334 91 38 46 33

Acknowledgements

This work was supported by the Brain and Language Research Institute (BLRI, ANR-11-LABX-0036), CNRS & Université d'Aix-Marseille. The authors wish to thank Christine Meunier for her assistance with audio recordings and analyses.

Abstract

Recent evidence suggests that auditory perception is an active process involving the interaction of the sensori-motor, attentional and auditory systems. Dynamic of attending theory proposes that, when the temporal structure is sufficiently regular, this interaction takes place in a hierarchical manner, enhancing the processing at the most salient moments in time. While previous studies explored top-down attentional effects at the cortical level, this study investigates active sensing, predictive coding and their interaction with musical expertise at the subcortical level. We recorded neural responses of nonmusicians and musicians listening passively or actively to the natural syllable [du]. Attention toward the stimuli was recruited by asking participants to tap along with one out of two stimuli which also allowed us to study whether attention, once entrained to the tapping pace, affects subcortical sound representations. We found that neural responses were enhanced by motor attention, with better spectral representation and more robust neural encoding of the fundamental frequency. Neural encoding also differed as a function of the phase of the tapping. Indeed, larger and less variable responses to the second harmonic were observed for the stimuli presented during the periods that preceded the tap. Interestingly, these effects were mainly driven by the nonmusician group while expert musicians were not affected by changes in attentional demands. These findings suggest that subcortical processing is modulated by attention at two time scales. At a slow time scale, motor-driven attention enhances subcortical encoding of natural speech sounds, supporting the idea of active sensing. At faster time scale, the rhythmic motor activity entrains cyclic fluctuations of attention also affecting the fine-grained coding of speech.

1. Introduction

The auditory system is classically viewed as distinct anatomical structures, hierarchically organized, relaying sensory inflow passively from the peripheral sensory organs to the central nervous system via bottom-up projections. In this framework, as sound is processed along the ascending pathway, neural structures' sensitivity to acoustic content decreases while the sensitivity to abstract features (syllables, words, intelligibility) increases (Davis & Johnsrude, 2003; Okada et al., 2010). However, in our everyday life we are often confronted to challenging listening conditions where the relevant information is merged into the noisy environment. To extract the relevant stimuli from the background noise, the auditory system must be able to balance the incoming acoustic flow by strengthening relevant and weakening non-relevant information. This suggests that the auditory perception is rather an active process, modulating incoming acoustic flow in a top-down manner.

In this view, the auditory system is considered as a distributed, but integrated, circuit, whereby auditory processing involves not only the sensory auditory system, but also other cognitive circuits such as the sensori-motor and attentional systems. This can be achieved via continuous interactions between the periphery and the central systems via afferent (bottom-up) and efferent (top-down) projections (Kraus & White-Schwoch, 2015). This is in line with the notion of active sensing which acknowledges the fact that information from the environment is sampled and structured via a continuous interaction between sensory and motor organs (Schroeder et al., 2010; Wachowiak, 2011). To explore the environment, the motor system orients sensory organs toward stimuli (finger squeezes, ocular saccades, olfactory sniffs) hence modulating incoming bottom-up information, and by sending efferent copies of motor commands to the sensory organs, the motor system shapes the processing of the incoming flow via top-down signals (Crapse and Sommer, 2008). This top-down mechanism helps to predict 'when' and 'where' sensory information is likely to occur (Sommer and Wurtz, 2008), thus improving its processing. In auditory perception, ears can be oriented toward the signal source

via head movements, and fluctuations of attention modulate the incoming acoustic flow in a top-down manner to strengthen target acoustic cues (Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010). Therefore, auditory perception can be viewed as an active process in which the auditory, motor and attentional systems operate in concert to process efficiently the acoustic information (Mottonen, van de Ven, & Watkins, 2014).

Several studies assessed whether attending to a stimulus modifies the responses of the auditory system. For instance, Teder-Sälejärvi & Hillyard (1998) showed that attending sounds in a specific place (right, middle or left speaker), affects response detection rate and neural responses (N1 and P3 amplitudes). Differences between cortical responses in passive and active listening conditions are larger in challenging listening condition, such as for instance, when presenting sounds in a noisy context (e.g. cocktail party, Zhang et al., 2016). Orienting attention during challenging listening conditions helps to identify the target talker's voice, that is to extract 'what' is particularly relevant for the listener. Similarly, when attention is oriented in time, both behavioral and neural responses to target/deviants sounds are improved. For instance, when participants are asked to attend deviant sounds within a flow of frequent sounds, target/deviant sounds are associated with novel-detection responses all along the auditory pathway, such as the mismatch negativity (100-200 ms) generated in the auditory cortex (Näätänen, Kujala, & Winkler, 2011), also visible in the Pa component (~30-40 ms) generated in the thalamus (Grimm, Escera, Slabu, & Costa-Faidella, 2011; Slabu, Escera, Grimm, & Costa-Faidella, 2010) and even in the early neural responses generated in the auditory brainstem (Slabu, Grimm, & Escera, 2012). Moreover, using fMRI, Rinne et al. (2008) showed an increased metabolic activity in both the auditory cortex and the inferior colliculus during attentive as compared to passive listening condition, confirming that auditory processing is modulated by attention in a top-down manner at the cortical and subcortical levels (Rinne et al., 2008). Overall, these studies suggest that allocating attentional resources via active listening increases the neuronal recruitment, which in

turn affects sound processing throughout the different relays of the auditory pathway in a top-down manner.

This dynamic view of the auditory perception is particularly beneficial when acoustic events unfold over time with temporal regularity. Indeed, natural stimuli are often organized in rhythmic or quasi-rhythmic streams (e.g. music, speech, walking, etc...), and extracting these temporal regularities is crucial to predict the occurrence of subsequent events and optimize their processing. The dynamic attending theory (DAT) suggests that attention oscillates in time as a function of the temporal regularity of relevant acoustic events (Jones, 1976; Jones and Boltz, 1989; Large and Jones, 1999), thus allowing for a more efficient sampling of information from the environment. For instance, in speech and music, temporal regularities are conveyed at different time scales : speech can be broken up into sentences, words and phonemes, and music into phrases, beats and notes. While speech and music do not have the same degree of temporal regularity both have nonetheless a temporal organization of acoustic events that unfold over time. Indeed, several studies confirm that language presents temporal regularities that play an important role to predict upcoming events, such as alternation of stressed and unstressed syllables (Schmidt-Kassow & Kotz, 2009). According to the DAT, attentional levels are entrained by the temporal structure of external events, such that events occurring in phase with high attentional levels are more efficiently encoded. Jones et al. (2002) showed that listeners were more accurate in evaluating tones' pitch when sounds were presented in a predictable stream (regular IOI) as compared to a less predictable stream (irregular IOI). Similar results were found in the visual modality : participants were more accurate to detect target visual stimuli embedded in rhythmic as compared to arrhythmic streams of noises patches (Cravo, Rohenkohl, Wyart, & Nobre, 2013; Rohenkohl, Cravo, Wyart, & Nobre, 2012). These results support the idea that temporal attention is influenced by the distribution of stimulus time intervals.

While the corpus of studies on the effect of temporal attention at a cortical level is growing quickly there is very little evidence on whether rhythmic fluctuations of attention can

modulate the processing of peripheral neural structures such as the auditory brainstem (Tierney & Kraus, 2013). The aim of this study was to assess whether the notion of active sensing and dynamic attending also hold at the subcortical level. The notion of active sensing was explored comparing auditory brainstem responses (ABR) under active and passive listening conditions. Attention toward the stimuli was recruited during the active condition by asking participants to tap along with one out of two stimuli. Because phasic movements, such as tapping, entrains attentional resources in phase with the movement (Morillon, Schroeder, & Wyart, 2014; Mottonen et al., 2014), this design also allowed us to study the effects of cyclic fluctuations of attention on speech ABR. We predicted that there would be an enhancement of spectral representation and inter-trial phase-coherence in the active condition compared to the passive condition. During the active condition, we predicted that neural entrainment at the tapping pace would modify spectral representation and phase-coherence of subcortical responses for stimuli immediately preceding compared to stimuli immediately following the tap. Finally, since previous studies have shown that musical expertise affects behavioral (sensori-motor) and neurophysiological responses at the subcortical level (Musacchia, Sams, Skoe, & Kraus, 2007; A. Parbery-Clark, Tierney, Strait, & Kraus, 2012; Repp, 2010; Strait & Kraus, 2014; Thompson, White-Schwoch, Tierney, & Kraus, 2015), we decided to compare a group of musicians and a group of nonmusicians, to see whether changes in attentional demands and fluctuations of attention would differently affect subcortical responses of one or the other population.

2. Material and method

2.1. Participants

Thirty-seven (22 females) French native speakers ranging in age from 18 to 34 years, participated in the study. Sixteen were musicians having started musical training before the age of 14 (mean = 8.4, SD = 0.75) and had consistently practiced for at least 8 years (mean = 13.5, SD = 1.13). The remaining twenty-one participants had received less than three years of musical

practice and were categorized as nonmusicians. Eleven of these nonmusicians had received no musical training at all. Inclusion criteria were a high-school level of education, click-evoked brainstem response latencies within lab-internal normal limits (5.41-5.97 ms; 100-µs click stimulus presented at 80 dB SPL). The two groups did not differ in term of age (p = 0.560), sex (p = 0.180), education (p = 0.168) or hearing thresholds (p = 0.845). Participants gave their informed consent and were paid for their participation. The study was approved by the local ethics committee.

2.2. Stimulus

The stimulus used was the natural speech syllable [du] recorded in an anechoic chamber by a French male speaker. This stimulus was chosen among other tokens of the same syllable by an experienced phonetician according to its spectral characteristics (clear voice-onset time and vowel) and its short duration (110 ms) necessary for an experiment with short inter-onset intervals. A fade-in and fade-out were applied to the initial and final 5 ms windows of the stimulus waveform. Since the stimulus was a natural speech sound, the fundamental frequency (*F*0) and its subsequent harmonics (*H*2-3) slightly varied over time and were thus estimated by computing a fast Fourier transform (FFT) on two time regions established according to the spectral changes by an experienced phonetician (consonant : 0-45 ms and vowel : 45-110 ms, see waveform, **Figure 1**). Spectral analysis revealed that from 0 to 45 ms (consonant), *F*0, *H*2 and *H*3 were equal to 103 Hz, 208 Hz and 318 Hz respectively, while from 45 to 110 ms (vowel), *F*0, *H*2 and *H*3 were equal to 108 Hz, 218 Hz and 326 Hz respectively (see FFT, **Figure 2**).

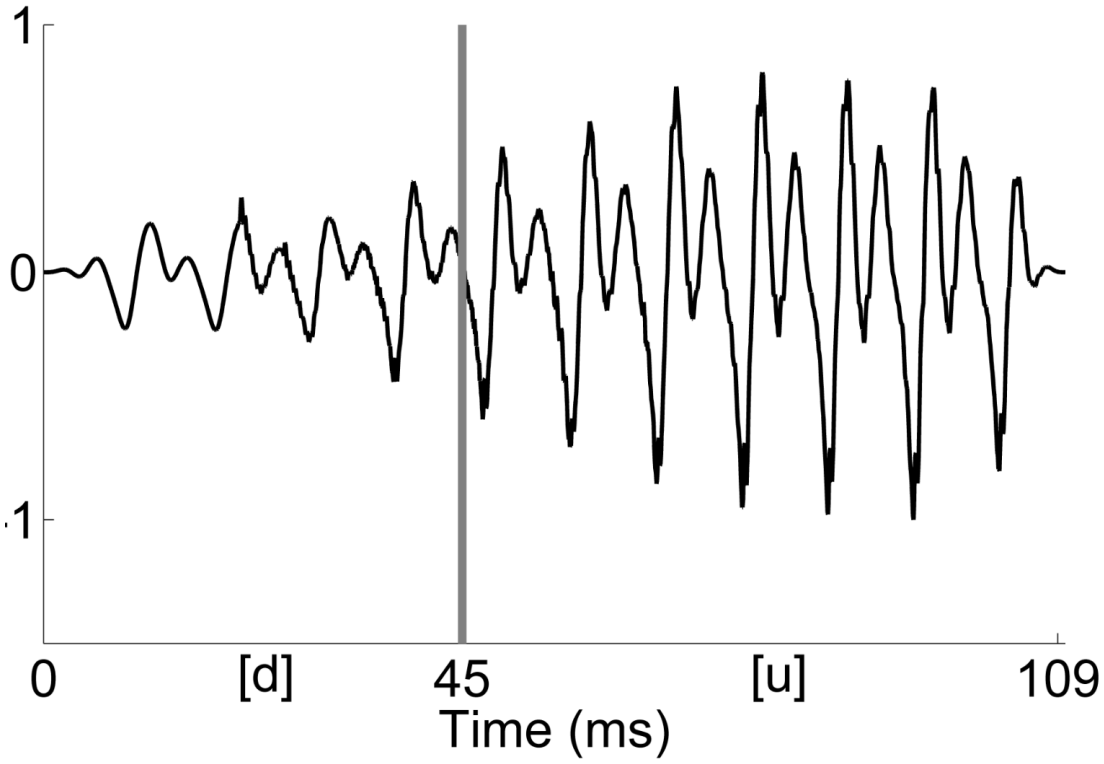


Figure 1. Waveform of the stimulus (normalized amplitude). The vertical gray line indicates the boundary between the consonant and the vowel, as established according to the spectral changes by an experienced phonetician.

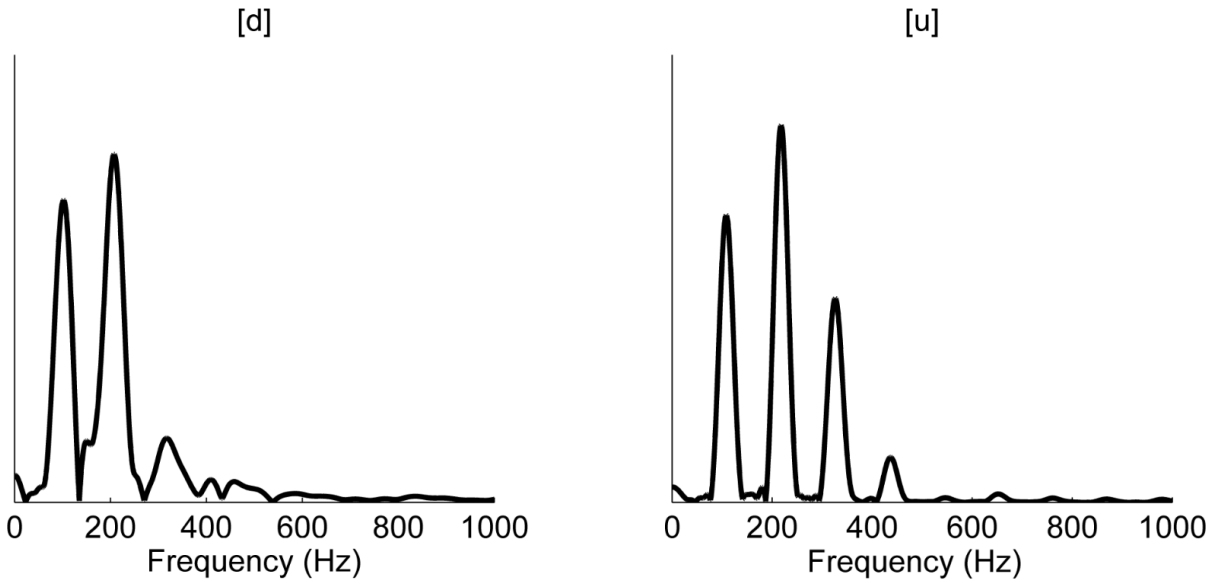


Figure 2. Fast Fourier transform of the stimulus computed on the consonant (left panel) and on the vowel (right panel). Amplitudes are normalized.

2.3. Electrophysiological recordings

In both conditions, the stimulus was presented monaurally to the right ear at 80 dB SPL at a rate of 3.3 Hz through magnetically shielded insert earphones (ER-3A, Etymotic Research) using the stimulus presentation software Microvitae (μV-ABR, Microvitae Technologies) while participants sat in a comfortable reclining chair in an electrically-shielded, sound-attenuated room. Brain responses were collected at 30 kHz sampling rate using Microvitae recording system (μV-ABR, Microvitae Technologies) with three Ag-AgCl scalp electrodes in a vertical montage (Cz active, forehead ground, and right earlobe reference). Electrode impedances were kept <5 kΩ. A total of six-thousand sweeps were collected for each condition (three thousands for each polarity). During the passive listening, the stimulus polarities were alternated across trials, while during the active listening they were alternated across blocks (see section 2.6 for more details), in order to minimize the contribution of stimulus artifact and cochlear microphonic (Aiken & Picton, 2008; Skoe & Kraus, 2010).

2.4. Procedure

To assess the global effect of attention on subcortical processing of speech sounds, brain responses were recorded under active and passive listening conditions. During the passive listening condition, stimuli were delivered to the right ear while participants were watching a subtitled movie of their choice to maintain relaxation and prevent drowsiness. During the active listening condition, attention toward the stimuli was recruited by asking the participants to tap along every two stimuli. The tapping experiment included a training session of 100 trials followed by six blocks lasting about 5 minutes each. During each block, one-thousand stimuli were delivered at a rate of 3.3 Hz (IOI : 303.3 ms) and participants were asked to tap along every two stimuli on a drum (target inter-tap interval : 606.6 ms).

2.5. Behavioral measurements

2.5.1. Apparatus

The apparatus used to record participants' taps was a custom-made drum, consisting of a plastic cylinder containing a microphone surrounded by acoustic foam to attenuate the sound produced by the taps. Participants listened to the stimulus delivered to the right ear and were instructed to tap along every two stimuli on the microphone in the drum box with the index finger of their dominant hand. Although tapping inside the acoustically isolated box only produced a minimal amount of acoustic signal, an ear plug was also inserted in the left ear. Synchronization of taps and stimuli recordings was made using two identical microphones (DPA d:vote 4099U, DPA Microphones) connected to the same sound card (Saffire Pro 24). One microphone recorded the taps while the other recorded the left channel of the insert. This experimental setup allowed us to record on the same stereo audio track the synchronized recordings of the taps (right track) and stimuli (left track) using Adobe Audition.

2.5.2. Taps detection

Adobe Audition was used to automatically detect the tapping onsets in the recorded session. Because of the high intra- and inter-subject variability in the acoustic amplitude of the taps, detection thresholds were adjusted for each recording block and participant. This resulted in approximately 98% of correct detection. Then, to ensure that no tap was omitted or marked twice, all markers were visually inspected and manually removed or added when appropriate.

2.5.3. Behavioral data processing

Tapping performances were assessed by calculating 1) the mean asynchrony between taps and stimuli, representing the tapping accuracy of the participant, 2) the standard deviation of asynchrony across trials, representing the tapping variability. A small asynchrony indicates a precise tapping, meaning that the subject's taps are close to the stimulus onsets. A negative value

indicates an anticipatory tap, while a positive value indicates a late tap, relative to stimulus onset. These values were calculated for each block using the Matlab toolbox CircStat Version 2012a (Berens, 2009), and were then averaged to form global measures of each subject's tapping performances. Asynchrony values were transformed into radians in order to use a circular representation.

2.6. EEG data processing

All analyses were performed using custom MATLAB scripts (Mathworks). First, electrophysiological recordings were bandpass filtered from 70 to 2000 Hz (12 dB/octave roll-off, zero phase-shift) using a Butterworth filter. Then, sweeps with activity exceeding ± 30 μV were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period (-30 to 0 ms). Neural responses were then averaged over a -30 to 149 ms window for each condition. The signal-to-noise ratio (SNR) was computed using the quotient of response root mean square (RMS) amplitude and pre-stimulus baseline RMS amplitude (Skoe & Kraus, 2010). SNR was first calculated across all conditions in order to have a global measure of the recording quality per participant. If the SNR was less or equal to 1.4 the participant was excluded which resulted in excluding 3 participants. After visual inspection of individual averages, one more participant was excluded because of a stimulus artifact at time zero. Then, to make sure that differences observed between conditions could not be attributed to different SNRs, the SNRs were calculated for each condition separately and compared between conditions. Statistical analyses showed that the SNR was equivalent across the active and passive listening conditions (p = 0.489) and between predictive and reactive conditions (p = 0.886).

Separating predictive and reactive brain responses. During the active listening condition, the participants had to tap along every two stimuli. Stimulus epochs preceding the tap onsets were categorized as predictive, and those following the tap onsets were categorized as reactive epochs.

Because participants (especially nonmusicians) did make some errors, in order to assign each epoch to a specific category, we first needed to assign each tap to the corresponding stimulus epochs. This was realized using custom Matlab scripts: every tap detected in a time window of +/- 100 ms surrounding a stimulus onset was assigned to that precise stimulus.

Since participants had to tap along one every two stimuli, stimulus polarities could not be alternated across trials because it would have resulted in tapping along with one polarity versus the other, and the procedure described above would have resulted in assigning one polarity to one category of epochs (e.g. positive polarities assigned only to predictive epochs, and negative polarities only to reactive epochs). To prevent this bias, stimulus polarity was kept constant within each block, but was alternated across the blocks. This design allowed us to average brain responses across negative and positive polarities for predictive and reactive epochs separately, thus amplifying the envelop response and reducing the stimulus artifact and the cochlear microphonic (Aiken & Picton, 2008; Skoe & Kraus, 2010). The order of the blocks was pseudo randomized across the participants.

Spectral amplitude. The spectral energy of the neural responses was analyzed by computing a Fast Fourier transform on two time regions of the response participants: consonant (20-55 ms) and vowel (55-120 ms). These time regions took into account a 10 ms neural delay in the response. For the consonant, the poor frequency resolution due to the short temporal window did not allow to analyze the different frequency components separately, thus the spectral amplitudes were averaged over 50-250 Hz wherein most of the spectral energy was present, and this was taken as a global measure of the *F*0 representation (Hornickel, Anderson, Skoe, Yi, & Kraus, 2012). For the vowel, the maximum spectral amplitudes of the fundamental frequency (*F*0) and its second and third harmonics (*H*2-3) were extracted in a bandwidth of 20 Hz surrounding the peak in the stimulus fast Fourier transform (e.g. for a peak at 108 Hz, values were extracted between 98 and 118 Hz (Alexandra Parbery-Clark, Anderson, Hittner, & Kraus, 2012). The *F*0 and its two

subsequent harmonics were chosen because the spectral energy was the most prominent at these frequencies in the FFT of the stimulus (**Figure 2**). This procedure was performed for each condition (active vs. passive listening) and response class (predictive vs. reactive responses).

Inter-trial phase-coherence. We used the same procedure described in (Tierney & Kraus, 2013). This technique measures the phase consistency across trials separately for each frequency component in the neural responses. More precisely, for the two time regions (consonant and vowel) a fast Fourier transform was performed on each trial that resulted in two values, the amplitude and the phase for each frequency component. Since we were interested in phase variability across trials, only the phase values were kept. The vector's length of each frequency was computed using the Matlab toolbox CircStat Version 2012a (Berens, 2009). The length of the resultant vector represents the phase-coherence across trials for each frequency component. This measure ranges from 0 (no phase coherence) to 1 (perfect phase coherence). Then, the same procedure described above for spectral analyses was used for the consonant and the vowel time regions.

Statistical analyses. All statistical analyses were performed using Statistica Version 7.1 (StatsSoft, Tulsa, OK). To investigate the global and local effect of attention, a repeated measure analysis of variance (RMANOVA) was performed on the complete dataset using group as a between subjects factor (musicians vs. nonmusicians) and conditions as a within subjects factor (active vs. passive). A second RMANOVA was performed on the active listening dataset using group as a between subjects factor and response class (predictive vs. reactive) as a within subjects factor. In both analyses the dependent variables were power spectral density as a measure of spectral representation and inter-trial phase-coherence as a measure of neural consistency across trials. *Post-hoc* tests were used when appropriate (Fisher LSD).

3. Results

3.1. Tapping performances

Most participants exhibited an anticipatory trend, meaning that they tapped before the onset of the stimulus (**Figure 3**). Musicians showed less variable (main effect of group, $F(1,31) = 29.86$, $p < 0.001$) and more accurate (main effect of group, $F(1,31) = 28.84$, $p < 0.001$) tapping across trials as compared to nonmusicians (**Figure 4**).

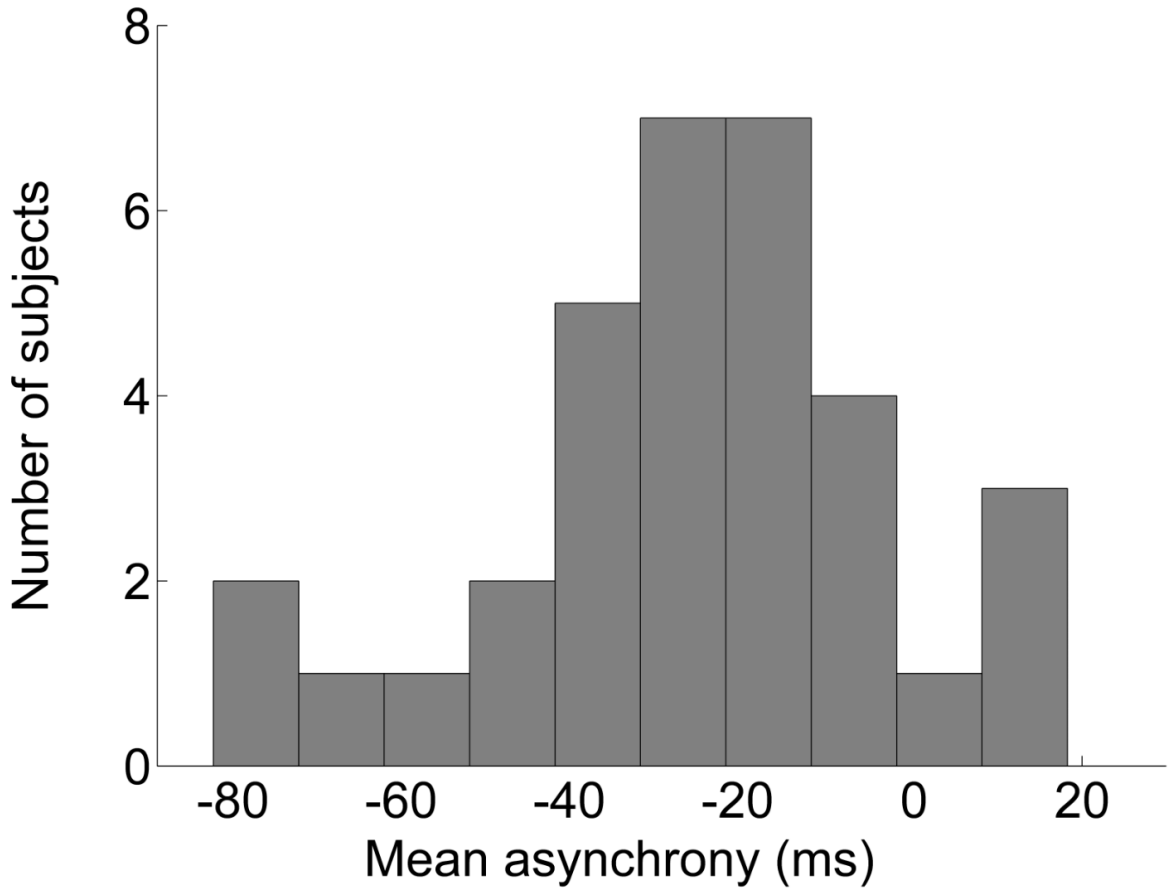


Figure 3. Histogram of mean asynchrony between tap and stimulus onset across subjects. Time zero represents the onset of the stimulus presentation. Negative values indicate anticipatory tapping, while positive values indicate reactive tapping.

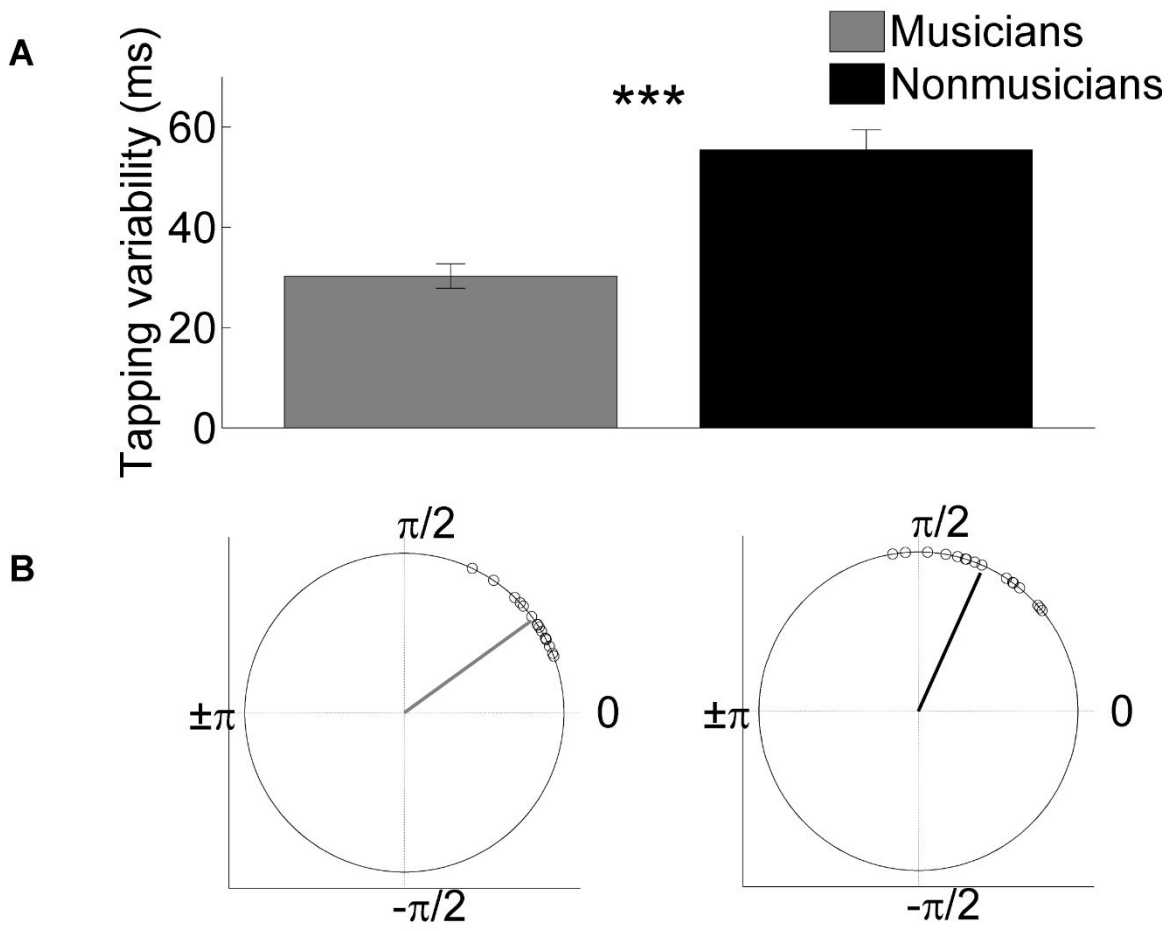


Figure 4. Tapping performances of musicians (grey) and nonmusicians (black). ***A***, Standard deviation of asynchrony between tap and stimulus across trials (in milliseconds). Lower tapping variability indicates better tapping performance. Error bars represent ± 1 standard error of the mean. ***B***, Circular representation of mean asynchrony of musicians (left panel) and nonmusicians (right panel). Zero represents the onset of the stimulus and dots represent the mean asynchrony of the participants (in radians). Musicians exhibit less variable and more accurate tapping across trials as compared to nonmusicians. *** $p < 0.001$.

3.2. Global effect of attention

3.2.1. Spectral representation

Fundamental frequency (F0). Neural responses were divided into two time regions, corresponding to the response to the consonant and the response to the vowel. Spectral analysis of the time region corresponding to the consonant revealed that participants exhibited significantly better spectral encoding of the fundamental frequency (*F*0) during the tapping task as compared to the passive listening task (main effect of task, $F(1,27) = 6.74$, $p = 0.015$, **Figure 5A**). There was no effect of

musicianship for the encoding of the *F*0 (main effect of group, $F(1,27) = 0.49$, $p = 0.491$), and no interaction between musicianship and task (group x task interaction, $F(1,27) = 0.56$, $p = 0.461$).

Analysis of time region corresponding to the vowel showed again a significant main effect of the task, with higher spectral representation of the *F*0 during the tapping task as compared to the passive listening task (main effect of task, $F(1,27) = 6.80$, $p = 0.015$, **Figure 5B**). There was no main effect of musicianship (main effect of group, $F(1,27) = 0.92$, $p = 0.346$). However, nonmusicians showed a stronger effect of the task revealed by a significant interaction between musicianship and task (group x task interaction, $F(1,27) = 6.88$, $p = 0.015$, **Figure 6A,B**). Indeed, *post-hoc* test showed that the task effect was highly significant for the nonmusician group ($p < 0.001$) while this effect was not significant for the musician group ($p = 0.992$), meaning that the main effect of the task was mainly driven by the nonmusician group.

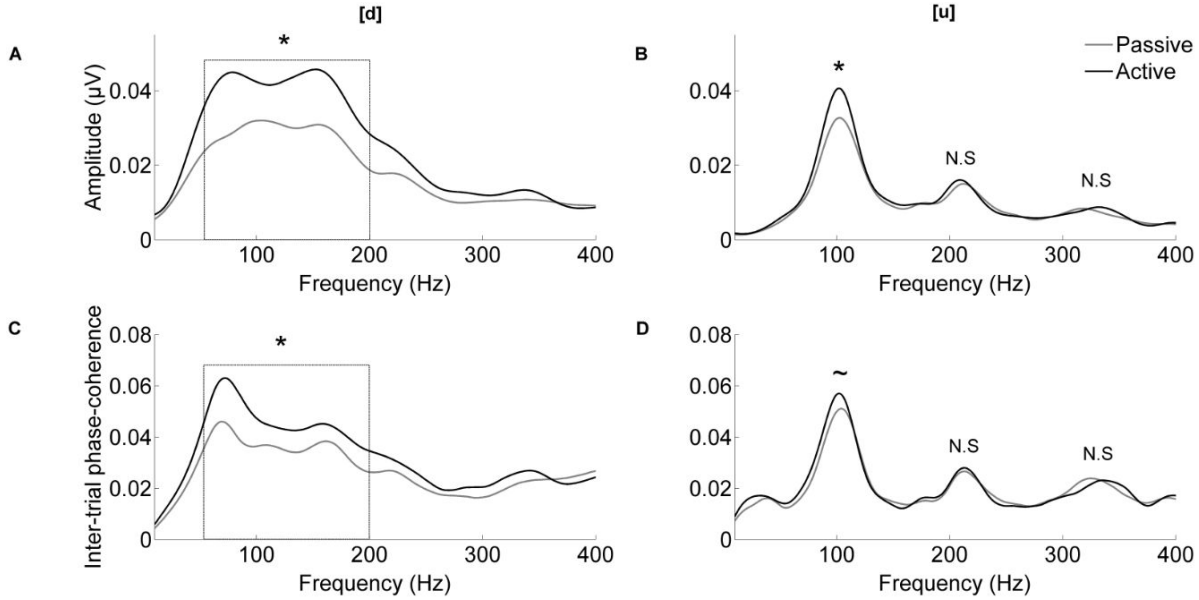


Figure 5. Global effect of attention. Spectral representation (top) and inter-trial phase-coherence (bottom) of the consonant [d] (left panels) and vowel [u] (right panels). ***A***,***C***, The results show that for the consonant, participants have enhanced spectral representation and more robust encoding of the pitch during the active listening (black) as compared to the passive listening condition (grey) . ***B***,***D***, Analyses of the response to the vowel show that participants have improved spectral representation of the *F*0 but not of the two subsequent harmonics (*H*2-3) during the active listening condition. The same effect is observed in the analyses of the inter-trial phase-coherence, though the difference is marginally significant. * $p < 0.05$.

Harmonics (H2-3). Spectral analysis of the time region corresponding to the vowel revealed that both groups of participants had equivalent spectral representation of the second (*H*2) and the third (*H*3) harmonics during the tapping task as compared to the passive listening task (main effect of task, *H*2 : $F(1,27) = 0.09$, $p = 0.754$; *H*3 : $F(1,27) = 0.01$, $p = 0.940$, **Figure 5B**). There was no effect of musicianship (main effect of group, *H*2 : $F(1,27) = 0.22$, $p = 0.640$; *H*3 : $F(1,27) = 0.06$, $p = 0.812$), and there was no interaction between musicianship and the task (group x task interaction, *H*2 : $F(1,27) < 0.01$, $p = 0.999$; *H*3 : $F(1,27) = 1.51$, $p = 0.230$, **Figure 6A,B**).

3.2.2. Inter-trial phase-coherence

Fundamental frequency (F0). The analysis of the response time region corresponding to the consonant showed that participants had more robust encoding of the fundamental frequency (*F*0) under active listening tapping task as compared to the passive listening task (main effect of the task, $F(1,27) = 6.84$, $p = 0.014$, **Figure 5C**). Both groups had equivalent representation of the *F*0 (main effect of group, $F(1,27) = 0.28$, $p = 0.603$) and there was no interaction between musicianship and task ($F(1,27) = 0.05$, $p = 0.817$). Analyses of the responses to the vowel showed a marginal effect of the task ($F(1,27) = 3.72$, $p = 0.064$, **Figure 5D**). There was no main effect of musicianship (main effect of group, $F(1,27) = 0.48$, $p = 0.493$). However, nonmusicians showed a stronger effect of the task revealed by a significant interaction between musicianship and task (group x task interaction ($F(1,27) = 4.33$, $p = 0.047$, **Figure 6C,D**). Indeed, *post-hoc* test showed that the task effect was highly significant for the nonmusician group ($p < 0.01$) while this effect was not significant for the musician group ($p = 0.921$), meaning that the main effect of the task was mainly driven by the nonmusician group.

Harmonics (H2-3). During the response to the vowel, inter-trial phase-coherence was equivalent during the tapping task and the passive listening task for both harmonics (main effect of task, *H*2 : $F(1,27) = 0.15$, $p = 0.701$; *H*3 : $F(1,27) = 1.47$, $p = 0.236$, **Figure 5D**). The variability of the

response was not affected by musicianship (main effect of group, *H*2 : $F(1,27) = 0.04$, $p = 0.835$; *H*3 : $F(1,27) = 0.24$, $p = 0.877$) and there was no interaction between group and task (*H*2 : $F(1,27) = 0.10$, $p = 0.750$; *H*3 : $F(1,27) = 0.003$, $p = 0.955$, **Figure 6C,D**).

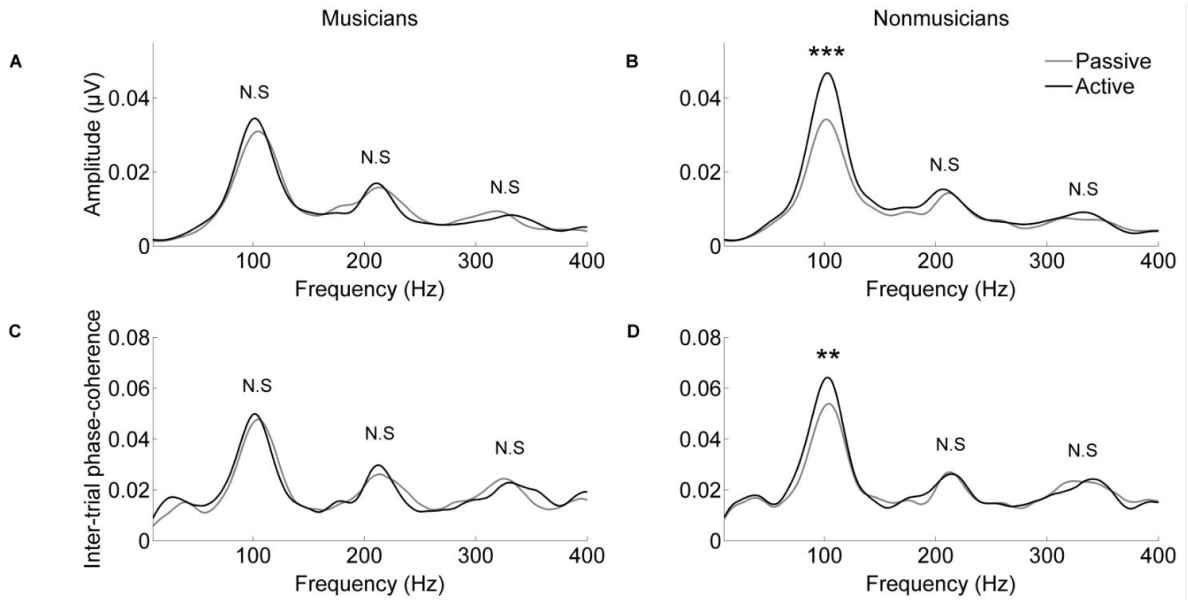


Figure 6. Global effect of attention in musicians and nonmusicians. Spectral representation (top) and inter-trial phase-coherence (bottom) of the vowel in musicians (left panels) and nonmusicians (right panels). ***A***,***C***, The results show that for the musician group the subcortical processing of the vowel was not affected by attention. ***B***,***D***, In the nonmusician group, the spectral representation of the *F*0 was significantly enhanced during the active listening condition. The same effects were observed in the analyses of the inter-trial phase-coherence. ** $p < 0.01$, *** $p < 0.001$.

3.3. Local effect of attention

3.3.1. Spectral representation

Fundamental frequency (F0). During the tapping task, spectral analyses of the *F*0 did not reveal any significant local effect of attention for both time regions (main effect of response class, consonant : $F(1,30) = 0.73$, $p = 0.399$; vowel : $F(1,30) = 1.07$, $p = 0.309$; main effect of group, consonant : $F(1,30) = 1.06$, $p = 0.312$; vowel : $F(1,30) = 3.92$, $p = 0.057$; group x response class

interaction : consonant : $F(1,30) = 1.91$, $p = 0.177$; vowel : $F(1,30) = 0.008$, $p = 0.930$, **Figure 7A,B**).

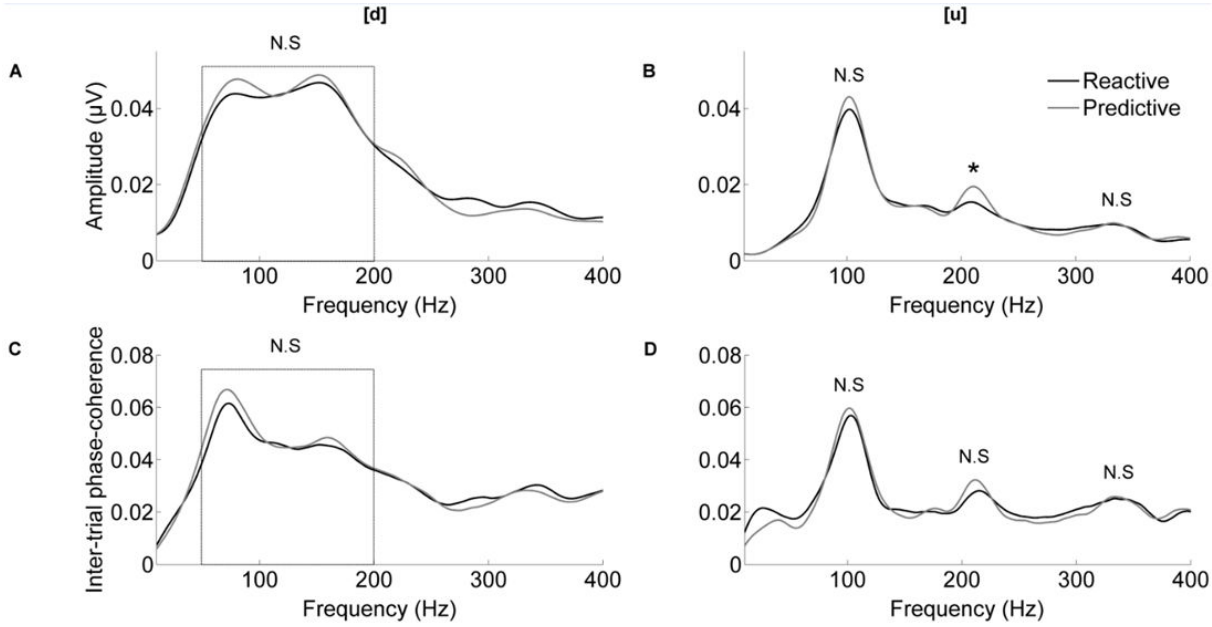


Figure 7. Local effect of attention. Spectral representation (top) and inter-trial phase-coherence (bottom) of the consonant [d] (left panels) and vowel [u] (right panels). ***A***,***C***, Spectral representation and neural coherence across trials in neural responses to the consonant were not modulated at fast time scale. ***B***,***D***, Analyses of the response to the vowel show that participants have a specific modulation of the spectral representation of the second harmonic (*H*2). The same tendency is observed in the analyses of the inter-trial phase-coherence, though the difference is not significant. * p <0.05.

Harmonics (H2-3). Participants exhibited significantly higher spectral representation of *H*2 during the predictive as compared to the reactive period in response to the vowel (main effect of response class, *H*2 : $F(1,29) = 4.44$, $p = 0.044$; *H*3 : $F(1,30) = 0.29$, $p = 0.867$). There was no effect of musicianship (main effect of group, *H*2 : $F(1,29) = 0.0007$, $p = 0.979$; *H*3 : $F(1,30) = 0.51$, $p = 0.482$) in both the predictive and reactive responses (group x response class interaction, *H*2 : $F(1,29) = 0.55$, $p = 0.464$; *H*3 : $F(1,30) = 0.45$, $p = 0.508$, **Figure 7B**). Although analyses did not reveal any interaction, visual inspection seems to show a stronger effect for the nonmusician group, suggesting that the main effect of local variation of attention was more strongly driven by the nonmusicians (**Figure 8A,B**).

3.3.2. Inter-trial phase-coherence

Fundamental frequency (F0). Analyses of phase-coherence did not reveal any significant effect of local variation of attention on *F*0 encoding for both time regions (main effect of response class, consonant : $F(1,30) = 2.18$, $p = 0.150$; vowel : $F(1,30) = 0.04$, $p = 0.840$, **Figure 8C,D**). There was no effect of musicianship (main effect of group, consonant : $F(1,30) = 0.24$, $p = 0.626$; vowel : $F(1,30) = 1.94$, $p = 0.174$), and this was true for both conditions (i.e. there was no group x response class interaction, consonant : $F(1,30) = 1.04$, $p = 0.316$; vowel : $F(1,30) = 1.09$, $p = 0.305$).

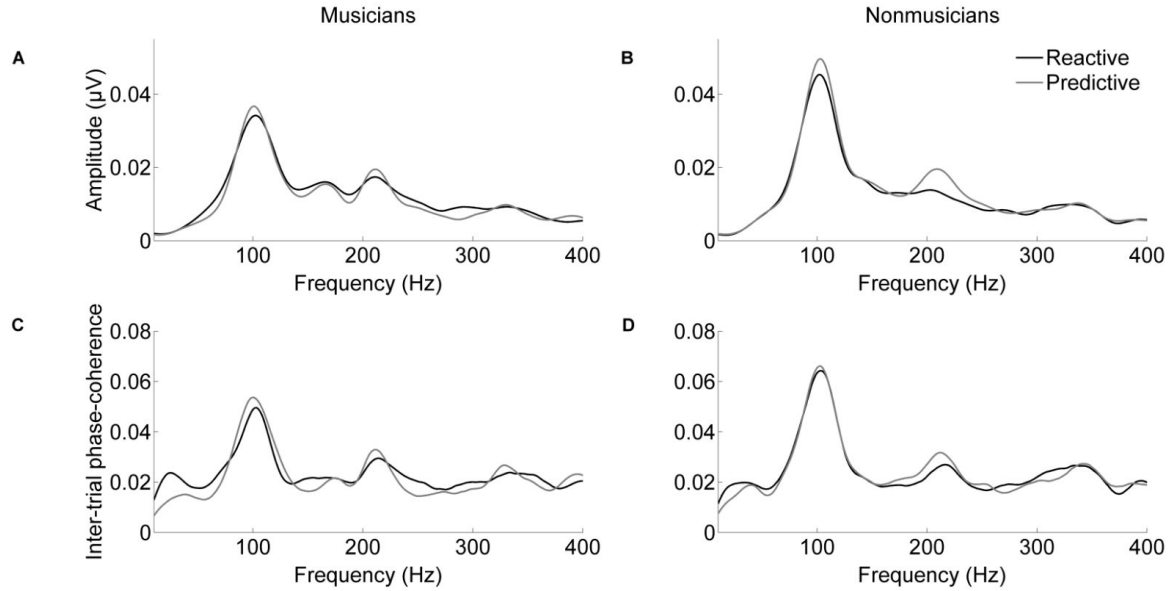


Figure 8. Local effect of attention in musicians and nonmusicians. Spectral representation (top) and inter-trial phase-coherence (bottom) of the vowel in musicians (left panels) and nonmusicians (right panels). ***A***,***C***, Spectral representation and neural coherence across trials in neural responses to the vowel in the musician group. ***B***,***D***, In the nonmusician group the effects on the second harmonic (*H*2) are visually larger compared to musicians, though the difference is not significant. Note that, because the interaction between group and response class was not significant, *post-hoc* comparisons were not run, explaining the lack of stars in the figure.

Harmonics (H2-3). During the tapping task, participants did not show any significant effect of local variations of attention for both harmonics (main effect of response class, *H*2 : $F(1,30) = 0.84$, $p = 0.367$; *H*3 : $F(1,30) = 0.04$, $p = 0.849$, **Figure 7D**). There was no effect of musicianship (main

effect of group, $H2$: $F(1,30) = 0.48$, $p = 0.495$; $H3$: $F(1,30) = 0.02$, $p = 0.885$) and this was the case for both the reactive and predictive responses (group x response class interaction, $H2$: $F(1,30) = 0.001$, $p = 0.980$; $H3$: $F(1,30) = 1.16$, $p = 0.289$, **Figure 8C,D**).

4. Discussion

The aim of this study was three fold. Firstly, we wanted to assess whether, compared to passive listening, active listening driven by motor attention modifies auditory brainstem responses. Secondly, we wanted to study the effects of fast period fluctuations of attention. Finally, we wanted to assess whether these variables would interact with musical expertise. To this aim, participants either listened passively to a speech phoneme or had to actively tap every two stimuli. We predicted more marked and consistent sound features representation in the active condition compared to the passive condition. During the active condition, we predicted different subcortical responses to stimuli immediately preceding compared to stimuli immediately following the taps. Results showed that directing attention toward the stimuli induces a global enhancement of the subcortical representation of the fundamental frequency visible in a greater spectral amplitude and a less variable phase-coherence across trials during the active compared to the passive listening condition. Interestingly, these effects were more strongly driven by the nonmusician group. At a faster time scale, the spectral representation at the second harmonic was significantly improved for stimuli preceding the taps, meaning that the fine-grained subcortical encoding of pitch was affected by rapid fluctuation of attention. We will discuss in turn the global and local effects of attention as well as potential mechanisms underlying top-down modulations of subcortical processing.

4.1. Global effect of attention

The effect of auditory attention is well documented at the cortical level, with enhanced amplitudes and/or shorter latencies to the attended stimuli as early as 50 ms post-stimulus onset (van Atteveldt et al., 2015). This support the idea of an early mechanism of selective attention that takes place possibly even before the cortical stage, by gating or filtering the incoming information in subcortical structures. By contrast, modulation of subcortical structures' activity by attention has been a long-lasting debate. Early studies have shown for instance that auditory

brainstem responses evoked by clicks are not affected by attention (Hackley, 1993; Hackley, Woldorff, & Hillyard, 1990). More recently, studies using more complex stimuli showed that the amplitude and the latency of the frequency following response (FFR) are modulated by attention, with shorter latency (Hoormann, Falkenstein, & Hohnsbein, 1994; Jörg Hoormann, Falkenstein, & Hohnsbein, 2000) and higher amplitude (G. C. Galbraith, Bhuta, Choate, Kitahara, & Mullen, 1998; G. C. C. Galbraith, Olfman, & Huffman, 2003; G. C. Galbraith & Doan, 1995; J. Hoormann et al., 1994; Jörg Hoormann et al., 2000) in response to the attended stimuli. Similarly, directing attention away from the stimuli will reduce the spectral amplitude of the FFR (Hairston et al., 2013). This suggests that under active listening, the processing of complex synthetic sounds is affected in a top-down manner by attention, by modifying the excitability threshold and/or the phase-locking properties of subcortical neurons underlying the generation of the FFR. Our results extend these findings by showing that the enhancement of the *F*0 representation by attention is also present using more ecological sounds such as natural speech syllables. More interestingly, the present study shows that attention can be recruited by motor demands (i.e. tapping) suggesting that the motor and attentional systems interact dynamically to enhance the sensory gain of the incoming acoustic flow at the subcortical level, which supports and extends the notion of active sensing to the subcortical processing. This suggests that modulation of subcortical processing by attention can play a crucial role in filtering relevant information before further cortical processing.

Our results also highlight the effect of musical practice on these top-down modulations by attention. By measuring spectral density and neural response consistency on a trial-to-trial basis, we found that directing attention to the stimuli enhanced the spectral representation and reduced the inter-trial variability of neural responses during the consonant and the vowel for both groups of participants when looking at the fundamental frequency. Surprisingly, in the response to the vowel this effect was mainly driven by the nonmusician group. This is likely due to the fact that the tapping task requires higher attentional resources for the nonmusicians as

compared to the expert musicians who are trained to perform similar tasks during musical practice. Indeed, behavioral analyses showed that musicians have less variable and more accurate tapping performances as compared to nonmusicians, suggesting that the benefit of musical practice is such that they do not need extra attentional resources to perform the tapping task. Tierney and Kraus (2013) reported that participants with lower tapping variability exhibited higher phase-coherence of neural responses across trials. In our study, both groups had equivalent phase-coherence of neural responses across trials. This lack of musicians' advantage may have different non exclusive explanations. First, in Tierney and Kraus study, participants had to tap along with every beat of a drum sound and this took place separately from speech ABR recordings. In our study participants had to tap along with one out two presentations of the natural speech syllable [du], which is possibly a more difficult task and the two measures were recorded simultaneously, which could differently affect brain responses. However, in our study we also had a passive listening condition, resembling to the speech ABR recording situation in Tierney and Kraus. Nonetheless, even on these data we did not find any group difference in neural response coherence, thus pointing to a lack of correlation between tapping performance and neural coherence. Another possible explanation relies in a difference in participants recruitment criteria. Indeed, in Tierney and Kraus study, participants were teenagers nonmusicians (mean age 15.8 years, SD 0.90 years; although musicianship was not precisely reported). As such, poor and proficient tappers were categorized as a function of performance, but the average accuracy of good tappers (not reported) was possibly different than in our musician group. Our participants were categorized as musicians or nonmusicians relative to the amount of musical practice and the age of practice onset. Moreover, they were young adults and not adolescents. This latter point may be crucial to explain a global lack of differences between the musician and nonmusician groups (no main effect of group). Indeed, Strait and Kraus (2014) reported that the difference between musicians and nonmusicians in speech ABR in response to sounds presented in quiet conditions only appears in school-aged children or in older adults, but

not in young adults. Moreover, differences between musicians and nonmusicians are usually found in challenging listening conditions, such as when sounds are presented in background noise (Parbery-Clark, Skoe, & Kraus, 2009) or in the presence of reverberation (Bidelman & Krishnan, 2010). Since our participants were young adults and stimuli were presented in a quiet condition, it is possible that this was not sufficient to elicit differences in subcortical representations between musicians and nonmusicians.

4.2. Local effect of attention

During the active listening, participants were asked to tap with their index finger every two stimuli. The comparison of sounds preceding and following the tap (predictive and reactive epochs, respectively) allowed us to study whether a rapid modulatory effect of attention could be visible at such a time scale (~300 ms), that is whether attention can be entrained by motor-driven rhythmic patterns (i.e. the tapping). Results show that sounds presented during the predictive periods are more efficiently encoded at the subcortical level, with higher spectral amplitudes. More specifically, these effects were observed only during the portion of the response corresponding to the vowel, and were restricted to the encoding of the second harmonic (*H*2).

Very few studies investigated whether attention can dynamically and rapidly modulate sounds processing at the subcortical level. For instance, Tierney and Kraus (2013, 2015) presented a target tone on or off the beat of popular music. They found enhanced cortical and subcortical responses to the sound when it was presented on the beat of the music. This is quite different from our study. In our study, the active listening task implies two levels of neural entrainment : one given *externally* by the rate of stimulus presentation (i.e. IOI of 303 ms) and another given *internally* by the rhythmic pattern of the tapping task (i.e. ITI of 606 ms). In the studies cited above, the level of entrainment is driven *externally* by the rate of stimulus presentation (music) and the contrast is done between a stimulus in phase with the musical structure and a stimulus that is not in phase. In our study the contrast is between two conditions

that equally fit the neural entrainment (predictive and reactive), but that differ in terms of internal predictive models generated by the tapping task. Thus, differences between our results and previous findings may also rely on different mechanisms when entrainment is driven externally or internally. Indeed, in order to predict and tap with accuracy on the onset of the upcoming target stimulus, participants have to create an internal model of the temporal regularity of the tapping. However, this internal model has to be updated continuously according to external temporal cues and the best temporal cue available is possibly the onset of the stimulus preceding the tap. The behavioral advantage of this strategy is that higher attentional levels during the predictive period would enhance the processing of timing information, thus giving a more accurate model of the temporal pattern of the tapping.

An alternative hypothesis to an enhanced attention during the preparatory period could be a reduction of global resources during motor response (i.e. reactive period). Several studies reported that auditory responses are suppressed during speech (Flinker et al., 2010; Houde, Nagarajan, Sekihara, & Merzenich, 2002). Moreover, Hairston et al. (2013) found that subcortical FFR are suppressed when participants are engaged in a perceptual discrimination task. Thus, it could be that the motor response was responsible for the reduced brainstem response observed over the reactive period, immediately following the tap.

While this rapid modulatory effect of attention was not significantly different in musicians and nonmusicians, it seems to be more strongly driven by the nonmusician group (**Figure 8**). This goes in line with the lack of global attentional effect in musicians observed when comparing active and passive listening conditions. Indeed, if the task was not challenging for the musicians, the modification of attentional resources may have been weaker than in the nonmusician group, both at the global and fast time scale. During musical training musicians need to coordinate and integrate precisely sensory and motor information. It is indeed important for musicians to be able to perceive accurately the sounds they are making while they are playing their instruments. Krause et al. (2010) have shown that expert musicians show an increased coherence between

motor (premotor cortex) and sensory regions (thalamus) as compared to nonmusicians while they are performing a tapping task (Krause et al., 2010). This network is involved in sensorimotor synchronization (Pollok, Krause, Butz, & Schnitzler, 2009) and more specifically in anticipatory motor control during auditory paced tapping task (Pollok, Gross, Kamp, & Schnitzler, 2008). The authors proposed that the increased functional interaction between auditory and motor areas in musicians would allow them to have a more accurate predictive model of the tapping task, thus possibly explaining higher tapping performances as compared to nonmusicians and a smaller cyclic modulation of attention.

To sum up, during the predictive period, an interaction between the auditory, motor and attentional systems would be at work: the focused attention would enhance timing information processing in order to update the internal predictive model of the upcoming sensory inputs; alternatively during the reactive period, the co-occurrence of a motor activity (tap) would reduce the energy resources allocated to the sound especially in nonmusicians while neural responses in musicians would be more robust to the motor activity. Overall, these findings raise the interesting question of what are the neural mechanisms underlying top-down modulation on subcortical processing (e.g. response enhancement, noise reduction and selection of sensory responses, Jääskeläinen et al., 2014; Serences, 2011). Finally, the differences in our results between predictive and reactive stimuli (separated by ~300 ms) show that adaptive changes may occur at a sub-second scale, and raise the question of how these fast dynamics of the corticofugal system may possibly be modulated. One good candidate seems to be the phase of slow cortical oscillations entrained at the pace of internal or external temporal patterns in order to improve their processing via sensory selection (Schroeder & Lakatos, 2009; Cravo, Rohenkohl, Wyart, & Nobre, 2011).

References

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1–2), 35–47. http://doi.org/10.1016/j.heares.2008.08.004

Berens, P. (2009). CircStat: A MATLAB Toolbox for Circular Statistics. *Journal of Statistical Software*, *31*(10), 1–21.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125. http://doi.org/10.1016/j.brainres.2010.07.100

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2011). Endogenous modulation of low frequency oscillations by temporal expectations. *Journal of Neurophysiology*, *106*(6), 2964–2972. http://doi.org/10.1152/jn.00157.2011

Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal Expectation Enhances Contrast Sensitivity by Phase Entrainment of Low-Frequency Oscillations in Visual Cortex. *The Journal of Neuroscience*, *33*(9), 4002–4010. http://doi.org/10.1523/JNEUROSCI.4675-12.2013

Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical Processing in Spoken Language Comprehension. *The Journal of Neuroscience*, *23*(8), 3423–3431.

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-Trial Speech Suppression of Auditory Cortex Activity in Humans. *Journal of Neuroscience*, *30*(49), 16643–16650. http://doi.org/10.1523/JNEUROSCI.1809-10.2010

Galbraith, G. C., Bhuta, S. M., Choate, A. K., Kitahara, J. M., & Mullen, T. A. (1998). Brain stem frequency-following response to dichotic vowels during attention: *NeuroReport*, *9*(8), 1889–1893. http://doi.org/10.1097/00001756-199806010-00041

Galbraith, G. C. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. [Miscellaneous Article]. *Neuroreport*, *14*(5), 735–738.

Galbraith, G. C., & Doan, B. Q. (1995). Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *International Journal of Psychophysiology*, *19*(3), 203–214. http://doi.org/10.1016/0167-8760(95)00008-G

Grimm, S., Escera, C., Slabu, L., & Costa-Faidella, J. (2011). Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. *Psychophysiology*, *48*(3), 377–384. http://doi.org/10.1111/j.1469-8986.2010.01073.x

Hackley, S. A. (1993). An evaluation of the automaticity of sensory processing using event-related potentials and brain-stem reflexes. *Psychophysiology*, *30*(5), 415–428. http://doi.org/10.1111/j.1469-8986.1993.tb02065.x

Hackley, S. A., Woldorff, M., & Hillyard, S. A. (1990). Cross-Modal Selective Attention Effects on Retinal, Myogenic, Brainstem, and Cerebral Evoked Potentials. *Psychophysiology*, *27*(2), 195–208. http://doi.org/10.1111/j.1469-8986.1990.tb00370.x

Hairston, W. D., Letowski, T. R., & McDowell, K. (2013). Task-Related Suppression of the Brainstem Frequency following Response. *PLoS ONE*, *8*(2), e55215. http://doi.org/10.1371/journal.pone.0055215

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (1994). Effect of Selective Attention on the Latency of Human Frequency-Following Potentials. *Neuroreport*, *5*(13), 1609–1612. http://doi.org/10.1097/00001756-199408150-00017

Hoormann, J., Falkenstein, M., & Hohnsbein, J. (2000). Early attention effects in human auditory-evoked potentials. *Psychophysiology*, *37*(1), 29–42.

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., & Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *Neuroreport*, *23*(1), 6–9. http://doi.org/10.1097/WNR.0b013e32834d2ffd

Houde, J. F., Nagarajan, S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Cognitive Neuroscience, Journal of*, *14*(8), 1125–1138.
Jääskeläinen, I., &#xe4, skel&#xe4, Inen, I. P., Ahveninen, J., J&#xe4, … Ahveninen. (2014). Auditory-Cortex Short-Term Plasticity Induced by Selective Attention, Auditory-Cortex Short-Term Plasticity Induced by Selective Attention. *Neural Plasticity, Neural Plasticity*, *2014, 2014*, e216731. http://doi.org/10.1155/2014/216731, 10.1155/2014/216731
Krause, V., Schnitzler, A., & Pollok, B. (2010). Functional network interactions during sensorimotor synchronization in musicians and non-musicians. *NeuroImage*, *52*(1), 245–251. http://doi.org/10.1016/j.neuroimage.2010.03.081
Kraus, N., & White-Schwoch, T. (2015). Unraveling the Biology of Auditory Learning: A Cognitive–Sensorimotor–Reward Framework. *Trends in Cognitive Sciences*, *19*(11), 642–654. http://doi.org/10.1016/j.tics.2015.08.017
Morillon, B., Schroeder, C. E., & Wyart, V. (2014). Motor contributions to the temporal precision of auditory attention. *Nature Communications*, *5*, 5255. http://doi.org/10.1038/ncomms6255
Mottonen, R., van de Ven, G. M., & Watkins, K. E. (2014). Attention Fine-Tunes Auditory-Motor Processing of Speech Sounds. *Journal of Neuroscience*, *34*(11), 4064–4069. http://doi.org/10.1523/JNEUROSCI.2214-13.2014
Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894–15898. http://doi.org/10.1073/pnas.0701498104
Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses: Auditory processing that leads to conscious perception. *Psychophysiology*, *48*(1), 4–22. http://doi.org/10.1111/j.1469-8986.2010.01114.x
Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., … Hickok, G. (2010). Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. *Cerebral Cortex*, *20*(10), 2486–2495. http://doi.org/10.1093/cercor/bhp318
Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience*, *4*. http://doi.org/10.3389/fnagi.2012.00030
Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound. *The Journal of Neuroscience*, *29*(45), 14100–14107. http://doi.org/10.1523/JNEUROSCI.3256-09.2009
Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience*, *219*, 111–119. http://doi.org/10.1016/j.neuroscience.2012.05.042
Pollok, B., Gross, J., Kamp, D., & Schnitzler, A. (2008). Evidence for Anticipatory Motor Control within a Cerebello-Diencephalic-Parietal Network. *Journal of Cognitive Neuroscience*, *20*(5), 828–840. http://doi.org/10.1162/jocn.2008.20506
Pollok, B., Krause, V., Butz, M., & Schnitzler, A. (2009). Modality specific functional interaction in sensorimotor synchronization. *Human Brain Mapping*, *30*(6), 1783–1790. http://doi.org/10.1002/hbm.20762
Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science*, *29*(2), 200–213. http://doi.org/10.1016/j.humov.2009.08.002
Rinne, T., Balk, M. H., Koistinen, S., Autti, T., Alho, K., & Sams, M. (2008). Auditory Selective Attention Modulates Activation of Human Inferior Colliculus. *Journal of Neurophysiology*, *100*(6), 3323–3327. http://doi.org/10.1152/jn.90607.2008

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal Expectation Improves the Quality of Sensory Information. *The Journal of Neuroscience*, *32*(24), 8424–8428. http://doi.org/10.1523/JNEUROSCI.0804-12.2012

Schmidt-Kassow, M., & Kotz, S. A. (2009). Attention and perceptual regularity in speech: *NeuroReport*, *20*(18), 1643–1647. http://doi.org/10.1097/WNR.0b013e328333b0c6

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. http://doi.org/10.1016/j.tins.2008.09.012

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–113. http://doi.org/10.1016/j.tics.2008.01.002

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of Active Sensing and Perceptual Selection. *Current Opinion in Neurobiology*, *20*(2), 172–176. http://doi.org/10.1016/j.conb.2010.02.010

Serences, J. T. (2011). Mechanisms of Selective Attention: Response Enhancement, Noise Reduction, and Efficient Pooling of Sensory Responses. *Neuron*, *72*(5), 685–687. http://doi.org/10.1016/j.neuron.2011.11.005

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302–324. http://doi.org/10.1097/AUD.0b013e3181cdb272

Slabu, L., Escera, C., Grimm, S., & Costa-Faidella, J. (2010). Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. *European Journal of Neuroscience*, *32*(5), 859–865. http://doi.org/10.1111/j.1460-9568.2010.07324.x

Slabu, L., Grimm, S., & Escera, C. (2012). Novelty Detection in the Human Auditory Brainstem. *The Journal of Neuroscience*, *32*(4), 1447–1452. http://doi.org/10.1523/JNEUROSCI.2557-11.2012

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research*, *308*, 109–121. http://doi.org/10.1016/j.heares.2013.08.004

Teder-Sälejärvi, W. A., & Hillyard, S. A. (1998). The gradient of spatial auditory attention in free field: an event-related potential study. *Perception & Psychophysics*, *60*(7), 1228–1242.

Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience. *PLoS ONE*, *10*(6), e0128839. http://doi.org/10.1371/journal.pone.0128839

Tierney, A., & Kraus, N. (2013). Neural responses to sounds presented on and off the beat of ecologically valid music. *Frontiers in Systems Neuroscience*, *7*. http://doi.org/10.3389/fnsys.2013.00014

Tierney, A., & Kraus, N. (2013). The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. *Journal of Neuroscience*, *33*(38), 14981–14988. http://doi.org/10.1523/JNEUROSCI.0612-13.2013

Tierney, A., & Kraus, N. (2015). Neural Entrainment to the Rhythmic Structure of Music. *Journal of Cognitive Neuroscience*, *27*(2), 400–408. http://doi.org/10.1162/jocn_a_00704

van Atteveldt, N., Musacchia, G., Zion-Golumbic, E., Sehatpour, P., Javitt, D. C., & Schroeder, C. (2015). Complementary fMRI and EEG evidence for more efficient neural processing of rhythmic vs. unpredictably timed sounds. *Frontiers in Psychology*, *6*, 1663. http://doi.org/10.3389/fpsyg.2015.01663

Zhang, C., Arnott, S. R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X., … Schneider, B. A. (2016). Attentional modulation of informational masking on early cortical representations of speech signals. *Hearing Research*, *331*, 119–130. http://doi.org/10.1016/j.heares.2015.11.002

Quatrième partie - DISCUSSION et PERSPECTIVES

Discussion générale et perspectives

12. Résumé des résultats

Dans notre première étude, en comparant l'encodage sous-cortical de sons natifs ou non-natifs chez un groupe de français et d'américains, nous avons constaté que l'effet d'expertise linguistique apparaît spécifiquement au niveau de l'encodage des fréquences formantiques (F1) du stimulus français, ce qui a confirme l'hypothèse selon laquelle l'expérience linguistique influence le traitement automatique de la parole, et que cette plasticité se produit en fonction de ce qui est important pour le locuteur. Un effet d'expertise linguistique plus global a été observé à travers les deux stimuli : chez les américains, pour qui la hauteur tonale joue un rôle fonctionnel important, l'encodage spectral de la fréquence fondamentale (F0) s'est avéré plus robuste par rapport aux français, et ce de manière indépendante au stimulus.

Plusieurs travaux ont montré que les effets d'expertise linguistique peuvent être en partie compensés par la pratique musicale (Chobert and Besson, 2013; Wong et al., 2007). Dans notre deuxième étude, nous avons voulu déterminer si la pratique musicale permettait de compenser les effets de l'expertise linguistique que nous avons observé précédemment entre Français et américains. En comparant l'encodage d'une syllabe anglaise chez des musiciens et non-musiciens français et des américains, nous avons démontré que les différences d'expertise linguistique sont réduites voire supprimées par la pratique musicale, puisque les musiciens français présentaient un encodage neuronal de la fréquence fondamentale et des fréquences formantiques aussi robuste que celui des américains natifs.

La spécificité des effets de plasticité dépendants de l'expérience que nous avons observés au cours de ces deux études suggère que des processus *top-down* seraient mis en jeu afin de renforcer de manière sélective l'encodage des informations les plus pertinentes (F0, formants). Nous avons postulé que l'attention pourrait induire à une échelle de temps plus courte des adaptations fonctionnelles rapides afin d'améliorer spécifiquement le traitement des indices acoustiques les plus importants. Dans notre troisième étude, nous avons donc testé l'effet du contexte attentionnel sur l'encodage sous-cortical des sons de parole. En comparant les réponses neuronales en condition d'écoute passive et active (tapping), les résultats ont révélé une

représentation de la fréquence fondamentale plus robuste lors de l'écoute active des sons, confirmant l'influence de processus *top-down* sur l'encodage sous-cortical des sons de parole. Le design expérimental de cette étude nous a aussi permis d'étudier les modulations dynamiques et rapides de l'attention. Plus précisément, en entraînant l'attention au rythme d'un stimulus sur deux via la tâche de tapping, nous avons observé une modulation significative de la représentation d'une harmonique en fonction de la phase du tapping. De manière intéressante, nous avons aussi montré, en séparant les sujets en fonction de leur niveau d'expertise musicale, que les modulations de l'encodage des sons en fonction du niveau attentionnel étaient principalement dues au groupe de sujets non-musiciens.

13. Discussion

Le but de cette thèse était d'une part de déterminer si l'expertise (linguistique et musicale) a une influence sur l'encodage sous-cortical des sons de parole, et d'autre part de déterminer si ces effets d'expertise se mettent en place à une échelle de temps plus courte via des modulations attentionnelles *top-down*.

Les effets de l'expertise linguistique on été principalement étudiés par l'équipe de Krishnan et collaborateurs, en comparant l'encodage sous-cortical de syllabes tonales chez un groupe de sujets natifs (chinois Mandarins) et non-natifs (anglais). Ces études ont tout d'abord révélé que la langue maternelle avait une influence sur le traitement des sons de parole au niveau sous-cortical, puisque les chinois Mandarins montraient un encodage plus robuste des variations de la F0 comparés aux anglais (Krishnan et al., 2005). Les résultats ont montré que l'avantage des chinois n'était pas global, mais s'observait de manière spécifique au niveau des variations les plus rapides de la F0, ce qui suggère que l'expertise linguistique confère un avantage aux locuteurs natifs au niveau de l'encodage des indices acoustiques qui véhiculent les informations linguistiques les plus importantes, c'est-à-dire celles qui permettent de distinguer les mots monosyllabiques en Chinois Mandarin. Dans une étude ultérieure, les auteurs ont voulu déterminer si l'amélioration du traitement de la F0 reflétait des effets dépendants d'un contexte langagier, ou plus généralement d'une exposition plus importante à des patterns de *pitch* par rapport aux non-natifs. En effet, des études précédentes avaient montré que chez les musiciens, qui sont exposés de manière intensive et répétée à différents patterns de *pitch* musicaux, les représentations sous-corticales du *pitch* sont plus robustes que chez des non-musiciens, et ce quelque soit le contexte, musical ou linguistique, des sons perçus (Musacchia et al., 2007). Afin de

déterminer si les effets d'expertise linguistique qu'ils avaient observés étaient dépendants d'un contexte linguistique ou non, l'équipe de Krishnan a donc utilisé des sons présentant les mêmes variations de *pitch* que les stimuli Mandarins testés dans leur étude précédente, mais dépourvus de contexte lexico-sémantique (*Iterated Rippled Noise*, *IRN* stimuli). Les résultats ont montré que les chinois présentaient toujours un encodage plus robuste du *pitch* par rapport aux anglais (Krishnan et al., 2009). En somme, ces études ont mis en évidence que les effets de plasticité sous-corticale observés étaient liés à une similarité des patterns de *pitch* présents dans les stimuli auxquels les locuteurs sont exposés plutôt qu'à un contexte (langagier ou non) ou à un domaine d'expertise (musical ou linguistique).

A partir de ce constat, nous avons voulu déterminer s'il existait des effets de plasticité sous-corticale spécifiques à l'expertise linguistique. Dans ce but, nous avons comparé l'encodage neuronal de sons natifs et non-natifs chez des locuteurs américains et français, pour lesquelles les différences d'expertise, plus subtiles, reposent essentiellement sur l'organisation de leurs répertoires phonémiques. Cette comparaison nous assurait ainsi que les effets d'expertise potentiels ne seraient pas nécessairement dépendants d'autres processus de traitement plus généraux tels que ceux nécessaires à l'encodage du *pitch*. En combinant les résultats de notre première et deuxième étude, nous pouvons observer d'une part qu'il existe des différences d'encodage de la F0 entre américains et français de manière indépendante à la langue du stimulus, ce qui rejoint les conclusions des études précédentes suggérant que la robustesse de l'encodage du *pitch* serait indépendante du domaine ou du contexte (linguistique ou non). Cet effet d'expertise sur la F0 serait donc lié à une exposition plus importante chez les américains à des patterns de *pitch* variés et complexes par rapport aux Français (Duanmu, 2004), mais aussi au rôle linguistique plus important que ces indices acoustiques jouent en Anglais, notamment au niveau suprasegmental, à la différence du Français qui est une langue non-stressée et pour laquelle les variations de F0 ne sont pas porteuses de sens (Braun et al., 2014; Vaissière, 1991). D'autre part, une découverte majeure de nos études est qu'il existe des différences d'encodage spécifiques au niveau des formants, qui jouent un rôle important dans la discrimination des phonèmes dans les langues non-tonales. Les locuteurs natifs ont un encodage du premier formant plus robuste que les locuteurs non-natifs, alors qu'au niveau des autres composantes fréquentielles (hormis la F0), aucune différence significative n'a été observée. Ces résultats révèlent qu'il existe des effets de plasticité dépendants du niveau d'expertise linguistique, et que cette plasticité se produit en fonction de ce qui est important pour le locuteur. De plus, étant donné que les réponses neuronales ont été recueillies de manière passive, ces résultats apportent la preuve que l'exposition et la pratique de langue maternelle induit sur le long terme des réorganisations

neuronales qui viennent renforcer le traitement automatique des indices acoustiques les plus pertinents pour le locuteur.

Notre deuxième étude avait pour but d'étudier les effets de transfert entre musique et langage au niveau sous-cortical. Les effets de l'expertise musicale sur le traitement des sons de la langue maternelle ont été particulièrement bien étudiés (Parbery-Clark et al., 2012; Strait et al., 2013). Cependant, nous en savons encore peu sur l'effet que peut avoir la pratique musicale sur la perception et la production des langues étrangères (Zeromskaite, 2014). De plus, le peu d'études disponibles sur le sujet se sont principalement focalisées sur les effets de transfert entre musique et langues tonales, puisque ces domaines d'expertise exigent tous deux une précision importante dans le traitement de la hauteur tonale (Wong et al., 2007). De ce fait, nous avons voulu déterminer si la pratique musicale pouvait compenser les différences d'expertise linguistique dans les langues non-tonales. Les résultats ont confirmé notre hypothèse puisque les musiciens français présentaient un traitement de la syllabe anglaise aussi robuste que celui des américains : d'une part, les musiciens présentaient un encodage de la F0 aussi robuste que les Américains, ce qui rejoint les résultats des études précédentes montrant un renforcement de l'encodage de la F0 chez les musiciens (Bidelman et al., 2011; Wong et al., 2007); d'autre part, l'effet compensatoire de la pratique musicale a aussi été observé au niveau de l'encodage des fréquences formantiques. Ces résultats s'inscrivent plus largement dans la littérature montrant un effet positif de la pratique musicale sur la perception et la production des langues étrangères (Chobert and Besson, 2013; Delogu et al., 2010; Lee and Hung, 2008; Marie et al., 2010), et apportent un nouvel éclairage sur les mécanismes neuronaux qui pourraient expliquer l'avantage des musiciens dans l'apprentissage de nouvelles langues. Il est possible que chez les musiciens la pratique intense et répétitive de la musique améliore leurs capacités d'encodage automatique des sons au niveau des dimensions acoustiques les plus importantes. Ces résultats sont concordants avec les études réalisées au niveau cortical, montrant une détection automatique plus efficiente des variations de *pitch* (Alexander et al., 2005; Delogu et al., 2010; Marie et al., 2010) et du VOT (Sadakata and Sekiyama, 2011) de sons étrangers chez les musiciens adultes.

Ces études ont mis en évidence un autre point important : les phénomènes de plasticité dépendants de l'expérience se produisent de manière spécifique au niveau des dimensions acoustiques les plus importantes (F0, formants...). Ces découvertes viennent aussi appuyer l'hypothèse selon laquelle cette plasticité serait dirigée par des processus *top-down* venant renforcer de manière sélective le traitement des indices acoustiques les plus importants. Récemment, Kraus & White-Schwoch (2015) ont proposé une nouvelle vision du système auditif comme un circuit distribué, mais intégré, dans lequel le traitement auditif met en jeu non seulement le système

auditif, mais aussi le système sensori-moteur et attentionnel. De plus, il existe au niveau anatomique de nombreuses projections descendantes du cortex vers les structures sous-corticales qui peuvent supporter les modulations *top-down* sur l'activité sous-corticale (Kral and Eggermont, 2007). Il est possible que des processus de haut niveau soient à l'origine de la spécificité des effets de plasticité dépendants de l'expérience observés au niveau sous-cortical. Nous avons postulé que l'attention pourrait moduler de manière *top-down* le flux d'informations acoustiques entrant au niveau des structures sous-corticales afin de renforcer sélectivement certains indices acoustiques. Cette approche rejoint aussi la notion d'*active sensing* qui soutient le fait que la perception est un phénomène actif par lequel l'information de l'environnement est échantillonnée et structurée via une interaction continue entre les systèmes sensoriels, attentionnels et moteurs (Schroeder et al., 2010). En effet, en général les sons de notre environnement ne nous parviennent pas de manière isolée, mais sont plutôt mélangés au bruit environnant. Il est donc important d'être capable d'extraire les informations acoustiques porteuses de sens. Dans ces conditions d'écoutes difficiles, l'attention jouerait un rôle important en venant renforcer le traitement précoces des sons au niveau cortical (Zhang et al., 2016). Nous avons postulé que l'attention pourrait intervenir avant même l'arrivée de l'influx nerveux au niveau cortical, c'est-à-dire au niveau des structures sous-corticales. Les résultats ont confirmé notre hypothèse puisque les représentations spectrales et la cohérence des réponses neuronales se sont avérées plus robustes au niveau de la F0 durant la phase d'écoute active. Bien que les effets de l'attention sur les réponses sous-corticales aient été sujets à débat pendant longtemps, les récentes études (Galbraith et al., 1998, 2003; Galbraith and Doan, 1995; Hoormann et al., 2000, 1994) et nos résultats semblent converger, et renforcent l'hypothèse selon laquelle les processus attentionnels modulent de manière précoce l'encodage des sons via le système corticofuge.

Le temps est une dimension particulièrement importante dans la perception des sons puisque, dans un environnement naturel, les signaux acoustiques sont transmis à travers plusieurs échelles temporelles et sont organisés sous forme de patterns rythmiques ou quasi-rythmiques (e.g. langage, musique). La *Dynamic Attending Theory* (*DAT*) propose que lorsque les évènements acoustiques se déroulent dans le temps avec régularité, l'occurrence des évènements à venir devient prédictible et leur traitement plus efficient. Aussi, ces phénomènes de prédiction reposeraient sur un entraînement des niveaux attentionnels en phase avec la régularité temporelle des évènements externes, de manière à ce que les évènements coïncidant avec des moments à hauts niveaux attentionnels soient traités de manière plus optimale (Jones et al., 2002; Large and Jones, 1999). Dans ce cadre théorique, nous avons fait l'hypothèse qu'en demandant aux sujets de

taper en phase avec la présentation d'un stimulus sur deux, leur attention serait entraînée au même rythme. Les résultats ont confirmé en partie nos attentes puisque nous avons observé une variation significative de la représentation de la deuxième harmonique en fonction de la phase du tapping. En résumé, notre étude montre d'une part que l'activité du tronc cérébral est affectée par l'attention portée au stimulus, et d'autre part que l'attention suit une dynamique oscillatoire lorsque celle-ci est entraînée par la commande motrice (tapping). Il est possible que dans des conditions d'écoute naturelle, les rythmes attentionnels se synchronisent à différentes échelles temporelles en phase avec la production de syllabes, de mots et de phrases, ce qui entraînerait des modulations de l'encodage des sons afin de renforcer les indices acoustiques les plus pertinents à chaque niveau d'intégration (e.g. le VOT et les formants pour les consonnes, les formants pour les voyelles, la F0 pour la position du stress syllabique, etc...).

A travers ces travaux, nous avons montré que les réponses du tronc cérébral sont sensibles au niveau d'expertise linguistique ou musicale. La spécificité des effets de plasticité suggère que ces réorganisations sont dépendantes de processus *top-down* permettant de renforcer sélectivement l'encodage des informations les plus pertinentes. Ces processus peuvent être modulés par une dynamique rapide et oscillatoire des niveaux attentionnel lorsque ces derniers sont entraînés par la commande motrice. Il est possible que les réorganisations du système sensoriel induites par l'expérience résultent d'une interaction dynamique entre les structures sous-corticales et les processus cognitifs de hauts niveaux via le système corticofuge (Banai et al., 2009; Krishnan et al., 2005; Musacchia et al., 2007; Russo et al., 2005; Song et al., 2008; Tzounopoulos and Kraus, 2009). Dans le chapitre suivant nous allons tenter d'expliquer comment ces interactions se produisent au cours de la vie d'un individu, et quels sont les mécanismes neuronaux sous-jacents.

14. Perspectives

14.1. Tentative d'unification des effets attentionnels à court terme et des effets d'expertises à long terme

Les effets de plasticité dépendants de l'expérience pourraient reposer sur des mécanismes *bottom-up* « guidés » par des mécanismes *top-down*. A la naissance, le répertoire phonémique du

nouveau-né n'est pas établit, les compétences linguistiques du bébé reposent alors essentiellement sur des processus *bottom-up* . Le bébé est aussi capable d'utiliser les propriétés statistiques de la parole afin de segmenter de manière automatique (*statistical learning*) un flux de parole continu en mots (Bortfeld et al., 2005; Saffran et al., 1996). A ce stade, les processus *bottom-up* pourraient être prédominants, et les processus *top-down* auraient une plus faible influence sur le traitement des sons de parole (**Figure 36**). Le répertoire phonémique du bébé est encore "en construction". A cela s'ajoute le fait que les phonèmes naturels présentent une grande variabilité acoustique, c'est-à-dire qu'une même syllabe peut être prononcée différemment par le même locuteur et par des locuteurs différents. Il en résulte un recouvrement important des catégories phonémiques. Par exemple, le /o/ et le /u/ étant très proches au niveau acoustique, la variabilité acoustique liée à la prononciation de ces voyelles par différents locuteurs induit un recouvrement important de ces phonèmes au niveau de l'espace vocalique, et par conséquent au niveau des représentations neuronales. De ce fait, il est difficile pour le bébé d'inférer les indices acoustiques qui sont importants pour la discrimination de ces phonèmes. Nous pouvons alors supposer que les réponses neuronales du tronc cérébral sont, à ce niveau de développement, peu ou pas modulées par le niveau attentionnel. En conséquence, les représentations spectrales des sons au niveau sous-cortical seraient identiques en condition d'écoute passive ou active. Le répertoire phonémique de l'enfant se développerait donc initialement de manière passive, par l'exposition répétée aux stimuli présents dans son environnement. Au cours du développent, l'enfant devient progressivement sensible aux phonèmes de sa langue maternelle, et insensible à ceux des langues étrangère auxquelles l'enfant est moins exposé (Cheour et al., 1998; Kuhl et al., 2006, 1992; Rivera-Gaxiola et al., 2005).

Au fur et à mesure que se construisent les représentations abstraites des phonèmes, l'enfant devient capable d'identifier les indices acoustiques qui vont l'aider à discriminer de manière plus efficiente les sons présents dans son environnement. Par exemple, la discrimination des voyelles /a/, /e/ et /i/ repose essentiellement sur des rapports formantiques (F1/F2) différents. Il devient donc important pour l'enfant d'encoder ces informations de manière plus fine et plus robuste. Progressivement, des mécanismes de modulation *top-down* se mettent en place via l'attention dirigée vers les composantes les plus pertinentes d'un point de vue phonologique (formants, VOT...), ceci dans le but de renforcer leurs représentations neuronales (corticales et sous-corticales). De manière intéressante, il a été proposé que les représentations de haut niveau auraient des influences *top-down* sur les traitements des informations sensorielles entrantes (Tenenbaum et al., 2011). Autrement dit, les connaissances linguistiques précédemment acquises créent des prédictions qui vont influencer l'encodage des sons de parole (Lew-Williams and

Saffran, 2012). Nous verrons à la fin de ce chapitre quels sont les mécanismes neuronaux potentiels qui sont à la base des interactions dynamiques entre les structures de haut et de bas niveau, et pourquoi les oscillations neuronales pourraient être un bon candidat pour expliquer les phénomènes de prédictions *top-down*.

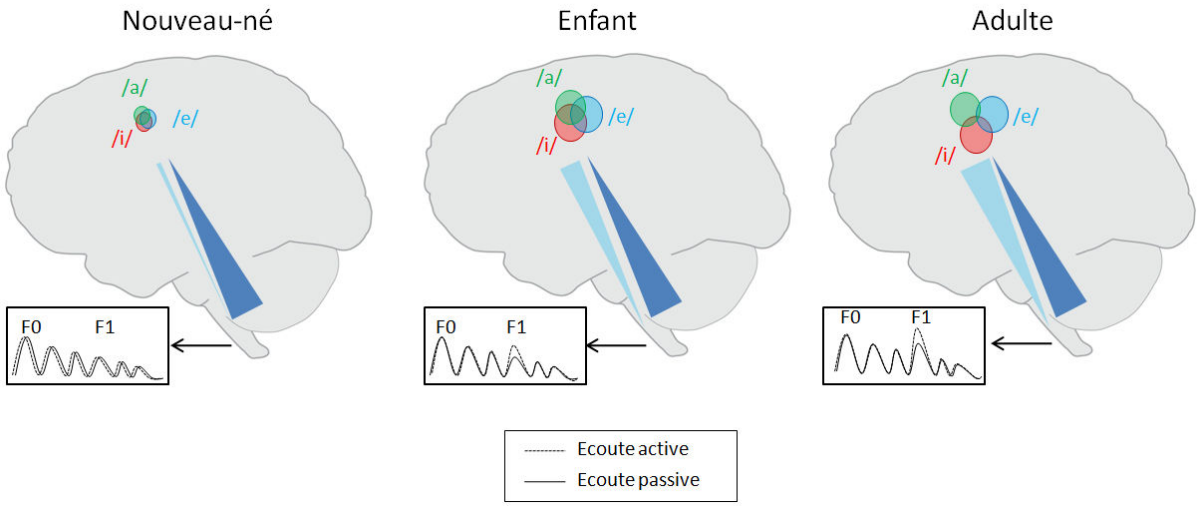


Figure 36. Schéma d'évolution développementale du répertoire phonologique et des voies ascendantes et descendantes. Le répertoire phonologique est représenté de manière simplifiée par les représentations chevauchantes des voyelles /a/, /e/ et /i/. Les voies ascendantes sont représentées par un triangle bleu foncé, et les voies descendantes par un triangle bleu clair. L'encadré illustre les représentations spectrales des sons au niveau sous-cortical à la façon d'une transformée de Fourier, en condition d'écoute passive (trait plein) et active (trait pointillé). Le répertoire phonémique se construit initialement par des processus *bottom-up*. Au fur et à mesure du développement des représentations phonémiques, les processus *top-down* prennent place afin de renforcer de manière dynamique et rapide les indices acoustiques importants (formants) en condition d'écoute active. Chez l'adulte, la répétition de ce processus a induit sur le long terme une réorganisation des structures sous-corticales qui deviennent alors plus efficientes dans le traitement des informations pertinentes même en condition d'écoute passive.

Au fur et à mesure de l'exposition répétée à la langue maternelle, le renforcement sélectif de l'encodage neuronal de certaines dimensions acoustiques via l'attention va entraîner une réorganisation des structures de bas niveau telles que le tronc cérébral, de telle manière que le traitement des indices acoustiques importants devient plus efficace, même en condition d'écoute passive, c'est-à-dire de manière automatique. Dans nos deux premières études, nous avons en effet montré que lors de l'écoute passive de sons de parole, l'encodage neuronal des indices acoustiques importants (F0 et formants) était amélioré chez le locuteur natif comparé au non-natif. Comme les réponses neuronales ont été recueillies dans des conditions d'écoute passive,

nous pouvons en déduire que les effets d'expertise observés reposent sur des réorganisations neuronales intrinsèques aux structures sous-corticales probablement initiées par des adaptions rapides telles que celles décrites précédemment.

14.2. Mécanismes neuronaux potentiels des modulations attentionnelles *top-down*

Dans le contexte de la perception et de la production de la parole, les modèles de prédiction interne apportent une perspective intéressante à nos résultats et aux futures études à mener. Ces modèles proposent qu'il est possible de prédire l'entrée sensorielle à venir ou les conséquences sensorielles d'une action motrice (Arnal and Giraud, 2012). Cela signifie qu'un locuteur peut prédire les prochains sons de parole ou les conséquences sensorielles d'une action motrice telle que la production d'un mot. Nous allons ici nous focaliser sur les modèles de prédictions appliqués à la perception de la parole puisqu'il s'agit du sujet de cette thèse.

Lorsque les évènements externes présentent une régularité temporelle, ce qui est le cas de la parole et de la musique, il est possible de prédire leurs occurrences, de ce fait leur encodage est optimisé. La *Dynamic of Attending Theory* (*DAT*) suggère que la prédiction temporelle est le résultat d'un entraînement des niveaux attentionnels par la structure rythmique des évènements acoustiques externes (Large and Jones, 1999). Les oscillations neuronales semblent être le mécanisme neuronal le plus effectif pour suivre la rythmicité des évènements acoustiques à différentes échelles temporelles (Giraud and Poeppel, 2012; Kraus and White-Schwoch, 2016). Etant donné que les oscillations neuronales sont supposées refléter les variations rythmiques de l'état d'excitabilité des populations de neurones (Schroeder and Lakatos, 2009), lorsque ces oscillations sont entraînées par un flux acoustique rythmique, l'attention peut utiliser les différents rythmes d'excitabilité neuronale afin d'optimiser de manière sélective les réponses neuronales lorsque les stimuli apparaissent au moment attendu (Lakatos et al., 2008). De nombreuses études rapportent que les oscillations neuronales lentes (delta, thêta) seraient le substrat neuronal des prédictions temporelles (Luo and Poeppel, 2007). Plus précisément, il a été proposé que le recalage de phase des oscillations corticales lentes (delta, thêta) durant la période qui précède l'apparition du stimulus constituerait le mécanisme permettant de moduler le gain sensoriel lorsque les stimuli sont prédictibles (Arnal and Giraud, 2012, **Figure 37**). D'autres études montrent que l'entraînement des oscillations neuronales n'est pas restreint aux basses fréquences, mais est aussi observé dans les gammes de fréquences supérieures, alpha, beta et gamma (Fujioka

et al., 2009; Gross et al., 2013). Par exemple, dans le contexte de la perception de la parole, les propriétés quasi-rythmiques de la parole telles que les phonèmes, les syllabes, les mots et les phrases sont suivies à différentes échelles temporelles par les rythmes delta, thêta et gamma, ce qui suggère que les oscillations neuronales servent de support à la ségrégation du flux de parole en unité discrètes (Giraud and Poeppel, 2012, **Figure 38**). Ces études montrent que l'information auditive est échantillonnée via les oscillations neuronales à plusieurs échelles temporelles, et que l'entraînement neuronal est le résultat d'interactions continues entre les informations entrantes *bottom-up* et les prédictions internes *top-down*, ces dernières étant induites par la structure rythmique des évènements acoustiques.

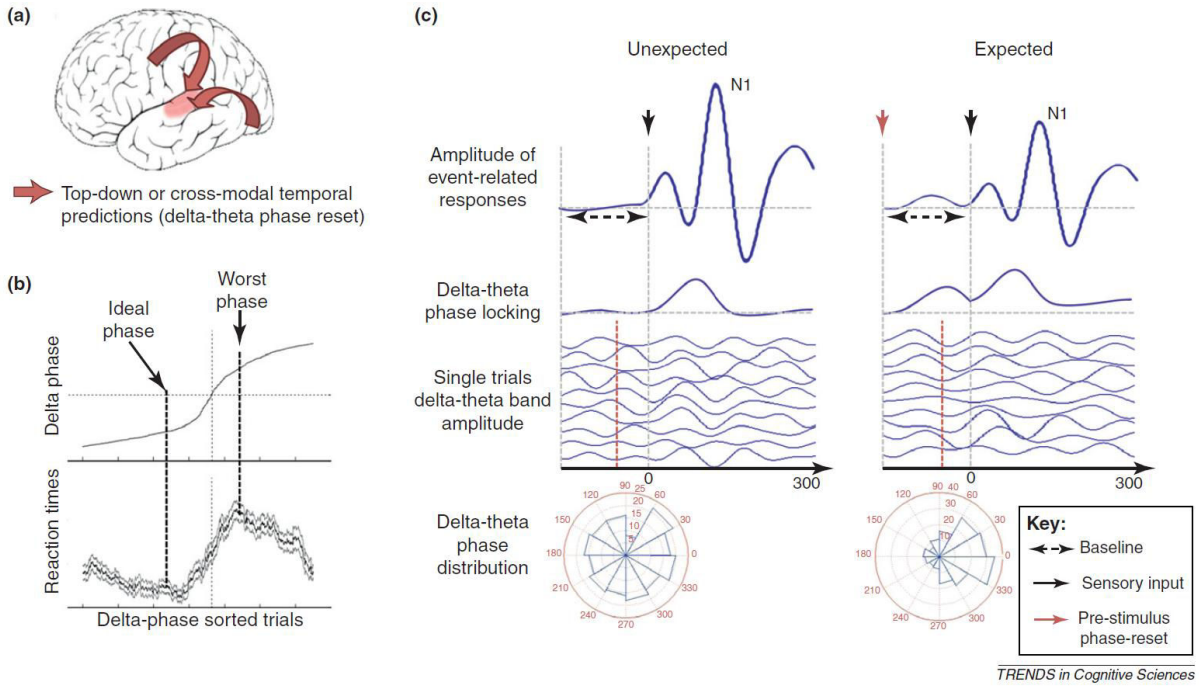


Figure 37. Oscillation neuronales et prédictions temporelles. (a) Les signaux temporels prédictifs modulent le traitement auditif via un recalage de phase des ondes corticales lentes. Les signaux temporels peuvent être générés de manière interne ou externe. (b) L'alignement de la phase des ondes delta module le traitement sensoriel et les réponses comportementales. (c) Ce schéma compare l'amplitude de la réponse évoquée N1 aux sons non-attendus (panneau gauche) et attendus (panneau droit). L'amplitude des réponses évoquées [*event-related potentials* (*ERPs*) et *event-related fields* (*ERFs*)] dépend du recalage de phase des ondes delta-thêta durant la période pré-stimulus. Quand un son est correctement anticipé (panneau droit) via la génération de prédictions rythmiques, la phase des ondes delta-thêta se recale juste avant l'apparition du stimulus, ce qui génère une réponse (*ERPs, ERFs*) réduite en amplitude. A l'inverse, lorsque le stimulus n'est pas prédictible, les réponses sont évoquées uniquement par l'apparition du stimulus et sont plus amples. Issue et adaptée d'Arnal et Giraud (2012).

En outre, Lakatos et al. (2005) ont montré que l'amplitude de chaque bande de fréquence neuronale est modulée par la phase de fréquences plus basse (**Figure 38**). Le couplage entre la phase des oscillations neuronales lentes (delta, thêta) et rapides (beta, gamma) est donc probablement le mécanisme neuronal sous-jacent des prédictions sensorielles (Womelsdorf et al., 2007). Le *phase-amplitude coupling* (*PAC*) est aussi connu comme un mécanisme de communication entre des régions neuronales distantes (Fries, 2005). En conséquence, la phase des ondes lentes entraînée par les patterns temporels externes ou internes, modulerait l'activité des régions sensorielles distantes de manière à optimiser l'encodage des informations acoustiques entrantes (Cravo et al., 2011).

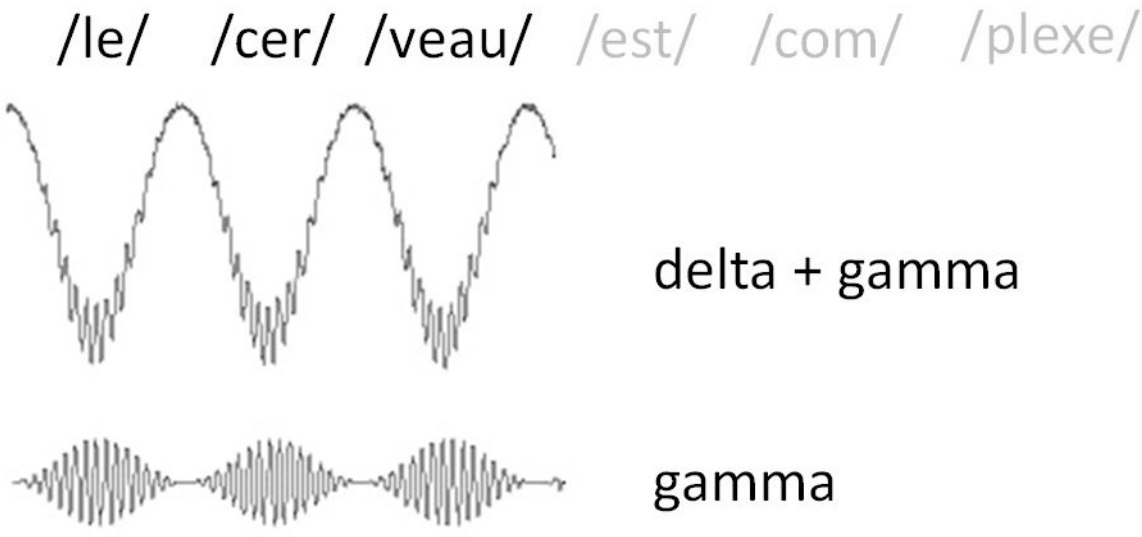


Figure 38. Couplage entre la phase des ondes lentes (delta) et l'amplitude des ondes rapides (gamma)**.** La phase des ondes lentes est entraînée par la cadence des syllabes prononcées (lettres noires), ce qui permet une ségrégation du flux de parole en unité discrètes, et de prédire l'occurrence des syllabes suivantes (lettres en gris clair). La phase des ondes delta est couplée à l'amplitude des ondes gamma, ce permet d'optimiser l'encodage des syllabes qui surviennent au moment attendu. Issue et adaptée de http://neuroimage.usc.edu/brainstorm/.

D'après ces travaux, nous pouvons supposer que les interactions dynamiques *top-down* entre structures corticales et sous-corticales reposeraient sur les mêmes mécanismes neuronaux. Par exemple, lors de l'écoute de la phrase "le cerveau est complexe", au niveau des régions associatives la phase des ondes corticales lentes (delta) serait entraînée à la cadence de prononciation des syllabes (voir **Figure 38**), ce qui permet d'une part de segmenter le flux de parole en unités distinctes, et d'autre part, de prédire l'occurrence des prochaines syllabes.

L'activité des régions sensorielles distantes (cortex auditif primaire, structures sous-corticales) serait modulée par la phase des ondes lentes, de telle manière que les syllabes dont l'occurrence coïncide avec des niveaux d'activité neuronale élevés soient mieux traitées. Il est possible que cette optimisation du traitement sensoriel se traduise au niveau sous-cortical par un encodage plus robuste des indices phonologiques les plus pertinents (F0, formants, etc...). Il serait intéressant dans de futures études d'étudier la relation de phase-amplitude entre les ondes corticales lentes (delta, thêta) entraînées au rythme de stimulations externes (fréquence syllabique, fréquence de présentation des stimuli) ou internes (tapping) et les différentes composantes spectrales des réponses sous-corticales (F0, formants). Par exemple, en comparant le couplage entre la phase des fréquences delta-thêta et l'amplitude des *cABRs*, selon différentes fréquences de présentation des stimuli (entraînement externe) et/ou différentes fréquences de tapping (entraînement interne). Nous pouvons prédire que le couplage de phase-amplitude se produira de manière spécifique en fonction de la fréquence de stimulation ou de tapping.

Conclusion

Les réponses complexes du tronc cérébral offrent une vue objective unique sur le fonctionnement dynamique du système auditif. Bien que l'enregistrement de ces réponses puisse être au premier abord fastidieux, une fois la technique maîtrisée, l'analyse de ces réponses neuronales permet d'apporter une nouvelle perspective sur la manière dont le cerveau traite l'information de son environnement. A travers ce travail de thèse, nous avons montré que la perception auditive n'est finalement pas un phénomène passif, indépendant de l'individu qui la perçoit. Au contraire, ces travaux ont mis en évidence l'interdépendance entre le monde extérieur et intérieur, entre l'environnement et le cerveau. Autrement dit, l'environnement acoustique affecte le fonctionnement de notre cerveau, ce qui en retour module notre perception des sons. Ces interactions continues entre le cerveau et son environnement se déroulent à plusieurs échelles temporelles : à court terme via des processus de prédiction attentionnelle, qui peuvent engendrer sur le long terme une réorganisation de l'architecture cérébrale à plusieurs niveaux.

Ces découvertes offrent également de nouvelles stratégies thérapeutiques chez les patients présentant des troubles de la perception de la parole. En démontant les bénéfices induits par la pratique musicale sur le traitement fin des sons de parole, la musique apparaît comme une méthode rééducative prometteuse.

Dans un cadre théorique plus large, l'ensemble de ces découvertes renforcent l'idée selon laquelle la réalité dans laquelle nous évoluons au quotidien n'est pas indépendante de celui qui perçoit, mais serait plutôt le résultat d'une interaction entre l'environnement et la projection de nos attentes.

Références bibliographiques

Abdul-Kareem, I.A., Stancak, A., Parkes, L.M., Al-Ameen, M., AlGhamdi, J., Aldhafeeri, F.M., Embleton, K., Morris, D., Sluming, V., 2011. Plasticity of the Superior and Middle Cerebellar Peduncles in Musicians Revealed by Quantitative Analysis of Volume and Number of Streamlines Based on Diffusion Tensor Tractography. The Cerebellum 10, 611–623. doi:10.1007/s12311-011-0274-1

Abrams, D.A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D.J., Menon, V., 2011. Decoding Temporal Structure in Music and Speech Relies on Shared Brain Resources but Elicits Different Fine-Scale Spatial Patterns. Cereb. Cortex 21, 1507–1518. doi:10.1093/cercor/bhq198

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464. doi:10.1016/j.tics.2004.08.011

Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47. doi:10.1016/j.heares.2008.08.004

Aiken, S.J., Picton, T.W., 2006. Envelope Following Responses to Natural Vowels. Audiol. Neurotol. 11, 213–232. doi:10.1159/000092589

Akhoun, I., Berger-Vachon, C., Collet, L., 2010. Approche empirique pour déterminer les générateurs neurophysiologiques sous-jacents des potentiels évoqués auditifs engendrés par des sons de parole (Speech ABR). IRBM 31, 242–256. doi:10.1016/j.irbm.2009.12.005

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clin. Neurophysiol. 119, 922–933. doi:10.1016/j.clinph.2007.12.010

Akhoun, I., Moulin, A., Jeanvoine, A., Ménard, M., Buret, F., Vollaire, C., Scorretti, R., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. Speech auditory brainstem response (speech ABR) characteristics depending on recording conditions, and hearing status: An experimental parametric study. J. Neurosci. Methods 175, 196–205. doi:10.1016/j.jneumeth.2008.07.026

Alexander, J.A., Wong, P.C., Bradlow, A.R., 2005. Lexical tone perception in musicians and non-musicians., in: Interspeech. pp. 397–400.

Amunts, K., Schlaug, G., Jäncke, L., Steinmetz, H., Schleicher, A., Dabringhaus, A., Zilles, K., 1997. Motor cortex and hand motor skills: structural compliance in the human brain. Hum. Brain Mapp. 5, 206–215.

Anderson, B.J., Eckburg, P.B., Relucio, K.I., 2002. Alterations in the Thickness of Motor Cortical Subregions After Motor-Skill Learning and Exercise. Learn. Mem. 9, 1–9. doi:10.1101/lm.43402

Anderson, B.J., Li, X., Alcantara, A.A., Isaacs, K.R., Black, J.E., Greenough, W.T., 1994. Glial hypertrophy is associated with synaptogenesis following motor-skill learning, but not with angiogenesis following exercise. Glia 11, 73–80. doi:10.1002/glia.440110110

Arnal, L.H., Giraud, A.-L., 2012. Cortical oscillations and sensory predictions. Trends Cogn. Sci. 16, 390–398. doi:10.1016/j.tics.2012.05.003

Atiani, S., David, S.V., Elgueda, D., Locastro, M., Radtke-Schuller, S., Shamma, S.A., Fritz, J.B., 2014. Emergent Selectivity for Task-Relevant Stimuli in Higher-Order Auditory Cortex. Neuron 82, 486–499. doi:10.1016/j.neuron.2014.02.029

Bajo, V.M., King, A.J., 2013. Cortical modulation of auditory processing in the midbrain. Front. Neural Circuits 6. doi:10.3389/fncir.2012.00114

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Reading and Subcortical Auditory Function. Cereb. Cortex 19, 2699–2707. doi:10.1093/cercor/bhp024

Bellier, L., Bidet-Caulet, A., Bertrand, O., Catenoix, H., Thai-Van, H., Caclin, A., in preparation. A cortical component in the Frequency-Following Response to Speech: evidence from intracranial EEG.
Benasich, A.A., Choudhury, N.A., Realpe-Bonilla, T., Roesler, C.P., 2014. Plasticity in Developing Brain: Active Auditory Exposure Impacts Prelinguistic Acoustic Mapping. J. Neurosci. 34, 13349–13363. doi:10.1523/JNEUROSCI.0972-14.2014
Berens, P., 2009. CircStat: A MATLAB Toolbox for Circular Statistics. J. Stat. Softw. 31, 1–21.
Bermudez, P., Lerch, J.P., Evans, A.C., Zatorre, R.J., 2009. Neuroanatomical Correlates of Musicianship as Revealed by Cortical Thickness and Voxel-Based Morphometry. Cereb. Cortex 19, 1583–1596. doi:10.1093/cercor/bhn196
Bernal, S., Dehaene-Lambertz, G., Millotte, S., Christophe, A., 2010. Two-year-olds compute syntactic structure on-line. Dev. Sci. 13, 69–76. doi:10.1111/j.1467-7687.2009.00865.x
Besson, M., Chobert, J., Marie, C., 2011. Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. Front. Psychol. 2. doi:10.3389/fpsyg.2011.00094
Besson, M., Schön, D., Moreno, S., Santos, A., Magne, C., 2007. Influence of musical expertise and musical training on pitch processing in music and. Restor. Neurol. Neurosci. 25, 399–410.
Bhatara, A., Tirovolas, A.K., Duan, L.M., Levy, B., Levitin, D.J., 2011. Perception of emotional expression in musical performance. J. Exp. Psychol. Hum. Percept. Perform. 37, 921–934. doi:10.1037/a0021922
Bidelman, G.M., 2015. Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. Hear. Res. 323, 68–80. doi:10.1016/j.heares.2015.01.011
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77, 1–10. doi:10.1016/j.bandc.2011.07.006
Bidelman, G.M., Gandour, J.T., Krishnan, A., 2009. Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. J. Cogn. Neurosci. 23, 425–434. doi:10.1162/jocn.2009.21362
Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125. doi:10.1016/j.brainres.2010.07.100
Bidelman, G.M., Weiss, M.W., Moreno, S., Alain, C., 2014. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. Eur. J. Neurosci. 40, 2662–2673. doi:10.1111/ejn.12627
Blood, A.J., Zatorre, R.J., Bermudez, P., Evans, A.C., 1999. Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. Nat. Neurosci. 2, 382–387. doi:10.1038/7299
Boersma, P., Weenink, D., 2001. Praat, a system for doing phonetics by computer.
Bortfeld, H., Morgan, J.L., Golinkoff, R.M., Rathbun, K., 2005. Mommy and Me Familiar Names Help Launch Babies Into Speech-Stream Segmentation. Psychol. Sci. 16, 298–304. doi:10.1111/j.0956-7976.2005.01531.x
Braun, B., Galts, T., Kabak, B., 2014. Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. Second Lang. Res. 30, 323–350.
Carcagno, S., Plack, C.J., 2011. Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. J. Assoc. Res. Otolaryngol. 12, 89–100. doi:10.1007/s10162-010-0236-1
Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47, 236–246. doi:10.1111/j.1469-8986.2009.00928.x

Chandrasekaran, B., Kraus, N., Wong, P.C.M., 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107, 1325–1336. doi:10.1152/jn.00923.2011

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1, 351–353.

Chobert, J., Besson, M., 2013. Musical Expertise and Second Language Learning. Brain Sci. 3, 923–940. doi:10.3390/brainsci3020923

Chobert, J., François, C., Velay, J.-L., Besson, M., 2014. Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. Cereb. Cortex 24, 956–967. doi:10.1093/cercor/bhs377

Chobert, J., Marie, C., François, C., Schön, D., Besson, M., 2011. Enhanced Passive and Active Processing of Syllables in Musician Children. J. Cogn. Neurosci. 23, 3874–3887. doi:10.1162/jocn_a_00088

Cravo, A.M., Rohenkohl, G., Wyart, V., Nobre, A.C., 2013. Temporal Expectation Enhances Contrast Sensitivity by Phase Entrainment of Low-Frequency Oscillations in Visual Cortex. J. Neurosci. 33, 4002–4010. doi:10.1523/JNEUROSCI.4675-12.2013

Cravo, A.M., Rohenkohl, G., Wyart, V., Nobre, A.C., 2011. Endogenous modulation of low frequency oscillations by temporal expectations. J. Neurophysiol. 106, 2964–2972. doi:10.1152/jn.00157.2011

Dajani, H.R., Purcell, D., W. Wong, Kunov, H., Picton, T.W., 2005. Recording Human Evoked Potentials That Follow the Pitch Contour of a Natural Vowel. IEEE Trans. Biomed. Eng. 52, 1614–1618. doi:10.1109/TBME.2005.851499

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical Processing in Spoken Language Comprehension. J. Neurosci. 23, 3423–3431.

Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. J. Cogn. Neurosci. 12, 635–647. doi:10.1162/089892900562390

Dehaene-Lambertz, G., Spelke, E.S., 2015. The Infancy of the Human Brain. Neuron 88, 93–109. doi:10.1016/j.neuron.2015.09.026

Delogu, F., Lampis, G., Belardinelli, M.O., 2010. From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. Eur. J. Cogn. Psychol. 22, 46–61. doi:10.1080/09541440802708136

Delogu, F., Lampis, G., Belardinelli, M.O., 2006. Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? Cogn. Process. 7, 203–207. doi:10.1007/s10339-006-0146-7

Don, M., Vermiglio, A.J., Ponton, C.W., Eggermont, J.J., Masuda, A., 1996. Variable effects of click polarity on auditory brain- stem response latencies: Analyses of narrow- band ABRs suggest possible explanations. J. Acoust. Soc. Am. 100, 458–466. doi:10.1121/1.415858

Duanmu, S., 2004. Tone and non-tone languages: An alternative to language typology and parameters. Lang. Linguist. 5, 891–923.

Dupoux, E., Pallier, C., Sebastian, N., Mehler, J., 1997. A destressing "deafness" in French? J. Mem. Lang. 36, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., Peperkamp, S., 2008. Persistent stress "deafness": The case of French learners of Spanish. Cognition 106, 682–706. doi:10.1016/j.cognition.2007.04.001

Edeline, J.-M., Manunta, Y., Hennevin, E., 2011. Induction of selective plasticity in the frequency tuning of auditory cortex and auditory thalamus neurons by locus coeruleus stimulation. Hear. Res., Central Auditory Pathways – A Tribute to Jeffery A. Winer 274, 75–84. doi:10.1016/j.heares.2010.08.005

Eerola, T., Himberg, T., Toiviainen, P., Louhivuori, J., 2006. Perceived complexity of western and African folk melodies by western and African listeners. Psychol. Music 34, 337–371. doi:10.1177/0305735606064842
Eggermont, J.J., 1995. Evoked potentials as indicators of auditory development. Int. J. Pediatr. Otorhinolaryngol. 32, Supplement, S183–S186. doi:10.1016/0165-5876(95)97408-Z
Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., Taub, E., 1995. Increased Cortical Representation of the Fingers of the Left Hand in String Players. Science 270, 305–307. doi:10.1126/science.270.5234.305
Flinker, A., Chang, E.F., Kirsch, H.E., Barbaro, N.M., Crone, N.E., Knight, R.T., 2010. Single-Trial Speech Suppression of Auditory Cortex Activity in Humans. J. Neurosci. 30, 16643–16650. doi:10.1523/JNEUROSCI.1809-10.2010
Fourtassi and Dupoux, 2014. A Rudimentary Lexicon and Semantics help Bootstrap Phoneme Acquisition. Proc. Eighteenth Conf. Comput. Lang. Learn. 191–200.
François, C., Chobert, J., Besson, M., Schön, D., 2013a. Music Training for the Development of Speech Segmentation. Cereb. Cortex 23, 2038–2043. doi:10.1093/cercor/bhs180
François, C., Chobert, J., Besson, M., Schön, D., 2013b. Music Training for the Development of Speech Segmentation. Cereb. Cortex 23, 2038–2043. doi:10.1093/cercor/bhs180
Francois, C., Schön, D., 2011. Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures. Cereb. Cortex 21, 2357–2365. doi:10.1093/cercor/bhr022
Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9, 474–480. doi:10.1016/j.tics.2005.08.011
Fritz, J., Elhilali, M., David, S., Shamma, S., 2007. Does Attention Play a Role in Dynamic Receptive Field Adaptation to Changing Acoustic Salience in A1? Hear. Res. 229, 186–203. doi:10.1016/j.heares.2007.01.009
Fritz, J., Elhilali, M., Shamma, S., 2005. Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. Hear. Res., 3rd Symp. on Molec. Mechanisms in Central Auditory Function, Plasticity and Disorder 206, 159–176. doi:10.1016/j.heares.2005.01.015
Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., Shamma, S.A., 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. 13, 1011–1019. doi:10.1038/nn.2598
Fujioka, T., Trainor, L., Ross, B., Kakigi, R., Pantev, C., 2004. Musical Training Enhances Automatic Encoding of Melodic Contour and Interval Structure. J. Cogn. Neurosci. 16, 1010–1021. doi:10.1162/0898929041502706
Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2012. Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations. J. Neurosci. 32, 1791–1802. doi:10.1523/JNEUROSCI.4107-11.2012
Fujioka, T., Trainor, L.J., Large, E.W., Ross, B., 2009. Beta and Gamma Rhythms in Human Auditory Cortex during Musical Beat Processing. Ann. N. Y. Acad. Sci. 1169, 89–92. doi:10.1111/j.1749-6632.2009.04779.x
Galbraith, G.C., 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr. Clin. Neurophysiol. Potentials Sect. 92, 321–330. doi:10.1016/0168-5597(94)90100-7
Galbraith, G.C., Arbagey, P.W., Branski, R., Comerci, N., Rector, P.M., 1995. Intelligible speech encoded in the human brain stem frequency-following response. Neuroreport 6, 2363–2367.
Galbraith, G.C., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen, T.A., 1998. Brain stem frequency-following response to dichotic vowels during attention: NeuroReport 9, 1889–1893. doi:10.1097/00001756-199806010-00041

Galbraith, G.C., Doan, B.Q., 1995. Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. Int. J. Psychophysiol. 19, 203–214. doi:10.1016/0167-8760(95)00008-G
Galbraith, G.C., Olfman, D.M., Huffman, T.M., 2003. Selective attention affects human brain stem frequency-following response. Neuroreport 14, 735–738.
Gao, E., Suga, N., 2000. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. Proc. Natl. Acad. Sci. 97, 8081–8086. doi:10.1073/pnas.97.14.8081
Gao, E., Suga, N., 1998. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. Proc. Natl. Acad. Sci. 95, 12663–12670. doi:10.1073/pnas.95.21.12663
Gaser, C., Schlaug, G., 2003. Brain Structures Differ between Musicians and Non-Musicians. J. Neurosci. 23, 9240–9245.
Gilbert, C.D., Sigman, M., 2007. Brain States: Top-Down Influences in Sensory Processing. Neuron 54, 677–696. doi:10.1016/j.neuron.2007.05.019
Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. doi:10.1038/nn.3063
Greenberg, S., 1980. WPP, No. 52: Temporal Neural Coding of Pitch and Vowel Quality. Dep. Linguist. UCLA.
Greenberg, S., Marsh, J.T., Brown, W.S., Smith, J.C., 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear. Res. 25, 91–114. doi:10.1016/0378-5955(87)90083-9
Gregersen, P.K., Kowalsky, E., Kohn, N., Marvin, E.W., 1999. Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component. Am. J. Hum. Genet. 65, 911–913.
Grimm, S., Escera, C., Slabu, L., Costa-Faidella, J., 2011. Electrophysiological evidence for the hierarchical organization of auditory change detection in the human brain. Psychophysiology 48, 377–384. doi:10.1111/j.1469-8986.2010.01073.x
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. PLOS Biol 11, e1001752. doi:10.1371/journal.pbio.1001752
Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol 6, e138. doi:10.1371/journal.pbio.0060138
Hackley, S.A., 1993. An evaluation of the automaticity of sensory processing using event-related potentials and brain-stem reflexes. Psychophysiology 30, 415–428. doi:10.1111/j.1469-8986.1993.tb02065.x
Hackley, S.A., Woldorff, M., Hillyard, S.A., 1990. Cross-Modal Selective Attention Effects on Retinal, Myogenic, Brainstem, and Cerebral Evoked Potentials. Psychophysiology 27, 195–208. doi:10.1111/j.1469-8986.1990.tb00370.x
Hairston, W.D., Letowski, T.R., McDowell, K., 2013. Task-Related Suppression of the Brainstem Frequency following Response. PLoS ONE 8, e55215. doi:10.1371/journal.pone.0055215
Harrison, J., Howe, M., 1974a. Anatomy of the afferent auditory nervous system of mammals, in: Auditory System. pp. 283–336.
Harrison, J., Howe, M., 1974b. Anatomy of the descending auditory system (mammalian), in: Auditory System. pp. 363–388.
Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97, 3099–3111. doi:10.1121/1.411872
Hoormann, J., Falkenstein, M., Hohnsbein, J., 2000. Early attention effects in human auditory-evoked potentials. Psychophysiology 37, 29–42.

Hoormann, J., Falkenstein, M., Hohnsbein, J., 1994. Effect of Selective Attention on the Latency of Human Frequency-Following Potentials. Neuroreport 5, 1609–1612. doi:10.1097/00001756-199408150-00017
Hoormann, J., Falkenstein, M., Hohnsbein, J., Blanke, L., 1992. The human frequency-following response (FFR): Normal variability and relation to the click-evoked brainstem response. Hear. Res. 59, 179–188. doi:10.1016/0378-5955(92)90114-3
Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., Kraus, N., 2012a. Subcortical representation of speech fine structure relates to reading ability. Neuroreport 23, 6–9. doi:10.1097/WNR.0b013e32834d2ffd
Hornickel, J., Knowles, E., Kraus, N., 2012b. Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. Hear. Res. 284, 52–58. doi:10.1016/j.heares.2011.12.005
Hornickel, J., Skoe, E., Kraus, N., 2008. Subcortical Laterality of Speech Encoding. Audiol. Neurotol. 14, 198–207. doi:10.1159/000188533
Houde, J.F., Nagarajan, S., Sekihara, K., Merzenich, M.M., 2002. Modulation of the auditory cortex during speech: an MEG study. Cogn. Neurosci. J. Of 14, 1125–1138.
Huffman, R.F., Henson Jr., O.W., 1990. The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. Brain Res. Rev. 15, 295–323. doi:10.1016/0165-0173(90)90005-9
Hutchinson, S., Lee, L.H.-L., Gaab, N., Schlaug, G., 2003. Cerebellar Volume of Musicians. Cereb. Cortex 13, 943–949. doi:10.1093/cercor/13.9.943
Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C., Schlaug, G., 2009. Musical Training Shapes Structural Brain Development. J. Neurosci. 29, 3019–3025. doi:10.1523/JNEUROSCI.5118-08.2009
Imfeld, A., Oechslin, M.S., Meyer, M., Loenneker, T., Jancke, L., 2009. White matter plasticity in the corticospinal tract of musicians: A diffusion tensor imaging study. NeuroImage 46, 600–607. doi:10.1016/j.neuroimage.2009.02.025
Intartaglia, B., Schön, D., in review. Active sensing and predictive coding in subcortical structures. J. Cogn. Neurosci.
Intartaglia, B., White-Schwoch, T., Meunier, C., Roman, S., Kraus, N., Schön, D., 2016. Native language shapes automatic neural processing of speech. Neuropsychologia 89, 57–65. doi:10.1016/j.neuropsychologia.2016.05.033
Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. 'ich, Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, B47–B57. doi:10.1016/S0010-0277(02)00198-1
Jääskeläinen, I., &#xe4, skel&#xe4, Inen, I.P., Ahveninen, J., J&#xe4, &#xe4, skel&#xe4, Inen, I.P., Ahveninen, J., 2014. Auditory-Cortex Short-Term Plasticity Induced by Selective Attention, Auditory-Cortex Short-Term Plasticity Induced by Selective Attention. Neural Plast. Neural Plast. 2014, 2014, e216731. doi:10.1155/2014/216731, 10.1155/2014/216731
Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., Lin, C.-D., 2011. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear. 32, 699–707.
Jenkinson, N., Brown, P., 2011. New insights into the relationship between dopamine, beta oscillations and motor function. Trends Neurosci. 34, 611–618. doi:10.1016/j.tins.2011.09.003
Jentschke, S., Koelsch, S., 2009. Musical training modulates the development of syntax processing in children. NeuroImage 47, 735–744. doi:10.1016/j.neuroimage.2009.04.090
Jewett, D.L., Williston, J.S., 1971. Auditory-Evoked Far Fields Averaged from the Scalp of Humans. Brain 94, 681–696. doi:10.1093/brain/94.4.681

Johnson, K.L., Nicol, T., Zecker, S.G., Bradlow, A.R., Skoe, E., Kraus, N., 2008. Brainstem encoding of voiced consonant–vowel stop syllables. Clin. Neurophysiol. 119, 2623–2635. doi:10.1016/j.clinph.2008.07.277
Johnson, K.L., Nicol, T., Zecker, S.G., Kraus, N., 2008. Developmental Plasticity in the Human Auditory Brainstem. J. Neurosci. 28, 4000–4007. doi:10.1523/JNEUROSCI.0012-08.2008
Jones, M. r., Moynihan, H., MacKenzie, N., Puente, J., 2002. Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays. Psychol. Sci. Sage Publ. Inc 13, 1313–1319.
Jones, M.R., Boltz, M., 1989. Dynamic attending and responses to time. Psychol. Rev. 96, 459–491. doi:10.1037/0033-295X.96.3.459
Jusczyk, P.W., Luce, P.A., 1994. Infants' Sensitivity to Phonotactic Patterns in the Native Language. J. Mem. Lang. 33, 630–645. doi:10.1006/jmla.1994.1030
Kilgard, M.P., 2012. Harnessing plasticity to understand learning and treat disease. Trends Neurosci. 35, 715–722. doi:10.1016/j.tins.2012.09.002
Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch Discrimination: Are Professional Musicians Better than Non-Musicians? J. Basic Clin. Physiol. Pharmacol. 12. doi:10.1515/JBCPP.2001.12.2.125
Kleim, J.A., Swain, R.A., Czerlanis, C.M., Kelly, J.L., Pipitone, M.A., Greenough, W.T., 1997. Learning-Dependent Dendritic Hypertrophy of Cerebellar Stellate Cells: Plasticity of Local Circuit Neurons. Neurobiol. Learn. Mem. 67, 29–33. doi:10.1006/nlme.1996.3742
Koelsch, S., Schroger, E., Tervaniemi, M., 1999. Superior pre-attentive auditory processing in musicians. Neuroreport 10, 1309–1313.
Kral, A., Eggermont, J.J., 2007. What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Res. Rev. 56, 259–269. doi:10.1016/j.brainresrev.2007.07.021
Kraus, N., 2011. Listening in on the listening brain. ww.physicstoday.org.
Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599–605. doi:10.1038/nrn2882
Kraus, N., Skoe, E., Parbery-Clark, A., Ashley, R., 2009. Experience-induced Malleability in Neural Encoding of Pitch, Timbre, and Timing. Ann. N. Y. Acad. Sci. 1169, 543–557. doi:10.1111/j.1749-6632.2009.04549.x
Kraus, N., White-Schwoch, T., 2016. Timescales of Auditory Processing. Hear. J. 69, 36–40.
Kraus, N., White-Schwoch, T., 2015a. Unraveling the Biology of Auditory Learning: A Cognitive–Sensorimotor–Reward Framework. Trends Cogn. Sci. 19, 642–654. doi:10.1016/j.tics.2015.08.017
Krause, V., Pollok, B., Schnitzler, A., 2010a. Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. Acta Psychol. (Amst.) 133, 28–37. doi:10.1016/j.actpsy.2009.08.003
Krause, V., Schnitzler, A., Pollok, B., 2010b. Functional network interactions during sensorimotor synchronization in musicians and non-musicians. NeuroImage 52, 245–251. doi:10.1016/j.neuroimage.2010.03.081
Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192–201. doi:10.1016/S0378-5955(02)00327-1
Krishnan, A., 1999. Human Frequency-Following Responses to Two-Tone Approximations of Steady-State Vowels. Audiol. Neurootol. 4, 95–103. doi:10.1159/000013826
Krishnan, A., Gandour, J.T., 2014. Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: Electrophysiological evidence. Acoust. Aust. Aust. Acoust. Soc. 42, 166–178.
Krishnan, A., Gandour, J.T., Bidelman, G.M., 2012. Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. NeuroReport 23, 498–502. doi:10.1097/WNR.0b013e328353764d

Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009a. Experience-dependent neural representation of dynamic pitch in the brainstem: NeuroReport 20, 408–413. doi:10.1097/WNR.0b013e3283263000
Krishnan, A., Swaminathan, J., Gandour, J.T., 2009b. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J. Cogn. Neurosci. 21, 1092–1105. doi:10.1162/jocn.2009.21077
Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Cogn. Brain Res. 25, 161–168. doi:10.1016/j.cogbrainres.2005.05.004
Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. Nat. Rev. Neurosci. 5, 831–843. doi:10.1038/nrn1533
Kuhl, P.K., 2000. A new view of language acquisition. Proc. Natl. Acad. Sci. 97, 11850–11857. doi:10.1073/pnas.97.22.11850
Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. 9, F13–F21. doi:10.1111/j.1467-7687.2006.00468.x
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.
Kumar, K., Bhat, J.S., D'Costa, P.E., Srivastava, M., Kalaiah, M.K., 2014. Effect of Stimulus Polarity on Speech Evoked Auditory Brainstem Response. Audiol. Res. 3. doi:10.4081/audiores.2013.e8
Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection. Science 320, 110–113. doi:10.1126/science.1154735
Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An Oscillatory Hierarchy Controlling Neuronal Excitability and Stimulus Processing in the Auditory Cortex. J. Neurophysiol. 94, 1904–1911. doi:10.1152/jn.00263.2005
Large, E.W., Jones, M.R., 1999. The dynamics of attending: How people track time-varying events. Psychol. Rev. 106, 119–159. doi:10.1037/0033-295X.106.1.119
Large, E.W., Palmer, C., 2002. Perceiving temporal regularity in music. Cogn. Sci. 26, 1–37.
Lee, C.-Y., Hung, T.-H., 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. J. Acoust. Soc. Am. 124, 3235. doi:10.1121/1.2990713
Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective Subcortical Enhancement of Musical Intervals in Musicians. J. Neurosci. 29, 5832–5840. doi:10.1523/JNEUROSCI.6133-08.2009
Levitin, D.J., Menon, V., 2003. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. NeuroImage 20, 2142–2152. doi:10.1016/j.neuroimage.2003.08.016
Lew-Williams, C., Saffran, J.R., 2012. All words are not created equal: Expectations about word length guide infant statistical learning. Cognition 122, 241–246. doi:10.1016/j.cognition.2011.10.007
Luo, H., Poeppel, D., 2007. Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. Neuron 54, 1001–1010. doi:10.1016/j.neuron.2007.06.004
Maess, B., Koelsch, S., Gunter, T.C., Friederici, A.D., 2001. Musical syntax is processed in Broca's area: an MEG study. Nat. Neurosci. 4, 540–545. doi:10.1038/87502
Marie, C., Delogu, F., Lampis, G., Belardinelli, M.O., Besson, M., 2010. Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. J. Cogn. Neurosci. 23, 2701–2715. doi:10.1162/jocn.2010.21585
Marques, C., Moreno, S., Luís Castro, S., Besson, M., 2007. Musicians Detect Pitch Violation in a Foreign Language Better Than Nonmusicians: Behavioral and Electrophysiological Evidence. J. Cogn. Neurosci. 19, 1453–1463. doi:10.1162/jocn.2007.19.9.1453

Martínez-Montes, E., Hernández-Pérez, H., Chobert, J., Morgado-Rodríguez, L., Suárez-Murias, C., Valdés-Sosa, P.A., Besson, M., 2013. Musical expertise and foreign speech perception. Front. Syst. Neurosci. 7, 84. doi:10.3389/fnsys.2013.00084

Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111. doi:10.1016/S0010-0277(01)00157-3

Mehler, J., Dupoux, E., Pallier, C., Dehaene-Lambertz, G., 1994. Cross-linguistic approaches to speech processing. Curr. Opin. Neurobiol. 4, 171–176. doi:10.1016/0959-4388(94)90068-X

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., Amiel-Tison, C., 1988. A precursor of language acquisition in young infants. Cognition 29, 143–178. doi:10.1016/0010-0277(88)90035-2

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219, 36–47. doi:10.1016/j.heares.2006.05.004

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P.A.A., Tervaniemi, M., 2008. Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. Brain Res. 1194, 81–89. doi:10.1016/j.brainres.2007.11.042

Milovanov, R., Pietilä, P., Tervaniemi, M., Esquef, P.A.A., 2010. Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. Learn. Individ. Differ. 20, 56–60. doi:10.1016/j.lindif.2009.11.003

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S.L., Besson, M., 2009. Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity. Cereb. Cortex 19, 712–723. doi:10.1093/cercor/bhn120

Morillon, B., Schroeder, C.E., Wyart, V., 2014. Motor contributions to the temporal precision of auditory attention. Nat. Commun. 5, 5255. doi:10.1038/ncomms6255

Mottonen, R., van de Ven, G.M., Watkins, K.E., 2014. Attention Fine-Tunes Auditory-Motor Processing of Speech Sounds. J. Neurosci. 34, 4064–4069. doi:10.1523/JNEUROSCI.2214-13.2014

Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr. Clin. Neurophysiol. 35, 665–667. doi:10.1016/0013-4694(73)90223-X

Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl. Acad. Sci. 104, 15894–15898. doi:10.1073/pnas.0701498104

Näätänen, R., Kujala, T., Winkler, I., 2011. Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses: Auditory processing that leads to conscious perception. Psychophysiology 48, 4–22. doi:10.1111/j.1469-8986.2010.01114.x

Oechslin, M.S., Imfeld, A., Loenneker, T., Meyer, M., Jäncke, L., 2010. The Plasticity of the Superior Longitudinal Fasciculus as a Function of Musical Expertise: A Diffusion Tensor Imaging Study. Front. Hum. Neurosci. 3. doi:10.3389/neuro.09.076.2009

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. Cereb. Cortex 20, 2486–2495. doi:10.1093/cercor/bhp318

Ortiz-Mantilla, S., Hämäläinen, J.A., Musacchia, G., Benasich, A.A., 2013. Enhancement of Gamma Oscillations Indicates Preferential Processing of Native over Foreign Phonemic Contrasts in Infants. J. Neurosci. 33, 18746–18754. doi:10.1523/JNEUROSCI.3260-13.2013

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., Hoke, M., 1998. Increased auditory cortical representation in musicians. Nature 392, 811–814. doi:10.1038/33918

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12, 169–174.
Pantev, C., Ross, B., Fujioka, T., Trainor, L.J., Schulte, M., Schulz, M., 2003. Music and Learning-Induced Cortical Plasticity. Ann. N. Y. Acad. Sci. 999, 438–450. doi:10.1196/annals.1284.054
Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4. doi:10.3389/fnagi.2012.00030
Parbery-Clark, A., Skoe, E., Kraus, N., 2009a. Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound. J. Neurosci. 29, 14100–14107. doi:10.1523/JNEUROSCI.3256-09.2009
Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009b. Musician enhancement for speech-in-noise. Ear Hear. 30, 653–661.
Parbery-Clark, A., Strait, D.L., Anderson, S., Hittner, E., Kraus, N., 2011. Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. PLoS ONE 6, e18082. doi:10.1371/journal.pone.0018082
Parbery-Clark, A., Tierney, A., Strait, D.L., Kraus, N., 2012. Musicians have fine-tuned neural distinction of speech syllables. Neuroscience 219, 111–119. doi:10.1016/j.neuroscience.2012.05.042
Patel, A.D., 2014. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hear. Res., Music: A window into the hearing brain 308, 98–108. doi:10.1016/j.heares.2013.08.011
Patel, A.D., 2012. The OPERA hypothesis: assumptions and clarifications. Ann. N. Y. Acad. Sci. 1252, 124–128. doi:10.1111/j.1749-6632.2011.06426.x
Patel, A.D., 2011. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Audit. Cogn. Neurosci. 2, 142. doi:10.3389/fpsyg.2011.00142
Patel, A.D., Xu, Y., Wang, B., 2010. The role of F0 variation in the intelligibility of Mandarin sentences, in: Proceedings of Speech Prosody. pp. 11–14.
Peelle, J.E., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front. Hum. Neurosci. doi:10.3389/fnhum.2010.00051
Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., Mauguière, F., Collet, L., 2006. Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cereb. Cortex 16, 941–948. doi:10.1093/cercor/bhj035
Picton, T.W., Alain, C., Otten, L., Ritter, W., Achim, A., 2000. Mismatch negativity: different water in the same river. Audiol. Neurotol. 5, 111–139.
Picton, T.W., Hillyard, S.A., Krausz, H.I., Galambos, R., 1974. Human auditory evoked potentials. I: Evaluation of components. Electroencephalogr. Clin. Neurophysiol. 36, 179–190. doi:10.1016/0013-4694(74)90155-2
Pollok, B., Gross, J., Kamp, D., Schnitzler, A., 2008. Evidence for Anticipatory Motor Control within a Cerebello-Diencephalic-Parietal Network. J. Cogn. Neurosci. 20, 828–840. doi:10.1162/jocn.2008.20506
Pollok, B., Krause, V., Butz, M., Schnitzler, A., 2009. Modality specific functional interaction in sensorimotor synchronization. Hum. Brain Mapp. 30, 1783–1790. doi:10.1002/hbm.20762
Raizada, R.D.S., Tsao, F.-M., Liu, H.-M., Kuhl, P.K., 2010. Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences. Cereb. Cortex 20, 1–12. doi:10.1093/cercor/bhp076
Rance, G., Rickards, F., 2002. Prediction of hearing threshold in infants using auditory steady-state evoked potentials. J. Am. Acad. Audiol. 13, 236–245.

Repp, B.H., 2010. Sensorimotor synchronization and perception of timing: Effects of music training and task experience. Hum. Mov. Sci. 29, 200–213. doi:10.1016/j.humov.2009.08.002
Rinne, T., Balk, M.H., Koistinen, S., Autti, T., Alho, K., Sams, M., 2008. Auditory Selective Attention Modulates Activation of Human Inferior Colliculus. J. Neurophysiol. 100, 3323–3327. doi:10.1152/jn.90607.2008
Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005. Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. Dev. Sci. 8, 162–172. doi:10.1111/j.1467-7687.2005.00403.x
Rohenkohl, G., Cravo, A.M., Wyart, V., Nobre, A.C., 2012. Temporal Expectation Improves the Quality of Sensory Information. J. Neurosci. 32, 8424–8428. doi:10.1523/JNEUROSCI.0804-12.2012
Rüsseler, J., Altenmüller, E., Nager, W., Kohlmetz, C., Münte, T.F., 2001. Event-related brain potentials to sound omissions differ in musicians and non-musicians. Neurosci. Lett. 308, 33–36. doi:10.1016/S0304-3940(01)01977-2
Russo, N.M., Nicol, T.G., Zecker, S.G., Hayes, E.A., Kraus, N., 2005. Auditory training improves neural timing in the human brainstem. Behav. Brain Res. 156, 95–103. doi:10.1016/j.bbr.2004.05.012
Sadakata, M., Sekiyama, K., 2011. Enhanced perception of various linguistic features by musicians: A cross-linguistic study. Acta Psychol. (Amst.) 138, 1–10. doi:10.1016/j.actpsy.2011.03.007
Saffran, J.R., Aslin, R.N., Newport, E.L., 1996. Statistical Learning by 8-Month-Old Infants. Science 274:1926–1928.
Salimpoor, V.N., van den Bosch, I., Kovacevic, N., McIntosh, A.R., Dagher, A., Zatorre, R.J., 2013. Interactions Between the Nucleus Accumbens and Auditory Cortices Predict Music Reward Value. Science 340, 216–219. doi:10.1126/science.1231059
Schlaug, G., Jäncke, L., Huang, Y., Staiger, J.F., Steinmetz, H., 1995. Increased corpus callosum size in musicians. Neuropsychologia, Neuropsychological And Developmental Studies Of The Corpus Callosum 33, 1047–1055. doi:10.1016/0028-3932(95)00045-5
Schlaug, G., Jancke, L., Huang, Y., Steinmetz, H., 1995. In-Vivo Evidence of Structural Brain Asymmetry in Musicians. Science 267, 699–701. doi:10.1126/science.7839149
Schmidt-Kassow, M., Kotz, S.A., 2009. Attention and perceptual regularity in speech: NeuroReport 20, 1643–1647. doi:10.1097/WNR.0b013e328333b0c6
Schneider, P., Scherg, M., Dosch, H.G., Specht, H.J., Gutschalk, A., Rupp, A., 2002. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nat. Neurosci. 5, 688–694. doi:10.1038/nn871
Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., Besson, M., 2010. Similar cerebral networks in language, music and song perception. NeuroImage 51, 450–461. doi:10.1016/j.neuroimage.2010.02.023
Schön, D., Magne, C., Besson, M., 2004a. The music of speech: Music training facilitates pitch processing in both music and language. Psychophysiology 41, 341–349. doi:10.1111/1469-8986.00172.x
Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18. doi:10.1016/j.tins.2008.09.012
Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., 2008. Neuronal oscillations and visual amplification of speech. Trends Cogn. Sci. 12, 106–113. doi:10.1016/j.tics.2008.01.002
Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of Active Sensing and Perceptual Selection. Curr. Opin. Neurobiol. 20, 172–176. doi:10.1016/j.conb.2010.02.010

Serences, J.T., 2011. Mechanisms of Selective Attention: Response Enhancement, Noise Reduction, and Efficient Pooling of Sensory Responses. Neuron 72, 685–687. doi:10.1016/j.neuron.2011.11.005
Shi, R., 2014. Functional Morphemes and Early Language Acquisition. Child Dev. Perspect. 8, 6–11. doi:10.1111/cdep.12052
Shi, R., Marquis, A., Gauthier, B., Bamman, D., Magnitskaia, T., Zaller, C., 2006. Segmentation and representation of function words in preverbal French-learning infants, in: Proceedings of the 30th Annual Boston University Conference on Language Development. Cascadilla Somerville, MA, pp. 549–560.
Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302–324. doi:10.1097/AUD.0b013e3181cdb272
Skoe, E., Nicol, T., Kraus, N., 2011. Cross-phaseogram: Objective neural index of speech sound differentiation. J. Neurosci. Methods 196, 308–317. doi:10.1016/j.jneumeth.2011.01.020
Slabu, L., Escera, C., Grimm, S., Costa-Faidella, J., 2010. Early change detection in humans as revealed by auditory brainstem and middle-latency evoked potentials. Eur. J. Neurosci. 32, 859–865. doi:10.1111/j.1460-9568.2010.07324.x
Slabu, L., Grimm, S., Escera, C., 2012. Novelty Detection in the Human Auditory Brainstem. J. Neurosci. 32, 1447–1452. doi:10.1523/JNEUROSCI.2557-11.2012
Slater, J., Skoe, E., Strait, D.L., O'Connell, S., Thompson, E., Kraus, N., 2015. Music training improves speech-in-noise perception: Longitudinal evidence from a community-based music program. Behav. Brain Res. 291, 244–252. doi:10.1016/j.bbr.2015.05.026
Slevc, L.R., Miyake, A., 2006. Individual Differences in Second-Language Proficiency Does Musical Ability Matter? Psychol. Sci. 17, 675–681. doi:10.1111/j.1467-9280.2006.01765.x
Song, J.H., Skoe, E., Wong, P.C.M., Kraus, N., 2008. Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training. J. Cogn. Neurosci. 20, 1892–1902. doi:10.1162/jocn.2008.20131
Starr, A., Picton, T.W., Sininger, Y., Hood, L.J., Berlin, C.I., 1996. Auditory neuropathy. Brain 119, 741–753. doi:10.1093/brain/119.3.741
Steinschneider, M., Nourski, K.V., Fishman, Y.I., 2013. Representation of speech in human auditory cortex: Is it special? Hear. Res., Communication Sounds and the Brain: New Directions and Perspectives 305, 57–73. doi:10.1016/j.heares.2013.05.013
Stillman, R.D., Crow, G., Moushegian, G., 1978. Components of the frequency-following potential in man. Electroencephalogr. Clin. Neurophysiol. 44, 438–446. doi:10.1016/0013-4694(78)90028-7
Strait, D., Kraus, N., 2011. Playing Music for a Smarter Ear: Cognitive, Perceptual and Neurobiological Evidence. Music Percept. Interdiscip. J. 29, 133–146. doi:10.1525/mp.2011.29.2.133
Strait, D.L., Kraus, N., 2014. Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. Hear. Res. 308, 109–121. doi:10.1016/j.heares.2013.08.004
Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. Hear. Res. 261, 22–29. doi:10.1016/j.heares.2009.12.021
Strait, D.L., Kraus, N., Skoe, E., Ashley, R., 2009. Musical experience and neural efficiency - effects of training on subcortical processing of vocal expressions of emotion. Eur. J. Neurosci. 29, 661–668. doi:10.1111/j.1460-9568.2009.06617.x
Strait, D.L., O'Connell, S., Parbery-Clark, A., Kraus, N., 2013. Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30. Cereb. Cortex. doi:10.1093/cercor/bht103
Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A 194, 169–183. doi:10.1007/s00359-007-0274-2

Swaminathan, S., Gopinath, J.K., 2013. Music Training and Second-Language English Comprehension and Vocabulary Skills in Indian Children. Psychol. Stud. 58, 164–170. doi:10.1007/s12646-013-0180-3
Teder-Sälejärvi, W.A., Hillyard, S.A., 1998. The gradient of spatial auditory attention in free field: an event-related potential study. Percept. Psychophys. 60, 1228–1242.
Tenenbaum, J.B., Kemp, C., Griffiths, T.L., Goodman, N.D., 2011. How to Grow a Mind: Statistics, Structure, and Abstraction. Science 331, 1279–1285. doi:10.1126/science.1192788
Terkildsen, K., Osterhammel, P., 1981. The Influence of Reference Electrode Position on Recordings of the Auditory Brainstem Responses. Ear Hear. 2, 9–14.
Theunissen, F.E., Elie, J.E., 2014. Neural processing of natural sounds. Nat. Rev. Neurosci. 15, 355–366. doi:10.1038/nrn3731
Thompson, E.C., White-Schwoch, T., Tierney, A., Kraus, N., 2015. Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience. PLoS ONE 10, e0128839. doi:10.1371/journal.pone.0128839
Tierney, A., Kraus, N., 2016. Getting back on the beat: links between auditory-motor integration and precise auditory processing at fast time scales. Eur. J. Neurosci. 43, 782–791. doi:10.1111/ejn.13171
Tierney, A., Kraus, N., 2015. Neural Entrainment to the Rhythmic Structure of Music. J. Cogn. Neurosci. 27, 400–408. doi:10.1162/jocn_a_00704
Tierney, A., Kraus, N., 2014. Auditory-motor entrainment and phonological skills: precise auditory timing hypothesis (PATH). Front. Hum. Neurosci. 8. doi:10.3389/fnhum.2014.00949
Tierney, A., Kraus, N., 2013a. Music Training for the Development of Reading Skills, in: Progress in Brain Research.
Tierney, A., Kraus, N., 2013b. The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. J. Neurosci. 33, 14981–14988. doi:10.1523/JNEUROSCI.0612-13.2013
Tierney, A., Kraus, N., 2013. Neural responses to sounds presented on and off the beat of ecologically valid music. Front. Syst. Neurosci. 7. doi:10.3389/fnsys.2013.00014
Tillein, J., Heid, S., Lang, E., Hartmann, R., Kral, A., 2012. Development of Brainstem-Evoked Responses in Congenital Auditory Deprivation. Neural Plast. 2012. doi:10.1155/2012/182767
Toscano, J.C., McMurray, B., 2010. Cue Integration With Categories: Weighting Acoustic Cues in Speech Using Unsupervised Learning and Distributional Statistics. Cogn. Sci. 34, 434–464. doi:10.1111/j.1551-6709.2009.01077.x
Trainor, L.J., Desjardins, R.N., Rockel, C., 1999. A Comparison of Contour and Interval Processing in Musicians and Nonmusicians Using Event-Related Potentials. Aust. J. Psychol. 51, 147–153. doi:10.1080/00049539908255352
Tzounopoulos, T., Kraus, N., 2009. Learning to Encode Timing: Mechanisms of Plasticity in the Auditory Brainstem. Neuron 62, 463–469. doi:10.1016/j.neuron.2009.05.002
Vaissière, J., 1991a. Rhythm, accentuation and final lengthening in French 108–120.
van Atteveldt, N., Musacchia, G., Zion-Golumbic, E., Sehatpour, P., Javitt, D.C., Schroeder, C., 2015. Complementary fMRI and EEG evidence for more efficient neural processing of rhythmic vs. unpredictably timed sounds. Front. Psychol. 6, 1663. doi:10.3389/fpsyg.2015.01663
Weinberger, N.M., 2004. Specific long-term memory traces in primary auditory cortex. Nat. Rev. Neurosci. 5, 279–290. doi:10.1038/nrn1366
Werker, J.F., Tees, R.C., 2002. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Dev., 25th Anniversary Special Issue 25, 121–133. doi:10.1016/S0163-6383(02)00093-0

Winer, J.A., Schreiner, C.E., 2005. The central auditory system: a functional analysis, in: The Inferior Colliculus. Springer, pp. 1–68.
Womelsdorf, T., Schoffelen, J.-M., Oostenveld, R., Singer, W., Desimone, R., Engel, A.K., Fries, P., 2007. Modulation of Neuronal Interactions Through Neuronal Synchronization. Science 316, 1609–1612. doi:10.1126/science.1139597
Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. 10, 420–422. doi:10.1038/nn1872
Worden, F.G., Marsh, J.T., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. Electroencephalogr. Clin. Neurophysiol. 25, 42–52. doi:10.1016/0013-4694(68)90085-0
Xiao, Z., Suga, N., 2002. Modulation of cochlear hair cells by the auditory cortex in the mustached bat. Nat. Neurosci. 5, 57–63. doi:10.1038/nn786
Xu, Y., Krishnan, A., Gandour, J.T., 2006. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17, 1601–1605. doi:10.1097/01.wnr.0000236865.31705.3a
Yan, W., Suga, N., 1998. Corticofugal modulation of the midbrain frequency map in the bat auditory system. Nat. Neurosci. 1, 54–58. doi:10.1038/255
Zeromskaite, I., 2014. The Potential Role of Music in Second Language Learning: A Review Article. J. Eur. Psychol. Stud. 5. doi:10.5334/jeps.ci
Zhang, C., Arnott, S.R., Rabaglia, C., Avivi-Reich, M., Qi, J., Wu, X., Li, L., Schneider, B.A., 2016. Attentional modulation of informational masking on early cortical representations of speech signals. Hear. Res. 331, 119–130. doi:10.1016/j.heares.2015.11.002
Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y. 'ichi, 2005. Effects of language experience: Neural commitment to language-specific auditory patterns. NeuroImage 26, 703–720. doi:10.1016/j.neuroimage.2005.02.040

Résumé

La perception auditive, longtemps considérée comme un phénomène passif, apparaît aujourd'hui comme un processus dynamique caractérisé par des interactions continues à double sens entre les systèmes corticaux et sous-corticaux, entre les systèmes cognitifs et sensoriels. Les récentes découvertes suggèrent que l'engagement actif pendant la perception des sons entraîne des réorganisations neuronales à plusieurs niveaux du système auditif et à plusieurs échelles temporelles. Les travaux réalisés durant cette thèse se sont focalisés sur les effets de plasticité induits au niveau sous-cortical : à long terme, par l'étude des effets de l'expertise (linguistique et musicale) et à court terme par l'étude des effets attentionnels. Notre hypothèse était que l'expertise linguistique renforce l'encodage automatique des sons natifs au niveau des indices linguistiques les plus importants.

Les réponses complexes du tronc cérébral (*complex Auditory Brainstem Responses* ou *cABRs*) se sont avérées plus robustes pour les sons natifs avec un renforcement spécifique de l'encodage des fréquences formantiques qui jouent un rôle important dans la discrimination des phonèmes, ce qui confirme l'hypothèse selon laquelle l'expertise linguistique induit une réorganisation des structures sous-corticales et que cette plasticité se produit en fonction de ce qui est important pour le locuteur. Un corpus d'études important montre des effets de transfert entre musique et langage. Nous avons montré que les différences d'expertise linguistique sont compensées par la pratique musicale grâce à un renforcement de l'encodage neuronal des indices linguistiques importants (formants). Puisque ces effets de plasticité se produisent de manière spécifique au niveau des dimensions acoustiques importantes, nous avons postulé qu'à plus court terme, des processus attentionnels *top-down* moduleraient les réponses sous-corticales de manière à induire un renforcement sélectif du codage neuronal. La comparaison des réponses neuronales recueillies en condition d'écoute passive et active (*tapping*) a révélé une représentation de la fréquence fondamentale plus robuste lors de l'écoute active des sons, confirmant l'influence des processus *top-down* sur l'encodage sous-cortical des sons de parole. Le design expérimental de cette étude nous a aussi permis d'étudier les modulations dynamiques et rapides de l'attention. Plus précisément, en entraînant l'attention au rythme d'un stimulus sur deux via une tâche de *tapping*, nous avons observé une modulation significative de la représentation du son en fonction de la phase du *tapping*, révélant ainsi les effets d'une dynamique oscillatoire (attentionnelle) au niveau de la voie auditive sous-corticale.

En conclusion, ce travail de thèse a permis de révéler la nature dynamique de la perception auditive, en montrant une modulation de l'encodage des sons au niveau sous-cortical à plusieurs échelles temporelles : à long terme via les effets de l'expertise et à court terme via les effets de l'attention. Les réponses neuronales sous-corticales sont probablement modulées via des boucles de rétrocontrôle entre le cortex et les structures sous-corticales. La communication entre ces deux niveaux de traitement pourrait se faire par des mécanismes de couplage entre la phase des ondes corticales lentes et l'amplitude des réponses du tronc cérébral. Néanmoins, de futures études sont nécessaires pour explorer cette hypothèse.