

UNIVERSITÉ DE LA MÉDITERRANÉE
FACULTÉ DE MÉDECINE DE MARSEILLE
THÈSE

En vue de l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ de la MÉDITERRANÉE

Formation doctorale : Neurosciences

Présentée et soutenue publiquement par

Clément FRANÇOIS

Le 31 Mai 2011

**Apprentissage implicite des structures linguistiques et
musicales: approche multi-méthodologique.**

Membres du Jury :

Dr. Barbara Tillmann, Lyon, France
Prof. Antoni Rodriguez-Fornells, Barcelone, Espagne
Dr. Mireille Besson, Marseille, France
Dr. Michel Hoen, Lyon, France
Dr. Daniele Schön, Marseille, France

Rapporteur
Rapporteur
Examineur
Examineur
Directeur

UNIVERSITÉ DE LA MÉDITERRANÉE
FACULTÉ DE MÉDECINE DE MARSEILLE
THÈSE

En vue de l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ de la MÉDITERRANÉE

Formation doctorale : Neurosciences

Présentée et soutenue publiquement par

Clément FRANÇOIS

Le 31 Mai 2011

**Apprentissage implicite des structures linguistiques et
musicales: approche multi-méthodologique.**

Membres du Jury :

Dr. Barbara Tillmann, Lyon, France
Prof. Antoni Rodriguez-Fornells, Barcelone, Espagne
Dr. Mireille Besson, Marseille, France
Dr. Michel Hoen, Lyon, France
Dr. Daniele Schön, Marseille, France

Rapporteur
Rapporteur
Examineur
Examineur
Directeur

RÉSUMÉ

Les objectifs de cette thèse sont multiples. Le premier objectif est de comparer, aux niveaux comportemental et électrophysiologique, l'apprentissage implicite de structures linguistiques et musicales après l'écoute d'un langage artificiel chanté. Alors qu'au niveau comportemental, seule la structure linguistique semble être apprise, les résultats électrophysiologiques révèlent un effet N400 pour les deux dimensions, linguistique et musicale. Le deuxième objectif de cette thèse est d'évaluer comment cet apprentissage est influencé par l'expertise musicale. Nous avons comparé un groupe d'adultes musiciens à un groupe de non musiciens. Alors qu'au niveau comportemental les musiciens sont à peine meilleurs que les non musiciens dans les deux dimensions, les données électrophysiologiques révèlent, via des différences précoces (N1/P2) et tardives (N400), une meilleure segmentation chez les musiciens. De plus, les analyses en potentiels évoqués et en temps-fréquences des données électrophysiologiques enregistrées pendant les phases d'apprentissage révèlent que les musiciens apprennent plus rapidement que les non musiciens. Cependant, un lien de causalité quant aux effets de l'apprentissage de la musique ne peut être mis en évidence qu'en réalisant une étude longitudinale. Nous avons mené une telle étude chez des enfants de 8 ans à qui l'on a fait suivre un apprentissage de la musique ou de la peinture pendant 2 années. Les résultats comportementaux et électrophysiologiques révèlent un large bénéfice de l'apprentissage musical comparé à celui de la peinture démontrant l'importance de la musique dans l'éducation des enfants.

ABSTRACT

The aims of the present thesis were two-folded. Firstly, we wanted to compare behavioral and electrophysiological measures related to the implicit learning of linguistic and musical structures contained within an artificial sung language. While behavioral measures suggest that only the linguistic structure was learned, electrophysiological data revealed similar N400 effects in both linguistic and musical dimensions, suggesting that participants did also learn the musical structure. The second goal was to evaluate to what extent musical expertise can affect speech segmentation. At this aim, we compared a group of adult musicians to a group of nonmusicians. While behavioral data showed that musicians had marginally better performance than non musicians in both dimensions, electrophysiological data revealed, via early (N1/P2) and late (N400) differences, a better speech segmentation in musicians than in non musicians. Moreover, event-related potentials and time-frequency analyzes during learning revealed a faster and more efficient learning process in musicians. However, the only way to unambiguously claim causality between expertise and the observed effects requires a longitudinal approach. At this aim, we conducted a study with 8 year-old children who followed either music or painting lessons over a period of 2 years. Behavioral and electrophysiological data revealed a larger benefit of musical compared to painting training, bringing evidences for the importance of music in childrens' education.

!!!!!!!!!!!!!!!!!!REMERCIEMENTS!!!!!!!!!!!!!!!!!!!!

Mes premiers remerciements vont à mon directeur de thèse Daniele Schön.

Merci pour m'avoir accueilli il y a quelques années et pour finalement ne jamais avoir cessé de me guider durant toute cette aventure. Merci pour ta patience, ta gentillesse, ta disponibilité, ton calme, ta sagesse, ton humour et pour tout ce que tu es. Tu m'auras appris énormément.

Mes seconds premiers remerciements, tout particuliers, iront ensuite à Mireille Besson, pour m'avoir accueilli dans son équipe et m'avoir permis de mener mes travaux de thèse tout en m'ayant confié la responsabilité partagée du projet MUSAPDYS si riche tant scientifiquement qu'humainement. Merci pour toutes ces discussions passionnantes, pour tous ces conseils éclairés, ces soirées si drôles et goutues grâce à Michel aux fournaux.

Merci à tous les membres de l'équipe Langage Musique et Motricité pour leur soutien, pour l'aide qu'ils m'ont apporté, pour toutes ces discussions scientifiques, philosophiques et parfois moins, toutes si enrichissantes les unes que les autres. Merci pour ces séjours inoubliables passés tous ensemble aux Seguin à Buoux. Ainsi, merci à : Jean-Luc Velay, Jean Vion-Dury, Mitsuko Aramaki, Marieke Longcamp, Vietmin, Jean Arthur, Michel, Aurélie, Yohanna, Nia, Dee, Nadia, Vera, Wiebke, Radou, Chi, Aline.

Un grand merci à tous les statutaires de l'INCM, les post-doc, les doc, la salle café, la machine à café, les grains de café et plus particulièrement à Bruno et Cécile, Fredo et Pascale, Thomas et Christine, Jérôme et Chloé, Biri et Anna, Lolo et Artémis, Ivo et Anne, Florent, GM, Alexa, Ivan, Joelle et Taarabte !!, Alex, Elisa, Romain, Andreia, Giacomo, Marina et sans oublier bien entendu Driss Boussaoud.

Un grand merci tout particulier à TakTak et Sandrine pour les week ends chantier, les soirées musique, les barbequuuu, le SAVIM, les sessions sur le toit à coté du prunier radioactif, tout ça quoi.

Un immense merci à tous les étudiants du Laboratoire de Mécanique Acoustique de Marseille, pour ces mémorables journées jeunes chercheurs en acoustique, pour ces apéros doc, pour ces parties de ping pong, pour ces heures passées dans les cabines à écouter 200 fois la même mélodie de clarinette, pour ces non répètes dans le studio, et pour tous ces moments à vous entendre discuter de comment arriver à faire la meilleure réverb du monde avec MaxMSP. Merci à Bastien, Fabrice, Françoise, Romain et Doris, Pyo et Alice, Jean-Francois, Doumé « le cachador » et Chen, Samy et Sophie, Charles, Adri Merer et Julie, Adri Sirdey, Charly la résille, Anaik et tous les autres !!

Il m'est impossible de ne pas consacrer un paragraphe entier pour remercier ma jumelle de thèse et véritable amie, Julie, car il me sera désormais impossible d'oublier ces si belles années durant lesquelles nous avons passé tout ce temps au côté des enfants dans les écoles, où nous avons passé tant de soirées à mourir de rire à Buoux, à Marrakech, à Dijon, Toulouse, à Aix avec Mam et Nan, Julien et Chizuru (merci au chat dans la goutière !;-)), Martin 'dafish' et Manon.

Merci pour toutes ces soirées qui n'en finissaient pas, avec Ben, Combo, et toute ta troupe à toi, Kzé Coline Pierre Marine et les autres. Merci pour tout le soutien que tu m'as apporté, ta joie et ta gaieté inébranlables. Merci d'avoir toujours été là pour me remonter le moral quand il n'était pas au plus haut. Dans un mois c'est ton tour... courage et puis t'inquiètes, tu vas envoyer du lourd !!

Je tiens également à remercier toutes les personnes qui ont contribué à la réalisation du projet MUSAPDYS. Ainsi, je remercie Jean-Jacques Gaubert, Murielle Gaïarsa et toute leur équipe pédagogique (Emeline et Cécile!!), Pierre-Yves et Ariane !! Merci aussi à Carine Verse pour son aide précieuse. Merci à tous ceux qui ont bien voulu jouer le rôle des assistants de recherche. Merci à tous les parents et surtout à tous les enfants qui ont bien voulu participer et passer des heures à cliquer sur les boutons !!!

Un grand merci à Clémence !!!! Bonne route et désolé pour ces derniers mois...

Au tour des musiciens...

à périgueux et un peu partout: Yo les copains, voilà bientôt 10 ans qu'on joue ensemble, on a fait un bel album pendant cette thèse, on a fait des concerts un peu partout en France....sauf à Marseille, il va falloir que ça change!!! Bon, juste pour vous dire merci les gars !! Merci pour tous ces tain de week end de ouf qui me font arriver dans des états pas possibles le dimanche soir à l'appart avec le clavier sur le dos mais la tête pleine de musique, j'adore!! Merci d'avoir ralenti le rythme sans perdre la motiv comme moi, j'espère qu'on va le faire ce deuxième!! Peu importe, MERCI à tous les membres du Ska pute orchestra pour leur soutien Roule Dessus: Martin aka « dafish » au micro, Tibo aka « sexydancer » au saxo, Gui aka « trompette bombetta », Fab aka « trombonito », Cédric aka « mentalo », Olivier aka « Yeti », Thomas aka « gratalakiss », Antoine aka « la tonguy », Seb aka « Tonton ». Merci aussi à tous les proches du groupe, les femmes de joueurs, les anciens !! Bref To lo moundo !!
à Marseille: Merci aussi à cette belle troupe qui m'accompagne depuis presque un an!! Merci à Adri M à la basse (bientôt toi aussi...), à Charly à la guitare radioreverb (revient vite!!), Ana au microreverb (reviens vite!!), Jaime le maître du temps (10 ans!!), Romain delay&reverbman, Adri S guitare flamnouch&lectric (vas y gamin!!). On fait l'album sans problème cet été !!

Merci enfin à tous les membres du collectif Massilia HiFi pour tous ces bon moments de musique et pour me faire découvrir votre univers, Merci à Jeff, Ed, Ben, Mougli, Goudjou, loubar mon pote !!

Merci à Marilyn Manson et Maceo parker pour les retours du labo en musique !!

Merci à mes très chers colocataires pour toute l'énergie positive qu'ils m'ont apportée pendant ces derniers mois difficiles, pour ces soirées d'enfer, merci pour leur bonne humeur, merci pour leurs animaux domestiques, merci pour les bonnes bouffes qu'on a partagé ensemble, merci pour m'avoir déménagé autant physiquement que moralement. Aurore Tomy Tomasao, Romain DJtrash Tabarique, Merci infiniment à tous les deux!!!

Finalement, mes derniers mots iront à ma famille,

Maman, Papa, Bruno, Claire, merci de m'avoir toujours soutenu dans mes projets tant universitaires que musicaux, merci d'avoir eu la bonne idée d'écouter de la musique, d'en faire, de me faire voyager et de me donner le goût d'apprendre.

Finalement, le plus grand des merci ira à ma tendre grand-mère,

Dédée, merci pour tout ce que tu m'as apporté depuis que je suis tout petit!

SOMMAIRE

Partie 1 : Introduction et cadre théorique.

Chapitre 1: Organisation structurelle du langage et de la musique.	2
1.1 Les sons du langage et de la musique.	2
1.1.1 La hauteur dans la musique.	2
1.1.2 La hauteur dans le langage: la prosodie.	4
1.2 Le timbre dans la musique.	5
1.3 Le timbre dans le langage.	5
1.4 Les structures rythmiques.	8
1.4.1 Les structures rythmiques dans la musique.	8
1.4.2 Les structures rythmiques dans le langage.	9
Chapitre 2: Corrélats électrophysiologiques de la perception du langage et de la musique.	10
2.1 Traitement attentif des paramètres acoustiques des stimuli: le complexe N1/P2.	10
2.2 Traitement pré-attentif des sons: la Mismatch Négativity ou Négativité de discordance.	12
2.2.1 Traitement du sens dans le langage et la musique: la N400.	12
2.3 Traitement de la syntaxe dans le langage et la musique.	14
Chapitre 3: Neuroplasticité et expertise musicale.	16
3.1 Traitement des sons non linguistiques et expertise musicale.	18
3.2 Traitement de sons non linguistiques contextualisés.	20
3.3 Traitement des sons linguistiques et expertise musicale.	23
3.4 Traitement de sons linguistiques isolés.	24
3.5 Traitement de sons linguistiques contextualisés.	25
Chapitre 4: Approche longitudinale des effets de la pratique musicale.	28
4.1 Principes de base.	28

4.2 Effets de l'apprentissage de la musique, approches longitudinales.	30
4.3 Le projet MUSAPDYS (Musique Apprentissage et Dyslexie) : Avant propos.	32
4.4 Caractéristiques des apprentissages.	33
Chapitre 5 : Apprentissage des structures linguistiques et musicales.	36
5.1 Apprentissage implicite de grammaires artificielles.	37
5.1.1 Apprentissage de grammaires artificielles dans le domaine linguistique.	37
5.1.2 Apprentissage de grammaire artificielle dans le domaine musical.	38
5.2 Apprentissage implicite des structures linguistiques et musicales.	40
5.3 L'apprentissage implicite, spécifique ou général ?	41
5.4 Bases Neuronales de l'apprentissage implicite de structures linguistiques.	42

Partie 2 : Méthodologie.

Chapitre 6: Méthodologie générale de Potentiels Evoqués.	49
6.1 Historique.	49
6.2 Principes Physiques.	49
6.3 Matériel et enregistrement.	50
6.4 Traitement des données.	51
6.4.1 Filtrage des données.	51
6.4.2 Rejet / Correction des artefacts.	51
6.4.2.1 Rejet automatique.	51
6.4.2.2 Analyses en Composantes Indépendantes.	52
6.4.3 Analyses Temps-Fréquence.	53
6.4.4 Frequency Tagging / Auditory Steady State Response (SSR)	54
Chapitre 7: Questions de méthodes autour de l'apprentissage statistique.	56
7.1 L'approche comportementale dans l'apprentissage statistique.	56
7.1.1 Description des protocoles utilisés chez le nourrisson, le singe et le rat.	56

7.1.2 Description des protocoles utilisés chez l'adulte et problèmes associés.	58
7.1.3 Analyses reportées dans la littérature.....	60
7.1.3.1 Performances moyenne d'une population.....	60
7.1.3.2 Performances moyennes des mots à Haute Probabilité vs. Faible Probabilité.....	62
7.1.4 Analyses effectuées au cours de la thèse... ..	62
7.1.4.1 Performances au cours du test : effet d'interférence versus effet d'apprentissage.....	62
7.1.4.2 Performances en fonction du contraste entre items de chaque paire de test.	63
7.1.4.3 Et l'apprentissage au niveau individuel ?.....	63
7.1.5 L'approche électrophysiologique dans l'apprentissage statistique.....	64

Partie 3 : Les expériences réalisées.

Etude 1: Learning of musical and linguistic structures: comparing event-related potentials and behavior.....	68
Etude 2: Musical expertise boosts implicit learning of both musical and linguistic structures.....	76
Etude 3: Musical training for language acquisition.....	86
Etude 4: Musical expertise and statistical learning of musical and linguistic structures.....	96
Etude 5: Making sense of continuous speech : electrophysiological correlates reveal faster word segmentation in musicians than in non musicians.....	124

Partie 4 : Discussion Générale.

Chapitre 8: Comparaison de l'apprentissage de structures linguistiques et musicales.	144
8.1 Hypothèse d'une plus grande familiarité vis à vis du langage que vis à vis de la musique.....	145
8.2 Hypothèse acoustico-perceptive.....	146
8.3 Hypothèse de résistance aux interférences.....	148
8.4 Perspectives et interprétations.....	149

Chapitre 9 : Comparaison des approches comportementales et électrophysiologiques.	152
9.1 Quel(s) statut(s) pour les items des tests ?.....	152
9.2 Les points forts de l'électro-encéphalographie.	156
9.3 Réconciliation des deux approches.	158
Chapitre 10: L'expertise musicale chez l'adulte et l'apprentissage de la musique chez l'enfant facilitent la segmentation de la parole.	162
10.1 Le cerveau musicien, rappel.	162
10.2 Le modèle d'apprentissage du langage proposé par Rodriguez-Fornells et collaborateurs.....	163
10.3 Vers une hypothèse de meilleur système perceptif des musiciens.....	166
10.4 Vers une hypothèse de meilleur fonctionnement du système moteur des musiciens.	167
10.5 Vers une hypothèse de meilleure communication entre les structures auditives et motrices.	169
10.6 Dynamique d'apprentissage : différences entre musiciens et non musiciens (étude 5).	169
10.6 Intérêt et précisions amenées par l'étude longitudinale.	170
10.7 Conclusions et Perspectives.	172

Partie 5 : Références Bibliographiques.

Première Partie

Introduction

Et

Cadre théorique

Chapitre 1 : Organisation structurelle du langage et de la musique.

Le langage et la musique sont deux systèmes de communication complexes, structurés et qui s'organisent sur plusieurs dimensions à partir d'un nombre limité d'éléments de bases, les notes pour la musique et les phonèmes pour le langage (Besson & Schön, 2003). Ces éléments de base peuvent être assemblés les uns aux autres de manière infinie selon des règles propres à chacun de ces systèmes: la syntaxe. Cependant, alors que ces deux systèmes reposent sur des paramètres acoustiques de bases qui sont comparables, il n'en est pas moins vrai que certaines spécificités propres à chaque domaine demeurent (Patel, 2008). Je présente dans la partie qui suit l'organisation structurelle des deux systèmes en mettant l'accent sur ces similarités et différences.

1.1 Les sons du langage et de la musique.

1.1.1 La hauteur dans la musique.

La musique occidentale est un système hiérarchisé et structuré sur les dimensions de hauteur et de temps. La hauteur est définie comme la propriété que possède un son à pouvoir le classer (l'ordonner) sur une échelle allant du plus grave au plus aigu (Randel, 1978). La hauteur se caractérise, au niveau physique, par sa fréquence exprimée en Hertz qui pour un son pur (son composé d'une seule sinusoïde) sera égale à sa hauteur. Dans la musique occidentale, la hauteur peut être considérée selon trois niveaux d'organisation que sont les notes, les accords et les tonalités. Ainsi, la musique occidentale est basée sur un ensemble fini d'unités que sont les 12 notes de la gamme chromatique et qui constituent l'octave (do, do#, ré, ré#, mi, fa, fa#, sol, sol#, la, la#, si). Ce système est construit de telle sorte que les 12 intervalles sont équivalents et correspondent à 12 demi-tons. Chacune de ces 12 notes représente une hauteur spécifique dont découlent les 12 tonalités majeures et les 12 tonalités mineures composées chacune d'un sous ensemble de 7 notes seulement.

Les accords correspondent au deuxième niveau d'organisation de la hauteur. Un accord se construit par la superposition de plusieurs notes jouées de façon simultanées et pouvant former chacune différents intervalles. Ainsi un accord de tonique dans la gamme de Do majeur correspond à la superposition du Do (qui est la tonique de la gamme de Do majeur) du Mi (qui est la tierce majeure du Do) et du Sol (qui est la quinte juste du Do). Le concept de tonalité est très important dans la perception de la musique car c'est la tonalité qui va permettre d'ordonner l'importance perceptive de chaque note et de chaque accord dans un contexte tonal donné. Ainsi, ce sont les tonalités qui vont permettre à l'auditeur de former des attentes structurelles au niveau harmonique (notion de tension et de résolution harmonique) et donc d'apprendre à catégoriser une note par rapport à un contexte tonal. Afin de tester l'influence du contexte sur le traitement d'une note ou d'un accord cible, de nombreuses études ont été réalisées et utilisaient un paradigme d'amorçage tonal. Classiquement, l'amorce peut être une séquence de 7 notes successives formant une gamme, une cadence ou une mélodie. Cette amorce est suivie d'une note cible dont le degré de concordance par rapport à l'amorce va pouvoir être manipulé. A chaque essai, on demande aux participants de juger le niveau de concordance entre cible et amorce. Un profil de stabilité perceptif est obtenu et révèle que les participants jugent la tonique comme la note la plus concordante (la plus stable) par rapport à l'amorce, suivie de la dominante puis de la médiane (Krumhansl & Shepard, 1979, Hébert et al., 1995). De manière intéressante, dans ce type de tâche, le profil de stabilité pour des participants non musiciens est quasiment identique que pour les musiciens, montrant que les non musiciens développent des attentes harmoniques et qu'ils perçoivent la musique comme étant organisée hiérarchiquement. Ce résultat suggère l'implication d'un mécanisme d'apprentissage implicite des tonalités qui se ferait par simple exposition passive à la musique (Bigand, 2004).

La succession de notes dans le temps va former des mélodies dont la perception implique la reconnaissance de deux types d'informations: les intervalles entre notes ainsi que le contour mélodique. Ce dernier correspond aux variations ascendantes et descendantes entre notes au cours du temps (Dowling et al., 1995).

1.1.2 La hauteur dans le langage: la prosodie.

Le terme prosodie vient du grec « prosodia » qui signifie littéralement la qualité d'une voix chantée. De manière intéressante, dans le cadre de la comparaison entre le traitement du langage et celui de la musique, il a été montré un parallèle entre les structures musicales et les structures prosodiques dans le langage (Nespor & Vogel, 1983). Lorsque l'on visualise le spectre d'une phrase parlée, on se rend tout de suite compte que le signal n'est pas uniforme au cours du temps mais plutôt qu'il contient des variations de hauteur, de durée et d'intensité. Ce sont ces variations qui sont regroupées sous le terme de prosodie du langage. Elles vont ainsi pouvoir servir d'indices acoustiques et vont pouvoir jouer une fonction linguistique, structurale, émotionnelle. Les variations de hauteur au sein d'une phrase vont notamment permettre de percevoir l'intonation d'une phrase: une variation de hauteur à la fin d'une phrase ou à la fin d'un groupe de mots va permettre à l'auditeur de percevoir une affirmation ou une interrogation. En effet, les phrases interrogatives sont caractérisées par une augmentation de la hauteur en fin de phrase alors que les phrases affirmatives sont plutôt caractérisées par une diminution de la hauteur (Liberman & Pierrehumbert, 1984). Il a été montré que les variations prosodiques sont importantes pour l'acquisition du langage chez les nouveaux nés. Lorsque l'on parle à un nouveau né, même pour la première fois, on parle avec une intonation particulière en utilisant une prosodie exagérée appelée Infant Directed Speech (IDS, discours dirigé vers le nouveau né). En comparaison avec le discours dirigé vers l'adulte (Adult Directed Speech, ADS), l'IDS se caractérise par un débit de parole plus faible, de plus grandes variations de hauteur et de contour, des pauses plus grandes, des structures de phrases plus simples (Fernald, 1992). Ainsi, il a été montré que les nouveaux nés segmentaient mieux un langage artificiel prononcé avec une prosodie de type IDS qu'un langage prononcé avec une prosodie de type ADS (Thiessen et al., 2005). Dans la même veine, l'IDS facilite la segmentation des derniers mots dans une phrase (Aslin, 2000). Dans la mesure où l'enfant ne comprend pas encore les mots du langage, l'exagération des variations prosodiques aurait une fonction émotionnelle, attentionnelle et linguistique car les régularités de contour prosodique peuvent être utilisées pour segmenter le flux

de paroles (Trainor et al. 2000). De manière intéressante, ces auteurs ont également montré que les nouveaux nés préféraient le chant de leur mère plutôt que le discours parlé de leur mère (Trainor et al., 1997).

1.2 Le timbre dans la musique.

Le timbre correspond au paramètre acoustique permettant de différencier un son en particulier, d'autres sons qui ont la même hauteur, la même durée et la même intensité. Le timbre est défini par ce qu'il n'est pas puisqu'il n'est ni intensité ni sonie ni hauteur ni durée. Ainsi le timbre est une combinaison de paramètres acoustiques complexes qui permet, par exemple, de différencier un « sol » joué par un violon d'un « sol » joué par un xylophone. De nombreuses études ont tenté de définir les paramètres acoustiques à la base du concept de timbre. Ces études consistaient à présenter différentes combinaisons de paires de stimuli contenant des variations des paramètres acoustiques responsables des modulations du timbre. Les participants devaient noter sur une échelle de 1 à 10 le degré de similarité entre les deux stimuli. Ces études semblent suggérer que le timbre est une propriété émergente d'une configuration spectrale et temporelle d'un signal acoustique et qu'il repose sur seulement quelques paramètres acoustiques de base tels que le temps d'attaque, le contenu spectral (spectrum centroïd) et plus particulièrement le contenu formantique du signal (nous verrons dans la partie suivante que cet aspect est essentiel pour le timbre dans le langage), la rugosité, l'amortissement. Ainsi, les modèles d'analyse et de synthèse du timbre font l'objet de nombreuses études en perception et acoustique musicale.

1.3 Le timbre dans le langage.

Alors que la perception et la mémorisation de la musique va se baser principalement sur la dimension de la hauteur avec un rôle important des intervalles et des contours mélodiques, la perception du langage est quant à elle plutôt basée sur les contrastes de timbres qui existent entre les

différents sons du langage. Ainsi, les phonèmes sont considérés comme les plus petites unités du discours permettant de différencier deux mots dans le discours. Ces phonèmes sont différenciables les uns des autres principalement sur la base de leurs propriétés spectrales (Patel 2008). Les modifications spectrales du signal de parole sont extrêmement rapides et proviennent des mouvements articulatoires et des turbulences de l'air dans l'appareil phonatoire. Ainsi, ce sont les mouvements d'air causés par les différentes configurations de mouvements coordonnés de la langue et des cordes vocales qui vont permettre la production de différents phonèmes dans le discours. C'est sur la base de ces configurations acoustico-articulatoires complexes que les phonèmes vont pouvoir être catégorisés en différentes classes de consonnes et voyelles. Par exemple, les consonnes sont classées selon la manière dont se fait l'articulation ainsi que selon l'endroit dans le conduit où se produit la réduction du conduit. Ainsi, les consonnes plosives (/p/, /b/, /d/, /t/, /k/, /g/) possèdent une manière caractérisée par une fermeture complète du conduit vocal suivie par une ouverture brutale ainsi qu'une place pouvant être bilabiale pour les /p/ et les /b/ ou bien alvéolaire pour les /t/ et les /d/. De nombreuses études s'intéressent ainsi à la perception catégorielle des sons du langage (consonnes, voyelles, syllabes). Par exemple, comme je le disais au sujet des consonnes, plusieurs paramètres participent à leur catégorisation et notamment le Voice Onset Time (temps de début de voisement). Ce paramètre permet aux auditeurs de percevoir les consonnes comme étant voisées (par exemple /b/) ou non voisées (par exemple /p/). Le VOT correspond à l'intervalle de temps présent entre l'explosion qui se produit au niveau du relâchement consonantique et le début de l'onde périodique provoquée par la vibration des cordes vocales (Lisker & Abramson, 1967). En français par exemple, les consonnes ayant un VOT négatif sont plutôt catégorisées comme étant voisées (donc comme des /b/) alors que les consonnes ayant un VOT positif sont plutôt catégorisées comme non voisées (donc comme des /p/).

De manière intéressante, la perception des phonèmes, des consonnes et des voyelles est modifiée par l'expérience des auditeurs. En effet, à l'âge de 6 mois, les enfants reconnaissent aussi bien les phonèmes de leur langue maternelle que les phonèmes d'une langue étrangère. Kuhl et

collaborateurs (2006) ont testé la discrimination de contrastes consonantiques (/r/ vs /l/) utilisés dans la langue anglaise et pas dans la langue japonaise chez des enfants américains et japonais de 6 mois et 12 mois. Alors qu'à 6 mois, les enfants japonais et américains ne sont pas sensibles au contraste phonémique ra-la, une dichotomie dans les performances des deux groupes d'enfants est observée à 12 mois. Les enfants américains présentent une augmentation de leurs performances entre 6 et 12 mois alors que les enfants japonais présentent un patron inverse. Cette étude illustre bien les concepts d'enculturation linguistique suggérant que le répertoire de phonèmes perçus est modulé en fonction de l'environnement linguistique dans lequel les enfants évoluent. Cette étude montre également que la sensibilité pour les contrastes phonémiques de la langue maternelle augmente entre 6 et 12 mois alors que la sensibilité pour les contrastes phonémiques des langues étrangères diminue pendant cette période. Cependant, dans le cadre de la comparaison entre langage et musique, ce sont plutôt les voyelles qui disposeraient d'un caractère musical. En effet, les voyelles sont porteuses d'une information de hauteur (la fréquence fondamentale de la voyelle f_0) provenant de la vibration des cordes vocales, vibration induite par le passage de l'air depuis les poumons jusqu'à la cavité buccale. Parallèlement à la vibration des cordes vocales, la position de la langue va moduler l'amplification du filtre résonateur que forme la cavité buccale. Ce résonateur va être à l'origine des premier et deuxième formants dont les variations déterminent le type de voyelle perçue. Les capacités de perception et de production du langage sont reliées à la conscience phonologique qui correspond à la capacité d'identification des différents sons des mots. Cette capacité est notamment importante pour les enfants en phase d'apprentissage de la lecture. En effet, le défi pour les enfants lorsqu'ils commencent à apprendre à lire est de mettre en correspondance les formes visuelles des mots avec leur forme auditive. Je ne m'attarderai pas sur ce point, cependant il semblerait que cette mise en correspondance de la forme acoustique des sons du langage avec la forme visuelle de ces sons soit affectée chez les sujets dyslexiques.

1.4 Les structures rythmiques.

1.4.1 Les structures rythmiques dans la musique.

Le terme de rythme concerne l'organisation des durées et des intervalles temporels entre évènements se succédant dans le temps. Le terme de rythme renvoie la plupart du temps au concept de périodicité caractérisant un patron d'évènements périodiques qui se répète dans le temps. Cependant, les structures rythmiques ne sont pas toutes périodiques et c'est notamment le cas dans le langage. Dans la plupart des musiques occidentales il est possible de distinguer plusieurs niveaux de périodicité rythmique tels que la pulsation et le mètre. Ainsi, les régularités temporelles des événements musicaux vont permettre aux auditeurs de synchroniser leurs mouvements avec la pulsation et vont également permettre aux musiciens de disposer d'une référence temporelle commune en vue de jouer de manière synchrone. Cependant, tous les événements musicaux n'arrivent pas forcément sur la pulsation et c'est cette alternance dans l'occurrence des événements musicaux par rapport à la pulsation qui va être à l'origine de la perception des temps forts et faibles dans la musique. Ainsi, dans une valse qui contient trois temps par mesure, les auditeurs perçoivent le premier temps de chaque mesure comme un temps fort et les deux suivants comme des temps faibles. Ainsi, la perception du rythme implique que les auditeurs possèdent une représentation mentale de patrons rythmiques leur permettant de développer des attentes temporelles des évènements musicaux (Jones, 1976). L'alternance des temps forts et faibles va induire la perception d'un niveau supérieur de périodicité, le mètre. Ce second niveau d'organisation rythmique va être à l'origine des structures de groupes référant à la capacité de grouper perceptivement des évènements musicaux successifs. Il semble donc exister une hiérarchie dans la perception des structures musicales qui trouve des points de comparaison avec le domaine linguistique.

1.4.2 Les structures rythmiques dans le langage.

Une des différences fondamentales entre langage et musique vis à vis des structures métriques est que la musique est basée sur une relative isochronie des événements alors que le langage ne l'est pas. Pourtant, il existe plusieurs niveaux de régularités temporelles au sein du signal de parole. Ces différents niveaux permettent aux auditeurs d'avoir une représentation mentale de la structure métrique du langage.

Plusieurs indices acoustiques vont ainsi permettre aux auditeurs de grouper les événements linguistiques entre eux tels que les patrons de durées et les patrons d'accentuation. Ces patrons d'alternances participent à la formation des indices prosodiques du discours. La prosodie va avoir une fonction dans la segmentation du discours grâce à la présence d'indices temporels que sont les patrons d'accentuation (stress patterns) et qui trouvent leur origine dans l'accent lexical. L'accent lexical est un indice métrique qui consiste la plupart du temps en l'augmentation corrélée de la durée, de la hauteur et de l'intensité de syllabes spécifiques dans les mots (Abercrombie., 1967). Cet indice métrique va contraindre l'alternance entre syllabes FORTES et syllabes faibles et ainsi être à l'origine de la perception de patrons métriques dans le discours. Ces patrons métriques vont finalement permettre de considérer différents types de langues en fonction du type de patron présent. Par exemple, en anglais, une étude de corpus de discours a révélé qu'environ 90 % des mots commencent par une syllabe accentuée (Cutler & Carter, 1987). Ce patron d'accentuation syllabique, de type trochaïque, n'est pas retrouvé dans certaines langues comme le français ou le polonais où le patron sera de type iambique (faible fort), et ce de part l'allongement de la dernière syllabe des mots.

Chapitre 2 : Corrélats électrophysiologiques de la perception du langage et de la musique.

Les Potentiels évoqués obtenus à partir de l'enregistrement continu de l'activité électrique du cerveau permettent de suivre à la milliseconde près le déroulement temporel de l'activation des zones corticales impliquées dans la perception de stimuli sensoriels. Dans le cas de la perception auditive du langage et de la musique, plusieurs ondes caractérisées par leur polarité, leur latence et leur distribution topographiques ont été révélées.

2.1 Traitement attentif des paramètres acoustiques des stimuli: le complexe N1/P2.

La composante N1 est une onde négative dont le maximum d'amplitude se situe en général 100 ms après la présentation du stimulus auditif. Sa distribution topographique trouve son maximum d'amplitude au niveau des électrodes fronto-centrales (Näätänen & Picton, 1987). Cependant, de par les phénomènes de conduction volumique, son générateur ne semble pas se trouver véritablement dans ces régions. En effet des études ont montré qu'il se situerait au niveau du cortex auditif secondaire (Hackett et al. 2001; Liégeois-Chauvel et al. 1994; Pantev et al., 1988, 1995; Yvert et al., 2005). L'amplitude de la N100 est sensible à différents paramètres acoustiques tels que, le temps d'attaque du stimulus, son intensité, sa fréquence de présentation. Typiquement, l'amplitude de la N1 est plus grande pour des sons ayant un temps d'attaque court (comme des sons de pianos ou des consonnes plosives) que pour des sons ayant un temps d'attaque long (comme des sons de violons ou des voyelles) (Loveless & Brunia, 1990). De la même manière, plus un son sera présenté à une forte intensité, plus l'amplitude de la N1 évoquée sera grande (Picton et al.,

1970) et finalement, plus l'intervalle de temps entre 2 stimulations sera grand, plus l'amplitude de la N1 sera grande (Hari et al., 1982). Il est également important de noter que l'amplitude de cette composante est affectée par le niveau d'attention des participants. Ainsi, l'amplitude de la N1 est plus grande lorsque les participants font attention aux sons (par exemple dans une tâche de détection de déviants) que lorsque les participants écoutent de manière passive (pendant qu'ils lisent un livre ou regardent un film muet) (Hillyard et al., 1973; Woldorff et al., 1993). J'exposerai dans le chapitre 3, l'effet de l'expertise musicale sur l'amplitude de la N1 et de la P2 dans des tâches perceptives. De plus, dans le chapitre 5 traitant des corrélats électrophysiologiques sous tendant l'apprentissage implicite de structures linguistiques, je reporterai une revue des travaux menés jusqu'à présent et qui rendent compte de modifications de l'amplitude de la N1 en fonction de la qualité de l'apprentissage.

La composante P2 est quant elle une déflexion positive qui trouve son maximum d'amplitude environ 200 ms après la présentation du stimulus. La littérature concernant la P2 est, à ma connaissance plus restreinte. Son amplitude est encore une fois sensible aux paramètres acoustiques des stimuli avec une amplitude plus grande pour des sons ayant un temps d'attaque court que pour des sons ayant un temps d'attaque long (Shahin et al., 2005). Son amplitude est, tout comme la N1, modulée en fonction du niveau d'attention des participants et, lors d'une tâche de ségrégation de 2 flux auditifs, son amplitude est plus grande lorsque les deux flux sont perçus comme 2 flux séparés que lorsqu'ils sont perçus comme un seul flux (Hillyard et al., 1987; Snyder et al., 2006). Finalement, il a été montré que le générateur de la P2 se situait non pas au niveau du cortex auditif primaire mais plus en aval de la chaîne de traitement de l'information à savoir au niveau du cortex auditif secondaire (Godey et al., 2001). De la même manière que pour la N1, je reporterai les travaux ayant montré des modifications de l'amplitude la P2 en fonction de l'apprentissage, dans le chapitre 5.

2.2 Traitement pré-attentif des sons: la Mismatch Negativity ou Négativité de discordance.

Il est possible d'étudier les processus de traitement pré-attentif des sons en utilisant le protocole de Mismatch Negativity (Näätänen et al., 1978; Kujala et al., 2007). Lorsque des sons déviants sur un paramètre acoustique (Intensité, localisation spatiale, fréquence, durée) sont introduits dans une séquence de sons standards, identiques et répétés, pendant que les participants ont leur attention focalisée sur un film muet ou un livre, on observe une composante négative de distribution fronto-centrale dont le maximum d'amplitude se situe entre 100 et 250 msec après la présentation du son déviant. Cette composante est obtenue en effectuant la différence entre l'activité évoquée par le déviant et l'activité évoquée par le standard. Elle est interprétée comme reflétant un processus de détection automatique d'un changement par rapport à la trace mnésique établie par la répétition du son standard (Näätänen et al., 1978) ou, plus récemment comme étant une N1 issue de l'activation de neurones non adaptés par le son standard (May & Tiitinen, 2009). Il est possible d'observer une MMN aussi bien pour des stimuli linguistiques que pour des stimuli non linguistiques et, de plus, il a récemment été montré que l'amplitude de la MMN n'était pas affectée par la présence de plusieurs types de déviants dans la séquence voire même qu'il était possible d'observer une MMN dans une séquence ne contenant pas de standard (Pakarinen et al., 2007, 2010). Ces dernières découvertes apparaissent comme étant un progrès en terme expérimental car il est ainsi possible de diminuer les temps d'expérimentation tout en obtenant des données de qualité suffisante auprès de populations difficiles à tester comme des bébés, des jeunes enfants ou des sujets pathologiques.

2.2.1 Traitement du sens dans le langage et la musique: la N400.

De nombreux travaux se sont consacrés à étudier précisément le décours temporel de processus cognitifs de haut niveau comme par exemple le traitement du sens. Il y a 30 ans, Kutas et

Hillyard montraient dans la modalité visuelle, que les potentiels évoqués par des mots incongrus par rapport au contexte établi dans une phrase, présentaient une négativité centro-pariétale plus ample aux alentours de 400 ms que les potentiels évoqués par les mots congruents. Cette différence était interprétée comme reflétant une plus grande difficulté d'intégration du sens du mot incongru par rapport au contexte (Kutas & Hillyard, 1980; 1984). La N400 s'observe également dans la modalité auditive avec une topographie plus fronto-centrale. De plus, elle est sensible à différents facteurs comme la familiarité du mot cible ou sa position dans la phrase (Kutas & Federmeier, 2000). Ainsi, la N400 est plus ample pour des mots peu fréquents (non familiers) que pour des mots fréquents (familiers) (Young & Rugg, 1992). La N400 est, dans ce dernier cas, interprétée comme reflétant un accès plus difficile à la mémoire sémantique pour les mots incongrus. Elle est également sensible à la facilité d'accès de l'information stockée en mémoire à long-terme (Neville et al., 1986) et, dans le cadre de l'apprentissage, elle prédit bien les performances de rappel immédiat suivant l'apprentissage (Elger et al., 1997). Concernant le traitement du sens dans la musique ou plus correctement le traitement des concepts musicaux, ce n'est que récemment que des études ont montré qu'une N400 pouvait être révélée dans un protocole d'amorçage musical (Koelsch et al, 2004; Daltrozzo & Schön, 2009a, 2009b). Ce protocole consiste en la présentation d'une amorce suivie d'une cible. L'amorce qui dans ce cas peut être un mot ou une phrase musicale est supposée créer un contexte qui va influencer le traitement de la cible. Cette amorce est suivie d'un mot ou d'une phrase musicale cible dont on manipule le degré de relation conceptuelle par rapport à l'amorce. Ces études ont permis de montrer que la N400 se développait même lorsque les cibles étaient des phrases musicales très courtes et donc que le traitement du sens dans la musique reposerait sur des processus cognitifs partagés avec le traitement du sens dans le langage. De manière intéressante, la composante N400 s'est vue impliquée dans des protocoles d'apprentissage statistique de langage artificiels. Dans la partie traitant de l'apprentissage implicite de langages artificiels, je reporterai les études qui ont suggéré son implication dans le processus de segmentation.

2.3 Traitement de la syntaxe dans le langage et la musique.

La syntaxe correspond à un ensemble de principes qui régissent la combinaison des unités de base du langage et de la musique que sont les mots et les notes (Jackendoff, 2002). De nombreux travaux ont été effectués afin de mieux comprendre les processus cognitifs impliqués dans le traitement de la syntaxe, tout d'abord au niveau du langage puis au niveau de la musique. Lorsque l'on introduit une violation de la structure syntaxique à l'intérieur d'une phrase et que l'on enregistre l'EEG pendant sa lecture, on peut observer trois composantes des Potentiels Evoqués plus amples dans la condition de violation syntaxique que dans la condition « normale ». Ces trois composantes sont observées chacune pour des violations syntaxiques spécifiques: la Early Left Anterior Negativity (ELAN) est observée pour des violations de syntaxe au niveau local (i.e. lorsque le dernier mot n'appartient pas à la catégorie du mot typiquement attendu, Friederici et al., 1993; Hahne & Friederici, 1999) et se développe rapidement à partir de 100 msec. La LAN est quant à elle observée pour des violations morphosyntaxiques (Friederici et al., 1993; Neville et al., 1991) et peut se développer de 100 à 300 msec, alors que la P600 est observée pour des phrases syntactiquement complexes et pour des phrases contenant une violations syntaxiques nécessitant un processus de révision syntaxique (Friederici & Mecklinger, 1996; Osterhout & Holcomb, 1992, 1993). La P600 a une topographie centro pariétale et se développe après la LAN, à partir de 600 msec.

Au niveau musical, le traitement de la syntaxe a été étudié au niveau harmonique en présentant une suite d'accords se terminant par un accord final plus ou moins relié à la structure harmonique créée précédemment (i.e. un accord plus ou moins attendu donc plus ou moins facile à traiter). Le potentiel évoqué par le dernier accord non relié présentait une Négativité maximale au niveau frontal droit (la ERAN) dont le maximum est à 200 msec (Koelsch et al., 2000, 2002). Cependant, une erreur au niveau de la dernière note d'une séquence peut également entraîner la formation d'une P600 plus ample pour l'incongruité harmonique que si le dernier accord ou la

dernière note sont harmoniquement corrects (Besson & Macar, 1987; Besson & Faïta, 1995). Ceci dit, à mon sens l'étude la plus importante dans la thématique de transfert d'apprentissage de la musique vers le langage reste celle de Patel et collaborateurs (1998). Chez les mêmes sujets, ces auteurs ont enregistré les Potentiels évoqués pour des violations syntaxiques dans le langage et dans la musique. Les deux types de violations étaient caractérisés par une P600 de latence et de distribution topographique similaires. En 2003, Patel proposait ainsi que l'intégration des structures linguistiques et musicales se basent sur les mêmes processus de traitement suggérant que les ressources neuronales dédiées à l'intégration des structures sont, en partie, communes aux deux systèmes linguistiques et musicaux (Shared Syntactic Integration Ressources Hypothesis, SSIRH, Patel, 2003).

Chapitre 3: Neuroplasticité et expertise musicale.

Dans le champ des neurosciences utilisant le modèle animal, de nombreuses études ont montré que des modifications d'efficacité du fonctionnement des synapses se produisaient suite à une stimulation répétée ou après un apprentissage associatif. Chez l'homme, de nombreuses études en neurosciences cognitives s'intéressent aux effets d'un apprentissage spécifique sur la réalisation d'une tâche cognitive. Typiquement, l'hypothèse de base est que la pratique intensive et répétée d'une activité spécifique va induire, en retour, des modifications anatomo-fonctionnelles du cerveau expert en comparaison avec celui du cerveau non expert. Ainsi, au niveau structural, il a été montré que la quantité de matière grise au niveau de l'hippocampe est plus grande chez les conducteurs de taxi que chez des personnes novices de même âge. De plus, la quantité de matière grise est positivement corrélée avec le nombre d'années passées à conduire un taxi (Maguire et al., 2000). La pratique professionnelle de la musique apparaît quant à elle comme une activité hyper complexe, tant au niveau sensoriel qu'au niveau moteur. En moyenne, les musiciens professionnels ont suivi une période d'une dizaine d'années d'apprentissage au conservatoire, correspondant à environ 10000 heures de pratique (Sloboda 2005; Krampe & Ericsson 1996). A terme, le musicien d'orchestre va devoir savoir placer ses mains sur son instrument, effectuer les bons mouvements au bon moment tout en lisant en temps réel une partition comprenant une quantité considérable d'indications sur le type de note à jouer et la manière de la jouer, tout cela en prenant en compte les ordres du chef d'orchestre et bien entendu, tout en écoutant les autres musiciens afin d'ajuster son propre jeu si besoin est. Dans le jazz et les musiques improvisées, les interactions entre systèmes perceptifs et moteurs sont très intenses puisque le musicien doit avant tout écouter ses partenaires afin qu'ensemble ils fassent émerger des structures harmoniques et rythmiques très complexes. Autant dire que la pratique d'un instrument est une des activités les plus complexes que l'homme sache maîtriser. Cependant, il est important de définir ce que l'on entend par musicien expert et par non musiciens. Les musiciens adultes qui ont participé aux expériences effectuées au cours de ma thèse pratiquaient tous au moins un instrument de musique de façon quotidienne. La majeure partie de ces

musiciens étaient issus de conservatoires nationaux et avaient validé leurs diplômes de fin d'étude théorique et pratique. Une grande partie de ces musiciens se destinaient à l'enseignement de la musique et étaient inscrits au Cefedem d'Aubagne (Centre de Formation Des Enseignants de la Musique) pour suivre une formation diplômante d'environ 1000 heures. Seulement deux ou trois participants n'étaient pas inscrits au Cefedem mais étaient musiciens professionnels et vivaient de la musique. Les participants non musiciens étaient issus de la même classe d'âge que les musiciens mais n'avaient, quant à eux, jamais suivi d'apprentissage de la musique en conservatoire ou en école de musique. Dans le champ des neurosciences cognitives, le cerveau musicien est considéré comme un modèle idéal pour étudier les phénomènes de neuroplasticité liés à la pratique intensive de la musique (Jäncke, 2002, Münte et al., 2002, Schlaug et al., 2001). De nombreuses études, tant au niveau structural que fonctionnel, ont permis de mettre en évidence des différences, la plupart du temps en faveur des musiciens. Dans la partie qui suit, je présente une revue de la littérature dont l'objectif est de cerner les différences anatomo-fonctionnelles qui existent entre musiciens et non musiciens.

Au niveau anatomique, plusieurs études utilisant l'IRM ont permis de montrer que les musiciens professionnels présentent de plus grandes aires corticales (i.e. une plus grande quantité de matière grise) impliquées dans le traitement des stimuli auditifs et dans le contrôle moteur que les non musiciens. Ces études ont généralement montré que les musiciens présentent une plus grande quantité et / ou densité de matière grise au niveau : du cortex auditif (Bermudez & Zatorre, 2005), du corps calleux (Schlaug et al., 1995a, Schmithorst et al., 2002), du planum temporale (Keenan et al., 2001, Luders et al., 2004, Schlaug et al., 1995b), du cortex moteur primaire et de certaines parties du cervelet (Hutchinson et al., 2003), de la partie antéro-médiale du gyrus de Heschl (Schneider et al., 2002), du gyrus frontal inférieur et du lobe temporal dans sa partie latérale inférieure (Gaser & Schlaug, 2009, Sluming et al., 2002).

3.1 Traitement des sons non linguistiques et expertise musicale.

Le traitement des sons non linguistiques peut s'étudier à différents niveaux en utilisant des indices comportementaux et/ou des indices électrophysiologiques. D'une part, au niveau psychophysique, il a été montré que les musiciens présentaient un seuil de détection fréquentiel plus petit que celui de non musiciens (Kishon Rabin et al. 2001; Micheyl et al., 2006). Concernant la dimension temporelle, les musiciens sont meilleurs que les non musiciens dans des tâches de détection de changement de tempo (Ehrle & Samson, 2005), de détection de variations de timing (Jones et al., 1995). D'autre part, il va être possible d'étudier plus finement (soit au niveau du décours temporel soit au niveau de la localisation des régions impliquées) le traitement des sons à l'aide des techniques de neuroimagerie. Ces techniques vont permettre notamment de mieux différencier les niveaux de traitement implicites ou explicites, des sons isolés ou dans un contexte musical. De plus, il est généralement admis qu'il est possible de différencier les traitements automatiques et contrôlés des sons. Ainsi, l'étude du traitement automatique des sons contextualisés ou non, se fait à l'aide de protocoles de type oddball grâce auquel il est possible d'observer la négativité de discordance (MMN), interprétée comme le corrélat électrophysiologique de la détection passive, automatique, d'un changement par rapport à la trace mnésique formée par un son répété (Näätänen et al., 1978).

Concernant la détection automatique de la fréquence, Koelsch et collaborateurs (1999) montraient que les musiciens présentaient une MMN plus ample que les non musiciens pour des sons déviants dont la fréquence était légèrement diminuée par rapport au standard. Tervaniemi et al. (2005), ne montraient pas de telle différence. Cependant, l'amplitude de la déviance par rapport au standard était grande et peut suggérer un effet plafond: lorsque les déviants sont trop faciles à détecter, les sujets non musiciens traitent le déviant de la même manière que les musiciens, ce qui se traduit par une MMN de même amplitude dans les deux groupes. Ainsi l'utilisation de fines variations de fréquence semble être plus judicieuse pour révéler des différences en faveur des musiciens.

Concernant la dimension temporelle, en comparaison avec les non musiciens, les musiciens

présentent une MMN plus ample en réponse à l'omission d'un son dans une séquence régulière (Rüsseler et al. 2001) ainsi que pour des déviants en durée (Marie et al., 2010b). Il a également été montré que les musiciens présentaient une MMN plus ample que les non musiciens pour des déviants en localisation spatiale (Tervaniemi et al., 2006). Ces résultats suggèrent ainsi que les musiciens disposent non seulement d'une plus grande fenêtre d'intégration temporelle que les non musiciens mais qu'ils disposent également d'une meilleure représentation des sources sonores probablement issues de la pratique musicale en groupe où la précision temporelle et la localisation du son sont primordiales.

Il est également possible d'étudier les composantes plus classiques des potentiels évoqués corticaux comme la N1 et la P2 qui reflèteraient le traitement des paramètres acoustiques des sons au niveau cortical. Ainsi, en utilisant la Magnéto Encéphalographie (MEG), il a été montré que les musiciens présentaient une N1m (pendant magnétique de la N1) plus ample que les non musiciens lors de l'écoute passive (pendant le visionnage d'un dessin animé) de sons de pianos (Pantev et al., 1998). Cette composante est plus ample pour l'instrument de pratique du musicien que pour un autre instrument: les trompettistes présentent une N1 plus ample pour des sons de trompettes que des violonistes, et inversement (Pantev et al., 2001). Shahin et al., ont montré que la N1 ainsi que la P2 étaient plus ample chez des musiciens professionnels que chez des non musiciens et ce, pour des sons purs et des sons d'instruments (piano et violon, Shahin et al., 2003, 2005). Ces résultats suggèrent que les musiciens ont développé de meilleures représentations des sons en général et que ces représentations seraient spécifiques de l'instrument de pratique. De plus, une étude récente montre que les différences retrouvées au niveau de l'amplitude de ces composantes ne semblent pas être dues à une plus grande attention sélective mais plutôt qu'elles reflètent de meilleures représentations des caractéristiques acoustiques des sons pour les musiciens (Baumann et al., 2008).

Concernant la détection active de sons, des différences corticales et sous-corticales ont été montrées. Au niveau cortical, il a été montré à l'aide d'un protocole similaire à celui de la MMN à la seule différence que les participants devaient appuyer le plus vite possible sur un bouton dès qu'ils

détection de déviants, que les musiciens présentaient une N2b et une P300 aux déviants plus amples que les non musiciens. Cette différence électrophysiologique était associée à de meilleures performances de discrimination des déviants reflétées par un plus faible pourcentage d'erreur et un Temps de Réaction plus court que les non musiciens (Koelsch et al., 1999; Tervaniemi et al., 2005; Tremblay et al., 2001). Si les différences au niveau cortical apparaissent désormais comme un fait établi, l'influence de l'expertise musicale sur la plasticité de régions sous corticales n'est que très récente. Le groupe de Nina Kraus est très actif dans ce domaine de recherche et s'intéresse à la plasticité induite au niveau du tronc cérébral. Généralement, la réponse du colliculus supérieur à des stimuli auditifs semble de meilleure qualité chez les musiciens que chez les non musiciens de par le fait que le signal de réponse du tronc cérébral des musiciens est plus similaire au signal du stimulus d'origine que chez les non musiciens (Wong et al., 2007; Musachia et al., 2007; Parbery-Clark et al., 2009; Bidelman & Krishnan, 2010). La plupart de ces études se base sensiblement sur le même protocole et je ne décrirai ainsi que l'étude la plus récente dans la partie sur le traitement des sons linguistiques isolés.

3.2 Traitement de sons non linguistiques contextualisés.

Si une MMN est observée pour des déviants sur les différents attributs acoustiques des sons, des chercheurs se sont posés la question de savoir si une MMN pouvait être observée à des déviations plus complexes. Ainsi il a été montré que les musiciens présentaient une MMN plus ample que les non musiciens en réponse à une déviation du contour mélodique ou de l'intervalle (Tervaniemi et al., 2001; Trainor et al., 2002; Fujioka et al., 2004, Habermeyer et al., 2009). Dans ce type de protocole, des séquences de plusieurs notes formant un contour stable ou un intervalle stable (e.g. 4 notes ascendantes et la 5ème descendante) étaient présentées de manière répétée et, des changements de ce contour mélodique ou de l'intervalle étaient introduits au niveau d'une des notes dans une séquence déviante (e.g. La 4ème note devient descendante). Une autre étude a montré que les musiciens présentent une MMN plus ample et plus précoce pour une variation de fréquence

présentée dans un contexte familier que dans un contexte non familier (Brattico et al., 2001). Cependant, alors que le traitement de la fréquence dans un contexte a été plutôt bien étudié, seulement peu d'études se sont intéressées à la dimension temporelle. Vuust et al. (2005) ont montré que la MMN en réponse à une incongruité rythmique introduite dans une séquence complexe était plus ample pour les musiciens Jazz que pour les non musiciens. De manière intéressante, cette MMN était plus latéralisée à gauche chez les musiciens que chez les non musiciens, suggérant un recouvrement entre les aires dédiées au traitement des indices acoustiques musicaux et celles dédiées au traitement des indices acoustiques linguistiques. Finalement, les musiciens présentent une MMN plus ample que les non musiciens lorsqu'une note supplémentaire est ajoutée à une séquence répétée contenant plusieurs notes, suggérant un meilleur traitement du nombre d'évènements perçus (Van Zuijen et al., 2005).

De nombreuses études ont également été menées pour mieux caractériser le fonctionnement du cerveau musicien vis à vis du traitement d'une fausse note ou d'un accord placé à la fin d'une phrase musicale. L'idée partait de la découverte de la composante N400 dans le domaine linguistique, composante censée refléter le traitement d'incongruités sémantiques vis à vis d'un contexte créé le long d'une phrase. Puisque le langage et la musique présentent certaines caractéristiques structurelles qui sont comparables, allait on observer une composante N400 en réponse à une incongruité contextuelle? Ainsi, au milieu des années 90, Besson et collaborateurs (1995) comparaient les PEs des musiciens professionnels à ceux de non musiciens dans une tâche de détection d'incongruités rythmiques et harmoniques. Les participants écoutaient des phrases musicales familières ou non familières qui se terminaient par une note congruente ou incongrue (au niveau rythmique ou au niveau harmonique). Ils devaient décider si la phrase musicale était familière ou pas et en même temps, devaient catégoriser la dernière note comme juste ou fausse. Cette étude était la première à montrer au niveau comportemental et électrophysiologique que les musiciens traitaient les incongruités musicales différemment des non musiciens. Les musiciens faisaient moins d'erreurs que les non musiciens seulement pour les phrases musicales non

familiales et présentaient une P600 suscitée par la note incongrue, plus précoce et plus ample que les non musiciens suggérant que le langage et la musique étaient indépendants au niveau de l'intégration d'incongruités contextuelles. Quelques années plus tard, la même équipe menait une série d'étude chez le musicien adulte et le musicien enfant, basée sur le même paradigme de détection d'incongruités dans la musique et le langage (Schön et al., 2004; Magne et al., 2006). Dans ces deux études, les incongruités harmoniques sur la dernière note d'une phrase musicale pouvaient être faibles (difficiles à détecter: 1/5 de ton) ou fortes (faciles à détecter: 1/2 de ton). Les adultes et les enfants musiciens avaient de meilleures performances que les non musiciens pour la détection des faibles variations de hauteur. De manière intéressante, les musiciens aussi bien adultes qu'enfants, présentaient une P600 en réponses aux incongruités plus précoce et plus ample que les non musiciens. Un autre groupe de recherche mené par Stefan Koelsch s'est intéressé à étudier le traitement de violations syntaxiques de l'harmonie dans la musique. Pour se faire, ils présentaient une séquence d'accords formant une cadence. La cadence est généralement utilisée pour conclure une pièce musicale et permet à l'auditeur de former une attente harmonique forte au niveau du dernier accord. Les résultats révèlent que la présentation d'un accord napolitain en dernière position de la séquence entraîne le développement d'une ERAN maximale à 200 ms au niveau frontal droit (Koelsch et al., 2002a). Lors de la comparaison d'un groupe de musiciens à un groupe de non musiciens, les auteurs reportaient une ERAN plus ample chez les experts que chez les non experts suggérant que la pratique musicale renforce les représentations des structures syntaxiques musicales (Koelsch et al., 2002b). Cependant, les deux critiques fortes faites à l'encontre de ce type d'études sont que les conclusions reposent sur des corrélations et non pas sur des causalités et que les effets différents observés chez les musiciens sont susceptibles de provenir de prédisposition génétiques et non pas d'un apprentissage intensif de la musique. Même si ces études montrent des effets positivement corrélés avec le nombre d'années de pratique et/ou l'âge de début de la pratique musicale, la seule manière de répondre à la question sur l'origine génétique ou expérimentale des différences observées chez les musiciens est de mener des études longitudinales chez l'enfant naïf et

de lui faire suivre un apprentissage intensif de la musique comparé à un autre type d'apprentissage tout aussi riche comme apprendre la peinture par exemple. Je reporterai dans la partie consacrée aux études longitudinales une revue des quelques études ayant utilisé cette méthode et présenterai l'approche que nous avons spécifiquement utilisé.

Si l'effet de l'expertise musicale sur le traitement des sons dans un contexte musical n'est pas si surprenant que ça, il n'en est pas de même quant au traitement du langage. En effet, depuis les travaux de Pinker soutenant l'idée que le langage est complètement hermétique vis à vis des autres fonctions cognitives, l'idée que l'expertise musicale puisse influencer le traitement du langage est encore mal acceptée par certains scientifiques. Pourtant un bon nombre de publications vont dans le sens d'un possible transfert de compétences de la musique vers le langage. Je reporte une revue de la littérature allant dans ce sens dans la partie qui suit.

3.3 Traitement des sons linguistiques et expertise musicale.

Sur quelles bases pouvons-nous penser qu'un transfert d'apprentissage entre musique et langage est susceptible d'exister?

D'une part, à un niveau général, langage et musique partagent des similarités en terme d'origine évolutive et d'organisation (Patel, 2008; Besson & Schön, 2001). De plus, les informations auditives linguistiques et non linguistiques vont passer par les différents relais de la chaîne de traitement dans les voies auditives. Dans la mesure où les musiciens présentent des caractéristiques différentes de celles des non musiciens sur des mesures électrophysiologiques et anatomiques au niveau du tronc cérébral et des cortex auditifs primaires et secondaires, il est possible d'émettre l'hypothèse que les paramètres acoustiques de stimuli linguistiques ne seront pas traités de la même manière chez les musiciens et les non musiciens. Cette hypothèse trouve un appui fort dans la méta-analyse de Vigneau et al. (2006) qui semble aller totalement dans le sens d'un partage des ressources impliquées dans le traitement du langage et de la musique. En répertoriant les sites d'activation obtenus dans des tâches linguistiques (phonologique, sémantique et syntaxique), ces auteurs ont pu

mettre en évidence un fort recouvrement des aires corticales impliquées dans le traitement du langage et de la musique. De plus, nous avons pu remarquer que les corrélats électrophysiologiques du traitement de la sémantique et de la syntaxe dans les deux domaines étaient similaires. Nous avons également vu que dans le domaine musical, les corrélats électrophysiologiques sous tendant le traitement de la sémantique et de la syntaxe étaient différents entre musiciens et non musiciens. Encore une fois, ceci peut nous amener à émettre l'hypothèse que la pratique musicale peut affecter positivement la façon de traiter les stimuli linguistiques.

3.4 Traitement de sons linguistiques isolés.

Comme nous avons pu le remarquer, les études sur l'effet de l'expertise musicale au niveau de la perception automatique des sons non linguistiques sont fleurissantes. Par contre, l'effet de l'expertise musicale sur le traitement automatique de mots isolés n'est pas un domaine encore très fourni. A ma connaissance, seulement quelques études ont tenté d'étudier cet effet au niveau automatique (Wong et al., 2007, Bidelman & Krishnan., 2010, Parbery-Clark et al, 2009). Dans l'étude de Wong et al, les potentiels évoqués du tronc cérébral étaient enregistrés en réponse à la syllabe « mi » présentée à l'oreille droite pendant que les participants regardaient et écoutaient un film, l'oreille gauche laissée libre. Cette étude était la première à montrer d'une part que la Frequency Following Response, correspondant à la Réponse du tronc cérébral, est plus robuste chez les musiciens que chez les non musiciens, et d'autre part que cette réponse est corrélée non seulement avec le nombre d'années de pratique de la musique mais également avec l'âge de début de la pratique musicale. Une autre étude utilisant cette fois le protocole MMN multi-feature avec des syllabes vient d'être acceptée (Chobert et al., in press) et compare le traitement automatique de déviants sur la fréquence, la durée et la phonologie chez des enfants musiciens et non musiciens de 8 ans. De manière intéressante, les résultats révèlent une MMN plus ample pour les déviants en durée et en VOT pour les musiciens que les non musiciens suggérant ainsi que la pratique musicale a pu développer la qualité des représentations de stimuli linguistiques. Ces études apportent un premier

faisceau d'arguments forts en faveur de l'hypothèse de transfert d'apprentissage de la musique vers le langage.

Le traitement de sons linguistiques a été largement abordé au niveau comportemental en utilisant différentes tâches sur différentes populations. L'étude d'Anvari et al (2002) a révélé sur un échantillon de 100 enfants de 4 et 5 ans que les aptitudes perceptives au niveau musical était un bon prédicteur des performances dans une tâche de conscience phonologique et de lecture, suggérant un partage des ressources d'intégration entre la perception de la musique et du langage. Slevc & Miyake (2006) ont par la suite montré chez l'adulte que les aptitudes musicales prédisaient les capacités phonologique en termes de production et de perception de l'apprentissage d'une seconde langue. Ces études ont été suivies par de nombreuses autres, notamment dans des tâches de perception du langage dans du bruit (Parbery-Clark et al., 2009, Kraus & Chandrasekaran, 2010) et semblent suggérer que la pratique musicale influence positivement la perception du langage et les capacités verbales (Milovanov et al., 2008, 2009; Forgeard et al., 2008). En utilisant une approche électrophysiologique, des études ont montré que les potentiels évoqués du tronc cérébral et du cortex auditif en réponse à une syllabe plus ou moins détériorée par du bruit ou de la réverbération sont plus ample chez des musiciens que des non musiciens dans des tâches de perception d'une syllabe. De plus, l'amplitude de ces réponses est positivement corrélée au nombre d'années de pratique musicale (Parbery-Clark et al., 2009; Musacchia et al., 2007, 2008).

3.5 Traitement de sons linguistiques contextualisés.

Toutes les études reportées dans la partie précédente, bien qu'importantes, se focalisaient sur des traitements perceptifs de bas niveau et pourraient s'expliquer par une plus grande capacité de focalisation de l'attention chez les musiciens que chez les non musiciens. De plus, l'utilisation de syllabes isolées n'apparaît pas très écologique d'un point de vue linguistique et finalement, l'intégration de mots dans une phrase repose sur des processus cognitifs de plus haut niveau que le simple traitement de paramètres acoustiques tels que l'extraction de règles structurelles, la capacité

de segmentation du flux de parole ou l'intégration du sens des mots dans une phrase.

Encore une fois, le groupe Marseillais a effectué une série d'études correspondant exactement au pendant de celles réalisées dans le domaine musical. Ainsi, ces auteurs ont étudié le traitement d'incongruités de hauteur sur le dernier mot d'une phrase. Ils ont mené ces études à la fois chez des adultes et des enfants musiciens et ont procédé à une manipulation paramétrique de la fréquence fondamentale du dernier mot d'une phrase simple. Les participants devaient décider si ce dernier mot était prononcé correctement ou pas pendant que l'activité EEG était enregistrée. Les résultats de ces études révèlent que les adultes et les enfants musiciens sont meilleurs que les non musiciens pour détecter les variations de hauteur et ce, particulièrement pour la faible variation (plus difficile). De plus, les PE évoqués par ces incongruités présentaient une positivité tardive entre 200 et 800 ms plus ample que pour les mots prononcés correctement et plus précoce chez les musiciens que les non musiciens (chez l'enfant, cette positivité était même absente) (Schön et al., 2004; Magne et al., 2006). Afin d'aller encore plus loin, cette équipe a mené une étude basée sur le même protocole de manipulation de la hauteur du dernier mot mais en utilisant cette fois une langue étrangère que les participants ne comprenaient pas. Des résultats similaires émergent de cette étude: les musiciens français sont meilleurs dans la détection de ces manipulations de hauteurs et développent une positivité tardive plus précoce que les non musiciens (Marques et al., 2007). Pour parfaire le tableau, une étude longitudinale chez l'enfant utilisant le même protocole expérimental a révélé que ces différences comportementales et électrophysiologiques provenaient spécifiquement de l'apprentissage de la musique et non pas de prédispositions génétiques (Moreno et al., 2009).

Comme nous venons de le voir, cette série d'étude se focalisait sur le traitement de la hauteur dans le langage. Une étude récente vient d'être publiée et étend ces découvertes au traitement du mètre dans le langage (Marie et al., 2011). Dans cette étude, les participants effectuaient deux blocks expérimentaux possédant chacun une tâche spécifique. Les stimuli étaient des phrases qui pouvaient contenir deux types de manipulations: des modifications de la longueur de l'avant dernière syllabe du dernier mot des phrase étaient introduites et, en parallèle, les derniers mots pouvaient être

congruents ou incongrus dans le contexte de la phrase. Ainsi, les participants devaient dans une première tâche, focaliser leur attention sur la structure métrique du langage alors que dans la deuxième tâche, ils devaient focaliser leur attention sur l'aspect sémantique du dernier mot. Des résultats intéressants émanent de cette étude et révèlent que les musiciens sont meilleurs que les non musiciens dans les deux tâches attentionnelles, ils font moins d'erreurs que les non musiciens. De plus les résultats électrophysiologiques confortent ces premiers résultats puisque les musiciens présentaient une P200 plus ample pour les mots métriquement incongrus quelle que soit la direction de leur attention. Les musiciens présentaient également une P600 aux incongruités métriques dans les deux tâches qui était présente uniquement dans la tâche métrique chez les non musiciens. Cependant, il n'y avait pas de différence entre les deux groupes concernant l'effet d'incongruité révélé par l'effet N400 suggérant que l'expertise musicale n'influence pas l'accès au sens des mots.

Chapitre 4: Approche longitudinale des effets de la pratique musicale.

4.1 Principes de base.

Dans le cadre de l'étude de l'influence de la pratique musicale sur le traitement du langage et de la musique, la littérature est riche. Ces études montrent des effets bénéfiques sur les performances dans des tâches cognitives basées sur des processus généralement directement influencés par la pratique musicale (perception, motricité). Tout un pan de recherche vise également à caractériser les effets de cette pratique musicale sur des compétences non musicales telles que l'intelligence générale ou les compétences linguistiques. Les études sur les effets de transfert de compétences ont été reportées de manière exhaustive dans le papier de Forgeard et collaborateurs (2008). Les auteurs reportent certains tests neuropsychologiques standardisés reportés pour être sensibles aux effets bénéfiques de la pratique musicale. Ainsi, la musique aurait des effets positifs sur les capacités verbales (conscience phonologique: Anvari et al., 2002; lecture: Butzlaff, 2000; Overy et al., 2003), les capacités mathématiques (Vaughn, 2000), le Quotient Intellectuel (Schellenberg, 2004, 2006). Cependant, alors que certaines études suggèrent un effet bénéfique de la pratique musicale, d'autres n'ont pas montré un tel effet (capacités spatiales notamment). Cependant, il n'en demeure pas moins qu'une grande partie de ces études se base sur une approche cross-sectionnelle (cross-sectional approach). Cette approche va consister à comparer des non musiciens à des musiciens déjà experts et appariés sur certaines variables comme la latéralité, le genre, l'âge, le niveau d'étude. Ces études sont très informatives mais restent fortement critiquées pour plusieurs raisons. D'une part, les conclusions tirées par les auteurs révèlent au mieux des corrélations positives entre le nombre d'années de pratiques de la musique ou même l'âge de début de pratique et les performances dans ces tâches. Cependant, il me semble que le caractère causal des hypothèses avancées est relativement ténu dans la mesure où un lien de corrélation entre deux variables n'implique pas forcément un lien de causalité. Pour ce qui est de la

musique, Glenn Schellenberg a fortement contribué au développement de l'approche longitudinale, qui, pour lui est la seule méthode valable pour clarifier le caractère causal des effets de la pratique musicale. Ses principaux sujets d'investigation sont celui du lien entre pratique musicale et intelligence générale ainsi que les effets spécifiques et généraux que peut avoir la pratique musicale. Pour Schellenberg, les études cross sectionnelles ne permettent pas de tirer des conclusions claires principalement de part la présence de différences non contrôlées entre le groupe de contrôle et le groupe d'experts. Pour lui, des différences telles que l'humeur, l'état d'éveil, la motivation, le nombre d'années d'études, d'activités extrascolaires, le revenus financiers des parents, le niveau d'éducation des parents, la pédagogie, la relation sociale vis à vis du maître de musique et même des différences d'ordre génétique, toutes ces différences pourraient être à l'origine voir expliquer de meilleures performances des musiciens par rapport aux non musiciens. Ainsi, la seule façon scientifiquement valide pour révéler une influence positive et directe de la pratique musicale sur des compétences musicales et/ou non musicales est donc de constituer deux groupes équilibrés et non différents, sur le plus de mesures possibles, avant le début de tout apprentissage. Au sein de chaque groupe, les participants vont suivre un apprentissage spécifique, défini par l'expérimentateur et équilibré du point de vue des stimulations cognitives proposées. Le point décisif dans la constitution des groupes est que la répartition des participants ne se fait pas de manière aléatoire mais plutôt de manière pseudo-aléatoire. C'est à dire qu'avant le début des apprentissages, les résultats de l'ensemble des tests effectués vont servir de base pour constituer les deux groupes d'apprentissage. Cela se réalise en montrant que les groupes ne sont pas statistiquement différents l'un de l'autre. De nombreux auteurs ont désormais souligné ces problèmes et tendent désormais à plébisciter l'approche longitudinale. Je propose dans la partie suivant une brève revue de la littérature à ce sujet.

4.2 Effets de l'apprentissage de la musique, approches

longitudinales.

Plusieurs études longitudinales visant à mettre en évidence les bénéfices que peut avoir la pratique musicale sur des capacités directement en relation avec la musique telle que les capacités motrices ou perceptives ont été récemment menées. Lappe et collaborateurs (2008) ont testé chez 32 adultes non musiciens l'effet différentiel de 2 semaines d'apprentissage unimodal ou multimodal. Pour ce faire, ils ont enregistré la MMNm (homologue de la MMN enregistrée en MEG) en réponse à une note déviante dans une séquence d'arpèges. Cette étude révélait que l'apprentissage multimodal (sensorimoteur) entraînait une plus grande augmentation de l'amplitude de la MMN que l'apprentissage unimodal, différence interprétée comme étant le reflet d'une plus grande plasticité induite par l'apprentissage multimodal (Lappe et al., 2008). Cependant, aucune mesure neuropsychologique n'était recueillie, les groupes étaient construits de manière aléatoire et la constitution des groupes ne semblait pas prendre en compte de nombreuses mesures neuropsychologiques. Bref, plusieurs critères décisifs semblent ne pas être remplis dans l'étude de Lappe et collaborateurs diminuant ainsi la valeur scientifique de cette étude. Le groupe de Schlaug a quant à lui lancé depuis quelques années toute une série d'études et s'efforce d'être le plus rigoureux dans le déroulement des protocoles ainsi que dans la constitution des groupes. Ainsi, plusieurs études longitudinales et cross sectionnelles sont parues (Forgeard et al., 2008) et vont sûrement paraître d'ici peu de temps. Dans l'étude de Hyde et collaborateurs (2009), les auteurs ont testé l'effet différentiel de 15 mois d'apprentissage collectif de la musique (groupe de contrôle) face à l'apprentissage du piano en cours personnel chez 32 enfants. Pour cela, ils ont effectué des mesures morphologiques de certaines régions du cerveau, avant et après les 15 mois d'apprentissage. Le point positif de cette étude est que les auteurs avaient pris la précaution de répartir les enfants de manière pseudo aléatoire sur la base des résultats de plusieurs tests neuropsychologiques. Alors qu'avant l'apprentissage, les deux groupes d'enfants ne présentaient pas de différences de morphologie, après 15 mois d'apprentissage du piano, les enfants du groupe musique présentaient

une augmentation de la taille du gyrus précentral droit, du planum tempore, du cortex auditif primaire droit relativement au groupe de contrôle. De plus, alors que l'évolution de la taille du gyrus précentral droit ainsi que du planum tempore était corrélée positivement avec l'évolution des performances dans un test de motricité de la main gauche, l'évolution de la taille du cortex auditif primaire droit était quant à elle positivement corrélée avec l'évolution des performances dans une batterie de tests de discrimination de rythme et de mélodies. Cependant, ils n'observaient aucune différence entre groupe musique et de contrôle au niveau des tests neuropsychologiques (Hyde et al., 2009).

La même année, une des premières études électrophysiologique rigoureuse à montrer un effet bénéfique de la pratique musicale sur le traitement d'incongruités prosodiques à la fin d'une phrase ainsi que sur les performances de lectures de mots irréguliers été publiée par le groupe de Mireille Besson (Moreno et al., 2009). Dans cette étude, 32 enfants Portugais étaient testés sur une batterie de tests comportementaux, issus du WISC III (QI verbal et QI de performance, digit span, mémoire de travail et capacités de lecture) ainsi que sur 2 tâches de détection d'incongruités de hauteur, une tâche dans le langage et une tâche dans la musique, ceci avant et après 9 mois d'apprentissage de la musique ou de la peinture. Au niveau des tests neuropsychologiques, seul le « groupe musique » améliorait significativement ses performances de lectures de mots irréguliers. En revanche, les autres mesures (digit span, QI) ne se révélaient pas modulées différenciellement en fonction du type d'apprentissage. Au niveau des tests de discrimination d'incongruités de hauteur dans la musique, les deux groupes d'enfants amélioraient leurs performances de manière quasi équivalente. Cependant, dans la tâche linguistique, les enfants du « groupe musique » amélioraient leurs performances dans la détection de faibles incongruités de hauteurs (35% d'augmentation de hauteur) alors que les enfants du « groupe peinture » ne présentaient pas une telle différence. De manière intéressante, ces changements comportementaux trouvaient également appuis dans les données électrophysiologiques avec une augmentation de l'amplitude de la composante N300 en réponse aux faibles incongruités uniquement pour les enfants du « groupe musique ». Cette étude avance un

premier faisceau d'arguments confortant l'hypothèse que les capacités musicales peuvent profiter également aux capacités linguistiques.

Cette brève revue de la littérature pointe le doigt sur la présence de plusieurs points décisifs relatifs à la procédure expérimentale, au type d'apprentissage, à sa durée, son intensité ainsi qu'à son contenu. Il semble également que les améliorations d'intelligence générale prônées par Schellenberg ne semblent pas aussi évidentes qu'il le prétend, même lorsque les expériences semblent irréprochables. Dans la partie qui suit je présente la procédure que nous avons utilisée pour réaliser notre projet d'étude longitudinale. Je commencerai par décrire brièvement la chronologie du projet puis reporterai les caractéristiques principales des apprentissages que les enfants ont suivis.

4.3 Le projet MUSAPDYS (Musique Apprentissage et Dyslexie) :

Avant propos.

Avant toute chose, il m'apparaît plus qu'important de resituer le projet MUSAPDYS (pour Musique Apprentissage et Dyslexie) dans son contexte aussi bien humain, logistique que scientifique. Il est ainsi important de savoir et de garder en tête que ce projet est toujours en cours à l'heure actuelle et qu'il a débuté il y a de ça presque 5 ans lorsque, sur la base de ses nombreux travaux dans le domaine de l'expertise musicale ainsi que sur la base des résultats de l'étude de Moreno et collaborateurs (2009), Mireille Besson demandait un important financement à l'Agence Nationale de Recherche visant à développer tout un programme de recherche sur les effets de l'apprentissage de la musique et ses éventuelles implications pour la remédiation à la Dyslexie. Julie Chobert (ma jumelle de thèse) et moi, finissions notre master avec l'espoir de continuer à mener nos travaux respectifs et en développer en commun. Le financement de l'ANR arriva, et Mireille Besson nous faisait suffisamment confiance pour nous faire participer à son projet! Après 2 mois de repos, le projet commençait avec Mireille Besson à la direction assistée de Michel Habib, Jean-Luc Velay, Daniele Schön, Julie et moi.

Les réflexions sur les protocoles expérimentaux, la passation de sujets pilotes, le

calibrage des stimuli étaient notre centre d'intérêt principal. Après très peu de temps, les difficultés, nombreuses et variées s'enchaînaient sans affecter ni notre humeur ni notre motivation, Mireille Besson se débattait pendant les réunions d'information pour les parents, les professeurs, les directeurs. Ces réunions s'accumulaient et il semblait difficile de trouver les écoles et les tests adéquats. Au bout de moins d'un an et après avoir lutté pour arriver à boucler tous les impératifs scientifiques et surtout administratifs, nous disposions d'une batterie de tests neuropsychologiques, d'une batterie de tests d'écriture sur palette graphique et de toute une batterie de tests électrophysiologiques visant à tester différents niveaux de traitement dans le langage. Nous avons surtout réussi à impliquer non pas une mais 2 écoles primaires, une à Aix et une à Marseille. Au total, 7 classes, 10 professeurs et 70 élèves ont été impliqués dans ce projet. Autant dire que ce projet était complexe. Concernant les enseignements artistiques, nous avons contacté Pierre-Yves Lawrence, futur professeur fraîchement diplômé du CFMI (Centre de Formation des Musiciens Intervenants) afin de pourvoir les apprentissages musicaux. Ariane Ponzio assurera le côté peinture. Le 11 septembre 2008, la passation des tests commençait pour finir le 23 octobre 2008. Nous ne disposions que de très peu de temps pour confectionner les groupes, il fallait donc analyser l'ensemble de ces données au plus vite pour communiquer les listes d'enfants aux écoles et professeurs afin que les apprentissages commencent au plus tôt.

4.4 Caractéristiques des apprentissages.

Tout d'abord, il est important d'avoir à l'esprit que les enseignements artistiques délivrés collectivement par Pierre Yves et Ariane n'étaient clairement pas de type classique et formels comme ils pourraient être donnés en conservatoire de musique ou aux beaux-arts pour la peinture. Les enseignements musicaux s'inspiraient des méthodes Kodaly et Orff. La méthode Kodaly porte le nom d'un compositeur, ethnomusicologue, philosophe Hongrois, Zoltan Kodaly (1882 – 1967) qui s'intéressa également à la pédagogie de la musique. Les disciples de Kodaly mirent en action certains principes qu'amena Zoltan Kodaly le long de sa vie. Cette méthode met ainsi l'accent sur

une pratique précoce du chant en parallèle avec la danse, elle insiste sur la place des chants folkloriques traditionnels pour commencer l'apprentissage en vue d'aboutir seulement un peu plus tard à la pratique instrumentale. Les enfants sont supposés apprendre la musique avec plaisir et non de manière forcée avec un intérêt pour la solmisation (chanter le nom des notes) en vue d'acquérir plus facilement les fonctions tonales des notes les unes par rapport aux autres. La nécessité de bien former les enseignants est donc également un point principal de cette méthode (<http://www.kodaly-inst.hu/>). La méthode développée par Carl Orff (1895 – 1982), compositeur et pédagogue est quant à elle principalement basée sur deux aspects que sont l'improvisation et le mouvement. L'enseignement est axé sur le rythme avec la pratique d'instruments percussifs inventés par Orff (glockenspiels, xylophones, métallophones, tambours maracas, tambourins, triangles...) et sur la mélodie avec notamment le principe du bourdon (<http://www.orff.de/>).

De manière plus scientifique, ces enseignements se focalisaient sur le rythme, la hauteur, le timbre par la pratique du chant et d'instruments en groupe. Ainsi, les enfants étaient amenés à manipuler des instruments de percussions pour la plupart, et ce, dans l'accompagnement de chansons ou lors de jeux rythmiques. A chaque fois, la priorité était donnée à l'écoute de la production des élèves avant de jouer avec les autres. Ainsi, les principales activités consistaient à : jouer avec leur voix, parcourir leur étendue dans la limite de leur tessiture, distinguer un petit intervalle d'un grand et savoir les chanter juste, distinguer les différents registres (grave, medium, aigüe), distinguer les différentes durées des sons et jouer avec, reconnaître les timbres de différents instruments, tenir leur place dans une polyrythmie, respecter la pulsation, répondre à des appels musicaux en entrant ou sortant du jeu en groupe, mémoriser et interpréter des chants, improviser en gardant des consignes, affiner l'écoute.

Les enseignements de peinture étaient quant à eux principalement basés sur la méthode développée par Arno Stern né en 1924, artiste peintre et directeur de l'Institut de Recherche de la Sémiologie de l'Expression (<http://www.arnostern.com/>). Son approche se base sur l'idée selon laquelle l'acte de Formulation (l'acte de laisser une trace sur un support) est un acte universel, structuré complexe et

original qui nécessite des conditions propices pour émerger. Il créa ainsi le « Jeu de Peindre », sa méthode d'apprentissage d'expression artistique. Les conditions propices pour que la Formulation émerge se retrouvent ainsi dans le clos-lieu, espace fermé dans le quel les élèves dessinent et peignent en groupe sur des supports de grand à très grand formats aidés d'un « praticien » qui a plus un rôle d'auxiliaire que de professeur. Le « Jeu de Peindre » dispose de ses propres règles: les peintures ne sortent pas du clos-lieu et ne sont pas destinées à être vues ou commentées par d'autres personnes, les élèves peignent ce qu'ils veulent et décident eux même si leur dessin est terminé. Ainsi, cette méthode permet de développer les capacités visuo-spatiales et motrices en travaillant sur le mélange des couleurs, les différentes textures.

Chapitre 5. Apprentissage des structures linguistiques et musicales.

Dans le chapitre 1, je décrivais l'organisation structurelle du langage et de la musique et montrais que ces deux systèmes sont régis par des règles d'organisation hiérarchique dont certaines sont communes. Si l'auditeur arrive à construire des attentes structurelles sur la base de ces règles, on peut se poser la question de l'existence d'un processus d'apprentissage commun aux deux systèmes. En effet, les questions sur le type d'apprentissage et les mécanismes sur lesquels il repose sont d'un intérêt grandissant dans les deux domaines d'étude. Dans le cas de la musique, nous avons vu que la perception de la tonalité et des phrases musicales ne nécessitent pas forcément de connaissances formelles de la musique (Bigand, 2004). Les participants non experts se comportent sensiblement de la même manière que des musiciens professionnels dans des tâches mnésiques et perceptives. Cela semblerait suggérer qu'il existe un mécanisme d'apprentissage suffisamment efficace pour ne pas nécessiter de supervision, apprentissage qui se ferait par simple exposition aux structures (Bigand, 2004). Du côté du langage, les enfants arrivent à apprendre les premiers mots de leur langue maternelle avec une facilité déconcertante sans pour autant avoir connaissance des règles syntaxiques et grammaticales qu'ils acquièrent plus tard à l'école (Mc Mullen & Saffran, 2004). Ce type d'apprentissage est classiquement appelé apprentissage implicite. Du point de vue cognitif, ce mécanisme reposerait sur la capacité à extraire des régularités de structure présentes dans l'environnement sans avoir quelconque intention d'apprendre (Cleeremans et al., 1998, Perruchet, 2008). Ce mécanisme permettrait d'apprendre des structures complexes sans qu'une verbalisation de ce qui a été appris ne soit possible. Cependant, est ce que ce mécanisme d'apprentissage est de même type que celui observé pour les structures musicales? De même, dans le cadre de l'hypothèse de partage de ressources neuronales du traitement des structures musicales et linguistiques (Patel, 2003), on peut se poser la question de savoir si cet apprentissage est susceptible d'impliquer les mêmes ressources neuronales, de la même manière et avec un décours temporel similaire? Dans la suite du chapitre, je présente une revue de la littérature visant à cerner le cadre théorique de l'apprentissage implicite des structures linguistiques et musicales.

5.1 Apprentissage implicite de grammaires artificielles.

5.1.1 Apprentissage de grammaires artificielles dans le domaine linguistique.

L'apprentissage implicite de structures linguistiques a commencé à être étudié en laboratoire avec des systèmes artificiels (des séquences de consonnes) de type « grammaires à état fini » par Reber en 1967. Le matériel de stimulation utilisé a la caractéristique d'être très réduit (5 lettres formant 20 séquences). Il a donc l'avantage d'être moins complexe que le langage réel mais très informatif pour étudier précisément les paramètres décisifs influençant l'intégration de règles structurales dans le langage. Le protocole expérimental utilisé comprend classiquement deux phases successives. Lors de la première phase, le matériel structuré (dans l'étude de Reber il s'agissait de séquences de consonnes affichées sur un écran) est présenté aux participants sans qu'ils ne soient avertis de l'existence des règles. Après cette phase de familiarisation, la présence des règles est révélée et le résultat de l'apprentissage est testé. Pour cela, les participants doivent établir un jugement de grammaticalité sur de nouvelles séquences. La moitié des séquences de test suivent les règles alors que l'autre moitié ne suit pas ces règles. Le résultat principal de l'étude princeps de Reber en 1967 est que les participants présentaient des performances plus grandes que le niveau de la chance (69% de réponses correctes), suggérant qu'ils avaient appris les règles structurelles de manière implicite.

5.1.2 Apprentissage de grammaire artificielle dans le domaine musical

Si dans le domaine linguistique les protocoles de grammaires artificielles ont été largement utilisés, ce n'est que plus récemment que ces protocoles ont été transférés au domaine musical. Ainsi, l'apprentissage implicite de grammaire artificielle a été montré avec des séquences de sinusoïdes, des séquences de timbres d'instruments, des notes de clarinette (Altmann et al., 1995, Bigand et al., 1998, Rohrmeier et al., 2010, Loui et al., 2008, 2010). Deux études semblent particulièrement intéressantes à détailler pour la qualité de leurs approches méthodologiques, novatrices, ainsi que leurs résultats. Dienes et Longuet-Higgins (2004) ont étudié l'apprentissage de structures musicales dodécaphoniques et l'effet de l'expertise musicale. Historiquement, la musique dodécaphonique a été créée par le compositeur Arnold Schoenberg au début des années vingt. Elle a la caractéristique d'être très différente de la musique occidentale tonale: elle n'est pas régie par la hiérarchie tonale mais plutôt par des règles d'enchaînement de notes, règles basées sur des motifs de transposition bien définis (rétrograde, inverse, inverse rétrograde, transposé). De part la forte complexité et l'originalité de sa grammaire, certains auteurs semblent douter que la musique sérielle soit apprenable voir accessible. Dans l'étude en question, des participants (des non experts, des étudiants en musicologie et un expert en musique dodécaphonique) étaient exposés à des séquences dodécaphoniques puis testés avec des nouvelles séquences respectant les différentes formes de transpositions. Alors que les non experts étaient incapables d'effectuer la tâche, les étudiants en musicologie arrivaient à reconnaître certaines transpositions et surtout, l'expert arrivait à reconnaître explicitement quasiment toutes les transpositions. Cependant, Rohrmeier et al. (2010), montraient récemment avec un protocole de grammaire à état fini que des musiciens n'étaient pas meilleurs que des non musiciens. Même si l'étude Rohrmeier et al., vient nuancer les interprétations de Dienes et Longuet-Higgins quant à l'effet de l'expertise musicale, le protocole utilisé par ces derniers est intéressant en comparaison avec ceux qui utilisent des jugements de grammaticalité. En effet, la prise en compte d'informations qualitatives subjectives, comme par exemple le niveau de confiance, semble permettre de mieux caractériser le contenu de conscience des participants donc de mieux

cerner les connaissances implicites des participants.

Dans la même veine, Tillmann et Poulin-Charonnat (2009) étudiaient l'apprentissage implicite de séquences de notes régies par une grammaire artificielle à l'aide d'un protocole expérimental original. Dans une première expérience, des participants devaient, pendant l'exposition à trois séquences de plusieurs notes, faire une tâche de mémoire (i.e. Dire à haute voix s'ils avaient déjà entendu la séquence présentée). Par la suite, ils étaient testés à l'aide d'un paradigme de priming consistant à détecter si une note de la séquence de test, différente à chaque essai, est dans le ton ou pas. Le groupe de contrôle ne faisait que la tâche de priming. Dans une seconde expérience, les participants devaient, pendant la phase d'exposition même, effectuer la tâche de détection puis étaient testés avec la tâche de priming. Dans les deux expériences les participants exposés aux séquences faisaient moins d'erreurs et répondaient plus vite pour les séquences qui respectaient les règles de la grammaire que pour celles qui ne la respectaient pas, alors que le groupe de contrôle ne présentait pas une telle facilitation. Cette étude est intéressante au moins pour deux raisons. D'une part, les connaissances implicites des participants semblent aller bien au delà de la structure globale des séquences, et vont jusqu'au développement d'attentes d'évènements uniques. D'autre part, encore une fois, le protocole expérimental novateur permet de mieux caractériser les connaissances implicites des participants. Ce dernier point est sujet à de nombreuses recherches et de nombreux auteurs ont tenté de créer de nouveaux paradigmes afin de mieux étudier, ou tout du moins d'apporter de nouvelles informations sur le fonctionnement des mécanismes d'apprentissage implicites (voir Cleermans & Jimenez, 2002).

5.2 Apprentissage implicite des structures linguistiques et musicales.

Par opposition au langage écrit, le signal de parole ne contient pas de pauses systématiques entre chaque mot. Certes, il existe des courtes périodes de temps pendant lesquelles le signal acoustique est « absent ». Cependant, d'une part, tous les mots ne commencent pas par une plosive et d'autre part, une segmentation qui aurait lieu à chaque plosive mènerait à énormément d'erreurs. Pourtant, l'apprentissage d'une langue nécessite d'arriver à procéder à l'extraction des mots du flux de paroles. Si l'auditeur ne peut pas se baser sur les indices acoustiques de manière fiable pour arriver à extraire du flux continu de paroles les unités signifiantes que sont les mots, c'est qu'il doit exister au sein du signal de parole un indice sur lequel se baser. Depuis une étude ayant suggéré l'importance de l'organisation structurelle des phonèmes dans le signal de parole (Hayes & Clark 1970), tout un pan de recherche s'est développé. L'étude la plus marquante dans ce domaine a été menée par Saffran et collaborateurs et publiée en 1996 dans la revue *Science*. Dans cette étude, les auteurs ont construit un langage synthétique artificiel (j'entends par synthétique artificiel, le fait que le langage est synthétisé par un ordinateur à l'aide d'un logiciel de synthèse vocale et artificiel dans le sens qu'il n'a pas réelle existence hormis dans le cadre expérimental en laboratoire) de telle façon que les seuls indices sur lesquels les auditeurs pouvaient se baser pour segmenter le flux continu de syllabes était les faibles probabilités transitionnelles (ou probabilités conditionnelles, PT) entre paires de syllabes présentes aux frontières des mots. En effet, le matériel était fait de telle façon que le signal de parole ne contienne pas d'indices prosodiques. De plus, les PT étaient élevées à l'intérieur des mots (PT grande pour les paires de syllabes qui sont à l'intérieur des mots) et basses aux frontières des mots (PT petites pour les paires de syllabes qui sont à cheval entre deux mots). Définissons ce que l'on entend par PT. D'un point de vue probabiliste, l'occurrence de deux syllabes successives X et Y dans le signal de parole peut se traduire par une équation permettant de quantifier la probabilité que la première syllabe X soit suivie de la deuxième syllabe Y.

Cette quantité se calcule grâce à la formule suivante:

$$PT = \frac{\text{Fréquence d'occurrence de la paire XY}}{\text{Fréquence d'occurrence de l'événement X}}$$

Ainsi dans la séquence JO LI FI LLE, la probabilité que la syllabe jo soit suivie de la syllabe li est forte (on pourrait également considérer cette PT élevée comme une forte familiarité car cette séquence, à savoir le mot « joli », sera également utilisée dans de multiples contextes et pourra être suivie de nombreux autres noms) et surtout plus forte que la probabilité que li soit suivie de fi.

5.3 L'apprentissage implicite, spécifique ou général ?

Au niveau auditif, l'apprentissage statistique a été révélé fonctionnel pour des séquences de notes (Saffran et al., 1999), des séquences de timbres d'instruments (Tillmann et al., 2004) suggérant que ce mécanisme n'est pas spécifique au domaine linguistique mais plutôt qu'il est « domain general ». Il a également été montré des possibilités d'apprentissage statistique pour des stimuli visuels (Fiser and Aslin, 2005), pour des séquences de mouvements ou de formes (Fiser et al., 2002). Il a été également montré dans la modalité somatosensorielle avec des stimulations tactiles (Conway and Christiansen, 2005). L'ensemble de ces résultats suggère que ce type d'apprentissage n'est donc pas spécifique à la modalité auditive mais plutôt qu'il est indépendant de la modalité (« modality independant »). Cet apprentissage a été révélé aussi bien chez l'adulte, que chez le jeune enfant, que chez le nourrisson (Saffran et al., 1996, 1996a; Aslin et al., 1998; Kuhl 2004; Gervain et al., 2008; Teinonen et al., 2009). Il semble donc qu'il soit indépendant du stade de développement. Finalement et de manière très intéressante, cet apprentissage est également retrouvé avec des stimuli linguistiques chez le singe tamarin et même le rat (avec néanmoins quelques modifications du protocole expérimental, Saffran et al., 2008; Toro et al., 2005). Ces dernières études suggèrent ainsi que l'apprentissage implicite n'est donc pas non plus spécifique à l'espèce

humaine. Si cet apprentissage est présent chez le primate non-humain, il pourrait certainement être possible d'étudier les corrélats électrophysiologiques de l'apprentissage implicite à l'échelle du neurone ou du réseau de neurones.

5.4 Bases Neuronales de l'apprentissage implicite de structures linguistiques.

L'ensemble des études que je viens de décrire ne se basaient que sur des indices comportementaux. Cependant, les méthodes de neuroimagerie permettent de mieux caractériser les régions cérébrales impliquées dans l'apprentissage implicite du langage. Ainsi, quelques études se sont intéressées à mettre en évidence des variations de marqueurs neurophysiologiques lors des phases de familiarisation, en utilisant la méthode des PEs, de l'IRMf ou de l'imagerie optique. Dans la partie qui suit, je décrirai les principales études en EEG et IRMf.

La première étude s'intéressant aux corrélats électrophysiologique de la segmentation d'un langage artificiel à été celle de Sanders et collaborateurs (2002). Pour se faire, les auteurs ont analysé la différence d'amplitude de la N100 évoquée par la première syllabe des mots trisyllabiques à la N100 évoquée par la syllabe médiale ainsi qu'à la N100 évoquée par la dernière syllabe lors de l'écoute du langage. Le protocole qu'ils ont utilisé était relativement original dans la mesure où les participants écoutaient plusieurs fois le langage et qu'ils suivaient également une phase d'entraînement consistant en la présentation visuelle des mots à apprendre. Ainsi, Sanders et collaborateurs montraient que cette différence (entre N100 évoquée par les syllabes initiales et N100 évoquée par les syllabes médiales) augmentait après apprentissage statistique de stimuli linguistiques uniquement chez des « bons apprenants ». Ils montraient également qu'indépendamment du niveau de performance (i.e. Pour les bons et pour les mauvais apprenants) une N400 pariétale était observée après apprentissage. Cependant, le protocole utilisé ne permettait

pas, à mon avis, de faire la différence entre un processus purement attentionnel (retrouvé au niveau de la N100) d'un éventuel processus de segmentation. En effet, la procédure d'entraînement utilisée entre les deux phases d'écoute aurait pu augmenter dramatiquement le niveau d'attention, favoriser la mise en place de stratégies, permettre d'explicitier le caractère trisyllabiques des mots menant à une augmentation de l'amplitude de la N100 évoquée par les syllabes initiales et une N400 pariétale. Suite à cette première étude ces mêmes auteurs se sont intéressés à étudier les corrélats électrophysiologiques de la segmentation du langage continu naturel (Sanders et Neville, 2003a, 2003b). Dans la première étude, chez des participants de langue maternelle Anglaise, les PEs par les syllabes initiales de mots naturels trisyllabiques étaient comparés aux PEs par les syllabes médiales. La N100 aux syllabes initiales était plus ample que la N100 aux syllabes médiales. Les auteurs interprétaient cette différence d'amplitude comme étant un index de segmentation du langage ou même comme indiquant un processus cognitif émergent (processus attentionnel) de la prédiction des débuts des mots. Dans leur seconde étude (Sanders & Neville, 2003b), ils montraient que des participants de langue maternelle Japonaise ayant appris tardivement l'Anglais ne présentaient pas cet effet typique de début de mot (i.e. Pas de N100 plus ample pour les syllabes initiales que pour les médiales), confortant leur interprétation de modulation de la N100 comme index de segmentation.

Abla et collaborateurs (2008) ont quant à eux étudié les corrélats électrophysiologiques sous tendant l'apprentissage statistique de séquences de sons au cours de 3 sessions successives d'apprentissage. Ils montraient des patrons différents le long des sessions d'apprentissage en fonction des performances, mesurées lors du test a posteriori, chez trois groupes. Le groupe de « bons apprenants » présentait un « word onset effect » maximal ainsi qu'une N400 fronto-centrale maximale dès la première session d'apprentissage. Le « groupe d'apprenants moyens » présentait de tels effets qui augmentaient au cours des sessions. Finalement, les « faibles apprenants » ne présentaient aucune modulation de ces effets.

Suite à ces travaux, plusieurs auteurs ont eu l'idée de comparer les potentiels évoqués par un

langage artificiel structuré à ceux évoqués par un langage artificiel déstructuré dans lequel les syllabes sont enchaînées de manière pseudo aléatoire (i.e. Sans information de PT, Cunillera et al., 2006, 2008, 2009, De Diego Balaguer et al., 2007). Le langage structuré évoquait généralement une N100 plus ample que le langage non structuré. D' autres études ont reporté des variations de la P200 et de la N400 sans montrer de différence au niveau de la N100. L'interprétation de la modulation de l'amplitude de la P200 va plus dans le sens de l'apprentissage des règles (la première syllabe d'un mot tri syllabique va prédire la dernière syllabe) ou pourrait également refléter la focalisation de l'attention grâce à l'insertion d'indices acoustiques tels que des pauses entre chaque mots ou un accent tonique (Cunillera et al., 2006, 2008, De Diego Balaguer et al., 2007).

L'étude de Cunillera et al. en 2009 est intéressante du point de vue méthodologique car elle combine deux méthodes. Dans ce papier, les auteurs ont comparé, chez les même participants, l'activité EEG et IRMf enregistrées séparément, lors de l'écoute de langages artificiels structurés à celle enregistrée lors de l'écoute de langages non structurés. Alors qu'ils n'ont pas observé de modification de l'amplitude de la N100 au cours de l'apprentissage, la N400 était plus ample pour le langage structuré que pour le langage non structuré après seulement deux minutes d'écoute. De plus, les données IRMf révélaient une plus grande activation au niveau du Gyrus Temporal Supérieur, du Gyrus Frontal Inférieur et du Cortex Moteur Primaire pour le langage structuré que pour le langage non structuré. Ces résultats confirment une étude précédente ayant également montré l'implication de ces structures dans une tâche sensiblement identique (McNealy et al., 2006). Dans cette dernière étude les auteurs effectuaient deux sessions IRMf. La première était faite pendant l'écoute de trois types de langages: un langage contenant uniquement les PT, un langage combinant les indices de PT et d'accent tonique, et finalement un dernier langage non structuré. Lorsque les auteurs comparaient la moyenne de l'activité pour les deux langages structurés par rapport au langage non structuré ($[(U\uparrow + S\uparrow) - R\uparrow]$), ils révélaient une plus grande augmentation de l'activité en fonction du temps d'écoute dans le Gyrus Temporal moyen de manière bilatérale, le Gyrus Supra Marginal gauche, le Gyrus Temporal Supérieur gauche. Dans la deuxième expérience, les participants devaient effectuer

une tâche de discrimination de mots. Lorsque les auteurs comparaient l'activité évoquée par les mots à celle évoquée par les non mots, ils révélaient une plus grande activation pour les mots dans le Gyrus frontal Inférieur Gauche et le Gyrus Frontal Moyen. Les activations pour le contraste Mots vs. Part-mot étaient similaires à celles révélées par le contraste précédent même si elles demeuraient tout de même légèrement moins fortes. De plus, l'activité dans le Gyrus Frontal moyen dans cette tâche était positivement corrélée avec les capacités de traitement auditif rapide (Rapid Auditory Processing Skills) révélés par un test comportemental, le Tallal Repetition Test. Les auteurs interprétaient l'augmentation de l'activation du Gyrus Temporal Supérieur au cours de la phase d'exposition comme reflétant le calcul en ligne des PT alors que l'activation dans le Gyrus Supra Marginal refléterait le développement des représentations phonologiques des mots du langage artificiel. Les activations révélées lors de la tâche de discrimination seraient quant à elles reliées à de multiples aspects du traitement phonologique et notamment les capacités de traitement de séquences temporelles.

De manière intéressante, ces mêmes auteurs ont récemment répliqué ce protocole chez des enfants de 10 ans (McNealy et al., 2010). Pendant la phase d'écoute, le contraste $((U\uparrow + S\uparrow) - R\uparrow)$ révélait une plus grande augmentation de l'activité en fonction du temps d'écoute dans le Gyrus Temporal Supérieur gauche et le Lobule Pariétal Inférieur gauche. Lors de la tâche de discrimination de mots, le contraste Mots vs. Non mots révélait une plus forte activation dans le Gyrus Frontal Inférieur gauche. A la différence des résultats de l'étude menée chez les adultes, le contraste Mots vs. Part mots ne révélait aucune différence d'activation et les auteurs interprétaient cette absence de différence en suggérant deux hypothèses alternatives impliquant dans un cas la notion d'interférence des part mots sur les mots et dans l'autre, la possibilité que les mots n'aient pas été suffisamment bien appris.

Concernant les analyses en temps fréquence et frequency tagging, je finis cette section par la description des deux études sur lesquelles nous nous sommes basés. Buiatti et collaborateurs (2009) ont étudié la phase d'apprentissage en utilisant l'analyse en frequency tagging. Pour cela, ils ont

comparé la puissance spectrale correspondant à la fréquence d'occurrences des syllabes, de groupes de deux syllabes ainsi qu'à la fréquence d'occurrence des mots lors de l'écoute de langage structurés ou aléatoires et contenant ou pas des pauses subliminales entre chaque mot. Les résultats révélaient un pic pour la fréquence d'occurrence des syllabes pour le langage aléatoire mais pas pour les langages structurés (avec ou sans pauses). Un pic à la fréquence d'occurrences des mots était présent uniquement pour le langage structuré contenant des pauses et finalement aucun pic n'était reporté pour le langage structuré ne contenant pas de pauses.

La deuxième étude particulièrement intéressante a été récemment menée par De Diego Balaguer et collaborateurs (2010). Dans cette dernière étude, les auteurs ont utilisé une approche Temps Fréquence essai par essai afin de suivre l'évolution des changements de puissance spectrale dans les bandes theta (4-8 Hz), alpha (8-12 Hz), beta (13-29 Hz) et gamma (30-40 Hz) en rapport avec l'apprentissage des mots et des règles dans un langage artificiel. Ils ont également étudié les cohérences de phases à large échelle (entre électrodes distantes) pour ces mêmes bandes de fréquence. L'analyse des puissances spectrales pour le groupe entier révélait une augmentation linéaire au cours de l'apprentissage pour les bandes theta et alpha. Pour la bande theta, ces modulations étaient réparties sur l'ensemble du scalp alors que les modulations pour la bande alpha apparaissaient plus largement au niveau centro-pariétal. Ces modulations linéaires étaient également accompagnées de modulations plus complexes pour les bandes beta et gamma avec des augmentations de type cubique au bout d'une minute d'écoute suivie d'une diminution. Cependant, il semble que ces points de mesures soient également les points qui présentent la plus grande variabilité en comparaison aux autres points, laissant penser que cet effet est mené par un ou deux sujets outlier. Différents patrons de modulations dans ces bandes de fréquences apparaissaient également en fonction des performances au test comportemental. Les mauvais apprenants présentaient une plus grande augmentation dans la bande gamma que les bons apprenants pendant la dernière partie de l'apprentissage au niveau frontal et pariétal gauche. Concernant les analyses de cohérence de phase entre régions distantes, elles révélaient des différences entre groupes à partir

d'une minute d'exposition. Les mauvais apprenants présentaient une augmentation de cohérence dans la bande theta entre les régions pariétales et temporales et entre les régions frontales et pariétales. Les bons apprenants présentaient une augmentation progressive de cohérence dans la bande gamma entre les régions frontales et temporales ainsi qu'entre les régions pariétales et temporales.

Deuxième Partie

Méthodologie

Chapitre 6 : Méthodologie générale de Potentiels Evoqués.

6.1 Historique.

C'est tout d'abord en 1790 que Luigi Galvani découvrit le lien entre électricité et système nerveux grâce à ses travaux de recherche menés sur la cuisse de grenouille. Il mit en évidence un phénomène longtemps dénommé « l'électricité animale » qui fut révélé par la mise en place d'un dispositif expérimental permettant d'observer la contraction des muscles de la cuisse d'une grenouille après stimulation électrique du nerf innervant ces muscles. Trente ans après, ce fut Hans Berger, considéré comme le pionnier de l'électrophysiologie humaine, qui en 1920 posa des électrodes sur le scalp d'un patient trépané. Il observa des fluctuations rythmique de l'activité électrique du cerveau à l'aide d'un galvanomètre à corde et montra notamment que le rythme « alpha » (entre 8 et 12 Hz) disparaissait lorsque le sujet était attentif et ouvrait les yeux. Enfin, en 1951, Dawson révéla une méthode d'analyse impliquant sommation et moyennage qui permettait d'isoler du signal les variations du potentiel électrique directement liées à la présentation d'un stimulus: les Potentiels Evoqués (PEs).

6.2 Principes Physiques.

Le signal électrique recueilli à la surface du scalp provient majoritairement de l'activité des neurones pyramidaux du cortex. En effet, ces neurones possèdent des dendrites organisées en macro colonnes parallèles entre elles et perpendiculaires à la surface corticale. Lorsqu'un neurone pyramidal reçoit une stimulation afférente, les canaux ioniques à la surface de la membrane cellulaire vont s'ouvrir. Cette ouverture des canaux ioniques va permettre le déplacement de particules chargées électriquement au niveau intracellulaire : les courants primaires, et également au niveau extracellulaire : les courants volumiques. Ainsi, la synchronisation et la sommation de

l'activité de plusieurs milliers de neurones pyramidaux va entraîner un courant volumique global qui sera récupéré comme une variation du potentiel enregistrée à la surface du scalp.

6.3 Matériel et enregistrement.

J'ai réalisé la totalité de mes expériences en utilisant 32 électrodes Ag₂AgCl placées à la surface du scalp selon le système international 10/20 (Jasper, 1958). Ce système a tout d'abord été créé pour obtenir une méthode standard de positionnement des électrodes au cours d'exams de routine en clinique. A partir de deux repères anatomiques que sont l'Inion et le Nasion, les électrodes sont positionnées en suivant des intervalles de 10 ou 20 %. Chaque électrode du système BIOSEMI est pré-amplifiée directement au niveau du scalp. Ceci permet de s'affranchir des problèmes d'impédances élevées d'une part et permet également de travailler correctement dans des environnements à fort bruit électrique (50 Hz) sans avoir de cage de Faraday. De plus, le système BIOSEMI est complètement transportable et tient entièrement dans une petite valise. Ceci m'a permis de réaliser les expériences avec les enfants directement au sein des écoles dans une salle de classe mise à notre disposition et d'aller proposer des séances de Travaux Pratiques d'acquisition et de traitement de données à la première école du GDRI France Maroc.

Lorsque l'on parle d'activité électrique enregistrée en surface, il s'agit toujours d'une différence de potentiel entre une électrode active et une électrode de référence (supposée ne pas enregistrer d'activité électrique corticale). Alors que le système BIOSEMI possède sa propre référence, nous procédons toujours à un re-référencage Offline. Ce dernier se fait par rapport à la moyenne d'activité de deux électrodes placées sur les apophyses mastoïdiennes derrière les deux oreilles. Ce choix est dicté par la volonté de diminuer les effets de latéralisation d'une part, ainsi que d'avoir une électrode la moins sensible possible à l'activité des sources corticales et à celle des sources d'artefacts physiologiques (battement du cœur et respiration) d'autre part (grâce à l'épaisseur de l'os mastoïdien). Ce choix étant largement répandu dans la littérature, il nous permet donc également de comparer les données acquises dans différents protocoles expérimentaux.

6.4 Traitement des données.

6.4.1 Filtrage des données.

Dans le cadre de la technique d'analyse en potentiels évoqués, les données brutes recueillies à la surface du scalp sont souvent contaminées par des artefacts. On entend par artefact, les perturbations du signal enregistré qui ne sont pas liées à la tâche et qui peuvent avoir une origine physiologique (clignements oculaire, activité des muscles masticateurs et des muscles du cou, activité cardiaque, transpiration) ou une origine externe comme notamment le bruit électrique. Au cours de ma thèse, j'ai utilisé principalement trois types de filtres. Le premier que l'on utilise généralement est un filtre « coupe-bas » qui va permettre d'enlever du signal toutes les fréquences inférieures à un seuil que l'on détermine (e.g. 0.01 ou 1 Hz). Ce type de filtre permet de corriger les dérives lentes du signal et permet dans le cas où l'on veut procéder à une Analyse en Composantes Indépendantes à préparer les données pour les analyses. Cependant il est nécessaire de bien faire attention à ce type de filtres car il peut affecter l'amplitude de certaines composantes des Potentiels évoqués. J'ai également utilisé un filtre « coupe-haut » qui permet quant à lui d'éliminer toutes les hautes fréquences qui sont supérieures au seuil déterminé. Généralement, on utilise un filtre à 30 Hz en vue d'éliminer le bruit électrique environnant (50 Hz) ainsi que l'activité musculaire (20-40 Hz). J'ai pu également utiliser un filtre « notch » qui permet de travailler plus finement au niveau fréquentiel. Ce type de filtre va permettre d'éliminer une bande de fréquence étroite et précise sans trop affecter les autres bandes de fréquences.

6.4.2 Rejet / Correction des artefacts.

6.4.2.1 Rejet automatique.

Avant de moyenner les données EEG au travers des essais pour chaque participant, il est nécessaire d'éliminer les données contenant des artefacts afin de ne garder que les essais de bonne qualité. Pour cela, les logiciels d'analyse de données EEG permettent d'éliminer les essais en fonction de

l'amplitude du signal. Ainsi on détermine un seuil d'activité en μV au dessus duquel les données sont automatiquement éliminées. Au cours de ma thèse j'ai généralement utilisé un seuil de $\pm 50 \mu\text{V}$ pour les expériences réalisées chez l'adulte et un seuil de $\pm 75 \mu\text{V}$ chez l'enfant. Le rejet automatique peut également se baser sur la puissance spectrale du signal. Par exemple, lorsqu'un participant commence à relâcher son attention, l'activité EEG se caractérise par une forte augmentation de la puissance spectrale dans la bande de fréquence « alpha » (8 – 12 Hz). De la même manière que précédemment, on détermine un seuil dans la bande alpha au dessus duquel les données seront éliminées. Ce type de pré traitement va être très utile lorsque l'on récolte des données avec un grand nombre d'essais car il permet d'aller assez vite. Cependant, je présente dans la partie suivante une méthode alternative au seul rejet des artefacts, méthode qui vise à corriger ces artefacts.

6.4.2.2 Analyses en Composantes Indépendantes

L'utilisation d'un rejet automatique des artefacts peut effectivement, dans certaines situations, ne pas être la solution appropriée pour obtenir des données exploitables. En effet dans la deuxième étude réalisée au cours de ma thèse, nous ne disposions que d'un nombre très limité d'essais (25). Si l'on rejette directement tous les essais artefactés, le rapport Signal sur Bruit (SNR) serait fortement déprécié et ne serait pas suffisant. Pour s'affranchir de ce problème, nous avons utilisé une méthode de traitement de signal un peu plus élaborée nous permettant de corriger les artefacts afin d'obtenir un SNR acceptable. Pour ce faire, j'ai utilisé l'Analyse en Composantes Indépendantes (ICA). Cette méthode permet d'effectuer la séparation de plusieurs sources spatialement indépendantes et mélangées dans un même signal recueilli à l'aide de plusieurs capteurs. Le principe est de décomposer linéairement le signal EEG continu en un produit de composantes spatialement indépendantes et chacune associée à un décours temporel. Dans la mesure où les clignements oculaires provoquent des variations de l'EEG qui ont une forme et une topographie stéréotypées, ils sont les candidats idéaux pour être séparés des sources cérébrales. Dans le logiciel, après avoir

effectué la décomposition en Composantes Indépendante, il suffit de repérer sur l'écran, la composante qui contient la forme et topographie typique des clignements oculaires et de reconstruire le signal sans la prendre en compte. En plus des artefacts de clignement oculaires, il apparaît relativement fréquent d'observer chez les participants une très forte activité des muscles masticateurs ou des muscles du front. Ces activités se retrouvent principalement au niveau des électrodes Temporales et Fronto Polaires. De la même manière, de part leur localisation spatiale fixe, ces activités peuvent être parfois corrigées à l'aide de cette méthode.

Moyennage

Après avoir obtenus des données EEG de qualité suffisante, on procède au moyennage des PE à travers les essais. Le moyennage se base sur l'hypothèse que l'activité cérébrale est synchronisée et identique lors des présentations successives du même stimulus (état stable). De plus, cette méthode présuppose que le bruit (par opposition au signal d'intérêt) est quant à lui aléatoire et non synchronisé avec la présentation du stimulus. Le résultat du moyennage est que les phénomènes stables émergent du bruit de fond qui s'annule de part sa nature aléatoire. Les potentiels évoqués étant obtenus par moyennage à travers les essais, tous les phénomènes transitoires qui ne sont pas en phase avec la présentation du stimulus sont annulés. Ainsi, les PE ne permettent de visualiser qu'une petite partie de l'activité cérébrale mais grâce aux progrès réalisés en mathématique et en traitement de signal, des méthodes récentes permettent d'étudier plus finement les activités oscillatoires qui ne sont pas tout le temps synchronisées avec la présentation des stimuli.

6.4.3 Analyses Temps-Fréquence.

Le signal enregistré à la surface contient différents types d'activités oscillatoires de plus ou moins hautes fréquences. Les analyses en temps fréquence vont permettre de quantifier ces activités en visualisant le signal sous la forme d'un graphique représentant la puissance dans une bande de fréquence choisie au cours du temps. Les graphiques obtenus sont communément appelées carte temps fréquences. Cependant, il est important de noter que l'on va pouvoir visualiser des

phénomènes oscillatoires différents selon que l'analyse temps fréquence aura été faite à chaque essai ou après le moyennage. En effet, il existe trois types d'activités oscillatoires. Les premières sont les activités évoquées. Elles sont synchronisées avec la présentation d'un stimulus externe et peuvent être mises en évidence en faisant l'analyse temps fréquence du potentiel évoqué moyenné au travers des essais. Le deuxième type d'activités oscillatoires correspond aux activités induites. Elles ont la particularité de ne pas être constamment en phase avec la présentation d'un stimulus. C'est à dire que leur latence d'apparition est variable à travers les essais. Cette caractéristique fait qu'elles s'annulent par simple moyennage alors qu'elles vont pouvoir être révélées si l'on effectue l'analyse temps fréquence essai par essai et qu'on moyenne les cartes temps fréquence. Afin de quantifier les variations de ces activités oscillatoires, nous avons effectué une décomposition temps fréquences en ondelettes. Cette méthode a notamment été reportée par Tallon-Baudry et Bertrand (1999) et consiste à convoluer le signal enregistré avec une famille d'ondelettes gaussiennes de Morlet. Cette méthode a l'avantage, comparativement à la transformée de Fourier, d'avoir une très bonne résolution aussi bien au niveau temporel qu'au niveau fréquentiel. Comme je le disais, cette méthode nous permet d'obtenir l'évolution temporelle de la quantité d'énergie pour chaque bande de fréquence étudiée présente dans le signal. Le résultat obtenu est représenté sous forme d'un diagramme temps fréquence dans lequel chaque pixel représente la puissance codée en couleur pour une fréquence précise et à un instant précis. Finalement nous avons utilisé une autre méthode d'analyse temps fréquence qui se base sur la propriété de réponse en état stable du cerveau. Je reporte le type d'analyses effectuées sur ce type d'oscillations dans la partie suivante.

6.4.4 Frequency Tagging / Auditory Steady State Response (SSR)

Lorsque l'on effectue des potentiels évoqués ou même des analyses temps fréquence de données électro physiologiques, il est nécessaire de disposer d'une bonne ligne de base. Cependant, le protocole expérimental que j'ai utilisé au cours de ma thèse impliquait une phase de stimulation répétitive et continue posant un certain nombre de problèmes méthodologiques dont celui de la

ligne de base. En effet, dans la mesure où la stimulation auditive est continue, il n'y a jamais de véritable moment pouvant faire office de ligne de base puisque le flux de syllabe est continu. Généralement on choisit les 100 ou 200 ms avant le stimulus comme étant la ligne de base, périodes pendant lesquelles les participants sont le plus souvent dans le silence (donc non stimulés). La méthode d'analyse par frequency tagging permet de s'affranchir de ce problème. Je détaillerai pourquoi cette technique semble plus appropriée que les précédentes après avoir expliqué les bases théoriques de cette méthode. Le frequency tagging se base sur la propriété que possède le système nerveux à se mettre à résonner avec des stimulations répétitives et continues. Typiquement, cette propriété a été observée lorsque l'on délivre une stimulation auditive répétée à la fréquence de 40 Hz. Le spectre de puissance de l'activité électrique de surface contient, comme celui du matériel de stimulation un pic aux alentours de 40 Hz. Ce type de réponse se retrouve pour des stimuli non linguistiques tels que des clics ou des sons complexes modulés sur l'amplitude (Picton et al., 2008). Pour ce qui me concerne directement, cette méthode permet de suivre le déroulement temporel de la réponse en état stable à la fréquence correspondant à l'occurrence d'une, de deux ou de trois syllabes. La SSR à la fréquence de trois syllabes (car les mots utilisés sont trisyllabiques) correspondrait à un des corrélats électro-physiologique du processus de segmentation (Buiatti et al., 2009). Nous avons donc ciblé une petite fenêtre autour des fréquences d'occurrence de chaque syllabe, de groupes de deux syllabes et finalement de groupes de trois syllabes. Afin d'obtenir la meilleure résolution fréquentielle possible, nous avons découpé le signal EEG en périodes de 18 mots soit 12.52 secondes car cela correspond à un multiple à la fois de 1 de 2 et de 3 syllabes. Nous avons ensuite effectué une transformée de Fourier sur cette fenêtre. Nous avons normalisé les données par rapport aux 3 voisins fréquentiels afin d'obtenir une visualisation correcte de contraste fréquentiel local.

Chapitre 7 : Questions de méthodes autour de l'apprentissage statistique.

Dans la section suivante, je commence par présenter les différents types de protocoles utilisés classiquement, je poursuis par une description des analyses comportementales réalisées la plupart du temps et termine par une section présentant les études électrophysiologiques qui me semblent importantes.

7.1 L'approche comportementale dans l'apprentissage statistique

7.1.1 Description des protocoles utilisés chez le nourrisson, le singe et le rat.

Chez le nouveau né, les études sur l'apprentissage statistique d'un langage artificiel comportent tout d'abord une phase d'écoute (phase de familiarisation). Pendant cette première phase, on présente dans la modalité auditive un flux continu de syllabes sans aucune instruction préalable. S'en suit une phase de test visant à mesurer le résultat de l'apprentissage. Cette phase de test se base sur la procédure Head Turn Preference Procedure (Colombo & Bundy, 1981). Le nourrisson, placé dans son siège-bébé est accompagné de sa mère dans une cabine insonorisée. Cette cabine comprend deux haut-parleurs cachés à côté de deux ampoules de couleur placées sur les murs de côtés. Une troisième ampoule de couleur différente est placée sur le mur en face du nourrisson. Lors d'un essai de test, l'expérimentateur caché derrière le mur central déclenche le clignotement de l'ampoule centrale. Dès que le nourrisson la regarde, l'expérimentateur l'éteint et déclenche le clignotement d'une des ampoules des murs de côté. Dès que le nourrisson tourne sa tête dans la direction de l'ampoule clignotante, l'expérimentateur déclenche la présentation du stimulus auditif via le haut parleur à côté de cette ampoule. Le stimulus est répété plusieurs fois jusqu'à ce que le nourrisson tourne sa tête dans une autre direction. Notons qu'à la

différence du protocole utilisé chez l'adulte (cf section suivante pour une description du protocole adulte), un seul type d'item (familier ou non familier) est présenté dans chaque essai de test. Le temps de regard (donc supposément d'écoute) est sommé pour chaque type d'essai (essais avec item familiers vs essais avec items non familiers) et permet d'obtenir une mesure de « préférence ». Le principal résultat lors de ce genre de protocole est que, le temps d'écoute des nourrissons est plus grand pour les items moins familiers que pour les items familiers. Ce résultat est interprété comme révélant un effet de préférence à la nouveauté, donc comme signifiant un apprentissage implicite (Saffran et al. 1996).

Chez le singe, la procédure est sensiblement la même que celle utilisée chez le nourrisson à la différence près que les participants sont dans des cages avec un seul haut parleur caché haut dessus de la cage. Les items sont présentés dès que le singe porte son regard à 180 degrés à la direction opposée au haut parleur (Saffran et al., 2008). La mesure de préférence est obtenue de la même manière que pour les nourrissons, c'est à dire en comparant les durées d'écoute sommées pour les items familiers aux durées d'écoute sommées pour les items moins familiers.

A ma connaissance, une seule étude a été menée chez le rat. Les rats sont familiarisés avec le langage artificiel alors qu'ils se trouvent dans une boîte de skinner (dispositif expérimental informatisé permettant de compter le nombre d'appuis sur un levier distribuant de la nourriture). Pendant cette période, la nourriture est délivrée à chaque appui sur le levier. Le jour suivant la familiarisation, les rats sont à nouveau placés dans la boîte de skinner. Cette fois ci, le levier est déconnecté et ne permet plus d'obtenir de nourriture. Cette phase de test permet de comparer le nombre d'appuis sur le levier pour chaque type d'item présenté de manière isolée et prédit sur une base associationniste que les mots plus fréquents devraient entraîner un taux d'appui plus grand que les part mots ou non mots moins fréquents (Toro et al. 2005).

7.1.2 Description des protocoles utilisés chez l'adulte et problèmes associés.

La plupart du temps, les protocoles d'apprentissage de langage artificiel utilisés chez l'adulte commencent, comme toujours, par une phase d'écoute pendant laquelle est présenté le flux continu de syllabes pendant plusieurs minutes selon l'étude. Suite à cette phase de familiarisation, le résultat de l'apprentissage est testé à l'aide d'une procédure en choix forcé (two alternative forced choice procedure). Lors du test, chaque essai correspond à la présentation successive de deux items, l'un familier (mot) et l'autre moins familier voir même jamais entendu (part mot, non mot ou classe mot) selon l'étude. L'ordre d'apparition des items est fait de manière aléatoire afin d'éviter tout biais d'ordre et la tâche des participants est alors de choisir lequel des deux items présentés, premier ou deuxième, ressemble le plus à ce qu'ils ont entendu lors de la phase d'écoute.

Il me semble important de noter que toutes les études menées chez le nourrisson, le singe et le rat, sont clairement implicites de par l'absence totale de consignes. Cependant, il n'en est pas de même pour les études menées chez l'adulte et l'enfant. Généralement, les participants sont au courant, avant même la phase de familiarisation, que le flux de syllabes qu'ils vont entendre contient des mots artificiels et qu'ils vont devoir en découvrir les frontières (e.g. Saffran 1996: « They were informed that their task was to figure out where the words began and ended »). Dans le meilleur des cas, les participants ne sont pas au courant que le langage contient des mots mais la consigne qui leur est donnée va clairement avoir pour effet de focaliser l'attention des participants sur le flux de paroles. Ce type de procédure est donc décidément bien moins implicite que celle utilisée avec les nouveau-nés.

Concernant la conception du test en lui même, certaines critiques sont à émettre notamment quant à la présentation par paire d'items familiers et non familiers. Cette présentation par paire peut avoir un effet d'interférence sur les représentations mentales des mots formées lors de la phase de familiarisation. En

effet, à la fin du test, les deux catégories d'items (mots: familiers et part-mots: moins familiers) auront été entendues par paire autant de fois l'une que l'autre. Or, les part-mots étant formés à partir des mêmes syllabes que les mots, leur présentation pourrait entraîner une diminution de la force des représentations des mots formées pendant l'apprentissage (interférence), tout en augmentant la force des représentations des part-mots (consolidation).

La procédure en choix forcé pose aussi la question du ou des type(s) de processus cognitifs mis en évidence grâce aux données comportementales. Dans notre cas, il me semble que cette procédure aurait tendance à révéler la capacité de segmentation certes, mais d'autre part et d'une manière difficilement dissociable, les capacités de résistance aux interférences et de rappel en mémoire des items appris. De plus, du point de vue des stratégies que les participants peuvent utiliser, il est impossible de savoir si, à chaque essai, les participants ont accepté le mot qui était dans le langage ou si bien au contraire, ils ont rejeté le part-mot. Grâce aux débriefings que j'ai fait passé systématiquement aux participants, j'ai pu constater que certains sujets tentaient consciemment et volontairement de chercher des mots qui existent réellement dans la langue française, d'autres participants procédaient par élimination et considéraient que les items qui contenaient des voyelles « y » ou « o » étaient des mots non familiers.

Aussi, il me semble qu'inclure une phase de violation avant la phase de test est une avancée méthodologique à prendre en considération dans le domaine de l'apprentissage implicite. Cette phase de violation utilisée dans l'étude de De Diego Balaguer et collaborateurs (2007), consiste à enregistrer l'activité EEG lors de la présentation du même langage structuré, mais à l'intérieur duquel (à la manière d'un Oddball) des non-mots (i.e. Jamais entendus du tout avant cette phase) sont inclus. Ce type de procédure va permettre de recueillir et d'analyser les données EEG qui sont moins contaminées par les processus de prise de décision que les réponses comportementales et va surtout permettre de s'abstraire de toute réponse comportementale, donc de jugement, et donc de biais stratégique. Ce paramètre est d'intérêt certain lorsque, par exemple, il sera envisagé d'étudier des populations pathologiques tels que

des patients parkinsonien qui ont de gros problèmes de coordination et de synchronisation motrice ou bien des patients dans le coma, dont les amplitudes de MMN et de P300 sont corrélées avec la probabilité de sortie de coma (Daltrozzo et al., 2007). Parallèlement à ce premier point, cette procédure va également permettre de baser les interprétations sur des indices électrophysiologiques connus, sujets aux prédictions et également non contaminés par tout ce qui est relatif à la prise de décision même si comme je le rapporte dans la partie sur les analyses ERPs on-line, cette approche n'est pas sans problèmes.

7.1.3 Analyses reportées dans la littérature.

7.1.3.1 Performances moyenne d'une population.

Dans une situation de choix forcé, il est possible d'analyser les performances de chaque sujet en termes de pourcentage de réponses correctes ainsi qu'en temps de réaction (TR). Les TR ne sont, à ma connaissance que rarement analysés car les consignes ne contiennent pas d'indication de vitesse. Ainsi, je ne parlerai que des analyses portant sur les pourcentages de réponses correctes des participants.

Excepté dans les études de Saffran et de Tillmann et collaborateurs, les auteurs ne comptent que le nombre de réponses correctes, à savoir le nombre d'essais au cours desquels le participant a bien choisi l'item familier. On effectue la somme des réponses correctes pour chaque participant puis on procède au moyennage à travers les participants. Finalement, on déduit d'un éventuel apprentissage lorsque la moyenne des performances de la population est significativement plus grande que le niveau de la chance (i.e. 50 % de Réponses Correctes). Or, les auteurs utilisent généralement un test de student. Rigoureusement, ce test statistique ne devrait s'effectuer que lorsque l'on compare deux échantillons de distributions gaussiennes et de variances homogènes. Ainsi, la comparaison des données observées à une série théorique composée uniquement de la valeur 50% répétée autant de fois que le nombre de sujets ne m'apparaît pas correcte au niveau théorique puisque la variance de la série de 50% est nulle.

Pour illustrer ce point, j'ai réalisé une simulation des performances obtenues chez 36 sujets virtuels dans un test contenant 36 paires d'items. Pour ce faire, j'ai considéré que chaque essai avait 50 % de chance d'être correct et j'ai donc construit une matrice de 36 colonnes par 36 lignes dont chaque cellule contenait une fonction mathématique permettant d'assigner les valeurs 1 ou 0 de manière aléatoire. J'ai reporté ma population classée par ordre croissant de performances sur la figure 1. Sur ce graphique, il est intéressant de noter deux caractéristiques. D'une part, la moyenne n'est pas de 50% et de manière décisive, la variance n'est pas nulle. En effet, on remarque clairement que certaines valeurs de performances sont assez éloignées du niveau de la chance (5 valeurs sur 36 supérieures à 60% de réponses correctes, représentées à l'extrémité droite et 2 valeurs sur 36 inférieures à 40% de réponse correctes à l'extrémité gauche du graphique).

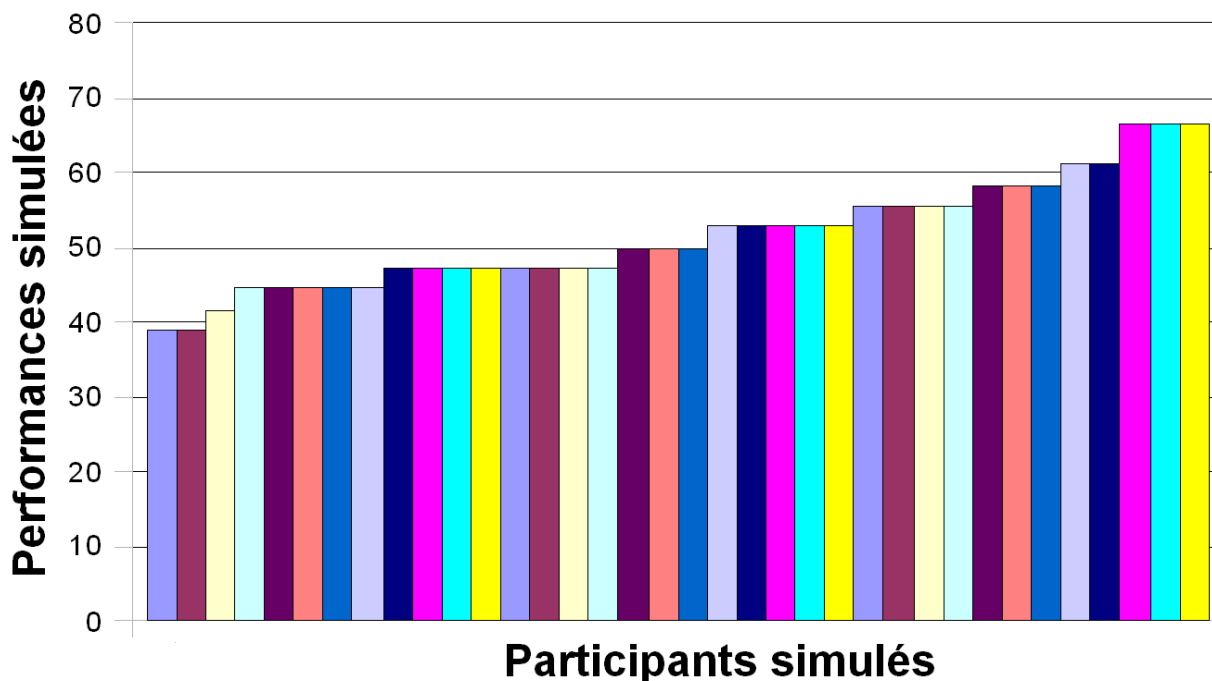


Figure 1 Simulation de la distribution des performances pour 36 sujets et 36 essais. Les participants sont classés de manière croissante de gauche à droite.

7.1.3.2 Performances moyennes des mots à Haute Probabilité vs.

Faible Probabilité.

Certaines analyses qui me semblent plus fines et informatives, vont prendre en compte le facteur Probabilité Transitionnelle (PT) en considération. Cette fois ci, la moyenne des performances de la population pour les N items à forte PT est comparée à la moyenne des performances de la population pour les N items à faible PT. Ce dernier type d'analyse va un peu plus loin que le précédent car il apporte un argument fort en faveur du rôle des PT dans la segmentation du langage. Ceci dit, il me semble que les deux analyses que je viens de décrire ne révèlent qu'une petite partie des informations disponibles. J'expose dans la partie suivante les analyses qui me paraissent complémentaires à toutes celles déjà effectuées dans la littérature.

7.1.4 Analyses effectuées au cours de la thèse.

7.1.4.1 Performances au cours du test : effet d'interférence versus effet d'apprentissage.

Nous sommes partis de l'idée que le test explicite pouvait entraîner des interférences entre les représentations des mots et celles des part-mots. Ainsi nous avons séparé le test en deux parties égales. Nous avons émis l'hypothèse de travail que l'effet d'interférence causé par la présentation explicite des part-mots au fur et à mesure du test devrait s'accompagner d'une diminution des performances entre la première et la deuxième moitié du test. Par opposition, si l'effet d'apprentissage au cours du test observé chez Rohrmeier et al., 2010 est consistant, les performances dans la deuxième moitié du test devraient être plus élevées que celles dans la première moitié.

7.1.4.2 Performances en fonction du contraste entre items de chaque paire de test.

Nous avons également eu l'idée de prendre en compte le facteur de contraste de PT entre item familier et item moins familier pour chaque essai des tests. Pour ce faire, nous avons séparé les essais présentant une forte différence entre la PT moyenne de l'item familier et la PT moyenne de l'item moins familier, des essais présentant une différence faible. Par exemple, dans un essai au cours duquel le mot gimysy (PT moyenne = 1) et le part-mot pygmi (PT moyenne = 0.32) sont entendus, le contraste de PT entre les deux items est de 0.68 ($1 - 0.32$). L'idée sous jacente est que, si les PT jouent un rôle dans la segmentation, l'essai devrait être plus facile lorsque ce contraste est élevé que lorsque ce contraste est faible. Ainsi, nous attendions de meilleures performances pour les essais à fort contraste que pour les essais à faible contraste. De plus, la sensibilité au contraste de PT pourrait être un paramètre important dans la comparaison et la caractérisation de différentes populations.

7.1.4.3 Et l'apprentissage au niveau individuel ?

La loi binomiale semble être une méthode de choix pour prouver statistiquement qu'un apprentissage a eu lieu. En effet, elle permet de comparer les données au niveau individuel (les données observées) à ce que prédit la chance (les données théoriques). Le niveau de chance peut être estimé avec un seuil d'erreur fixe en prenant en compte le nombre total d'items ainsi que le nombre total de réponses correctes. Typiquement, pour un test en choix forcé comprenant 36 essais, le niveau de la chance est classiquement considéré à 18 réponses correctes. Cependant, la loi binomiale est plus nuancée. En effet, à partir de moins de 23 réponses correctes, la loi binomiale prédit que les performances ne sont pas différentes du niveau de la chance avec 5% de chance de se tromper. Ainsi, au niveau de la comparaison de deux groupes, plutôt que de réaliser une analyse de variance comme je l'ai fait lors de cette thèse, il pourrait être envisagé d'effectuer ce test pour chaque participant, de compter le nombre de sujets dont les performances sont supérieures au niveau de la chance dans chaque groupe et finalement de comparer le nombre de sujets ayant appris dans les deux populations avec un test du Chi 2. Finalement et certainement de manière cruciale, cette méthode statistique permettrait de mieux

caractériser l'apprentissage au niveau individuel en la couplant avec l'approche électrophysiologique. Pour conclure, il me semble que les limites de l'approche comportementale résident en deux points. Le premier point est qu'elle ne permet pas d'isoler le processus d'apprentissage per se puisque l'on teste le résultat de l'apprentissage a posteriori. Il nous est ainsi impossible de dissocier les processus d'apprentissage de ceux reliés à la récupération des informations mises en mémoire ni même à la résistance aux interférences. Le deuxième point est plus dicté par une volonté de transférer ce type de protocole au domaine clinique, donc au niveau individuel. En effet, l'utilisation du test est impossible dans certaines pathologies et le caractère implicite du processus étudié se voit fortement déprécié par la simple présence des tests.

7.1.5 L'approche électrophysiologique dans l'apprentissage statistique.

Après m'être posé la question sur les problèmes liés à la nature explicite des tests comportementaux, la question sur l'évaluation de la sensibilité des approches utilisées pour mettre en évidence des phénomènes d'apprentissage m'est apparue, au fur et à mesure de ma thèse, comme une question importante. Dans une étude sur l'apprentissage d'une seconde langue, Mc Laughlin et al (2002) montraient que les performances des participants, mesurées par un d' , ne changeaient quasiment pas au cours des sessions d'apprentissage. De manière intéressante, les PE révélaient quant à eux, des modifications importantes au cours des sessions. Sur la base de ces résultats, les auteurs suggéraient que l'approche électrophysiologique serait plus sensible que l'approche comportementale pour observer et mettre en évidence les changements subtils qui sous tendent un apprentissage implicite. Ils suggéraient également que l'approche comportementale pouvait sous estimer ces changements. D'autres études ont révélé des différences électrophysiologiques sous tendant un apprentissage sans qu'elles ne soient accompagnées par une modification des performances comportementales (Tremblay et al., 1998). Dans un cadre plus clinique, une étude chez des patients amusiques (patients présentant un

déficit sélectif de la perception de la hauteur dans la musique) a révélé des effets électrophysiologiques en réponse à des changements de hauteur sans pour autant que les patients ne puissent détecter ce changement (Peretz et al., 2009). Finalement, les études que j'ai menées pourraient être également mise en correspondance avec certaines recherches dont les sujets d'étude sont la conscience, les contenus de conscience, les niveaux de consciences, la nature consciente (ou pas) des représentations acquises implicitement. Seth et collaborateurs (2008) présentent une revue intéressante dans laquelle ils affirment que les études combinant plusieurs mesures, à la fois comportementales et neurophysiologiques seront les plus informatives, et ceci en vue de répondre notamment à la question de comment plusieurs mesures peuvent être combinées pour mieux isoler les mécanismes neuronaux de la conscience.

Un autre point fort de l'approche électrophysiologique est qu'elle permet d'obtenir une information supplémentaire afin de différencier ou en tous les cas de mieux caractériser deux types de populations ou deux sous populations. Des exemples dans le domaine de l'apprentissage implicite commencent à se répandre et trouvent leurs fondations dans les études de Sanders et Ablat (2002; 2008). Dans l'étude de Sanders et collaborateurs, lorsque le groupe de participant est considéré dans sa totalité, l'amplitude de la N1 à la première syllabe ne semble pas être modulée au cours de la phase d'apprentissage. De manière intéressante, lorsque les groupes sont séparés en deux sur la base des résultats comportementaux obtenus a posteriori, l'augmentation de la N1 à la première syllabe au cours de l'apprentissage est observée uniquement chez les « bons apprenants » et pas chez les « mauvais apprenants ». Ablat et collaborateurs révélaient chez le groupe de « bons apprenants » une diminution de la N1 et de la N400 au cours de trois sessions d'apprentissage, un patron inverse pour les « apprenants moyens » (i.e. Augmentation de l'amplitude de ces mêmes composantes) et aucune variation pour les « faibles apprenants ». Ces résultats laissent penser que des patrons de variations temporelles des indices électrophysiologiques différents pourraient représenter les dynamiques temporelles des processus de segmentation (per se) et post segmentation (pré lexicalisation) différentes

entre individus. Ce dernier point est très intéressant car il offrirait la possibilité de mieux caractériser les déficits cognitifs observés chez certaines populations pathologiques, notamment par le fait qu'elle permettrait, à terme, de s'affranchir de la phase de test. En effet, dans certaines pathologies de type maladie de Parkinson, la simple sortie comportementale comme un appui sur un bouton de réponse est fortement dépréciée. Dans d'autres pathologies telles que les troubles autistiques, la communication avec le sujet est dans certains cas presque impossible posant donc la difficulté d'arriver à faire comprendre au participant la tâche à effectuer lors de la phase de tests.

Troisième Partie

Les expériences réalisées

Etude 1

Apprentissage de structures linguistiques et musicales: comparaison des mesures comportementales et électrophysiologiques.

Clément François & Daniele Schön

Publié dans *Neuroreport* en 2010 -

Résumé:

L'apprentissage d'une langue (maternelle ou seconde) peut se décomposer en plusieurs étapes dont la première nécessite d'arriver à extraire les mots du flux de paroles, c'est à dire à segmenter les mots les uns des autres. Par opposition au langage écrit, le signal de parole ne contient pas de pauses systématiques entre chaque mot suggérant que des indices, autres qu'acoustiques, sont susceptibles d'être utilisés. Ainsi, il a été montré que le signal de parole peut être segmenté de manière implicite sur la seule base de la structure statistique du langage (Kuhl et al., 1996). Les statistiques utilisées sont notamment les probabilités conditionnelles entre syllabes adjacentes, à savoir, la probabilité qu'une syllabe X soit suivie d'une syllabe Y. Ces probabilités transitionnelles sont connues pour être fortes à l'intérieur des mots et faibles aux frontières des mots. Ainsi, adultes, nouveaux nés, nourrissons et enfants arrivent à segmenter un flux continu de syllabes lorsque les seules informations disponibles sont ces probabilités transitionnelles (Saffran et al., 1996; 1996a; Teinonen et al. 2009; Gervain et al. 2008; Kuhl 2004; Aslin et al. 1998). Si ce type d'apprentissage a été tout d'abord démontré dans le domaine linguistique en modalité auditive, il a également été démontré dans le domaine non linguistique avec des séquences de timbres et de sons purs (Tilman et al., 2000; Abl et al., 2008), ainsi que dans la modalité visuelle pour des séquences de formes ou de mouvements (Fiser and Aslin, 2002). Cependant, avec l'objectif de caractériser les relations entre le traitement du langage et de la

musique, il semble difficile de tirer une conclusion claire et nette de ces travaux dans la mesure où l'apprentissage des dimensions linguistiques et musicales était étudié de manière séparée. Dans ce contexte, l'utilisation d'un langage chanté, contenant les deux types d'informations dans le même signal, semble être une solution de choix afin d'étudier une éventuelle interaction des deux dimensions chez les mêmes participants. Ainsi, Schön et collaborateurs (2008) montraient que des participants adultes apprenaient plus vite et mieux un langage chanté qu'un langage parlé. Cependant, cette étude ne s'intéressait qu'à la dimension linguistique et l'apprentissage de la dimension musicale n'était pas évalué. De plus, Mc Laughlin et collaborateurs (2004) montraient dans un étude s'intéressant à l'apprentissage d'une langue étrangère, que les potentiels évoqués peuvent être plus appropriés que les mesures comportementales pour révéler les changements fins et subtils qui sous tendent un apprentissage implicite. Finalement, il existe une littérature riche concernant l'intégration familiarité des mots utilisés dans le langage naturel

Sur la base de ces résultats, nous avons mené une étude couplant les approches comportementales et électrophysiologique et visant à comparer l'apprentissage des dimensions musicales et linguistiques au travers de l'écoute d'un langage artificiel chanté chez des participants non musiciens. Pour se faire, les participants écoutaient pendant plusieurs minutes un langage artificiel chanté et effectuaient deux tests en choix forcé, l'un pour la dimension linguistique et l'autre pour la dimension musicale. Au niveau comportemental, nous nous attendions à un meilleur apprentissage de la dimension linguistique que de la dimension musicale, reflété par des meilleures performances comportementales pour la dimension linguistique que pour la dimension musicale. Au niveau électrophysiologique, nous nous attendions à observer dans les deux dimensions une composante fronto-centrale négative aux alentours de 400 msec, sensible au degré de familiarité des items.

Alors que l'analyse des données comportementales révélait des performances significativement plus grande que le niveau de la chance pour la dimension linguistique, ce n'était pas le cas pour la dimension musicale suggérant que cette dernière n'était pas apprise. Néanmoins, les potentiels évoqués par les

items familiers et non familiers présentaient, dans les deux dimensions cette fois ci, une composante fronto-centrale négative piquant à 600 msec plus ample pour les items non familiers que pour les items familiers. D'une part ces derniers résultats confirment la plus grande sensibilité de l'approche électrophysiologique par rapport à l'approche comportementale dans l'étude des processus sous tendant un apprentissage implicite. D'autre part, ces résultats suggèrent que la dimension musicale est susceptible d'avoir tout de même été apprise même si la représentation des items musicaux ne semble pas être suffisamment robuste pour permettre une reconnaissance au niveau comportemental.

Learning of musical and linguistic structures: comparing event-related potentials and behavior

Clément Francois and Daniele Schön

To learn a new language, it is necessary for the learner to succeed in segmenting the continuous stream of sounds into significant units. Previous behavioral studies have shown that it is possible to segment a language or musical stream based only on probabilities of occurrence between adjacent syllables/tones. Here we used a sung language and tested participants' learning of both linguistic and musical structures while recording electroencephalography. Although behavioral results showed learning of the linguistic structure only, event-related potential results for both dimensions showed a negative component sensitive to the degree of familiarity of items. We discuss this component as an index of lexical search, also pointing to the greater

sensitivity of the event-related potentials compared to the behavioral responses. *NeuroReport* 21:928–932
© 2010 Wolters Kluwer Health | Lippincott Williams & Wilkins.

NeuroReport 2010, 21:928–932

Keywords: event-related potentials, language, music, N400, segmentation, song, statistical learning

Institut de Neurosciences Cognitives de la Méditerranée, CNRS - Aix-Marseille Universities, France

Correspondence to Clément Francois, CNRS-INCM, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
Tel: +33 4 91 16 41 13; fax: +33 4 91 16 44 98;
e-mail: clement.francois@incm.cnrs-mrs.fr

Received 31 May 2010 accepted 4 July 2010

Introduction

The very first steps in language acquisition require the ability to segment speech into separate words. Speech being a continuous stream of linguistic sounds, the challenge is to be able to segment it into separate (and possibly) meaningful units or words. Several studies have been conducted to better understand the cognitive mechanisms underlying language learning, and more precisely, its segmentation. These studies have shown the importance of regularities and statistical structure in word segmentation and confirm the observation that conditional probabilities between adjacent syllables, namely the probability of a syllable Y given the occurrence of a syllable X, are typically higher within words than between words [1].

The importance of conditional probabilities in language acquisition in infants and adults has been shown by Saffran and colleagues [1,2]. In these studies, the authors built an artificial language in such a way that the only cues to word segmentation were conditional probabilities between syllable pairs, which were high within words and lower across word boundaries. The design typically used in such studies comprises a passive listening session followed by a two-alternative forced choice procedure (or head turn procedure when studying infants) in which participants have to choose the most familiar of the two items. Thereafter, this type of design also proved effective in showing implicit learning in the visual modality [3] and in the auditory modality for nonlinguistic stimuli, such as tone sequences [4] or timbre sequences [5]. Recently we showed that performance in the test with spoken items is facilitated when the learning material is a sung language compared with a spoken language [6]. Overall, these

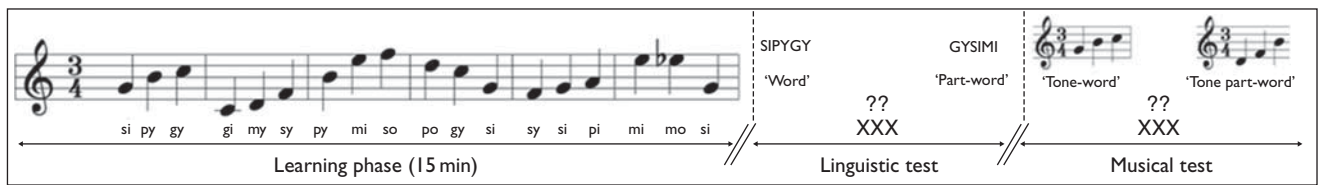
results point to the fact that a similar learning mechanism might be at work independent of the modality (auditory or visual) or stimulus type (linguistic or musical) of the input.

In this study, we used a sung artificial learning language and we collected behavioral and electrophysiological responses during two separate tests on linguistic and musical dimensions. The aim was two-fold. First, we wanted to compare the sensitivity of the behavioral and electrophysiological measures and the extent to which they correlate. Second, we wanted to compare event-related potentials (ERPs) with linguistic and musical items with ERPs described in the literature on the familiarity effect.

Concerning the first aim, we hypothesized that ERP measures might be more sensitive than behavioral measures to the subtle changes that may undergo implicit learning [7]. Indeed, while behavior reflects the combined effects of several processing stages, ERPs, thanks to their continuous time resolution, can (sometimes) show differences in the brain activity even when this is not backed by an overt behavioral response [8,9].

Concerning the second aim, few earlier studies have attempted to investigate statistical learning using electroencephalography (EEG) [10–14], reporting learning-related modifications of two well-known ERP components, the N100, related to perceptual processing, and the N400 whose amplitude is sensitive to lexical/semantic memory processes [15]. With one exception [13] the above cited studies recorded (or reported) ERPs during the learning phase only. Overall, while earlier studies investigated the

Fig. 1



The experimental design. A learning phase using a sung language was followed by two tests on linguistic and musical structures. Stimuli were presented in the auditory modality.

emergence of ‘lexical units’, we investigated the contrast between more or less stable/familiar (lexicalized) units directly after the learning phase. Earlier research showed that unfamiliar words give rise to a larger N400 than familiar words possibly reflecting the greater difficulty in accessing semantic memory for unfamiliar words [16]. In this study, as in general in artificial language learning studies, participants are familiarized to a language during passive listening. This seems to allow lexicalization of the learning output [17]; also enabling participants to choose, during the test session, the more familiar out of the two presented items. Therefore, we hypothesized that unfamiliar items would yield greater N400-like responses than familiar items for both linguistic and musical dimensions.

Methods

Participants

Fourteen native French participants (mean age 25 years, 13 right handed, 11 males, nonmusicians) were paid for their participation in the experiment (approved by the national ethical committee).

Material

The sung language was that used by [6]. Eleven syllables were combined to give rise to six trisyllabic sung words. Each syllable was associated with a distinct tone (C3, D3, F3, G3, A3, B3, C4, Db4, D4, E4, and F4); therefore, each word was always sung on the same melodic contour (gimysy C3 D3 F3, mimosi E4 Db4 G3, pymiso B3 E4 F4, pogysi D4 C3 G3, sipygy G3 B3 C4, sysipi F3 G3 A3).

Transitional probabilities ranged from 0.31 to 1.0 within words and from 0.1 to 0.2 across word boundaries. This was obtained by combining 216 repetitions of each of the six words in pseudorandom order (same word was never repeated twice in a row). The language stream was synthesized using the Mbrola speech synthesizer (<http://tcts.fpms.ac.be/synthesis/mbrola.html>).

Procedure

During the learning phase, participants were asked to listen carefully to the sound stream. During the linguistic test, participants had to press one of two buttons to indicate which of the two strings (first or second spoken

item) resembled the most to what they just heard in the stream (Fig. 1). In each of the 36 trials, while one item was a ‘word’ from the nonsense language, the other was not (hereafter, part-word). The part-words consisted of either the last syllable of a word plus the first syllable pair of another word, or the last syllable pair of a word plus the first syllable of another word (gysimi, mosigi, pygyimi, pysipi, sogimy, and syogy). All items were pronounced using a flat contour, that is, without melodic information. During the musical test, the stimuli were ‘midi’ sequences of three notes generated with a piano sound. The six sequences, called tone-melodies, were built with the same pitches defining the melodic contour of the words heard during the learning phase (C5 D5 F5, E6 Eb6 G5, B5 E6 F6, D6 B5 G5, G5 B5 C5, F5 G5 A5). The sequences, called part-tone melodies, were built using the pitches of the part-words. The instructions were the same as in the linguistic test. Stimuli were presented using loudspeakers. Both the linguistic and musical tests lasted 5 min each and their order was counter balanced across participants (Fig. 1).

Data acquisition and analysis

The EEG was recorded from 32 scalp electrodes located at standard positions (international 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, Af3, Af4, F3, F4, C3, C4, P3, P4, P7, P8, Po3, Po4, O1, O2, F7, F8, T7, T8, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). The data were then re-referenced offline to the algebraic average of the mastoids. Four participants were discarded because of major EEG artifacts. Artifacts were detected by eyeballing and using a 75 μ V maximum amplitude criterion (in approximately 10% of trials). The EEG was amplified by Biosemi amplifiers (Amsterdam University) with a band pass of 0–102.4 Hz, digitized at 512 Hz and analyzed using Brain Vision Analyzer (Brain Products, Munich, Germany).

ERPs were analyzed by computing the mean amplitude, starting at the onset of the stimulus presentation and ending after 1500 ms, using latency windows of 50 ms. Analyses were performed for correct trials only and for all trials (correct and incorrect trials). A repeated measures analysis of variance was used for statistical assessment. To test the distribution of the effects, six regions of interest

(ROIs) were selected: frontal, central, and parietal (left and right). P values were adjusted using the Greenhouse–Geisser correction, when appropriate, and Sidak tests were used in post-hoc comparisons. Owing to the increased likelihood of type I errors associated with the large number of statistical tests, only effects that reached significance ($P < 0.05$) in at least two consecutive time windows were considered significant.

Results

Behavioral data

The participants' level of performance in the linguistic test was significantly above chance (62.2% of correct responses, $P = 0.007$, Wilcoxon test). By contrast, performance in the musical test (54.5%) fell between performances with linguistic material ($P = 0.06$, Wilcoxon test) and chance level ($P = 0.2$, Wilcoxon test). We also ran nonparametric one-way analyses of variances using the items as a factor (six levels). Items analysis showed no effect of item type in the musical task ($P = 0.1$, Friedman test). By contrast, there was a significant item effect in the linguistic task: two items with low-transitional probabilities ($P = 0.5$ and 0.41) were not learned as well as the others ($P = 0.002$, Friedman test).

Electrophysiological data

Language

A negative ERP component starting 200 ms after stimulus onset and lasting until 1400 ms is visible at fronto-central sites (Fig. 2). The amplitude of this component is modulated by the degree of familiarity: part-words elicited a larger negativity between 450 and 800 ms than words. The main effect of familiarity was significant between 600 and 750 ms [$F(1,9) = 7.5$, $P = 0.03$; $2.4 \mu\text{V}$ of effect size: unfamiliar–familiar]. Interestingly, an even stronger

difference was found in the same latency band when using all trials rather than correct trials only [$F(1,9) = 9.4$, $P = 0.01$; $2.6 \mu\text{V}$ of effect size].

The correlation between the effect size (main effect of Familiarity in the 600–750 ms latency range) and performance was significant ($r = 0.77$, $P = 0.018$). Thus, the higher the participants' performance, the greater is the difference between part-words and words.

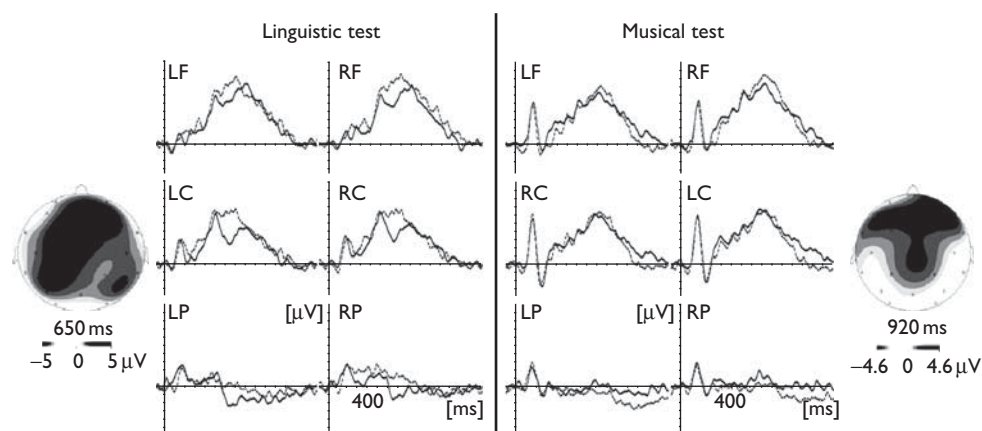
Music

A negative ERP component starting 250 ms after stimulus onset and lasting until 1300 ms is visible at fronto-central sites (Fig. 2). The amplitude of this component was modulated by the degree of familiarity, with part-tone words eliciting a larger negativity between 600 and 1000 ms than tone-words. The interaction of familiarity and ROIs was significant between 750 and 900 ms [$F(2,18) = 5.1$, $P = 0.03$]. Indeed part-tone words elicited a larger negativity over fronto-central regions than tone-words. Interestingly, an even stronger difference was found in the same latency band when using all trials rather than correct trials only [$F(2,18) = 6.3$, $P = 0.01$]. This effect size did not significantly correlate with performance. Finally, a main effect of familiarity was also significant between 1250 and 1400 ms, with part-tone words eliciting a larger positivity than tone-words [$F(1,9) = 5.7$, $P = 0.04$, $1.81 \mu\text{V}$ of effect size].

Discussion

The aim of this study was two-fold. First, we wanted to compare the sensitivity of behavioral and electrophysiological responses. Second, we wanted to compare ERPs to linguistic and musical items to ERPs described in the literature on familiarity effect. Participants were asked to

Fig. 2



Grand average across all participants for the linguistic and musical sessions. Solid lines illustrate ERPs to words/tone sequences; dashed lines illustrate ERPs to part-words/part-tone sequences. Each trace represents the mean of the three electrodes included in the ROI (LC, left central; LF, left frontal; LP, left parietal; RC, right central; RF, right frontal; RP, right parietal). Maps illustrate the topographical distribution of the peak of the familiarity effect (unfamiliar–familiar).

listen to a continuous stream of sung language, built using a pseudo-random concatenation of six sung words and were then tested on both the linguistic and musical dimensions using a two-alternative forced choice task.

Concerning the sensitivity of the methods, while the performance in the musical test was not significantly different from chance at a group level, ERP data revealed a significant familiarity effect. Interestingly, this familiarity effect was even stronger when using all trials rather than correct trials only. This seems to point to the fact that participants, on average, learned the musical structure better than they actually performed in the behavioral test. It seems that, after learning has taken place, tone and part-tone sequences do have distinct, although probably, poor quality representations. These poor quality representations might be enough to yield different ERP responses, but might be hidden by the noise surrounding the decisional and response-related processing required by the task (see Ref. [18] for a graded perspective on conscious and unconscious cognition).

Results in the linguistic test also point in the same direction. Although there was quite a robust positive correlation between participants' performance and the amplitude of the ERP effect (part-words-words) in the 600–750 ms time windows, ERPs seem to be more sensitive than behavior. In fact, all participants, but one had the expected ERP effect. Thus, even participants whose performance was barely above chance (e.g. two participants with 55% of correct responses) show the typical familiarity effect (-2.2 and -3.2 μV).

Concerning the comparison of ERPs with linguistic and musical items, electrophysiological data from both the sessions showed a larger negativity at fronto-central sites in the 400–900 ms latency band for part-words (or part-tone sequences) than for words (or tone-sequences). To interpret this ERP difference it is important to clarify the respective 'linguistic' status of the two conditions. In fact, because each word/tone-sequence has been repeated 216 times, they have 'popped out' of the continuous sung stream, thanks to statistical learning or associative mechanisms. By contrast part-words/part-tone sequences are heard five times less frequently during learning. Therefore, the linguistic status of the items used in the test is that of more or less familiar item (words and part-words, respectively).

Earlier ERP studies have already shown that, compared with high-frequency (familiar) words, low-frequency (unfamiliar) words elicit a larger negativity peaking around 400 ms [16,19]. This negative component is often referred to as N400 and its amplitude seems to be sensitive to the ease of retrieving long-term memory traces [20]. In our study, only sufficiently reinforced (repeated) items may 'survive' concurrent interference and end up being stored in a long-lasting manner [21]. The negative ERP components found in the linguistic and musical tests

might be an index for the search of the memory traces that have been stored during implicit learning [17]. It is interesting to note that the topographic distribution of this ERP component is similar in the linguistic and musical tests (only one significantly different ROI out of six). Although this ERP component peaks later in the musical test, such a difference might be because of the fact that participants needed more time in the musical test than in the linguistic test to 'search' whether an item was familiar or not.

Conclusion

In the framework of implicit language learning, the electrophysiological approach seems to be more revealing than the behavioral approach. Finally, our findings showed that superposition of two identical linguistic and musical structures in song facilitates segmentation of both linguistic and musical information.

Acknowledgements

The authors thank Benedicte Poulin-Charronnat, Deirdre Bolger, and Mireille Besson for constructive comments on a previous version. This study has been supported by the French National Research Agency (ANR 05-BLAN-0214 'Music and Memory' to D. Schön). The authors have no competing financial interests to declare.

References

- Saffran JR, Newport EL, Aslin RN. Word segmentation: the role of distributional cues. *J Mem Lang* 1996; **35**:606–621.
- Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month old infants. *Science* 1996a; **274**:1926–1928.
- Fiser J, Aslin RN. Encoding multielement scenes: statistical learning of visual feature hierarchies. *J Exp Psychol Gen* 2005; **134**:521–537.
- Saffran JR, Johnson EK, Aslin RN, Newport EL. Statistical learning of tones sequences by human infants and adults. *Cognition* 1999; **70**:27–52.
- Tillmann B, McAdams S. Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *J Exp Psychol Learn Mem Cogn* 2004; **30**:1131–1142.
- Schön D, Boyer M, Moreno S, Besson M, Peretz I, Kolinsky R. Song as an aid for language acquisition. *Cognition* 2008; **106**:975–983.
- McLaughlin J, Osterhout L, Kim A. Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat Neurosci* 2004; **7**:703–704.
- Sergent C, Baillet S, Dehaene S. Timing of the brain events underlying access to consciousness during the attentional blink. *Nat Neurosci* 2005; **8**:1391–1400.
- Tremblay K, Kraus N, McGee T. The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport* 1998; **9**:3557–3560.
- Abla D, Katahira K, Abla OK. On-line assessment of statistical learning by event-related potentials. *J Cogn Neurosci* 2008; **20**:952–964.
- Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res* 2006; **1123**:168–178.
- Cunillera T, Cámara E, Toro JM, Marco-Pallares J, Sebastián-Gallés N, Ortiz H, et al. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* 2009; **48**:541–553.
- De Diego Balaguer R, Toro JM, Rodríguez-Fornells A, Bachoud-Lévi AC. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE* 2007; **2**:e1175.
- Sanders LD, Newport EL, Neville JH. Segmenting nonsense: an event-related-potential index of perceived onsets in continuous speech. *Nat Neurosci* 2002; **5**:700–703.
- Kutas M, Hillyard SA. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 1980; **207**:203–204.

- 16 Young MP, Rugg MD. Word frequency and multiple repetition as determinants of the modulation of event-related-potentials in a semantic classification task. *Psychophysiology* 1992; **29**:664–676.
- 17 Fernandes T, Kolinsky R, Ventura P. The metamorphosis of the statistical segmentation output: lexicalization during artificial language learning. *Cognition* 2009; **112**:349–366.
- 18 Cleeremans A. Conscious and unconscious cognition: a graded, dynamic, perspective. In: Jing Q, Rosenzweig MR, d'Ydewalle G, Zhang H, Chen H-C, Zhang K, editors. *Progress in psychological science around the world. Vol 1. Neural, cognitive and developmental issues*. Hove: Psychology Press; 2006. pp. 401–418.
- 19 Van Petten C, Kutas M. Interactions between sentence context and word frequency in event-related brain potentials. *Mem Cogn* 1990; **18**:380–393.
- 20 Federmeier KD, Kutas M. It's about time. *Brain Lang* 2000; **71**:62–64.
- 21 Perruchet P, Vinter A. The self-organizing consciousness. *Behav Brain Sci* 2002; **25**:297–330.

Etude 2

L'expertise musicale facilite l'apprentissage implicite des structures linguistiques et musicales.

Clément François & Daniele Schön
Publié dans *Cerebral Cortex* en 2011

Résumé:

L'expertise musicale est connue pour entraîner des modifications au niveau de la perception auditive et de l'organisation corticale sous-jacente. Dans cette étude, nous montrons que les modifications liées à l'expertise musicale sont susceptibles d'affecter des fonctions cognitives de plus hauts niveaux et que ces modifications se généralisent au traitement de la parole. Des études précédentes ont montré que les adultes et les nouveau-nés peuvent segmenter des flux continus de stimuli linguistiques et non linguistiques sur la seule base des probabilités d'occurrence entre syllabes adjacentes ou notes adjacentes. Dans cette étude, nous avons utilisé un protocole d'apprentissage implicite de langage artificiel (chanté) tout en le couplant avec une approche électrophysiologique.

Alors que les données comportementales ne montrent pas d'effet d'expertise musicale clair, les données des Potentiels évoqués montrent que comparés aux non-musiciens, les musiciens ont mieux appris les deux dimensions contenues dans le langage chanté. Nous discutons ces résultats en termes de modifications induites par l'expertise musicale au niveau du fonctionnement du système auditif, des processus de segmentation de flux et des processus de mémoire.

Musical Expertise Boosts Implicit Learning of Both Musical and Linguistic Structures

Clément Francois and Daniele Schön

Institut de Neurosciences Cognitives de la Méditerranée CNRS and Aix-Marseille Universities, Marseille 13402, France

Address correspondence to Clément Francois, Institut de Neurosciences Cognitives de la Méditerranée CNRS and Aix-Marseille Universities, 31, Chemin Joseph Aiguier, 13402 Marseille, France. Email: clement.francois@incm.cnrs-mrs.fr.

Musical training is known to modify auditory perception and related cortical organization. Here, we show that these modifications may extend to higher cognitive functions and generalize to processing of speech. Previous studies have shown that adults and newborns can segment a continuous stream of linguistic and nonlinguistic stimuli based only on probabilities of occurrence between adjacent syllables or tones. In the present experiment, we used an artificial (sung) language learning design coupled with an electrophysiological approach. While behavioral results were not clear cut in showing an effect of expertise, Event-Related Potentials data showed that musicians learned better than did nonmusicians both musical and linguistic structures of the sung language. We discuss these findings in terms of practice-related changes in auditory processing, stream segmentation, and memory processes.

Keywords: brain plasticity, language, music, musical expertise, N400, statistical learning

Introduction

Musicians often undergo an intensive formal training period that can last 10–15 years and that can imply up to 10000-h practice by early adulthood (Krampe and Ericsson 1996). Moreover, they keep practicing several hours a day during their whole career. Thus, comparing musicians with nonmusicians allows studying the effects of extensive audiomotor training on the functional and structural organization of the brain. At a perceptual level, musicians have a lower frequency threshold than do nonmusicians (Kishon-Rabin et al. 2001). These differences might be related to functional differences in the auditory pathway. Musicians' brainstem responses show more robust encoding of sound pitch processing (Musacchia et al. 2007; Wong et al. 2007). Moreover, when listening to synthetic or instrumental sounds, musicians show larger N1 and P2 event-related potentials than nonmusicians do (Shahin et al. 2003, 2005). Moreover, they are more sensitive to sound spectral structure: For instance, musicians present a larger N1m to piano sounds than to pure tones, while nonmusicians are not sensitive to this contrast (Pantev et al. 1998). When deviant stimuli are introduced in a regular stream of identical sounds, participants, typically watching a silent movie (i.e., not paying attention to the sound stream), show a mismatch negativity (MMN) to deviant sounds. Interestingly, it has been shown that musicians have a larger MMN and MMNm than nonmusicians do when deviant chords are introduced in the stream (Koelsch et al. 1999; Brattico et al. 2009) as well as when a sound is omitted (Rüsseler et al. 2001). These differences can be interpreted as reflecting a greater efficiency of musicians' auditory system in processing sound features and can be accompanied by morphological differences, showing that musicians compared with nonmusicians have a different gray-

matter concentration in the auditory cortex (Bermudez and Zatorre 2005) and a larger planum temporale (Schlaug et al. 1995; Keenan et al. 2001).

The fact that musicians perceive some sound features more accurately than nonmusicians do is not so surprising. After all, they spend hours and hours of their life focusing on sounds and the way they are generated, paying particular attention to pitch, timber, duration, and timing. However, what seems less evident to us is whether or not this intensive musical practice can affect nonmusical abilities. Several recent studies seem to confirm this possibility.

For instance, both adult and child musicians have a better performance than matched controls do when asked to detect fine contour modifications in the prosody of an utterance (Schön et al. 2004; Magne et al. 2006). These results are supported by other findings showing a possible correlation between musical and linguistic aptitudes in children (Anvari et al. 2002; Slevc and Miyake 2006; Milovanov et al. 2008, 2009). Indeed Slevc and Miyake (2006) have shown that musical aptitude is a rather good predictor of both receptive and productive phonological abilities in second language learning.

In this study, we took the challenge of focusing on a rather high cognitive function: word segmentation, namely, the ability to extract words from continuous speech. Because the speech stream does not systematically carry consistent acoustic cues such as pauses or accents at word boundaries, there must be other ways to learn how to segment words. According to several publications, speech might be segmented in a rather implicit manner based on one source of information: the statistical structure of the language (Saffran et al. 1996b; Kuhl 2004).

For instance, in English, the syllable pre can be followed by a small number of syllables (like ty, pare, clude); thus, the probability that pre is followed by ty is quite high in regular speech. By contrast, the syllable ty mostly occurs words finally and can thus be followed by any word initial syllables, making the probability that ty is followed by ba, as in "pretty baby," very low (Saffran 2003). In general, "syllables that are part of the same word tend to follow one another predictably, whereas syllables that span word boundaries do not" (Saffran et al. 2001).

The role of transitional probabilities in speech segmentation (viz., the probability of syllable X given syllable Y) has been elegantly shown by Saffran and collaborators with infants and adults and has also been extended to neonates (Saffran et al. 1996a, 1996b; Aslin et al. 1998; Kuhl 2004; Gervain et al. 2008; Teinonen et al. 2009). Throughout this series of studies, the authors showed that listening to an artificial language without acoustic cues at word boundaries yields correct word segmentation. Indeed, in a test following the listening phase, participants are able to discriminate words that are part of the language from similar words that are not part of the language. This learning paradigm has also been replicated using non-linguistic stimuli such as sounds with different pitches (Saffran

et al. 1999) and timber (Tillmann and McAdams 2004), as well as in the cotton top tamarin and in rats (Hauser et al. 2000, 2002; Ramus et al. 2000; Toro and Trobalón 2005).

We previously showed (Schön et al. 2008) that a sung language allowed a better word segmentation than a spoken language, and we interpreted this as a beneficial effect of the structuring and motivational properties of music on speech segmentation. Recently, we used the same sung material, and we investigated both melodic and word segmentation while recording electroencephalography (EEG). Nonmusician participants listened to a sung artificial language and were then tested using a 2-alternative forced-choice test (familiar/unfamiliar) (Francois and Schön 2010). While the linguistic dimension was learned better than the melodic dimension, at a behavioral level, a negative N400-like component was larger for unfamiliar than for familiar words in both linguistic and musical dimensions. This well fits the literature showing a larger N400 component to unfamiliar than familiar words (Young and Rugg 1992).

In the present study, we used the same learning paradigm as in the study by Francois and Schön (2010) but comparing 2 groups. Participants listened to an artificial sung language (wherein music and language dimensions are highly intertwined) and were then tested with a 2-alternative forced-choice task on pairs of words and melodies (familiar vs. unfamiliar). The main goal of this study was to test whether musical expertise can facilitate word segmentation. With this aim, we compared 2 groups, one group with formal musical training and one without. At the behavioral level, we hypothesized that musicians may perform better than nonmusicians do on both musical and linguistic dimensions, because they would benefit the most from the musical information carried by the sung language. At the electrophysiological level, we hypothesized that the N400 familiarity effect with a frontocentral distribution would be larger for musicians than for nonmusicians for both linguistic and musical dimensions.

Materials and Methods

Participants

Two groups participated in the experiment. Sixteen professional musicians (mean age 27 years, 15 right-handed, 10 males, normal hearing, no known neurological problems, >12 years of formal musical learning and from 3 to 7 h of daily practice, 5 of them reported absolute pitch) and 20 nonmusicians (mean age 25 years, 20 right-handed, 11 males, normal hearing, no known neurological problems, not >2 years of formal musical training and no instrument practice). All participants were native French speakers and listened to 5.5 min of a continuous speech stream resulting from the concatenation of 5 3-syllable nonsense words (hereafter words) that were repeated in a pseudorandom order. All participants were paid 20 Euros. Informed consent was obtained from all participants, and the data were analyzed anonymously. This study was approved by the CNRS, Mediterranean Institute for Cognitive Neuroscience, and was conducted in accordance with national norms and guidelines for the protection of human subjects.

Materials

The language consisted of 4 consonants and 3 vowels, which were combined into a set of 11 syllables. The average syllable length was 230 ms, standard deviation was 16 ms. These syllables were then combined to give rise to 5 trisyllabic sung words (gimysy, mimosi, pogysi, pymiso, and sipyy). Each of the 11 syllables was associated with a distinct tone (C3, D3, F3, G3, A3, B3, C4, Db4, D4, E4, and F4). Therefore, each word was always sung on the same melodic contour (gimysy C3 D3 F3, mimosi E4 Db4 G3, pymiso B3 E4 F4, pogysi D4 C3 G3, and sipyy G3

B3 C4). The mean pitch interval within words was not significantly different from the mean interval between words ($P = 0.4$). Moreover, pitch-contour changes could not be used to segment the stream because they took place 50% of the time only at word boundaries.

Transitional probabilities within words ranged from 0.5 to 1.0. Transitional probabilities across word boundaries ranged from 0.1 to 0.5. While we used a 6-word language stream in our previous experiment (Francois and Schön 2010), in the present study, the language stream was built by a random concatenation of the 5 words (only constraint: no repetition of the same item twice in a row) and synthesized using Mbrola (<http://tcts.fpms.ac.be/synthesis/mbrola.html>). No acoustic cues were inserted at word boundaries. Each word was repeated 100 times in the stream.

Design and Procedure

During the learning phase, participants were told that they would listen to a continuous stream of sung syllables for several minutes, and they were asked to carefully listen to these sounds. During the test, participants had to choose, by pressing 1 of 2 buttons on a computer keyboard, which of 2 strings (first or second item) most closely resembled what they had just heard in the stream. Items had a flat contour ("spoken" version) in the linguistic test while they were played with a piano sound in the musical test (Fig. 1). In each test trial, one item was a "word" (linguistic test) or "melody" (musical test) from the artificial language (hereafter familiar word/melody); the other item was built by putting together the end of a word/melody with the beginning of another. Thus, these partial items were legal items of the language but had been heard 4 times less than the words/melodies (hereafter unfamiliar words/melodies). More precisely, these items contained either the last syllable of a word plus the first syllable of another word or the last syllable pair of a word plus the first syllable of another word. For instance, the last 2 syllables of the word pogysi were combined with the first syllable of mimosi to create the unfamiliar word gysimi (see Fig. 1). The 5 unfamiliar words used during the test were gysimi, mosigi, pygyimi, sogimy, and syogy. Each familiar word of the language was presented with each unfamiliar word, making up 25 pairs. During the musical test, the unfamiliar melodies were built with the same pitches defining the melodic contour of the unfamiliar words described above. Stimuli were presented via loudspeakers. Linguistic and musical tests lasted 5 min each, and their order was counter-balanced across participants.

Data Acquisition and Analysis

Participants were comfortably seated in a Faraday booth. The EEG was recorded from 32 scalp electrodes (Biosemi ActiveTwo system, Amsterdam University) located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (International 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F3, F4, C3, C4, P3, P4, P7, P8, Po3, Po4, O1, O2, F7, F8, T7, T8, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). The bandpass was of 0–102.4 Hz, and sampling rate was 512 Hz. The data, acquired during the testing phase, were then referenced offline to the algebraic average of the left and right mastoids. Raw data containing movement artifacts or amplifier saturation were excluded. Signal containing ocular artifacts was corrected using Independent Component Analysis decomposition by removing the component containing the blink (Makeig et al. 1996). The EEG was then epoched for each item (familiar and unfamiliar) and the 200-ms baseline (before the stimulus onset) zero-mean normalized using Brain Vision Analyzer software (Brain Products, Munich, Germany).



Figure 1. Illustration of the experimental design used in the present experiment. Stimuli were presented auditorily.

Analyses on the N1 and the P2 components were performed on peak amplitudes. Peak amplitudes were computed for each subject and defined as the mean value around the peak latency (i.e., the mean in a window of ± 10 ms around the peak). Later components were analyzed by computing the mean amplitude using latency windows of 50 ms. Analyses were performed for correct trials only and for all trials (correct and incorrect trials). Repeated-measure analysis of variance (ANOVA) was used for statistical assessment, using expertise (musicians and nonmusicians) and familiarity (familiar and unfamiliar items) as factors. This differed from behavioral analyses wherein familiarity could not be taken into account (i.e., 1 response/trial in behavior vs. 2 separate items in EEG). Moreover, the topographical distribution of the effects was modeled by 2 additional factors (Hemisphere, left and right and Location, frontal, central, and parietal) defined as follows: left (AF3, F3, F7) and right (AF4, F4, F8) frontal, left (Fc1, C3, Cp1) and right (Fc2, C4, Cp2) central, and left (Po3, P3, P7) and right (Po4, P4, P8) parietal. All *P* values reported below were adjusted using the Greenhouse-Geisser correction for nonsphericity, when appropriate, and Sidak tests were used in post hoc comparisons. Because of the increased likelihood of type I errors associated with the large number of statistical tests, only effects that reached significance ($P < 0.05$) in at least 2 consecutive time windows were considered significant.

Results

Behavioral Data

Results of a 2-way ANOVA (Expertise [as between factor with 2 levels] and Dimension [as within factor with 2 levels]) showed a main effect of dimension ($F [1,35] = 12.52$; $P = 0.001$): The linguistic dimension was learned better than the musical one for both groups. Although not significant, there was a trend for musicians to have a higher level of performance than non-musicians in both tested dimensions (main effect of Expertise: ($F [3,35] = 2.7$; $P = 0.11$) (see Fig. 2). This trend was equally distributed across the 2 dimensions (Expertise by Dimension interaction: [$F < 1$]). Comparison of performance with chance level (here 50%) showed that while the participants' level of performance in the linguistic test was above chance for both nonmusician and musician groups (56% and 60% of correct responses $P = 0.03$, $P = 0.007$, respectively, Wilcoxon test) this was not the case in the musical test (48% of correct responses for nonmusicians and 52% for musicians). Finally, an additional 3-way ANOVA including expertise (as between factor with 2 levels), dimension (as within factor with 2 levels) and "transitional probability (TP)" (as within factor with 2 levels) was conducted. With this aim, the average performances for the 3 words with the highest average TP (mean TP = 0.83) were compared with the average performances for the 2 words with the lowest average TP (mean TP = 0.5). The dimension by TP interaction was significant ($F [1,35] = 8.6$; $P = 0.005$). Post hoc analyses revealed that in the musical test, high TP items were better learned (52%) than low TP items (42%; $P = 0.02$) for the 2 groups. By contrast, in the linguistic test, no significant differences were found between performances for high TP items (54%) and performances for low TP items (60%; $P = 0.10$). Neither the main effect of Expertise was significant nor the interactions involving this factor (all $P > 0.2$).

Electrophysiological Data

In the following section, electrophysiological data are presented separately for each dimension for 2 reasons. First, at the behavioral level, musical dimension was not learned as well as the linguistic dimension, thus possibly leading to smeared ERP results because of confounding and/or overlapping cognitive

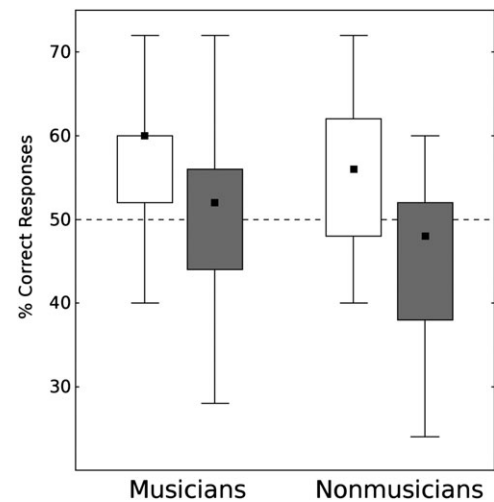


Figure 2. Percentage of correct responses: box plot of performances in the linguistic (white) and musical tests (gray) for musicians (left side) and nonmusicians (right side) (dashed line = chance level). Black squares indicate the medians.

processes engaged. Second, the acoustic characteristics of musical and linguistic stimuli are quite different; piano tones (as used for the musical dimension) have shorter attack time than syllables, thus leading to early differences in the ERPs (larger N1-P2 complex for piano tones).

Language Test

In order to compare ERPs to familiar items and to unfamiliar items, we performed a 4-way ANOVA including Expertise as between factor, Familiarity (Familiar and Unfamiliar), Location (Frontal, Central, and Parietal), and Hemispheres (Right and Left) as within factors on correct trials only.

Results of the ANOVA performed on the N1 peak amplitude showed a significant main effect of Expertise ($F [1,35] = 5.81$; $P = 0.02$): The N1 component was larger for musicians ($-2.6 \mu\text{V}$) than for nonmusicians ($-1.3 \mu\text{V}$). The familiarity effect was marginally significant ($F [1,35] = 2.35$; $P = 0.13$). No differences were significant on the P2 component (see Fig. 3).

In the 550- to 750-ms latency range, the Familiarity by Location interaction was significant ($F [2,70] = 3.82$; $P = 0.04$): ERPs were more negative for unfamiliar ($-3.3 \mu\text{V}$) than for familiar words ($-2.1 \mu\text{V}$; $P < 0.001$) over frontal regions. This familiarity effect spreads over the scalp in the 600- to 850-ms range, as revealed by the significant main effect of Familiarity ($F [1,35] = 11.68$; $P = 0.004$).

Most importantly, the Expertise by Familiarity interaction was significant in the 750- to 850-ms latency band ($F [1,35] = 6.16$; $P = 0.01$): Post hoc analyses showed a significant familiarity effect for musicians ($2 \mu\text{V}$ of effect size; $P = 0.02$) but not for nonmusicians ($0.5 \mu\text{V}$ of effect size; $P = 0.5$, see Figs. 4 and 5; see also the footnote for additional information concerning all trial analyses¹).

¹All trials analyses showed a significant main effect of Familiarity in the 600-850 ms and Expertise by Familiarity interaction in the 750- to 850-ms latency bands ($P = 0.001$ and 0.02 , respectively). Only the *P* value for the Familiarity by Location interaction in the 550- to 750-ms latency band was no more significant ($P = 0.23$), pointing to a greater sensitivity of correct trials analyses.

Musical Test

The same analyses described for the linguistic test were performed but on all trials. Indeed, because in this test behavioral responses did not significantly differ from chance, an analysis on correct trials would not be appropriate. Moreover, participants' correct trials were not enough to yield an acceptable signal-to-noise ratio. This was also the reason why we decided not to pool together the data from the linguistic and musical tests.

Results of the ANOVA on the N1 peak amplitude showed a main effect of Expertise ($F [1,35] = 4.24; P = 0.04$): The N1 amplitude was larger for musicians ($-4.8 \mu\text{V}$) than for non-musicians ($-3.2 \mu\text{V}$, see Fig. 6). The familiarity effect was not significant ($F [1,35] = 1.1; P = 0.3$). Analyses of the P2 component also showed a significant main effect of Expertise ($F [1,35] = 9.73; P = 0.003$): the P2 amplitude was larger for musicians ($4.4 \mu\text{V}$) than for nonmusicians ($2.2 \mu\text{V}$). Most interestingly, this analysis also revealed a significant Expertise by Familiarity by Hemisphere interaction ($F [1,35] = 4.13; P = 0.04$). Post hoc analyses showed a significant Familiarity effect size for musicians but not for nonmusicians ($1 \mu\text{V}$ vs. $0.1 \mu\text{V}$). Musicians' Familiarity effect was larger over the left hemisphere ($1.2 \mu\text{V}$) than over the right ($0.7 \mu\text{V}$).

Between 350 and 500 ms, the Expertise by Familiarity by Location by Hemisphere interaction was significant ($F [2,70] = 4.98; P = 0.02$). Post hoc analyses revealed a significant familiarity effect for musicians but not for nonmusicians (0.75 vs. $0.15 \mu\text{V}$). Musicians' Familiarity effect was largest over central and left frontal regions ($0.9\text{-}\mu\text{V}$ effect size, $P < 0.001$, see Figs. 7 and 8).²

Between 600 and 700 ms, results of the ANOVA showed a significant main effect of Familiarity ($F [1,35] = 3.99; P = 0.05$). In this latency range, unfamiliar melodies ($-1.2 \mu\text{V}$) elicited more negative ERPs than familiar melodies ($-0.7 \mu\text{V}$).

Between 700 and 800 ms, results of the ANOVA showed a significant Expertise by Familiarity by Location by Hemisphere interaction ($F [2,70] = 3.42; P = 0.05$). Post hoc analyses revealed a significant familiarity effect for musicians but not for nonmusicians (0.4 vs. $0.1 \mu\text{V}$). Musicians' Familiarity effect was significant over right frontal regions only ($0.4 \mu\text{V}$ effect size, $P < 0.001$; see Fig. 8).

Discussion

The main goal of this study was to test whether artificial language learning (segmentation) may benefit from formal

²We interpret this negativity as an MMN-like component to the second tone. One way to test this is to analyze separately the 2 types of unfamiliar items. Type 312 are built using the last syllable of a word plus the first syllable pair of another word whereas type 231 are built using the last syllable pair of a word plus the first syllable of another word). Because in type 312 items the "illegal" transition takes place at the second note, whereas for type 231 it does not, we expected a larger negativity to type 312 than to type 231. We conducted additional analyses in the 350- to 500-ms latency band aiming at comparing mean amplitudes elicited by these 2 different types of unfamiliar items. The Expertise by Item by Hemisphere was significant ($F [1,35] = 6.41; P = 0.01$). Post hoc analyses revealed that, for musicians only, ERPs for 312 unfamiliar items were significantly more negative than for 231 items over the right ($P = 0.04$) but not over the left Hemisphere ($P = 0.99$). No significant differences were found for nonmusicians.

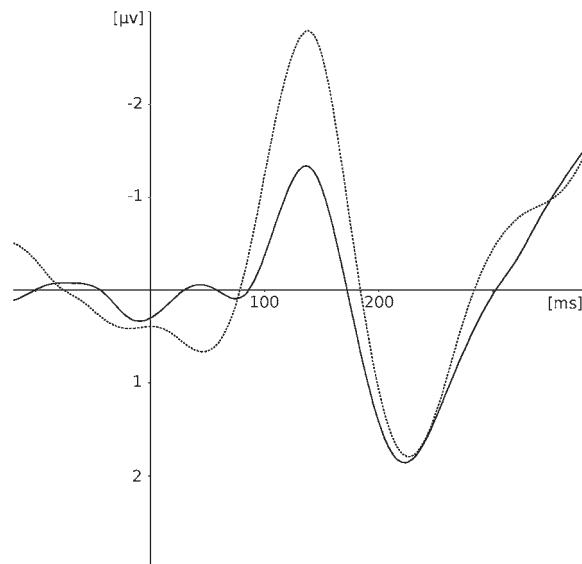


Figure 3. Linguistic test: modulation of the N1-P2 complex as a function of expertise at Cz electrode (dashed line = musicians, solid line = nonmusicians).

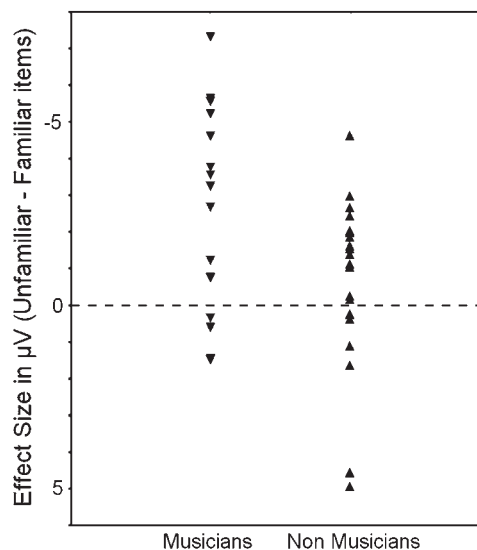


Figure 4. Familiarity effect size in the linguistic dimension for musicians and nonmusicians in the 750-850 latency band at Cz electrode (negativity is up).

music training. With this aim, we recorded both behavioral and electrophysiological measures during the testing sessions in an artificial language learning paradigm and we compared musician and nonmusician participants.

Music Training Shapes the Auditory System

In both tested dimensions, electrophysiological data revealed an effect of expertise on the N1 component with larger amplitude for musicians than for nonmusicians. Previous studies described a similar effect of expertise on the N1 in response to musical sounds (Pantev et al. 1998; Shahin et al. 2003). This effect is even stronger for timbers of the played instrument (e.g., trumpet sound for trumpeters, Pantev et al. 2001). These functional changes may rely on anatomical changes induced by musical training in the auditory regions

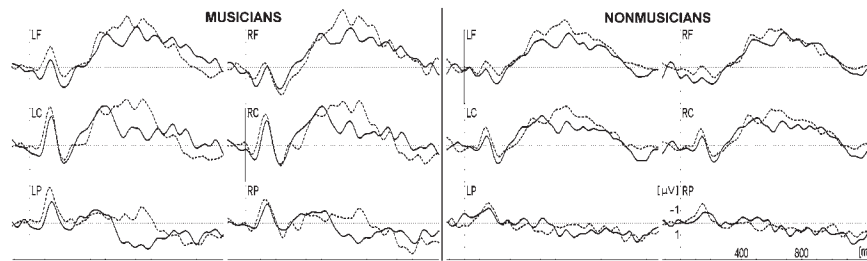


Figure 5. Musicians (Left side) and nonmusicians (right side), ERPs to familiar (solid) and unfamiliar words (dashed) in the linguistic session. Each site represents the mean of the 3 electrodes included in the ROI (LF: left frontal, RF: right frontal, LC: left central, RC: right central, LP: left parietal, RP: right parietal).

(Schlaug et al. 1995; Keenan et al. 2001; Schneider et al. 2002). Moreover, recent studies using musical or linguistic stimuli and focusing on an earlier stage of auditory processing (Musacchia et al. 2007; Wong et al. 2007; Lee et al. 2009; Parbery-Clark et al. 2009; Bidelman and Krishnan 2010), found a larger frequency-following response in the brainstem of musicians compared with that of nonmusicians. Our results show that these differences can also be found at later stages of processing, such as the N1, probably originating in the belt and parabelt auditory regions (Liégeois-Chauvel et al. 1994; Hackett et al. 2001). However, care must be taken in the interpretation of the present results, as one cannot exclude a top-down effect such as an increase of selective attention to sounds for musicians. Nonetheless, recent results indicate that the effect of music expertise on the N1-P2 complex is not primarily caused by selective attention. Rather, they support the view that increased auditory-evoked potentials in musicians reflect an enlarged neuronal representation for specific sound features of these tones (Baumann et al. 2008).

Enhanced Categorization of Tonal and Interval Information in Musicians

In the musical test, only musicians showed a significant familiarity effect on the P2 component. The P2 is known to be sensitive to the level of attention and to the level of musical expertise (Hillyard et al. 1987; Shahin et al. 2003; Kuriki et al. 2006). P2 amplitude modulations have also been reported in an auditory stream segregation paradigm: The P2 amplitude is greater when a stream of 2 sounds can be perceptually segregated in 2 independent streams (Snyder et al. 2006). Interestingly, a P2 modulation has been reported using artificial language learning paradigms, using linguistic stimuli presented in the auditory modality: The P2 amplitude increases along the listening phase for good learners but not for poor learners (De Diego Balaguer et al. 2007). In our case, the P2 observed in musicians was larger for unfamiliar than for familiar melodies. This is probably due to an increase in short-term memory load (Conley et al. 1999) or to a more difficult categorization of the unfamiliar than familiar items (Liebenthal et al. 2010). Interestingly, this familiarity effect took place 200 ms after the first tone onset that is before the beginning of the second tone. Because the musical structure of the sung language contained a rather strong tonal center (10 notes of 11 were in C major), it is possible that musicians kept in memory a representation of this tonal center that influenced the processing of the first tone of each item (Krumhansl and Kessler 1982). In fact, once a tonal center is established, each individual pitch can be processed and categorized relative to the tonal center (with no need of absolute pitch). Also note that, because the pitches of

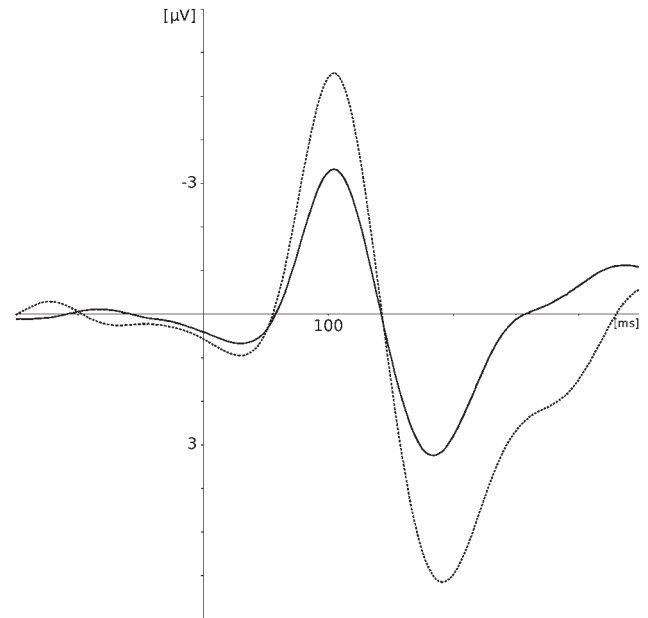


Figure 6. Musical test: modulation of the N1-P2 complex as a function of expertise at Cz electrode (dashed line = musicians, solid line = nonmusicians).

the first tone of familiar and unfamiliar items all differed (except for one), a good performance was theoretically possible on the sole basis of the information carried by the first pitch. Taken together, these results suggest that musicians might be able to discriminate familiar and unfamiliar musical items based on the tonal function of the first pitch.

The familiarity effect on the P2 was followed by another familiarity effect on a negative component, only significant for musicians in the music test. This component, larger for unfamiliar than familiar items had its peak around 450 ms. Due to its narrow width, rather than interpreting it as a late negative component, we consider it as an MMN in response to the second tone of the melody (see Fig. 7). This interpretation is supported by the fact that this effect was stronger for the unfamiliar items, wherein the second tone was an illegal transition (i.e., “312” type: last syllable of a word plus the first syllable pair of another word) compared with items, wherein the second tone was a legal transition (i.e., “231” type: last syllable pair of a word plus the first syllable of another word).

Although the MMN is typically elicited by simple changes in a regular sequence (Näätänen et al. 1978), it has also been shown for more abstract rules (Tervaniemi et al. 1994). Interestingly, Pantev and collaborators (Herholz et al. 2008) compared musicians and nonmusicians in a musical imagery

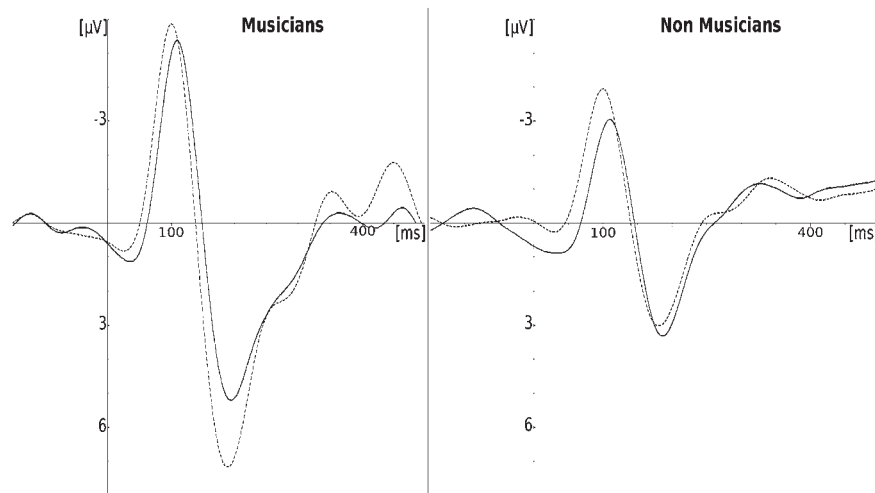


Figure 7. Musicians (left side) and nonmusicians (right side), ERPs to familiar (solid) and unfamiliar melodies (dashed) in the musical session, at Cz electrode.

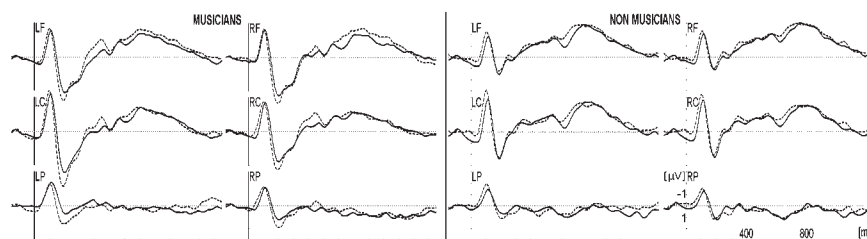


Figure 8. Musicians (left side) and nonmusicians (right side): ERPs to familiar (solid) and unfamiliar melodies (dashed) in the musical session. Each site represents the mean of the 3 electrodes included in the ROI (LF: left frontal; RF: right frontal, LC: left central, RC: right central, LP: left parietal, RP: right parietal).

task with familiar melodies. Participants listened to the incipit of the melodies, continued them in their mind and then heard a tone that fitted in or not with the melody. Only musicians showed an MMN to incorrect continuations. These results suggest that a long-term memory trace (familiar melody) can induce, in musicians, a sufficiently strong representation to yield an MMN to an incorrect item. This is in line with our results. First, participants stored in memory the representations of melodies segmented during learning; then they heard, during the test, familiar (correct) or unfamiliar (incorrect) items. The larger MMN to the second tone of unfamiliar items can thus be interpreted as a mismatch of the first interval with respect to a memory trace stored during the learning session. Indeed, the second tone allows the building of the first interval of the melodies, also giving, in theory, all the necessary information to discriminate familiar from unfamiliar items.

Overall, both familiarity effects on the P2 and MMN-like components point to the fact that musicians did learn the musical structure better than nonmusicians, and this in turn affected the way they processed the musical items presented during the test. This confirms and extends previous findings showing that implicit learning of 12-tone serialist music is facilitated by experience with this music style. Indeed, while participants without a specialized training in atonal music cannot (implicitly) perceive the transforms of serialist music, participants with routine exposure to atonal music do (implicitly) perceive the distinction between different types of transforms (Dienes and Longuet-Higgins 2004). Again, as stated for the results on the N1, one cannot exclude that

musicians pay more attention to sounds in general and to the musical material in particular, even when the experimental tasks do not require them to do so. Indeed, also in our experiment, attentional processes during the passive listening may differ between musicians and nonmusicians in terms of selective attention or in terms of switching attention from one dimension to the other, and this may in turn affect behavioral and/or electrophysiological data beyond direct influences related to expertise.

Familiarity Effect and Memory Retrieval

In both tested dimensions, we found a larger frontal negativity for unfamiliar words than for familiar words (600- to 850-ms interval for the linguistic dimension and 600- to 800-ms interval for the musical one). This result is interesting as it confirms our previous results with nonmusicians showing a similar fronto-central negativity modulated by the degree of familiarity in both dimensions (Francois and Schön 2010). Due to the topography and the sensitivity to word familiarity, we previously interpreted this component as an N400-like component, possibly reflecting the memory trace stored during the listening phase. Although the latency is delayed with respect to the typical N400, one should consider that the items used here were sung words without any meaning. Interestingly, previous findings showed that the latency of the N400 to pseudowords is around 600 ms (O'Rourke and Holcomb 2002). Moreover, the involvement of N400-like components has been reported in various studies focusing on artificial language learning (Sanders

et al. 2002; Cunillera et al. 2006, 2009; De Diego Balaguer et al. 2007). Its modulation is often reported during the listening phase and is interpreted as an index of word prelexicalization due to speech segmentation processes. To our knowledge, only one study briefly analyzed and discussed the ERP data obtained during the testing phase, reporting a larger N400 for words than for nonwords (De Diego Balaguer et al. 2007). However, 2 important differences with the present design might explain this discordance. First, in the De Diego Balaguer experiment, participants underwent 2 successive learning phases, one with a speech stream similar to ours, and one with a stream including violations. Second, the testing phase contrasted words to illegal words (i.e., not present in the speech stream), while we contrasted familiar and unfamiliar words. While De Diego Balaguer and collaborators interpreted the N400 reduction to nonwords as reflecting the detection of impossible linguistic items, we interpret the N400 increase to unfamiliar words in terms of a classic N400 familiarity effect. Moreover, an alternative interpretation of the De Diego Balaguer results could also be that illegal words may lead to a quickly abandoned search resulting in smaller N400 amplitude for illegal compared with legal words. Indeed, Young and Rugg (1992) showed that unfamiliar words elicit a larger N400 than familiar words do, probably due to a longer lexical search.

Overall, the presence of this familiarity effect points to an ability of both musicians and nonmusicians to segment both linguistic and musical structures. As we previously found in this type of paradigm for nonmusicians (Francois and Schön 2010), the electrophysiological measures (on a N400-like component) seem to be more sensitive than behavioral measures. While accuracy measures pointed to a lack of learning of the musical dimension by nonmusicians and musicians, ERP results are clear cut in showing that both groups process differently familiar and unfamiliar items.

Musical Training Facilitates Implicit Learning of Both Linguistic and Musical Structures

The most important finding of this experiment is that such a familiarity N400-like effect was significantly greater in musicians compared to nonmusicians in both dimensions. This greater N400-like effect can be interpreted in terms of more distinguished memory representations for familiar and unfamiliar items in musicians. Indeed, the N400 amplitude seems to be sensitive to the ease of accessing information from long-term memory (Neville et al. 1986) and predicts very well immediate recall performance after learning (Elger et al. 1997).

These results reveal that musicians do have more “robust” representations of musical and linguistic structures shaped during the listening phase than nonmusicians. Even though the behavioral data are not clear-cut concerning an advantage of musicians over nonmusicians, several electrophysiological indices confirm the trend observed in the behavioral results. First, the familiarity effects present on the P2 to the first tone and on the MMN to the second tone show that musicians process differently familiar and unfamiliar musical items quite early on. Second, the familiarity effect present on a later negative component is larger for musicians than nonmusicians in both tested dimensions, pointing to a greater distinctiveness between familiar and unfamiliar items in musicians.

The most straightforward explanation of such an effect of Expertise is that musicians may be better in using the musical

structure contained in song language than nonmusicians and that this might be due to a better perception and/or memory for pitch. Another explanation might be that musical expertise facilitates regularity extractions and sequence learning in general. It is known that musicians can organize a sound sequence according to a number regularity, for instance, implicitly distinguishing segments containing 4 tones from the segments containing 5 tones, while such a perceptual organization of sound according to number is less relevant for nonmusicians (Van Zuijen et al. 2005). Moreover, recent findings showed that deaf children with cochlear implants are impaired in visual sequence learning, suggesting that a period of auditory deprivation may have a major impact on cognitive processes that are not specific to the auditory modality (Conway et al. 2011). Therefore, sound seems to provide a cognitive scaffolding for the development of serial-order behavior: whether sound processing is impaired or whether it is extensively practiced have opposite effects. Interestingly, Sluming et al. (2002) reported for musicians, an increased gray matter density and volume in the left-inferior frontal gyrus, Broca’s area. Moreover, neuroplastic development throughout musicians’ life seems to promote the retention of cortical tissue. This region is known to be involved in on-line speech stream segmentation as well as in music perception (Tillmann et al. 2003; McNealy et al. 2006). Overall, our results support an “auditory scaffolding hypothesis” (Conway et al. 2009), as we presently show that increased exposure to sounds leads to a benefit for implicit learning, putatively via anatomical and/or functional modifications going beyond the auditory regions. This second explanation makes the prediction that musicians should also have an advantage in learning an artificial spoken language, that is, without any specific musical cue. Further work will be necessary to test this possibility.

Finally, it is important to note that the differences due to expertise were smaller in the musical (familiarity effect musicians – familiarity effect nonmusicians = 0.4 μ V) than in the linguistic dimension (1.5 μ V). This is not surprising insofar as musicians, as well as nonmusicians, had troubles in learning the musical structure. While one may want to interpret such an advantage in speech compared with music processing as a general advantage of language learning over musical learning or as a greater resistance to interference of verbal items than musical items (Schendel and Palmer 2007), care must be taken in comparing results across dimensions because it is very difficult to control and balance the perceptual saliency of the linguistic and musical structures.

To conclude, these results contribute new evidence to the positive effects of musical practice on language segmentation. Although more research is needed to disentangle whether this effect is a general effect or more specific to the sung material we used, the results described above show that musicians are better able to segment both linguistic and musical structures. Because we focused here on the data acquired during tests taking place straight after learning, future studies will need to address whether these differences are due to an advantage in stream segmentation or in lexical storage or both.

Funding

French National Research Agency (ANR 05-BLAN-0214 “Music and Memory” to D.S.).

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Notes

We thank Julie Chobert and Jérôme Daltrozzo for assistance in EEG recordings. We thank Deirdre Bolger, Nia Cason, Mireille Besson, and 2 anonymous reviewers for constructive comments on a previous version. *Conflict of Interest*: None declared.

References

- Anvari SH, Trainor LJ, Woodside J, Levy BA. 2002. Relations among musical skills, phonological processing, and early reading ability in preschool children. *J Exp Child Psychol*. 83(2):111-130.
- Aslin RN, Saffran JR, Newport EL. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychol Sci*. 9:321-324.
- Baumann S, Meyer M, Jäncke L. 2008. Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J Cogn Neurosci*. 20(12):2238-2249.
- Bermudez P, Zatorre RJ. 2005. Differences in gray matter between musicians and nonmusicians. *Ann New York Acad Sci*. 1060:395-399.
- Bidelman GM, Krishnan A. 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res*. 1355:112-125.
- Brattico E, Pallesen KJ, Varyagina O, Bailey C, Anourova I, Järvenpää M, Eerola T, Tervaniemi M. 2009. Neural discrimination of non-prototypical chords in music experts and laymen: an MEG study. *J Cogn Neurosci*. 21(11):2230-2244.
- Conley EM, Michalewski HJ, Starr A. 1999. The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. *Clin Neurophysiol*. 110(12):2086-2093.
- Conway CM, Pisoni DB, Anaya EM, Karpicke J, Henning SC. 2011. Implicit sequence learning in deaf children with cochlear implants. *Dev Sci*. 14(1):69-82.
- Conway CM, Pisoni DB, Kronenberg WG. 2009. The importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. *Curr Dir Psychol Sci*. 18(5):275-279.
- Cunillera T, Càmarà E, Toro JM, Marco-Pallares J, Sebastián-Galles N, Ortiz H, Pujol J, Rodríguez-Fornells A. 2009. Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*. 48(3):541-553.
- Cunillera T, Toro JM, Sebastián-Gallés N, Rodríguez-Fornells A. 2006. The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Res*. 1123(1):168-178.
- De Diego Balaguer R, Toro JM, Rodríguez-Fornells A, Bachoud-Lévi AC. 2007. Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS ONE*. 2(11):e1175.
- Dienes Z, Longuet-Higgins C. 2004. Can musical transformations be implicitly learned? *Cogn Sci*. 28:531-558.
- Elger CE, Grunwald T, Lehnertz K, Kutas M, Helmstaedter C, Brockhaus A, Van Roost D, Heinze HJ. 1997. Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia*. 35(5):657-667.
- Francois C, Schön D. 2010. Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport*. 21(14):928-932.
- Gervain J, Macagno F, Cogoi S, Peña M, Mehler J. 2008. The neonate brain detects speech structure. *Proc Natl Acad Sci U S A*. 105:14222-14227.
- Hackett TA, Preuss TM, Kaas JH. 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol*. 441(3):197-222.
- Hauser MD, Newport EL, Aslin RN. 2000. Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition*. 75:1-12.
- Hauser MD, Weiss D, Marcus G. 2002. Rule learning by cotton-top tamarins. *Cognition*. 86(1):B15-B22.

- Herholz SC, Lappe C, Knief A, Pantev C. 2008. Neural basis of music imagery and the effect of musical expertise. *Eur J Neurosci*. 28(11):2352-2360.
- Hillyard SA, Woldorff M, Mangun GR, Hansen JC. 1987. Mechanisms of early selective attention in auditory and visual modalities. *Electroencephalogr Clin Neurophysiol*. 39(Suppl):317-324.
- Keenan JP, Thangaraj V, Halpern AR, Schlaug G. 2001. Absolute pitch and planum temporale. *Neuroimage*. 14(6):1402-1408.
- 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(2 Suppl):125-143.
- Koelsch S, Schröger E, Tervaniemi M. 1999. Superior pre-attentive auditory processing in musicians. *Neuroreport*. 10(6):1309-1313.
- Krampe RT, Ericsson KA. 1996. Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J Exp Psychol*. 125:331-359.
- Krumhansl CL, Kessler EJ. 1982. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychol Rev*. 89:334-368.
- Kuhl PK. 2004. Early language acquisition: cracking the speech code. *Nat Rev Neurosci*. 5:831-843.
- Kuriki S, Kanda S, Hirata Y. 2006. Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J Neurosci*. 26(15):4046-4053.
- Lee KM, Skoe E, Kraus N, Ashley R. 2009. Selective subcortical enhancement of musical intervals in musicians. *J Neurosci*. 29(18):5832-5840.
- Liebenthal E, Desai R, Ellingson MM, Ramachandran B, Desai A, Binder JR. 2010. Specialization along the left superior temporal sulcus for auditory categorization. *cerebral cortex*. 20(12):2958-2970.
- Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P, Chauvel P. 1994. Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol*. 92(3):204-214.
- Magne C, Schön D, Besson M. 2006. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J Cogn Neurosci*. 18(2):199-11.
- Makeig S, Bell AJ, Jung TP, Sejnowski TJ. 1996. Independent component analysis of electroencephalographic data. *Adv Neural Inform Proc Syst*. 8:145-151.
- McNealy K, Mazziota JC, Dapretto M. 2006. Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci*. 26:7629-7639.
- Milovanov R, Huotilainen M, Esquef PA, Alku P, Välimäki V, Tervaniemi M. 2009. The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neurosci Lett*. 460(2):161-165.
- Milovanov R, Huotilainen M, Välimäki V, Esquef PA, Tervaniemi M. 2008. Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Res*. 1194:81-89.
- 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 U S A*. 104(40):15894-15898.
- Näätänen R, Gaillard AW, Mäntysalo S. 1978. Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol*. 42(4):313-329.
- Neville HJ, Kutas M, Chesney G, Schmidt AL. 1986. Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *J Mem Lang*. 25(1):75-92.
- O'Rourke TB, Holcomb PJ. 2002. Electrophysiological evidence for the efficiency of spoken word processing. *Biol Psychol*. 60:121-150.
- Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. 1998. Increased auditory cortical representation in musicians. *Nature*. 392(6678):811-814.
- Pantev C, Roberts LE, Schulz M, Engelien A, Ross B. 2001. Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*. 12(1):169-174.

- 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(45):14100-14107.
- Ramus F, Hauser MD, Miller C, Morris D, Mehler J. 2000. Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*. 288:349-351.
- Rüsseler J, Altenmüller E, Nager W, Kohlmetz C, Münte TF. 2001. Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci Lett*. 308(1):33-36.
- Saffran JR. 2003. Statistical language learning: mechanisms and constraints. *Curr Direc Psychol Sci*. 12(4):110-114.
- Saffran JR, Aslin RN, Newport EL. 1996a. Statistical learning by 8-month old infants. *Science*. 274:1926-1928.
- Saffran JR, Johnson EK, Aslin RN, Newport EL. 1999. Statistical learning of tones sequences by human infants and adults. *Cognition*. 70:27-52.
- Saffran JR, Newport EL, Aslin RN. 1996b. Word segmentation: the role of distributional cues. *J Mem Lang*. 35:606-621.
- Saffran JR, Senghas A, Trueswell JC. 2001. The acquisition of language by children. *Proc Natl Acad Sci U S A*. 98:12874-12875.
- Sanders LD, Newport EL, Neville JH. 2002. Segmenting nonsense: an event-related-potential index of perceived onsets in continuous speech. *Nat Neurosci*. 5(7):700-703.
- Schendl ZA, Palmer C. 2007. Suppression effects on musical and verbal memory. *Mem Cogn*. 35(4):640-650.
- Schlaug G, Jäncke L, Huang Y, Staiger JF, Steinmetz H. 1995. Increased corpus callosum size in musicians. *Neuropsychologia*. 33(8):1047-1055.
- Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A. 2002. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat Neurosci*. 5:688-694.
- Schön D, Boyer M, Moreno S, Besson M, Peretz I, Kolinsky R. 2008. Song as an aid for language acquisition. *Cognition*. 106(2):975-983.
- Schön D, Magne C, Besson M. 2004. The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*. 41(3):341-349.
- Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci*. 23(13):5545-5552.
- Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B. 2005. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*. 16(16):1781-1785.
- Slevc LR, Miyake A. 2006. Individual differences in second-language proficiency: does musical ability matter? *Psychol Sci*. 17(8):675-681.
- Sluming V, Barrick T, Howard M, Cezayirli E, Mayes A, Roberts N. 2002. Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage*. 17(3):1613-1622.
- Snyder JS, Alain C, Picton TW. 2006. Effects of attention on neuro-electric correlates of auditory stream segregation. *J Cogn Neurosci*. 18(1):1-13.
- Teinonen T, Fellman V, Näätänen R, Alku P, Huotilainen M. 2009. Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci*. 13:10-21.
- Tervaniemi M, Maury S, Näätänen R. 1994. Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport*. 5(7):844-846.
- Tillmann B, Janata P, Bharucha JJ. 2003. Activation of the inferior frontal cortex in musical priming. *Cogn Brain Res*. 16(2):145-161.
- Tillmann B, McAdams S. 2004. Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis-)similarities. *J Exp Psychol Learn Mem Cogn*. 30:1131-1142.
- Toro JM, Trobalón JB. 2005. Statistical computations over a speech stream in a rodent. *Percept Psychophys*. 67(5):867-875.
- Van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M. 2005. Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. *Brain Res Cogn Brain Res*. 23(2-3):270-276.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N. 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*. 10(4):420-422.
- Young MP, Rugg MD. 1992. Word frequency and multiple repetitions as determinants of the modulation of event-related-potentials in a semantic classification task. *Psychophysiology*. 29(6):664-676.

Etude 3

L'apprentissage de la musique facilite la segmentation de la parole.

Clément François, Julie Chobert, Mireille Besson and Daniele Schön

Soumis à *Nature Neuroscience* en 2011

Résumé:

Nous reportons dans cet article les résultats d'une étude longitudinale de deux ans, conduite chez l'enfant de 8 ans et utilisant les approches comportementales et électrophysiologiques. L'objectif de cette étude est de tester l'influence de l'apprentissage de la musique sur la segmentation de la parole. Deux groupes d'enfants ont été constitués de manière pseudo-aléatoire et ont suivi soit un apprentissage de la musique soit un apprentissage de la peinture pendant deux années. Les données comportementales et électrophysiologiques montrent clairement une amélioration de la segmentation de la parole au fur et à mesure des sessions pour le groupe d'enfant ayant suivi l'apprentissage de la musique uniquement. Ces résultats soulignent l'importance de la musique pour la perception de la parole et de manière plus générale pour l'éducation des enfants.

Music training for the development of language acquisition

Clément François, Julie Chobert, Mireille Besson and Daniele Schön

INCM, CNRS-Mediterranean University, Marseille, France

Abstract

We describe the results of a longitudinal study over 2 years using both behavioral and electrophysiological methods to test for the influence of music training on speech segmentation in 8 year-old children. Children followed either music or a painting training and were tested on their ability to extract words from a continuous flow of syllables. Both behavioral and electrophysiological measures clearly showed improved speech segmentation across testing sessions for the music group only. Our results point to the importance of music for speech perception and more generally for children's education.

Word count (abstract, main text, references and legends, max 1200): 1168

Music engages a wide range of processing mechanisms, from sound encoding to higher cognitive functions such as sequencing, attention, memory, and learning. These functions, shared with several other human abilities (e.g. language), might in turn be shaped by music training. Thus, musicians provide a privileged population for studying brain plasticity as well as investigating the intriguing possibility that musical expertise transfers to other domains such as language.

It is now established that music training induces functional and structural changes in the auditory and sensori-motor systems, making musicians more efficient and more sensitive in music-related tasks than nonmusicians. Moreover, there is growing evidence that music training benefits linguistic skills such as dynamic acoustic analysis, pitch and lexical stress processing, phonological awareness, reading and second-language proficiency (1, 2). Some of these findings have also been extended to children. Musically trained children better detect pitch changes in speech (3) and have a better vocabulary and reading ability than children who did not receive music training (2), thereby providing evidence from music to language transfer effects.

Previous findings on adults (4) showed that music training facilitates speech segmentation (i.e. word extraction from a continuous flow of sounds). Here we describe a longitudinal study spanning over 2 years and using a test-training-retest procedure with 8 year-old children to follow the developmental dynamics of music to speech transfer effect.

There is evidence that infants, children and adults can use the statistical properties of auditory input to discover words and sound patterns (5, 6). The conditional probability of syllable Y happening given syllable X will be higher for syllables that follow one another within a word than for those at word boundaries (e.g. in “pretty music”, the probability of “ty” given “pre” is higher than that of “mu” given “ty”).

In Experiment 1, twenty-four 8-year old children listened to 5 min of an artificial sung language (7) built by random concatenation of 4 trisyllabic words. Syllables were always sung using a fixed syllable-pitch mapping (Fig1 A). After this familiarization phase, children were presented two items (one with high and the other with low conditional probabilities) and had to decide which sounded more familiar (32 trials). Importantly, all items in the test were spoken, not sung. EEG and behavioral responses were recorded. Children were then pseudo-randomly assigned to two homogeneous training groups (controlling for age, school level, sex, musical expertise and the level of performance in several neuropsychological tests assessing reasoning, memory and attentional processing). One group took music and the other painting classes (45 minutes), twice a week in year 1 and once a week in year 2.

Experiments 2 and 3 were identical to Experiment 1 and took place after one and two years. The level of performance steadily increased for the music group across testing sessions while it did not change for the painting group [Training by Session interaction $F(2, 44) = 3.4, p = .04$] (Fig 1B), remaining at chance level (0.5) in all three experiments (p always $> .4$). By contrast, while the level of performance in the music group was also at chance before music training ($p = .4$), it was higher than chance after one and two years of music training ($p = .02$ and $p < .001$, respectively). Moreover, the benefit due to the second year of training was similar to the benefit of the first year ($p = .8$).

We also examined the neurophysiological basis of these behavioral findings. Before training, and in both groups, low probability (unfamiliar) words elicited a larger negativity than high probability (familiar) words between 450-550 ms over frontal regions [$F(2, 40)=10.5, p=.001$]. However, and most importantly this familiarity effect was larger, after two years of music training than of painting training (400-550 ms: [$F(1, 20)=5.5, p=.02$]). The frontal scalp distribution of this effect is in line with previous findings showing activity in the inferior and

middle frontal gyri that positively correlates with participant's rapid auditory sequencing skills (8).

Music training thus facilitates speech segmentation and fosters brain plasticity. This facilitation may result from various (but not exclusive) processes. Music training may improve general auditory encoding abilities that, in turn, facilitate speech segmentation (1, 2). Alternatively, music training may facilitate the emergence of more stable memory traces via a more efficient working memory system integrating pitch and syllabic structures. Finally, music training may reduce the effect of interference of adjacent syllables/items, possibly via more efficient temporal dynamic processing and focusing of attention.

Importantly, the longitudinal approach controls for possible pre-existing genetic differences and ascertains that music is the cause of the observed changes (3). Moreover, the effect described here was not a general effect due to higher motivation or arousal in the music class, since children in both groups improved equally well in reasoning and visual attention tasks. Our findings provide new evidence that music training can play an important role in children's education by facilitating speech segmentation, a building block of language acquisition: by developing both perceptual and cognitive functions, music training shapes individual development (9).

References and notes:

- (1) Tallal, P and Gaab, N. *Trends in Neurosci.* 29 (7), 382-390 (2006)
- (2) Nina Kraus, N, Chandrasekaran, B. *Nat. Rev. Neurosci.* **11 (8)**, 599 (2010).
- (3) Moreno, S et al. *Cereb. Cortex.* **19 (3)**, 712 (2009)
- (4) François, C and Schön, D. *Cereb. Cortex*, in press
- (5) Jenny R. Saffran, JR, Aslin, RN, Newport, EL. *Science* **274 (5294)**, 1926 (1996).

(6) Saffran, JR, Johnson EK, Aslin, RN, Newport, EL. *Cognition* **70 (1)**, 27 (1999).

(7) Schön, D et al. *Cognition* **106 (2)**, 975 (2008).

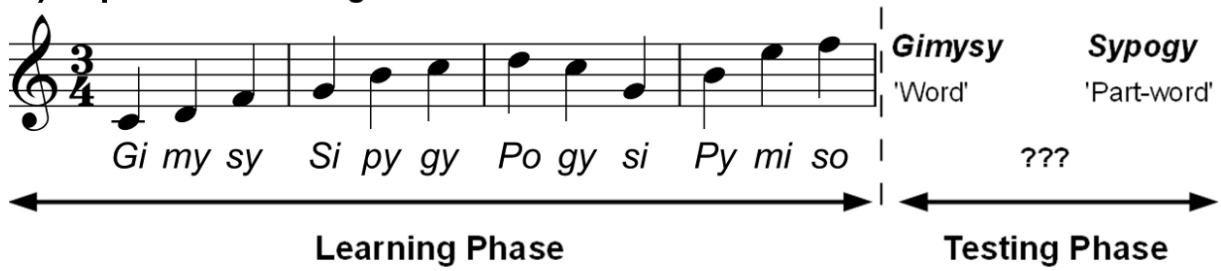
(8) McNealy, K, Mazziotta, J and Dapretto, M. *J. Neurosci*, 26(29): 7629-7639 (2006)

(9) A. D. Patel, in *Emerging Disciplines*, M. Bailar, Ed. (Rice Univ. Press, Houston, 2010), pp. 91-144.

(10) This research was conducted at the INCM, CNRS & Université de la Méditerranée, Marseille, France and was supported by a grant from the ANR-Neuro (#024-01)

Figure 2: Event-Related Potentials to familiar (solid) and unfamiliar (dashed) words in the music and painting groups after 2 years of training (F3 electrode). The map shows the average distribution of the negativity in the 400-550 ms range window.

A) Experimental Design



B) Performances

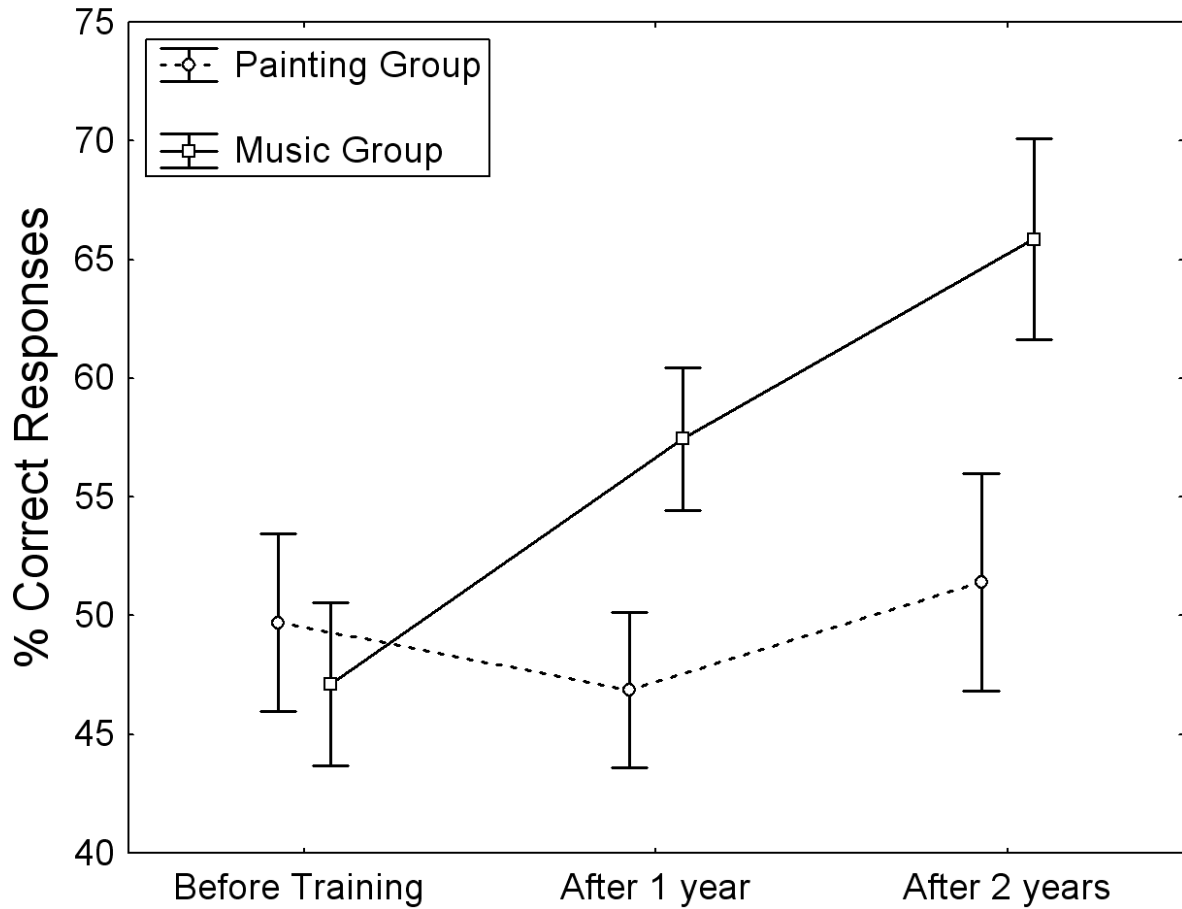


Figure 1: Effects of 1 and 2 years of music and painting training on speech segmentation. A) Experimental design used in the 3 experiments. B) Percentage of correct responses in the 3 experiments for the music (solid line) and painting (dashed line) groups. Error bars indicate the standard error of the mean.

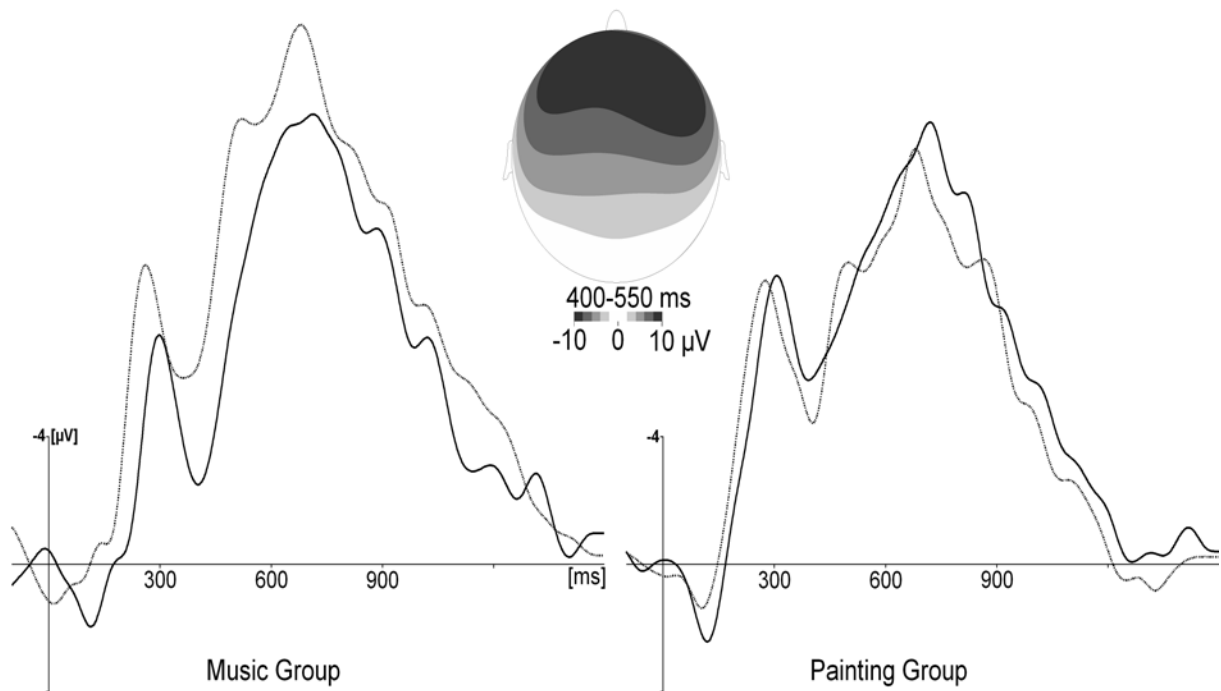


Figure 2: Event-Related Potentials to familiar (solid) and unfamiliar (dashed) words in the music and painting groups after 2 years of training (F3 electrode). The map shows the average distribution of the negativity in the 400-550 ms range window.

Supporting Online Material

Participants

Twenty-four 8-years-old nonmusician children were enrolled in these experiments (mean age=8, sd=.45, 19 right-handed, 14 boys, normal hearing, no known neurological problems). All children were French native speakers. Parental informed consent was obtained for each child and the data were analyzed anonymously. This study was approved by the CNRS and was conducted in accordance with national norms and guidelines for the protection of human subjects.

Material

The language was built using 9 syllables combined to give rise to 4 trisyllabic words (gimysy, pogysi, pymiso, sipygy). Each of the 9 syllables was associated with a distinct tone. Therefore each word had a unique melodic contour (gimysy C3 D3 F3, pymiso B3 E4 F4, pogysi D4 C4 G3, sipygy G3 B3 C4). The language stream was built by a random concatenation of the four words (without repetition of the same item twice in a row) and synthesized using Mbrola (<http://tcts.fpms.ac.be/synthesis/mbrola.html>). Each word was repeated 100 times in the stream.

Experimental Design and Procedure

During the learning phase, children were asked to listen carefully to a continuous stream of sounds. During the following test, children had to choose, by pressing one of two response buttons, which of two items (first or second) most closely resembled what they just heard in the stream. Items in the test were spoken (i.e. flat contour). In each test trial, one item was a “word” from the artificial language while the other was built by putting together the end of a word with the beginning of another (partial words: Gysigi, Pygygi, Sisipy, Sypymi). The mean Transitional probabilities were 0.8 for words and 0.4 for partial words. Each word of the language was presented with each partial word, making up 16 pairs and repeated twice in a quasi-random order (32 trials). Stimuli were presented via headphones. Learning phase and test lasted 5 minutes each.

Children were tested before training (Exp. 1), after one year (Exp 2) and after 2 years (Exp 3). The assignment to music or painting training group was based on the results of several neuropsychological tests issued from the WISC-IV and NEPSY as well as age, school level, sex, and socioeconomic background. None of the children and none of their parents had formal training in music or painting.

EEG was recorded before training (Exp 1) and after 2 years (Exp 3) from 32 scalp electrodes located at standard positions (International 10/20 system sites). The data were then re-referenced offline to the algebraic average of the mastoids. Trials containing artifacts were excluded. Two participants were discarded due to major EEG artifacts. The EEG was amplified by Biosemi amplifiers with a band-pass of 0-102.4 Hz and was digitized at 512 Hz.

ERP data were analyzed by computing the mean amplitudes between 0 and 1000 ms post-stimulus onset in 50 ms windows. Repeated measure ANOVA was used for statistical assessment. To test the distribution of the effects the model included the anterior-posterior (frontal, central and parietal) and a laterality factors (left and right). P-values were adjusted using the Greenhouse-Geisser correction. Because of the increased likelihood of type I errors, only effects that reached significance ($p < 0.05$) in at least two consecutive time windows were considered significant.

Two teachers professionally trained in music or painting, were specifically hired for this project. Music training was based on a combination of Kodaly and Orff approaches (<http://www.iks.hu/>; <http://www.orff.de/en.html>). Painting training was based on the approach developed by Arno Stern (<http://www.arnostern.com/>). The teaching activity was coordinated by the research group and care was taken to ascertain that both groups were similarly motivated and stimulated.

Etude 4

L'expertise musicale et l'apprentissage de structures musicales et linguistiques

Daniele Schön & Clément François

Soumis à *Frontiers In Auditory Cognitive Neurosciences* en 2011

Résumé:

Les adultes et les nouveaux nés peuvent utiliser les propriétés statistiques de séquences de syllabes pour extraire les mots du flux continu de parole. Dans cet article, nous présentons une revue fondée sur une série d'études électrophysiologiques ayant visé à comparer : 1) la segmentation d'un flux de syllabes parlées par rapport à la segmentation d'un flux de syllabes chantées, 2) l'extraction des informations linguistiques et musicales contenues dans un flux chanté ; 3) des musiciens avec des non musiciens. Les résultats montrent que, en comparaison avec des syllabes parlées, les performances sont améliorées avec des syllabes chantées . De plus, les résultats montrent qu'avec un matériel de stimulation chanté la structure linguistique est mieux segmentée que la structure musicale, et que l'expertise musicale facilite l'extraction des deux structures linguistiques et musicales. Finalement, cet article montre que l'approche électrophysiologique semble être plus sensible que l'approche comportementale.

Musical expertise and statistical learning of musical and linguistic structures

Daniele Schön and Clément François

Institut de Neurosciences Cognitives de la Méditerranée, Marseille, France

Abstract

Adults and infants can use the statistical properties of syllable sequences to extract words from continuous speech. Here we present a review of a series of electrophysiological studies comparing: 1) learning based on speech sequences versus learning based on sung sequences, 2) extracting linguistic versus musical information from a sung stream, 3) musicians versus nonmusicians. Results show that performance is better with sung compared to spoken material. Moreover, using sung material, the linguistic structure is better learned than the musical structure, and the musical expertise facilitates learning of both linguistic and musical structures. Finally the electrophysiological method seems to be more sensitive than the behavioral one.

Language and music are both highly complex and articulated systems. They involve combining a small number of elements according to rules that allow the generation of unlimited numbers of utterances or musical phrases. Strikingly, most infants acquire their mother tongue in the very first years of their life with disconcerting ease and without formal instructions despite the complexity of the input. This ability to learn new languages continues, though to a lesser extent, during adulthood. Similarly, music learning also takes place in the very first years of life, and most children acquire without apparent effort several musical skills such as tapping along with music and singing. This similarity between speech and music competences acquisition as well as the extensive training period that can be found in musicians let us make several predictions. First, learning of linguistic and musical structures may be similar. Second, learning linguistic structure may be influenced by the musical structure, and vice versa. Third, musical expertise may **transfer** to language learning.

Statistical learning and speech segmentation

In this paper we will focus on one of the first major difficulties that infants as well as second language learners encounter when facing speech perception: the ability to segment speech into separate units, words. This turns out to be a rather challenging process because the speech stream does not systematically carry consistent acoustic cues such as pauses or stresses marking word boundaries. In other words, speech is more like a continuous stream of sounds and the challenge for the learner is to be able to segment it into separate units, words, that will then be mapped onto conceptual representations. According to several publications, an important source that would allow speech segmentation in a rather implicit manner is the statistical structure of the language (Saffran et al. 1996; Kuhl 2004; Rodriguez-Fornells et al. 2009). In general, “syllables that are part of the same word tend to follow one another predictably, whereas syllables that span word boundaries do not” (Saffran et al. 2001). The design typically used in this kind of studies comprises a passive listening phase lasting few minutes. During this phase, participants listen to a synthetic and artificial language built in such a way that the only cues to segment words are conditional probabilities between syllable pairs (high within words and lower across word boundaries). A

two-alternative forced choice test follows: one item is a word from the language while the other is not. Participants have to choose which of the two items sounds as the most familiar.

Implicit learning: specific or general?

8-months old infants can learn a simple artificial language in 2 minutes (Saffran et al., 1996a). Similar results have been reported for new-borns and adults (Teinonen et al., 2009; Aslin et al., 1998; Gervain et al., 2008), pointing to a certain degree of independence from the developmental stage. **Statistical learning** has also been described in non human primates and rats (Saffran et al., 2008; Toro et al., 2005), suggesting that it is not specific to the human species. Statistical learning also seems to be « modality independent » insofar as several results showed efficient learning using visual stimuli (Fiser and Aslin, 2005), but also visual movement or shape sequences (Fiser and Aslin, 2002) as well as sequences of tactile stimulations (Conway and Christiansen, 2005). Finally, learning takes also place when replacing syllables with tones (Saffran et al., 1999). On a similar vein, Tillmann and collaborators (2004) replicated previous findings using complex nonverbal auditory material (sounds of different musical instruments). Moreover, they showed that listeners are sensitive to statistical regularities despite acoustical characteristics in the material that are supposed to affect grouping. These results reveal that this type of learning is not specific to speech (thus not **modular**), but seems to be a more general domain process.

Prosody and segmentation

In the context of the comparison between music and language processing, it is striking to note some parallels between the **hierarchy of musical structures** and the one of prosodic structures in speech (Nespor & Vogel, 1983). The term prosody comes from the greek and literally indicates the quality of a singing voice. Prosody refers to the patterns of rhythms, stress and intonation in speech. Prosodic variations have different functions and are crucial cues in communicating emotions, disambiguating syntax but also at a segmental level. Metrical cues are very powerful and widespread cues in **word segmentation**: for instance,

lexical stress typically implies increased duration, pitch and loudness of specific syllables in a word, thus creating different perceptual rhythmic patterns (Abercrombie, 1967). These metrical cues constrain the ordering of stressed and unstressed syllables and are at the origin of the perception of strong and weak points similarly to what is found in music (strong and weak beats). When prosodic cues are inserted, even subliminally, within the speech stream, performance improves (Peña et al., 2002; Cunillera et al 2006, 2008; De Diego Balaguer et al., 2007). An interesting case is that of infant-directed (ID) speech, that is this specific intonation that we use when talking to infants. Compared to adult directed (AD) speech, ID speech is characterized by a slower speech rate, larger pitch and contour variations, longer pauses (Fernald 1992). Moreover, ID speech is easier to segment than AD speech both using a stream of nonsense words and utterance-final words (Thiessen et al., 2005; Aslin 2000). If ID speech is a form of speech with exaggerated prosody features, singing comes next.

Better learning of linguistic structures in song

Most of the studies investigating the relationship between music and language have used language and music separately and this is also the case for statistical learning. This is somewhat problematic insofar as the comparison of results issued from different tasks, subjects and types of analyses is not straight forward. With this respect, singing is particularly well-suited to the study of the relation between language and music. The advantage of singing is that both linguistic and musical information are merged into one acoustic signal with two salient dimensions, allowing for a direct comparison within the same experimental material. Moreover, this allows asking subjects to perform a task while manipulating the relation between linguistic and musical dimensions, thus studying the potential interferences from one dimension on the other. Of course, care must be taken because, in song, the phonological and metrical structures of language are strongly influenced by the type of melody that is used to sing. For instance, vowels sung in a very high register have the tendency to be deformed and less recognizable (Scotto di Carlo, 1994). We recently run a series of studies using a sung language. While the test following the learning always used spoken words, we compared learning with a spoken language to learning with a sung language, wherein each syllable was always associated to a unique pitch (see Figure 1).

While 7 minutes of listening were not enough to learn a spoken language, they were sufficient to learn the same language when it was sung. These results point to the fact that redundancy in statistical musical and linguistic structures benefits the learning process even if more general factors linked to motivation and arousal could not be excluded. Indeed the sung version might have well been more « arousing » than the spoken version. To gain a finer understanding into the mechanisms responsible for the differences between spoken and sung languages we ran a third experiment using a sung language wherein linguistic and musical statistical structures mismatched. Performances laid in between the spoken version and the original sung version obtained in the two first experiments. These results point to a beneficial effect of both structural (“statistical”) and motivational properties of music in the very first step of language acquisition.

Comparing learning of linguistic and musical structures

However, because the test we used only contained spoken items, we could not know whether learning of the musical structure took place. At this aim and in order to better understand how music information facilitated speech segmentation, we designed a similar experiment that tested both learning of musical and linguistic structures contained in the sung language, also recording event-related potentials during the tests (ERPs). Thus, two tests followed the learning phase, one using spoken items at a fixed pitch, and another using piano tone sequences. First we run a study with a random sample (François & Schön, 2010), then we compared a group of musicians and nonmusicians (François & Schön, in press). At the behavioral level, results showed that linguistic structures were better learned than musical ones (though significance is reached only when pooling the data across the 2 studies). Comparison of performance with chance level (0.5) showed that while the participants' level of performance was above chance in the test using spoken items, it was not in the test using tones. Indeed, as can be seen on Figure 2, the distribution of the performances in the musical test were highly heterogeneous across participants and only half of the subjects exhibited performances higher than chance level. Several non-exclusive explanations can account for this difference.

First, linguistic storage and/or retrieval of “lexical” items might have been more efficient or less sensitive to interferences caused by part-words/tone part-words. Schendel and Palmer (2007) recently showed that, overall, verbal items were better recalled than musical items. Moreover, they found a larger interference effect on musical recall than on verbal recall. Researches investigating memory representation of song have also shown an advantage for words over melodies (Crowder et al., 1990; Hébert & Peretz, 2001; Morrongiello & Roes, 1990; Peretz et al., 2004; Serafine et al., 1984). All together, these results seem to suggest that linguistic information may be more resistant to interferences than musical information.

The second explanation is more stimulus related. Indeed, while the statistical structure of melodic and phonologic information was the same in our material, other important factors may come into play that make melodic segmentation more difficult than word segmentation. For instance, phoneme discrimination is mostly influenced by transient information in the spectre (eg “py” vs “gy”), while pitch discrimination is influenced by more stationary information (eg. fundamental frequency). Moreover, phoneme transitions in our language were probably sensitive to the implicit phonotactic knowledge humans have in their mother tongue, stating which sound sequences are allowed to occur within the words of a language (Friederici and Wessels, 1993). This is hardly transposable to music, wherein the implicit knowledge affecting music perception concerns tonal relationships between notes (Tillmann et al., 2000). Overall, it is very difficult to control and balance the perceptual saliency of the linguistic and musical structures and care must be taken in comparing and interpreting results across dimensions.

Nonetheless, when pooling the data across the 2 studies, we observe a significant positive correlation between performances in music and linguistic tests: the higher the performance in the linguistic test, the higher in the music test (Figure 3). This fits quite well with results showing a correlation between linguistic and musical abilities. For instance, some studies reported a positive correlation between phonological awareness and music perception (Butzlaff, 2000; Anvari et al, 2002; Foxtan et al., 2003), suggesting that these competences may share some of the same auditory mechanisms. Interestingly, phonological awareness requires to be able to segment speech into its component sounds, and to extract

the phonological invariants. Similarly, the perception of music also requires the listener to be able to segment the stream of tones into relevant units, and to be able to recognize these units when played with different timbres tempos, keys and styles.

Statistical learning: comparing ERPs and behavior

Overall, behavioral data alone could not clear cut with respect to whether participants could or could not segment the musical structure. Indeed, the lack of significance might simply be related to a lack in sensitivity of the testing procedure and of the dependent variable. ERPs, seem to be more sensitive to the subtle mechanisms underlying implicit learning. Indeed, McLaughlin and colleagues (2004), in a study focusing on second language learning, showed that ERPs can be more sensitive in reflecting implicit learning than do explicit categorical judgments and that behavioral assessment might underestimate the extent to which learning has taken place. While behavior reflects the combined effects of several processing stages, ERPs, thanks to their continuous time resolution, can (sometimes) show differences in brain activity even when this is not backed by an overt behavioral response. Several examples of ERP-behavior dissociations have been reported pointing the greater ERP sensitivity, for instance in the field of subliminal perception (Sergent et al., 2005) or learning (Tremblay et al., 1998).

In our studies, ERPs were also more sensitive than behavior in that they revealed a significant difference between familiar and unfamiliar items in both linguistic and musical test. Very few subjects had an inverse effect (*i.e.* smaller negativity for unfamiliar than familiar items). Indeed, even subjects with a performance that could be difficultly interpreted as higher than chance (e.g. subjects with 55% of correct responses), did show the typical familiarity effect in the ERPs. These results seem to point to a greater sensitivity of the ERP data compared to the behavioral data in implicit learning designs.

Lexicality, Familiarity and the N400

Electrophysiological data of the linguistic test showed a negativity at fronto-central sites in the 450-800 ms latency band larger for part-words than for words (see Figure 4). In order to interpret this ERP difference it is important to clarify the respective “linguistic” status of the two conditions. What researchers in field of statistical learning typically call “words” are in fact nonsense words that are repeated in a random order for N minutes during the learning phase. Therefore, “words” do not have an entry in the lexicon. However, because they are repeated so many times (typically > 100 each), they “pop out” of the continuous stream of syllables, thanks to statistical learning or associative mechanisms. Thus, to a certain extent, “words” become familiar, and as such they can be distinguished from “part-words” acquiring a sort of proto-lexical status (Fernandes et al., 2009; Rodriguez-Fornells et al., 2009). Indeed, in an experiment using X words, if, during the learning phase, “words” are heard N times each, “part-words” are heard $N/(X-1)$ times (e.g., 216 and 44 times, respectively). This is because “part-words” are built by merging the end of a word with the beginning of another (and each word can be followed by X-1 possible words). Therefore, the linguistic status of the items used in the test is that of pseudo-words, more or less familiar (words and part-words, respectively).

The most common use of pseudo-words in the ERP literature is found in studies concerned with lexicality. These studies most often compare a pseudo-word to a word condition. For instance, O'Rourke and Holcomb (2002) used a lexical decision task and found a larger negativity for pseudo-words than for words, most prominent at anterior sites. Moreover, the negative component to pseudo-words was peaking around 600 ms, that is 200 ms later than the classical N400 effect described for words. The authors claimed that pseudo-word processing may undergo an additional time consuming step, consisting of a top-down checking process. Interestingly, in our studies, the negative component to both familiar and unfamiliar items was also peaking around 600 ms.

Turning to familiarity effects, previous ERP studies have already shown that the amplitude of the N400 is sensitive to familiarity. Indeed, low frequency words elicited a larger negativity peaking around 400 ms compared to high frequency words (Van Petten & Kutas, 1990; Young & Rugg, 1992). Thus, the N400 amplitude seems to be sensitive to the ease of accessing the information from long-term memory (Federmeier & Kutas, 2000). In

statistical learning experiments, during the passive listening phase, only sufficiently reinforced items may “survive” concurrent interference and end up being stored in a long lasting manner (Perruchet & Vinter, 2002). Therefore, the observed differences over the negative component between words (familiar) and pseudo-words (unfamiliar), might reflect the difficulty of accessing and/or retrieving item representations.

Electrophysiological data of the musical test revealed a similar ERP component whose amplitude was larger for unfamiliar than for familiar items: a fronto-central negative component peaking around 900 ms after stimulus onset.

In the studies described, insofar as the sung language was built by concatenating trisyllabic words, participants are (implicitly) familiarized with the three-pitch melodies corresponding to the melodic contour of the trisyllabic words. Therefore the same issues on lexical representation and familiarity described for linguistic items also apply when testing with musical sequences. Interestingly, we found a fronto-central negative component larger for unfamiliar than familiar musical items, with a topographic distribution quite similar to the one observed in the linguistic test for the same contrast (unfamiliar – familiar, see Figure 4). However, although sensitive to the same factor (familiarity) and with a similar topography, this negative component is peaking around 600 ms to linguistic items and 900 ms to musical items possibly due to the fact that participants found the musical test harder than the linguistic test. Thus, they probably needed more time to “search” whether an item was familiar or not in the musical test than in the linguistic test. Although less commonly described in the literature, late negative components have been observed in response to musical stimuli, for instance following unexpected harmonies (Steinbeis & Koelsch, 2008) but also following musical excerpts that do not match the concept of a previous verbal context (Daltrozzo & Schön, 2009) or at the familiarity emergence point of a melody (Daltrozzo et al., 2009).

Musical expertise affects behavioral and brain indices.

Comparing experts to non experts in a given domain is an elegant way to study the effect of extensive training on brain plasticity. The musicians' brain is considered since many

years as a model of plasticity (Münte et al., 2002). Thus, comparing musicians to nonmusicians allows studying the effects of extensive audio-motor training on the functional and structural organization of the brain (even if causation is only showed using a longitudinal design). At a perceptual level, it has been shown that musicians outperform non musicians on a variety of music related task: musicians have a lower frequency threshold (Spiegel & Watson, 1984; Kishon Rabin et al., 2001; Micheyl et al., 2006) and have a finer rhythm processing than nonmusicians (Jones et al., 1995; Yee et al., 1994). Compared to nonmusicians, musicians also show enhanced cortical attentive and pre-attentive processing of linguistic and musical features as reflected by larger amplitude and/or shorter latencies of many ERP components such as the N1, P2, MMN, P300 (Shahin et al., 2003; Pantev et al., 1998; Koelsch et al., 1999; Van Zuijen et al., 2005). Also in our study on statistical learning the N1 component was larger for musicians than nonmusicians in both the linguistic and the musical tasks. Interestingly recent results support the view that increased auditory evoked potentials in musicians (N1-P2) reflect an enlarged neuronal representation for specific sound features of these tones rather than selective attention biases (Baumann et al. 2008). Even at the sub-cortical level, musicians show more robust encoding of linguistic and musical features as reflected by earlier and larger brainstem responses compared to non musicians (Kraus & Chandrasekaran, 2010). Overall, these differences can be interpreted as reflecting a greater efficiency of musicians' auditory system in processing sound features and can be accompanied by morphological differences, showing that musicians have a larger grey matter concentration in the auditory cortex (Bermudez & Zatorre, 2005), an increased gray matter density and volume in the left inferior frontal gyrus, Broca's area (Slumming et al., 2002) and a larger planum temporale (Schlaug et al. 1995; Keenan et al. 2001). The fact that musicians perceive some sound features more accurately than non musicians is not so surprising. After all, they spend hours and hours of their life focusing on sounds and the way they are generated, paying particular attention to pitch, timbre, duration and timing. It would be rather surprising and even deceiving if this did not affect the way they hear, and therefore their brain. What seems to us less evident is the fact that such an intensive musical practice seems to affect also non musical abilities. Both adults and children musicians have a better performance than matched controls when asked to detect fine contour modifications in the prosody of an utterance (Magne et al. 2006; Schön et al. 2004). Adult musicians have better

performances and larger ERP component to metric incongruities at the end of an utterance (Marie et al., 2010). There is also evidence for a possible correlation between musical and linguistic aptitudes in children (Anvari et al. 2002; Milovanov et al. 2008, 2009) and adults (Foxton et al., 2003; Slevc and Miyake 2006) as well as a benefit of music training on linguistic skills (Butzlaff, 2000; Gaab et al., 2005; Overly et al., 2003 Forgeard et al., 2008; Moreno et al., 2009; Parbery-Clark et al., 2009; Tallal & Gaab, 2006).

Better segmentation of the melodic structure in musicians

When using a sung language in artificial language learning, one needs to choose a series of pitches that are mapped onto syllables. The choice of pitches is an important one, insofar as it determines the tonal or atonal character of the stream. Although one may guess that a tonal structure will benefit more language learning than an atonal one, further work is needed to understand how and to what extent the tonality of the musical structure influences learning of the phonological one. In the experiments described here we used a musical structure with a rather strong tonal center (ten notes out of eleven were in C major). Thus, it is possible that participants, during the test following learning, keep in memory a representation of a tonal center influencing the processing of the musical items rather early on. In fact, once a tonal center is established, each individual pitch can be processed and categorized relative to the tonal center, and this does not require having absolute pitch. Indeed, musicians, in the musical test, showed a significant effect of familiarity 200 ms after the first tone onset and before the beginning of the second tone on the P2 component (larger to unfamiliar than familiar items), possibly due to a more difficult categorization of the unfamiliar than familiar items (Liebenthal et al. 2010). Thus, although musicians' performance was not good, they were sensitive to the tonal structure of the language that in turn influenced their perception of the first tone of the items in the test.

This benefit of musical expertise in segmenting the musical structure was also visible on later ERP components. For instance, compared to familiar items, unfamiliar items yield a larger mismatch negativity (MMN) in response to the second tone of the melody. Overall, both familiarity effects on the P2 and MMN-like components point to the fact that musicians did

learn the musical structure better than nonmusicians and this in turn affected the way they processed the musical items presented during the test. This is in line with previous findings showing that implicit learning of 12-tone music is influenced by expertise. Indeed, only participants with routine exposure to atonal music do (implicitly) perceive the distinction between different types of transforms (Dienes & Longuet-Higgins, 2004).

Musical training facilitates implicit learning of both linguistic and musical structures

The most striking finding of this series of experiments reviewed here is that musicians seem to have more “robust” representations of both musical and linguistic structures shaped during the listening phase, compared to nonmusicians. Indeed, musicians showed a familiarity N400-like effect larger than nonmusicians, in both dimensions.

The way we interpret these findings is that musical expertise may facilitate regularity extractions and sequence learning in general (Janata & Grafton, 2003). For instance, it is known that musicians can organize a sound sequence according to a number regularity, implicitly distinguishing segments containing four tones from the segments containing five tones. By contrast such a perceptual organization of sound in terms of number is less relevant for nonmusicians (Van Zuijen et al. 2005). Moreover, recent findings showed that deaf children with cochlear implants are impaired in visual sequence learning, suggesting that a period of auditory deprivation may have a major impact on cognitive processes that are not specific to the auditory modality (Conway et al. 2010). Therefore, sound seems to provide a cognitive scaffolding for the development of serial-order behavior: whether sound processing is impaired or whether it is extensively practiced have opposite effects. Interestingly, Sluming and colleagues reported an increased gray matter density and volume in the left inferior frontal gyrus, Broca’s area. Moreover, neuroplastic development throughout musicians’ life seems to promote the retention of cortical tissue (Sluming et al 2002). This region is known to be involved in on-line speech stream segmentation as well as in music perception (Cunillera et al., 2009; McNealy et al. 2006; Tillmann et al. 2003). Overall our results support an “auditory scaffolding hypothesis” (Conway et al 2009) as we presently

show that increased exposure to sounds leads to a benefit for implicit learning, putatively via anatomical and/or functional modifications going beyond the auditory regions.

Future Questions

Learning Process versus Learning Outcome

Most studies on word segmentation and language learning rely on behavioral measures. These are typically obtained using the head-turning procedure in infants and a familiarity two-alternative forced choice procedure in adults. These procedures are used after the listening phase, wherein learning takes place. Thus, these procedures test rather the result of learning than the learning process itself. Also the data presented here comparing language and music segmentation in song or musicians and nonmusicians segmenting skills suffer from the same limitation. Recently, in the field of implicit learning, researchers realized the importance of better defining the status of the representations derived from the learning processes and promote the use of combined both subjective behavioral measures (guessing criterion and zero correlation criterion) and EEG or fMRI measures (Seth et al., 2008). Studies measuring EEG during learning are promising in this sense, pointing to several possible electrophysiological indices, ERPs or oscillations that may reflect how learning takes place in time (De Diego Balaguer et al., 2011, Rodriguez-Fornells et al., 2009). The challenge for the coming years is to access to a temporally detailed learning curve. Ideally one would want to know how each single word presented during the learning has been processed. As some studies seem to point out, learning curves may differ across individuals (Abla et al., 2008). This might be crucial in studying pathology, also because similar behavioral results (null effects) may rely upon different learning curves.

Extending music and language comparisons

The literature comparing music and language processing is mostly focused on the perceptual side, the production side having received little attention (also in studies on language learning). While the statistical learning framework is a brilliant way of studying language

learning in a laboratory, it lacks by construction the rich context and complex behaviors that are typically found in language acquisition. For instance the natural context typically implies a word to world mapping: once a unit/word is identified, it is mapped onto one or several concepts. Moreover, learning acquisition outside the laboratory relies on coupling listening and imitation. Thus, further efforts need to be done to elaborate new experimental paradigms comparing language and music using a richer learning context as well as linguistic and musical unit production. Also, while the direct comparison between music and language is certainly a privileged one it should also be extended to other domains and modalities (eg visual sequence learning, number sequences) in order to better understand to what extent the underlying mechanisms are domain general and provide support to the auditory scaffolding hypothesis (Conway et al., 2009).

Music Training and Brain Dynamics

The study of the effects of music training on language learning poses also a challenge in terms of brain plasticity. Language learning probably relies on a distributed network involving the superior temporal gyrus (phonological representation), the dorsal stream and premotor cortex for mapping sounds to movement (phonological-articulatory trace) and the medial temporal lobe (lexical trace, Rodriguez-Fornells et al., 2009; Hickok & Poeppel, 2007). Thus, changes due to music training may take place at each of these different processing steps. Moreover, it is highly likely that music training influences the connectivity between these regions (Oechslin et al., 2010). Thus the future challenge is to track to what extent music training influences the temporal evolution of the brain dynamics of this complicated network as learning takes place.

Key concepts

Modularity

A module is a computational system dedicated to the processing of a specific input-output function. Its most important feature is informational encapsulation (compute its own function independently of any other information). Local systems can be modular, while global systems are not modular because informational encapsulation does not hold. If a linguistic function is thought to be modular, then it must be independent of any other non linguistic information.

Word Segmentation

In natural speech, words are not systematically separated by silences. “How are you?” is an example of three words pronounced without any interruption. In order to understand an utterance the listener must be able to segment, that is “separate” the different words in order to retrieve their meaning in the lexicon and build a syntactic structure.

Statistical learning

It has been shown that, in artificial situations, learning can implicitly take place on the sole bases of the statistical properties of the language. The statistics used are the conditional probabilities between syllables. The higher the probability that syllable Y follows syllable X, the more likely they form a word [XY]. In more ecological conditions, statistics are only one cue for speech segmentation, the others being prosodic, lexical, syntactic and pragmatic (contextual) cues.

Transfer of learning

One can speak of transfer when learning a specific operation A has an impact on another operation B. It requires the operation B not be encapsulated. Thus if there is transfer from A to B, B cannot be modular. When A and B are quite similar one speaks of close transfer, when they are quite different (from a cognitive point of view) one speaks of far transfer.

Music and hierarchical temporal structures

In music, tone durations most often have simple ratios (2:1, 3:1, 4:1, 2:3). Depending upon durations (as well as pitch, timbre and tonal context), tones are grouped together and perceived as rhythmical patterns rather than as a series of separated sounds. This gives rise to the emergence of a further time-frame, meter, that allows us to synchronize with music in a succession of strong and weak beats in a rather hierarchical fashion.

REFERENCES:

- Abercrombie, D. (1967). Elements of general phonetics. Edinburgh University Press.
- Abla, D., Katahira, K., Abla, O. K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neurosciences*. 20, 952.
- Anvari, S. H., Trainor, L. J., Woodside, J., Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*. 83 (2), 111.
- Aslin, R. N., Saffran, J. R., Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*. 9, 321.
- Aslin, R. N. (2000, July). Interpretation of infant listening times using the headturn preference technique. Paper presented at the International Conference on Infant Studies, Brighton, England.
- Baumann, S., Meyer, M., Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*. 20 (12), 2238.

Bermudez, P., Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*. 1060, 395.

Butzlaff, R. (2000). Can music be used to teach reading? *J Aesthet Educ*. 34, 167.

Conway, C. M., Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 31(1), 24.

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., Henning, S. C. (2010). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*. 11 (1), 69.

Conway, CM, Pisoni, DB, Kronenberg, WG. (2009). The importance of sound for cognitive sequencing abilities: the auditory scaffolding hypothesis. *Curr Dir Psychol Sci*. 18 (5), 275.

Crowder, R. G., Serafine, M. L., Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory and Cognition*. 18 (5), 469.

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Research*. 1123 (1), 168.

Cunillera, T., Gomila, A., Rodríguez-Fornells, A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neuroscience*. 18, 9. Cunillera, T., Càmarà, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., Pujol, J., Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*. 48 (3), 541.

Daltrozzo, J., Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*. 21 (10), 1882.

Daltrozzo, J., Tillmann, B., Platel, H., Schön, D. (2010). Temporal aspects of the feeling of familiarity for music and the emergence of conceptual processing. *Journal of Cognitive Neuroscience*. 22 (8), 1754.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., Bachoud-Lévi, A. C. (2007). Different Neurophysiological Mechanisms Underlying Word and Rule Extraction from Speech. *PLoS ONE*. 2 (11), e1175.

De Diego Balaguer, R., Fuentemilla, L., Rodriguez-Fornells, A. (2011). Brain Dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.2011.21636

Dienes, Z., Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*. 28, 531.

Federmeier, K. D., Kutas, M. (2000). It's about time. *Brain & Language*. 71(1), 62.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, England: Oxford University Press.

Fernandes, T., Kolinsky, R., Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition*. 112 (3), 349.

Fiser, J., Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 28(3), 458.

Fiser, J., Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General*. 134 (4), 521.

Forgeard, M., Winner, E., Norton, A., Schlaug, G. (2008). Practicing a Musical Instrument in Childhood is Associated with Enhanced Verbal Ability and Nonverbal Reasoning. *PlosONE*. 3(10), e3566.

Foxton, J. M., Talcott, J.B., Witton, C., Brace, H., McIntyre, F., Griffiths, T. D. (2003). Reading skills are related to global, but not local, acoustic pattern perception. *Brief Communication Nature Neuroscience*. 6, 343.

Francois, C., Schön, D. (2010). Learning of musical and linguistic structures: comparing event-related potentials and behavior. *Neuroreport*. 21 (14), 928.

Francois, C., Schön, D. (2011). Musical expertise boosts Implicit Learning of Both Musical and Linguistic Structures. *Cerebral Cortex*. In press.

Friederici, A. D., Wessels, J. M. (1993) Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception & Psychophysics*. 54 (3), 287.

Gaab, N., Tallal, P., Kim, H., (2005). Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. *Ann N Y Acad Sci*, 1060, 82.

Gervain, J, Macagno, F, Cogoi, S, Peña, M, Mehler, J. (2008). The neonate brain detects speech structure. *Proc Natl Acad Sci U S A*. 105, 14222.

Hébert, S., Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain and Cognition*. 46 (1-2), 169.

Hickok, G., Poeppel, D. (2007). The cortical organization of speech processing. *Nature Review Neuroscience*. 8, 393.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*. 7 (2), 153.

Janata, P., Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat Neurosci*. 6, 682.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology – Human Perception and Performance*. 21, 293.

Keenan, J. P., Thangaraj, V., Halpern, A. R., Schlaug, G. (2001). Absolute pitch and planum temporale. *Neuroimage*. 14 (6), 1402.

Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y. (2001). Pitch discrimination: are professional musicians better than non-musicians? *Journal of Basic and Clinical Physiology and Pharmacology*. 12 (2 Suppl), 125.

Koelsch, S., Schröger, E., Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*. 10 (6), 1309.

Kraus, N., Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Review Neuroscience*. 11, 599.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Review Neuroscience*. 5, 831.

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., Binder, J. R. (2010). Specialization along the Left Superior Temporal Sulcus for Auditory Categorization. *Cerebral Cortex*. [Epub ahead of print].

McLaughlin, J., Osterhout, L., Kim, A. (2004). Neural correlates of second -language word learning: minimal instruction produces rapid change. *Nature Neuroscience*. 7 (7), 703.

McNealy, K., Mazziota, J. C., Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *Journal of Neuroscience*. 26, 7629.

Magne, C., Schön, D., Besson, M. (2006). Musician children detect pitch violations in both music and language better than non musician children. *Journal of Cognitive Neuroscience*. 18, 199.

Marie, C., Magne, C., Besson, M. (2011). Musicians and the Metric Structure of Words. *Journal of Cognitive Neuroscience*. 23 (2), 294.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research*. 1194, 81.

Milovanov, R., Huotilainen, M., Esquef, P. A., Alku, P., Välimäki, V., Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience letters*. 460 (2), 161.

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. *Cerebral Cortex*. 19, 712.

Morrongiello, B. A., Roes, C. L. (1990) Children's memory for new songs: integration or independent storage of words and tunes? *Journal of Experimental Child Psychology*. 50 (1), 25.

Münste, T. F., Altenmüller, E., Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Review Neuro*. 3 (6), 473.

Nespor, M., Vogel, I. (1983). Prosodic structure above the word. In: A. Cutler & D. R. Ladd (Eds), *Prosody: Models and Measurements*. Berlin, Germany: Springer-Verlag.

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.* 8, 3.

O'Rourke, T. B., Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of spoken word processing. *Biological Psychology.* 60, 121.

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Ann N Y Acad Sci.* 999, 497.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature.* 392 (6678), 811.

Parbery-Clark, A., Skoe, E., Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience.* 29 (45), 14100.

Peña, M., Bonatti, L., Nespors, M., Mehler J. (2002). Signal-Driven Computations in Speech Processing, *Science.* 298, 604.

Peretz, I., Radeau, M., Arguin, M. (2004) Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs. *Memory and Cognition.* 32 (1), 142.

Perruchet. P., Vinter. A. (2002) The self-organizing consciousness. *Behavioral and Brain Sciences.* 25 (3), 297.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., De Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Phil. Trans. R. Soc. B.* 364, 3711.

Saffran, J. R., Newport, E. L., Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal Of Memory And Language.* 35, 606.

Saffran, J. R., Aslin, R. N., Newport, E. L. (1996a). Statistical learning by 8-month old infants. *Science.* 274, 1926.

Saffran, J. R., Senghas, A., Trueswell, J. C. (2001). The acquisition of language by children. *Proc Natl Acad Sci U S A.* 98,12874.

Saffran, J.R. (2003). Statistical language learning: Mechanisms and Constraints. *Current Directions in Psychological Science*. 12 (4), 110.

Saffran, J. R., Hauser, M., Seibel, R. L., Kapfhamer, J., Tsao, F., Cushman, F. (2008). Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition*. 107, 479.

Schendel, Z. A., Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory and Cognition*. 35 (4), 640.

Schlaug, G., Jäncke, L., Huang, Y., Staiger, J. F., Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*. 33 (8), 1047.

Schön, D., Magne, C., Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*. 41 (3), 341.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., Kolinsky, R. (2008). Song as an aid for language acquisition. *Cognition*. 106 (2), 975.

Scotto di Carlo, N. (1994). L'Intelligibilité de la voix chantée. *Médecine des Arts*. 10, 2.

Serafine, M. L., Crowder, R. G., Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*. 16 (3), 285.

Sergent, C. Baillet, S., Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*. 8 (10), 1391.

Shahin, A., Bosnyak, D. J., Trainor, L. J., Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*. 23 (13), 5545.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *TICS*. 12 (8), 314.

Slevc, L. R., Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychological Science*. 17 (8), 675.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., Roberts, N. (2002). Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca's Area in Male Symphony Orchestra Musicians. *Neuroimage*. 17 (3): 1613-1622.

Steinbeis, N., Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE*. 3, e2226.

Spiegel, M. F., Watson, C. S. (1984). Performance on frequency discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*. 76 (6), 1690.

Tallal, P., Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neuroscience*. 29(7), 382.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*. 13, 10.

Thiessen, E. D., Hill, E., Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*. 7, 53.

Tillmann, B., Bharucha, J. J., Bigand, E. (2000) Implicit learning of tonality: a self-organizing approach. *Psychological Review*. 107 (4), 885.

Tillmann, B., Janata, P., Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*. 16 (2), 145.

Tillmann, B., McAdams, S. (2004). Implicit Learning of Musical Timbre Sequences: Statistical Regularities Confronted With Acoustical (Dis)Similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 30, 1131.

Toro, J. M., Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*. 67 (5), 867.

Tremblay, K., Kraus, N., McGee, T. (1998) The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport*. 9 (16), 3557.

Van Petten, C., Kutas, M. (1990) Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*. 18(4), 380.

Van Zuijen TL, Sussman E, Winkler I, Näätänen R, Tervaniemi M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity: a mismatch negativity study comparing musicians and non-musicians. *Brain Research, Cognitive Brain Research*. 23 (2-3): 270-276.

Young, M. P., Rugg, M. D. (1992). Word frequency and multiple repetition as determinants of the modulation of event-related-potentials in a semantic classification task. *Psychophysiology*. 29 (6), 664.

Yee, W., Holleran, S., Jones, M. R. (1994). Sensitivity to event timing in regular and irregular sequences – influences of musical skill. *Perception and Psychophysics*. 56, 461.



Figure 1: Illustration of the experimental design used in our 3 experiments: a learning phase is administered to participants and is followed by one (Schön et al., 2008) or two tests (François & Schön, 2010; In press) depending on the experiment. The linguistic test uses spoken items while the musical test uses piano tones.

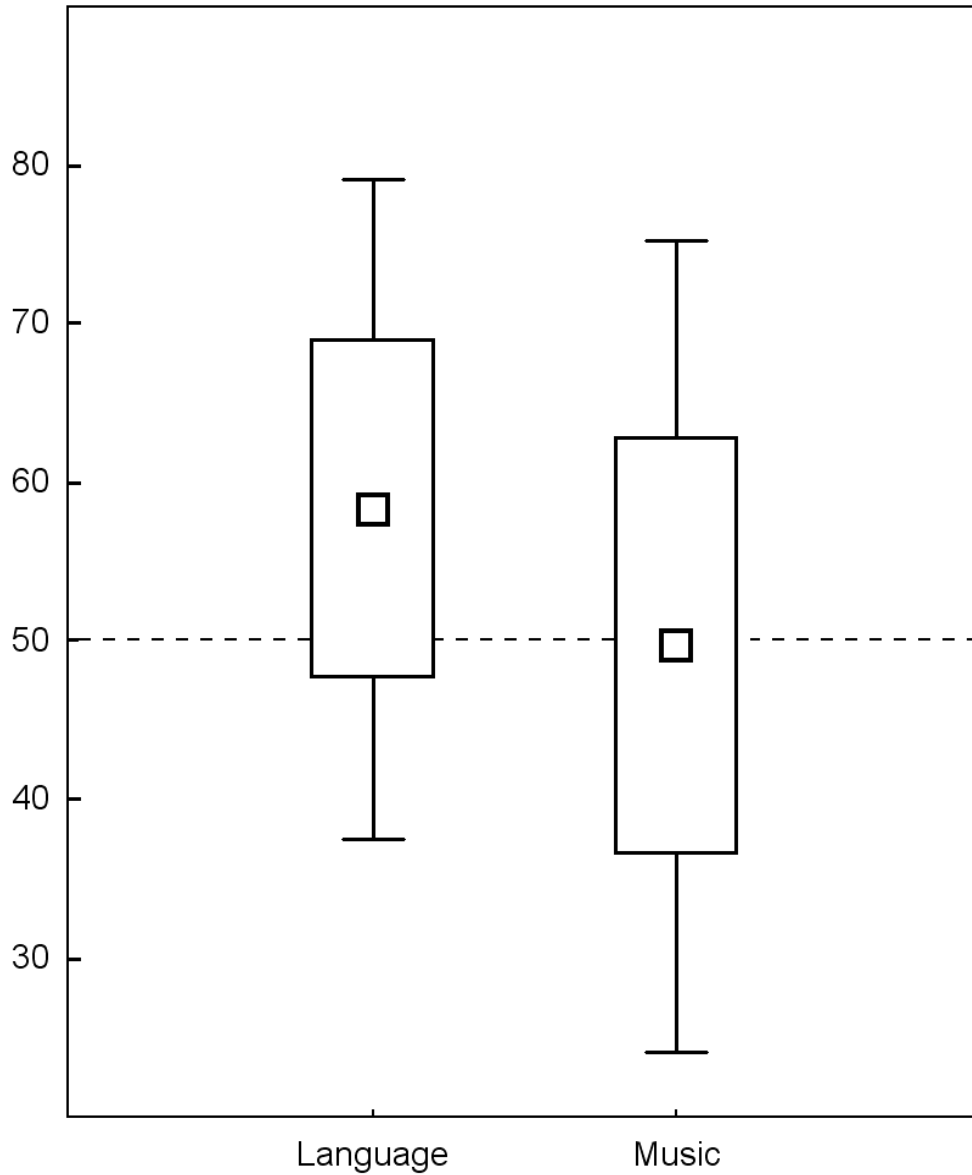


Figure 2: Percentage of correct responses: box plot of performances in the linguistic (left) and musical tests (right). Data from 2 experiments (Francois and Schön, 2010 and in press), are pooled together leading to N = 50 subjects (dashed line = 50%, chance level).

Correlation between performances in the musical and linguistic tests

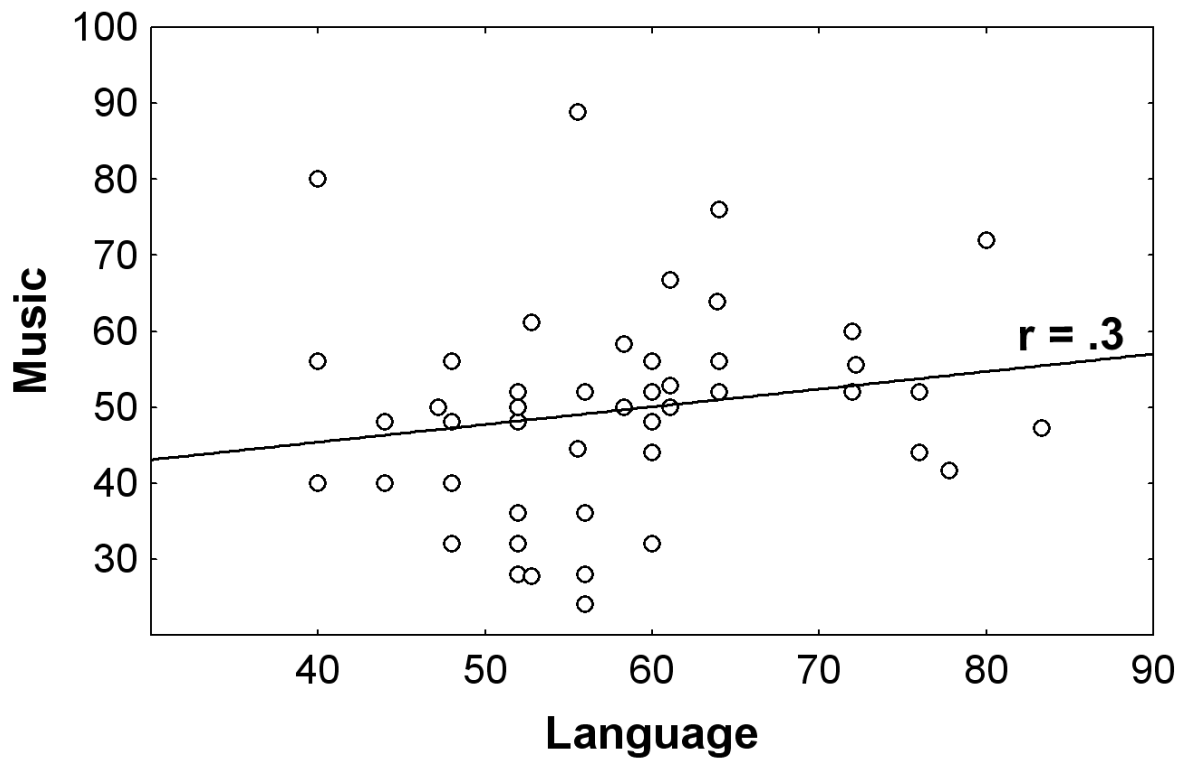


Figure 3: between performances in the linguistic and musical tests for 50 participants (Francois and Schön, 2010 and in press). The correlation is significant ($p = .05$), that is the higher participants performance in the musical test, the higher are performances.

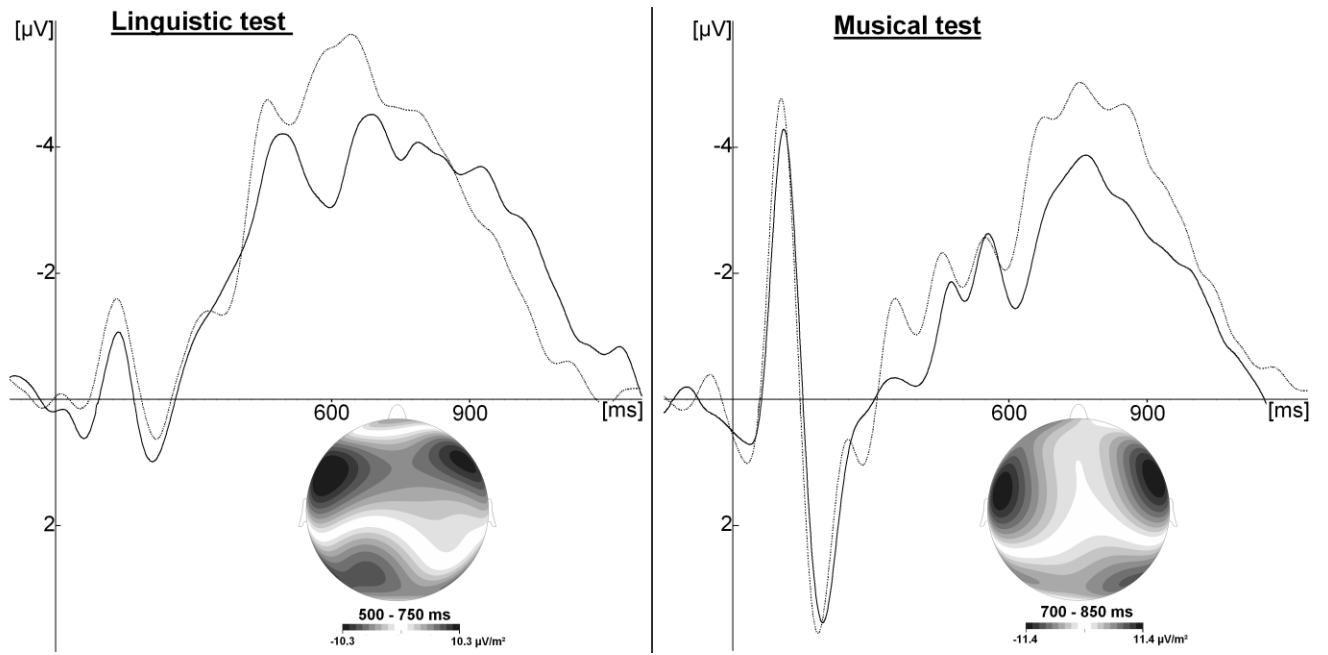


Figure 4: Grand average across 2 experiments (François & Schön, 2010 and in press; N= 50), for the linguistic (left) and musical test (right) at Cz electrode. Solid lines illustrate ERPs to familiar words/tone-words; dashed lines illustrate ERPs to unfamiliar part-words/tone-part-words. Current Source Density Maps illustrate the topographic distribution of the familiarity effect in the significant time windows.

Etude 5

Making sense of continuous speech: electrophysiological correlates reveal faster word segmentation in musicians than in nonmusicians.

Clément François, Florent Jaillet, Sylvain Takerkart et Daniele Schön

Article en préparation

Making sense of continuous speech: electrophysiological correlates reveal faster word segmentation in musicians than in nonmusicians.

Clément **François**, Florent **Jaillet**, Sylvain **Takerkart** & Daniele **Schön**

Institut de Neurosciences Cognitives de la Méditerranée CNRS and Aix-Marseille
Universités, Marseille, France.

31, chemin Joseph Aiguier

13402 Marseille, France

corresponding author:

Clement **Francois**

Email: clement.francois@incm.cnrs-mrs.fr

Tel: +33 4 91 16 41 13; fax: + 33 4 91 16 44 98;

INTRODUCTION

Performing music at a professional level implies a tremendous amount of daily instrumental practice and implies a mastery in handling frequency, timbre and duration variations in the auditory input. Thus, comparing musicians to nonmusicians allows studying the effects of intensive multimodal training on brain plasticity. Psychophysical studies have shown that musicians have lower perceptual thresholds than nonmusicians for frequency and temporal changes (Kishon Rabin et al. 2001; Micheyl et al., 2006; Ehrle & Samson, 2005). These differences might be underpinned by functional and structural differences in the auditory neural circuitry. Indeed, it is now well established that musical practice does induce brain changes as reflected by cortical and sub-cortical electrophysiological responses. At the brainstem level, musicians show more robust encoding of auditory stimuli (Wong et al. 2007; Musacchia et al. 2007; Parbery-Clark et al., 2009; Bidelman & Krishner, 2010; Kraus & Chandrasekaran, 2010 for a review). At the cortical level, musicians show larger N1 and P2 event-related potentials than nonmusicians when listening to synthetic or instrumental sounds (Shahin et al. 2003, 2005). Musicians are also more sensitive to sound spectral complexity (or richness): for instance musicians show larger N1m to piano sounds than to pure tones, while nonmusicians are not sensitive to this contrast (Pantev et al. 1998). Interestingly, it has also been shown by using passive listening paradigms that compared to non musicians, musicians have larger Mismatch Negativity to unfrequent deviant chords introduced in a stream of repeated standard chord (Koelsch et al. 1999; Brattico et al. 2009) as well as when a sound is omitted in the stream (Russeler et al. 2001). These differences can be interpreted as reflecting a greater efficiency of musicians' auditory system in processing sound features and can be accompanied by morphological differences, showing that musicians compared to nonmusicians have a different grey matter concentration in the auditory cortex (Bermudez and Zatorre 2005) and a larger planum temporale (Schlaug et al. 1995; Keenan et al. 2001).

If these music related benefits were primarily observed in music related tasks, some studies have

also shown that musical practice extends to the processing of speech sounds. For instance, both adults and children musicians better detect fine contour modifications in the prosody of an utterance than matched controls (Magne et al. 2006; Schön et al. 2004). Recent studies have shown that musical practice also refines the sensitivity to durational changes in both speech sounds and utterances (Marie et al., 2011; Chobert et al., submitted). These findings are supported by other studies showing a correlation between musical and linguistic aptitudes in children (Anvari et al. 2002; Slevc and Miyake 2006; Milovanov et al. 2008, 2009). Taken together these results led us to focus on a rather high cognitive function: word segmentation, namely the ability to extract words from continuous speech. Because the speech stream does not systematically carry consistent acoustic cues such as pauses or accents at word boundaries, there must be other ways to learn how to segment words. According to several publications, speech might be segmented in a rather implicit manner based on one source of information: the statistical structure of the language (Saffran et al. 1996; Kuhl 2004). In general, “syllables that are part of the same word tend to follow one another predictably, whereas syllables that span word boundaries do not” (Saffran et al. 2001). The role of transitional probabilities in speech segmentation (namely the probability of syllable X given syllable Y) has been elegantly shown by Saffran and collaborators with infants, adults and has been also extended to neonates (Saffran et al. 1996, 1996a; Teinonen et al. 2009; Gervain et al. 2008; Kuhl 2004; Aslin et al. 1998). Throughout this series of studies the authors typically showed that listening to an artificial language without acoustic cues at word boundaries yields correct word recognition in a consecutive test as revealed by participants’ ability to discriminate words that are part of the language from similar words that were not part of the language. This learning paradigm has also been replicated using sung syllables (Schön et al., 2008), non linguistic stimuli such as sounds with different pitches (Saffran et al. 1999) and timbre (Tillmann and McAdams 2004), as well as in the cotton top tamarin and in rats (Saffran et al., 2008; Toro et al. 2005).

We recently found that ERPs recorded during the tests show a fronto central negative component

larger for unfamiliar items than for familiar items. We interpreted this larger component for unfamiliar items as reflecting a difficulty in accessing items representations (François & Schön, 2010). Moreover, in a further experiment (François & Schön, 2011) we collected behavioral and electrophysiological responses in two groups, one with musical experts and one with laymen. We showed that even if musicians barely outperform non musicians in the recognition of newly learned nonsense words, ERP traces revealed a clear increased sensitivity to unfamiliar items in musicians compared to non musicians. This was reflected by a larger N400 familiarity effect over fronto central regions in the musical experts compared to the laymen. However, care must be taken in interpreting the data collected during the tests. While data recorded during the listening phase theoretically reflect, the learning process *per se*, at least at the very beginning, this does not hold true anymore for the data collected during the tests. Indeed, these data are more related to the result of the learning. Furthermore, these data may be more contaminated by decisional processes thus blurring the interpretation. In this context, some studies reported electrophysiological data recorded during the learning phase. The authors typically analysed ERPs during the learning and reported N100 and/or P200 and/or N400 amplitude modulations as a function of exposure to the stream (Cunillera et al., 2006, 2009, De Diego Balaguer et al., 2007). Interestingly, they found different pattern of N400 amplitude modulations (learning curve) depending on the performance in the consecutive test. For instance, two previous studies reported similar inverted U-shape N400 learning curves for the high learners groups. Middle learners did present more linear N400 learning curves whereas Low learners did not show N400 modulation as a function of exposure (Abla et al., 2008; Cunillera et al., 2009). These two converging results seem to give an insight into the interpretation of the N400 component being an index of learning. Moreover, Buiatti and colleagues recently conducted a study in which electrophysiological data were analyzed by using the Steady State Response approach. The Steady State Response is an electrophysiological property of the brain that consists in a EEG power spectrum peak at the frequency of the stimulation. The authors

analysed the electrophysiological data acquired during the learning phase and elegantly showed that the explicit extraction of trisyllabic words in the speech stream was associated by a peak in the EEG power spectrum at the frequency of word occurrence.

In the present study, we used the same learning paradigm as in François and Schön, 2011 but analysing electrophysiological data collected during the learning phase. Participants listened to an artificial sung language (wherein music and language dimensions are highly intertwined) and were then tested with a two-alternative forced-choice task on pairs of words and melodies (familiar vs unfamiliar). The aim of this study was two-folded. Firstly, we wanted to test whether musical expertise can facilitate the word segmentation process per se as measured by comparing the N400 learning curves obtained in the 2 groups. Secondly, by using the frequency tagging approach, we wanted to test the sensitivity of this technique in dissociating the 2 groups. With these aims we compared two groups, one group with formal musical training and one without. At the N400 level, we hypothesized that musicians may present a clear inverted U-shape N400 learning curve and a rather linear one for nonmusicians. Concerning the frequency tagging approach, we hypothesized that musicians should display a larger increase in the steady state power response at the frequency of occurrence of trisyllabic words than non musicians.

Materials and Methods

Participants

Two groups participated in the experiment. 13 professional musicians (mean age 27, 15 right-handed, 10 males, normal hearing, no known neurological problems, more than 12 years of formal musical learning and from 3 to 7 hours of daily practice, 5 of them reported absolute pitch) and 13 nonmusicians (mean age 25, 20 right-handed, 11 males, normal hearing, no known neurological problems, no more than 2 years of formal musical training and no instrument practice). All participants were French native speakers and listened to 5 and a half minutes of a continuous speech

stream resulting from the concatenation of five three-syllable nonsense words (hereafter words) that were repeated in a pseudo-random order. All participants were paid 20 Euros. Informed consent was obtained from all participants, and the data were analyzed anonymously. This study was approved by the CNRS - Mediterranean Institute for Cognitive Neuroscience and was conducted in accordance with national norms and guidelines for the protection of human subjects.

Material

The language consisted of four consonants and three vowels, which were combined into a set of 11 syllables. The average syllable length was 230 ms, standard deviation was 16 ms. These syllables were then combined to give rise to five trisyllabic sung words (gimysy, mimosi, pogysi, pymiso, sipygy). Each of the 11 syllables was associated with a distinct tone (C3, D3, F3, G3, A3, B3, C4, Db4, D4, E4, and F4). Therefore each word was always sung on the same melodic contour (gimysy C3 D3 F3, mimosi E4 Db4 G3, pymiso B3 E4 F4, pogysi D4 C3 G3, sipygy G3 B3 C4). The mean pitch interval within words was not significantly different from the mean interval between words ($p = .4$). Moreover, pitch-contour changes could not be used to segment the stream because they took place 50% of the time only at word boundaries.

Transitional probabilities within words ranged from 0.5 to 1.0. Transitional probabilities across word boundaries ranged from 0.1 to 0.5. While we used a 6 words language stream in our previous experiment (Francois and Schön 2010), in the present study, the language stream was built by a random concatenation of the five words (only constraint: no repetition of the same item twice in a row) and synthesized using Mbrola (<http://tcts.fpms.ac.be/synthesis/mbrola.html>). No acoustic cues were inserted at word boundaries. Each word was repeated 100 times in the stream.

Design and Procedure

During the learning phase, participants were told they would listen to a continuous stream of sung

syllables for several minutes and they were asked to carefully listen to these sounds.

Data Acquisition and Analysis

Participants were comfortably seated in a Faraday booth. The electroencephalogram (EEG) was recorded from 32 scalp electrodes (Biosemi ActiveTwo system, Amsterdam University) located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (International 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F3, F4, C3, C4, P3, P4, P7, P8, Po3, Po4, O1, O2, F7, F8, T7, T8, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). The band-pass was of 0-102.4 Hz and sampling rate 512 Hz.

ERP analyses:

Six participants were discarded due to major artifacts, thus yielding two groups of 10 participants each for the statistical analyses. The data, acquired during the learning phase, were then re-referenced offline to the algebraic average of the left and right mastoids. Raw data containing movement artifacts, or amplifier saturation were excluded. Signal containing ocular artifacts was corrected using ICA decomposition by removing the component containing the blink (Makeig et al. 1996). The EEG was then epoched for each word and a 100 to 800 ms baseline zero-mean normalization was applied using Brain Vision Analyzer software (Brain Products, Munich). Based on the literature and on visual inspection of the data, analyses on the N1 and P2 components were performed on the mean amplitude computed in the 100-170 and 200-300 ms latency bands respectively (ranges were centered on the mean peak latency). Analyses on the N400 component were performed on the mean amplitude in the 350-500 ms latency band. Analyses were performed for 4 blocks of 1.20 minute as well as for 8 blocks of 40 seconds. Repeated measure Analysis of Variance (ANOVA) was used for statistical assessment, using expertise (musicians and nonmusicians) and time bins (4 blocks, 8 blocks) as factors. Topographical distribution of the effects was modeled by 2 additional factors (Hemisphere, left and right and Location, frontal,

central, and parietal) defined as follows: left (AF3, F3, F7) and right (AF4, F4, F8) frontal, left (Fc1, C3, Fc5) and right (Fc2, C4, Fc6) central, and left (Po3, P3, P7) and right (Po4, P4, P8) parietal. All P values reported below were adjusted using the Greenhouse-Geisser correction for nonsphericity, when appropriate, and Fisher tests were used in post hoc comparisons.

Exploratory N400 analyses were also performed using 4 blocks of 20 seconds each, covering the first 80 seconds of learning (over the 300-400 ms range).

Frequency Tagging Analysis:

This analysis was adapted from Buiatti et al., 2009. The preprocessing of the EEG data (re-referencing, artifacts rejection, ocular artifacts correction) was similar to the one applied for ERP analyses. For each electrode, the power spectrum of each block was estimated using EEGLAB “spectopo” function. The choice of the analysis window ensured to obtain one frequency bin centered at each target frequency corresponding to one, two, and three-syllable occurrences. At this aim, an 18 words window length was used (12.528s), thus providing an integer number of one-syllable, two-syllable or three-syllable occurrences within each time window. This processing resulted in an interval of ca 0.08 Hz between two frequency bins. Based on Buiatti processing, the temporal overlap between two successive windows was fixed at 5/6 of their length. The three target frequencies were selected as the inverse of the duration of one syllable ($f = 4.31$ Hz), two syllables ($f = 2.16$ Hz) and three syllables occurrences ($f = 1.44$ Hz). In order to control for the effect of the typical $1/f$ -like low-frequency background in EEG spectrum, each point of the power spectrum was corrected by multiplying its value by the corresponding frequency. Normalized power (NP) at each target frequency was then calculated as the ratio between the power at the target frequency and the mean corrected power spectrum for the 11 frequency bins centered on each target frequency excluding the 3 frequency bins previously used to compute the power of the target frequency.

RESULTS (in progress)

ERP analyses:

N100 component:

In order to test the evolution of the N100 component, we performed an ANOVA on the mean amplitude in the 100-170 ms latency band for the 4 time bins. The Main effect of Time Bin was significant ($[F(3,60) = 3.57; p = .02]$, Figure 1). Post-hoc analyses revealed that the N100 mean amplitude decreased as a function of exposure, that is, the N100 was maximal in the first time bin (-0.4 μV) and then significantly decreases in the second (0.1 μV ; $p = .005$), third (0.2 μV ; $p = .001$) and fourth time Bins (0.3 μV ; $p = .0002$). No between groups differences were found in these analyses. In order to have a finer level of analysis regarding the temporal evolution of the N100 mean amplitude, we divided the learning phase in 8 time bins. The main effect of Time bins was again significant [$F(7,133) = 5.3; p < .001$]. Post hoc analyses revealed a significant decrease in the N100 amplitude starting from the fourth time bin. No between groups differences were found in these analyses.

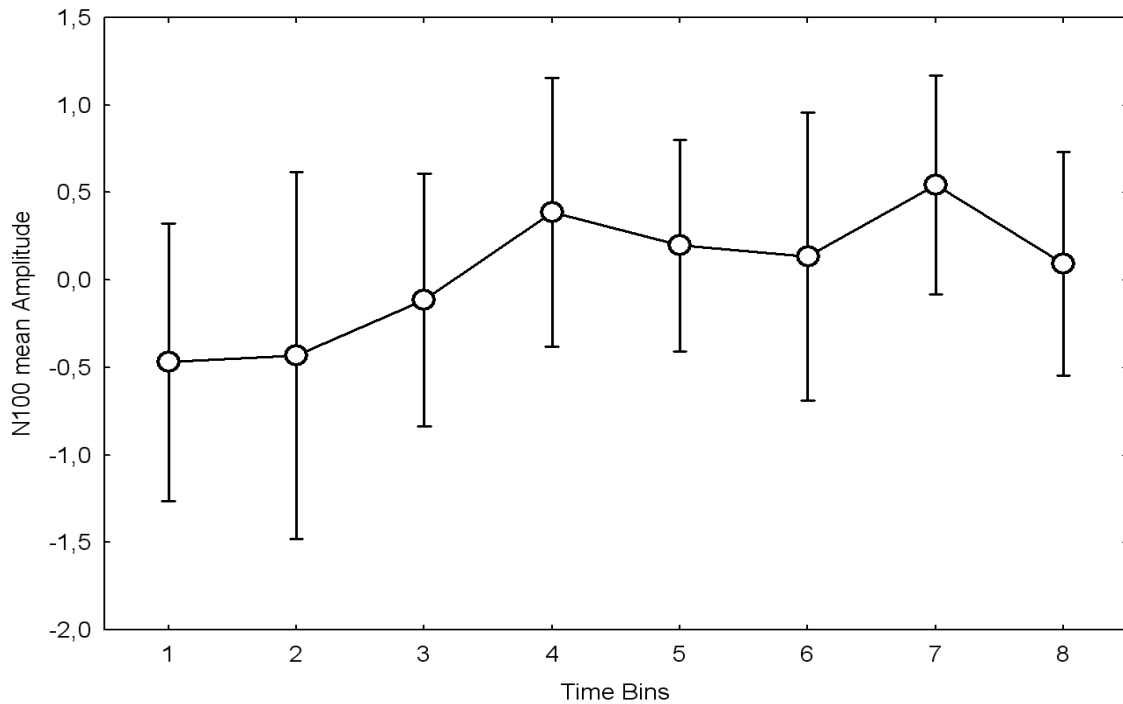


Figure 1: N100 mean amplitude (y axis) for each time Bins (x axis) averaged across all participants. Negativity is down. Error bars indicate the between subject variability

N400 component:

4 Time bins analyses:

The Main effect of Bin was significant [$F(3,60) = 3.20$; $p = .03$]. Most interestingly, the Expertise by Bin by Antero-posterior interaction was significant [$F(6,120) = 3.62$; $p = .01$] (Figures 2 and 3). For musicians the negativity increased from the first to the second bin over frontal regions ($p = .07$). From the first to the third bin, the negativity significantly increased over both Frontal and Central regions (all p 's $< .001$). Finally, from the third to the fourth bin, the negativity significantly decreased over frontal and central regions (all p 's $< .001$). By contrast, for non musicians, the negativity did not increase from the first to the second bin ($p > .9$). From the first to the third bin, the negativity increased although these differences did not reach significance ($p = .09$ at frontal site; $p > 0.3$ elsewhere). Finally, during the fourth bin the negativity increased and it was significantly larger than during the first and second bins over frontal and central regions ($p < .004$).

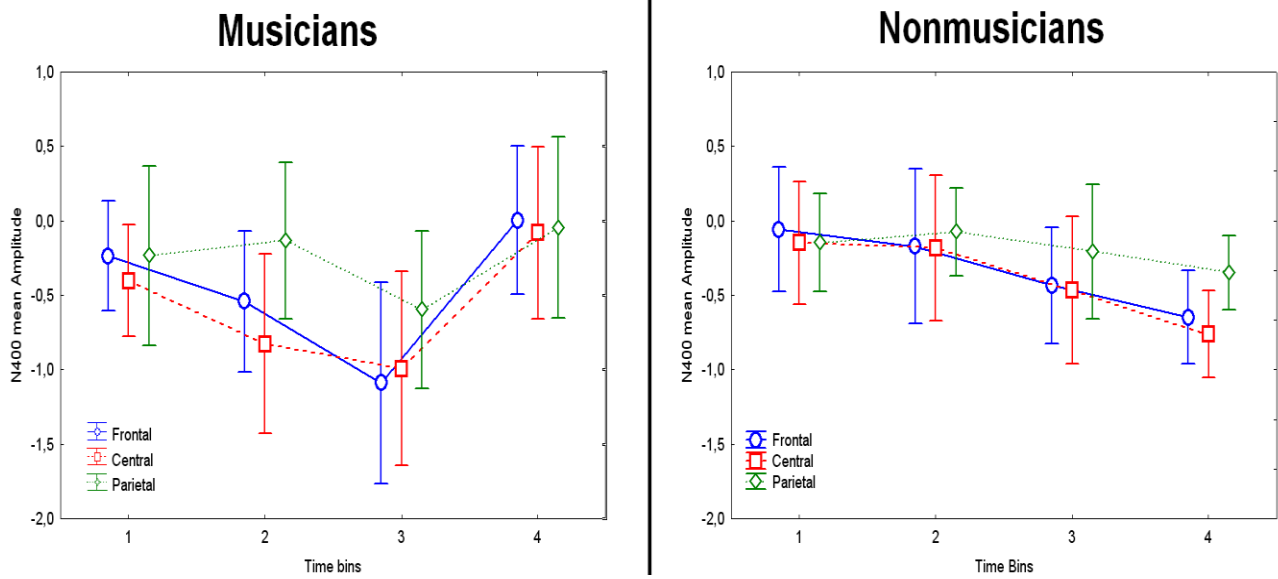


Figure 2: Illustration of the Expertise by Time Bin by Antero-Posterior interaction. N400 mean amplitude as a function of Time bins for the Musicians (left panel) and for the Nonmusicians (right panel). Error bars indicate the between subject variability

8 Time bins analyses:

Interestingly, more temporally fine-grained analyses confirm that the evolution of this negativity was different between the two groups (Expertise by Bin interaction [F (7, 133) = 2.58 ; p = .02]). The U-shape is now clearly visible for musicians, while nonmusicians show a rather linear learning trend (Figures 3 and 4).

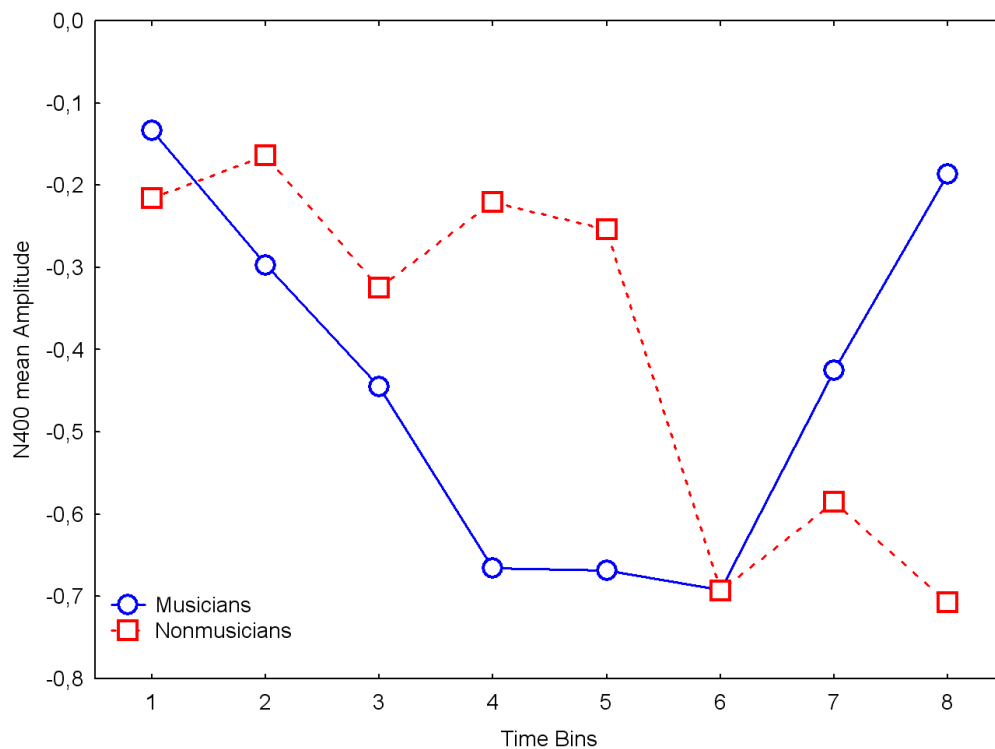


Figure 3: N400 mean amplitude (y axis) for each Time Bins (x axis) in both groups (Musicians in blue, Nonmusicians in red)

p values evolution for comparison against the 1st Time bin

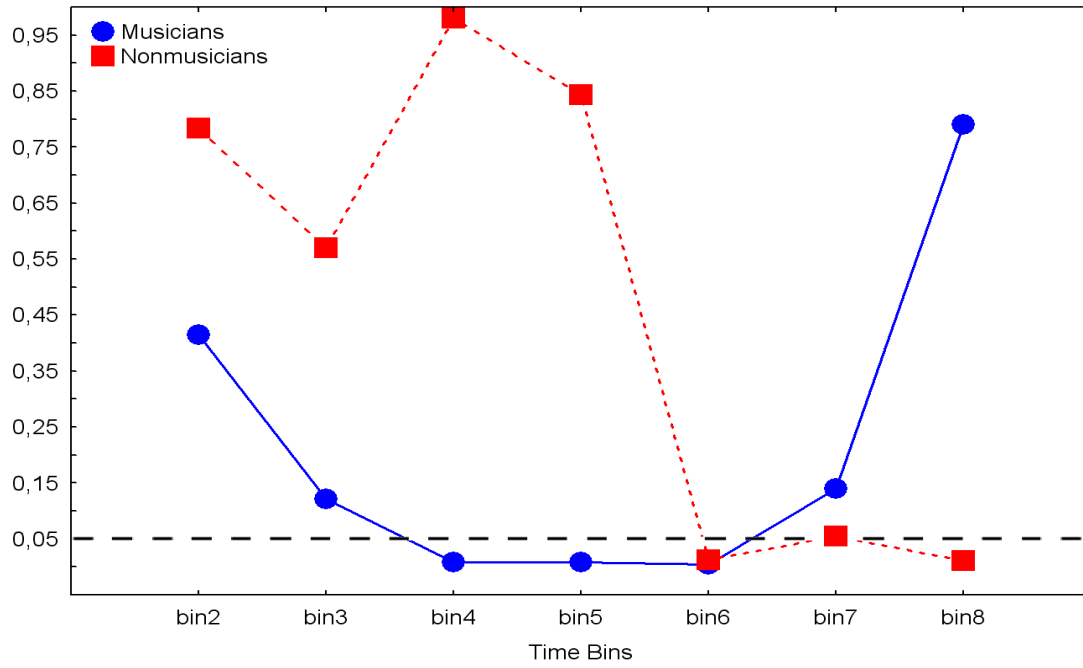
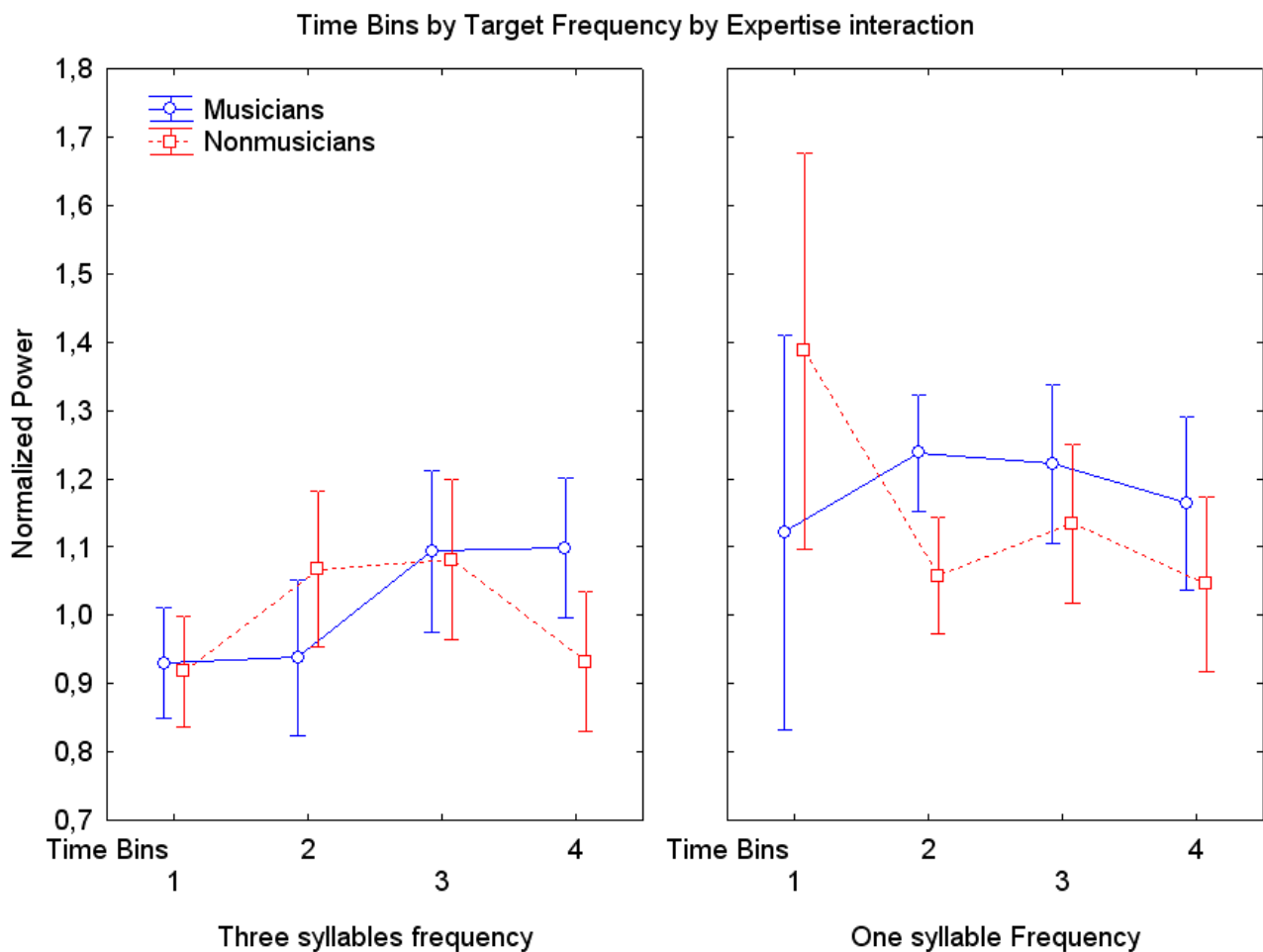
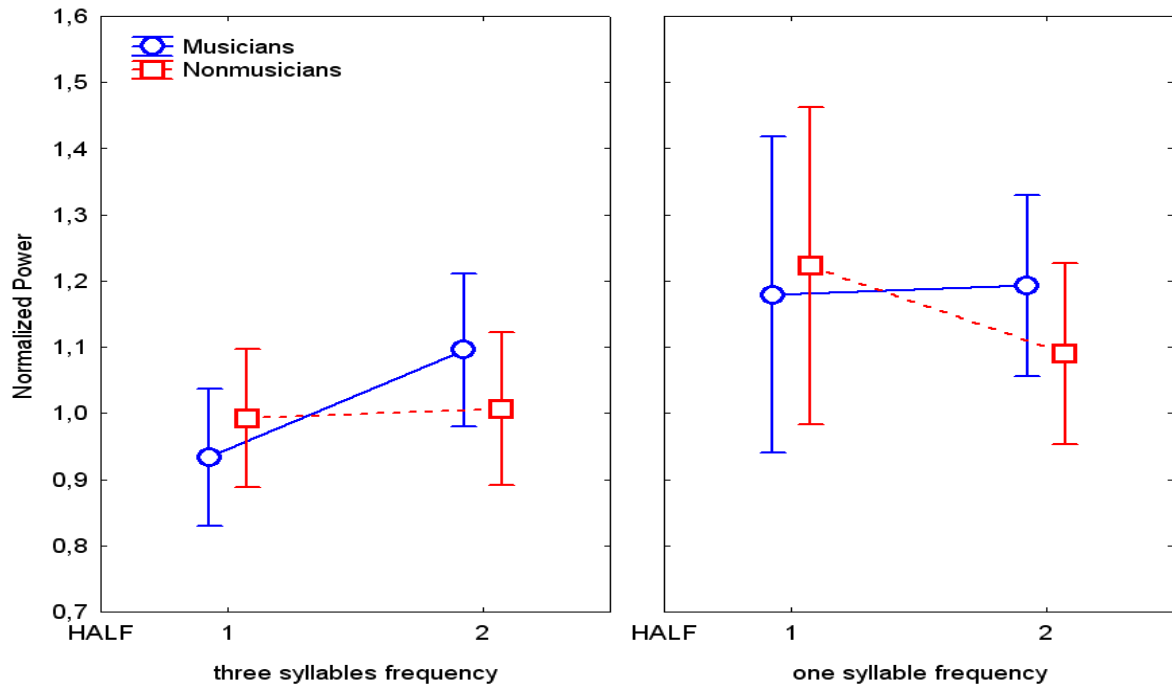


Figure 4: Illustration of the p values for the the comparison of each Time Bin against Time Bin 1, in each group.

Frequency-Tagging Analyses:

The present ANOVA was conducted on the normalized power spectrum at the frequency of three syllables and one syllable occurrences exclusively. This analysis included both Frontal and Central midline electrodes for the 4 Time Bins condition. Results of the ANOVA revealed no significant interaction involving the factor Expertise. However, when conducting the ANOVA with an additional factor consisting of modeling a 2 time Bins condition (instead of dividing the learning phase in 4, we divide it in 2) we found a significant Half by Time bins by Expertise interaction [$F(1, 16) = 4.3; p = .05$].





DISCUSSION (in progress)

The goal of this study was two-folded. Firstly, we wanted to test whether musical expertise can facilitate the word segmentation process *per se* by comparing the online dynamics of the N400 component through learning (N400 learning curves) obtained in two groups, one with musical expertise and one without formal musical training. Secondly, we wanted to test the sensitivity of the frequency tagging approach in distinguishing these two groups. At these aims, we recorded EEG during the learning of an artificial sung language and compared musician and non musician participants by using both ERP and frequency tagging analyses. Two main findings emerged of this study. Firstly, musicians needed a shorter exposure than non musicians to segment the stream. This was revealed by an inverted U-shape N400 learning curve for the musician group while a more linear N400 learning curve was observed for non musicians. Secondly, the present study confirms that frequency tagging analyses are a powerful tool to study the ongoing process of segmentation. Indeed, musicians did show a larger increase in the steady state power response at the frequency of word occurrence while non musicians did not show such an increase. These results will be discussed in turn in terms of more efficient stream segmentation abilities and more efficient phonemic template pattern matching process in musicians than in non musicians.

ERP correlates of online speech segmentation in musicians and non musicians:

The results of the ERP analyses are clear cut in showing that after a very short exposure to the speech stream, segmentation was achieved in both groups. This was revealed by significant increase of mean amplitude in the 350-500 ms latency band in the two groups. These results are in line with both previous behavioral and electrophysiological studies of artificial language learning. Indeed, Saffran and colleagues (1996a) showed that 2 minutes of exposure with a similar stream were sufficient for participants to recognize words from partial items in a subsequent behavioral test. Moreover, recent studies focusing on the on-line segmentation process and using rather similar

experimental paradigms reported N400 modulations as a function of exposure. For instance, Cunillera and colleagues (2009) compared the evolution of ERP traces elicited by random or structured language streams. They reported a learning related increase on a fronto-central negative component in the 350-550 ms latency range for structured streams when compared to random streams. It is interesting to note that the negative component observed in the present study shows similar topographical distribution and similar learning related time-course. The emergence of similar N400 components has been shown to be involved in several language learning experiments (McLaughlin et al., 2004; De Diego Balaguer et al., 2007; Mestres-Misse et al., 2007; Sanders et al., 2002). This component and his associated learning related modulations have been interpreted as reflecting the degree of implicit learning (Abla et al., 2008) or as an index of participants ability to extract structural information contained in the steam (Cunillera et al., 2009). Moreover, in the present study, we observed different N400 learning curves in the two groups: musicians presented an inverted U-shape N400 learning curve whereas non musicians presented a more linear N400 learning curve. Again, these results are clearly in line with two previous studies in which differential N400 learning curves were reported depending on performance on a subsequent behavioral test. For instance, in Cunillera et al and Abla et al., the authors observed in both studies that the group of participants with a good behavioral performance presented an inverted U-shape N400 learning curve. That is, the N400 amplitude increased in the first part of the learning and then decreased in the second part. The group of middle learners presented a rather linear N400 learning curve. That is, the N400 amplitude increased and reached a maximum at the end of the learning period. According to Cunillera et al., the increase in N400 amplitude may reflect the segmentation process per se while the decrease in N400 amplitude may rather reflect the recognition process. Moreover, Cunillera et al., revealed in an additional fMRI experiment that, compared to random stream, listening to the language stream elicited more activation in a large network involving the STG, IFG, PMC. Using NIRS, Abla et al reported increased BOLD signal in the IFG when

segmenting a stream of tones. Interestingly, these regions are known to be part of the dorsal stream in the Hickok and Poeppel model of language processing (2007). This dorsal stream is believed to play a role of auditory motor interface in speech perception and had been elegantly implemented in a model of word learning (Fornells et al., 2009). This model is interesting as it makes a good and convenient framework to formulate and test hypotheses on why musicians may segment better and faster than non musicians.

Quatrième Partie

Discussion générale

DISCUSSION GENERALE

Dans la partie suivante je discute les principaux résultats obtenus au cours de mes travaux de thèse. Pour ce faire je présente trois parties principales dont la première se focalisera sur une discussion autour de la comparaison entre l'apprentissage des structures linguistiques et musicales dans un langage artificiel chanté. Je poursuis par une discussion autour de la comparaison des approches comportementales et électrophysiologiques ainsi que leurs implications relatives dans la mise en évidence de processus d'apprentissage implicite. Finalement, je discuterai de l'origine des effets bénéfiques que peuvent avoir l'apprentissage de la musique chez l'enfant et l'expertise musicale chez l'adulte sur ce type d'apprentissage.

Chapitre 8 : Comparaison de l'apprentissage de structures linguistiques et musicales.

Les études réalisées au cours de cette thèse se basaient sur l'étude de Schön et collaborateurs (2008) et avaient notamment pour objectif de comparer le niveau d'apprentissage des deux dimensions: linguistique et musicale, contenues dans un langage artificiel chanté. Schön et collaborateurs montraient que l'apprentissage des structures linguistiques était facilité lors de l'écoute d'un langage artificiel chanté par rapport à celui engendré par l'écoute d'un langage artificiel parlé. Cependant, la question de savoir si la dimension musicale également présente dans le matériel de stimulation était aussi bien apprise que la dimension linguistique restait en suspens. Pour tenter de répondre à ces questions nous avons ajouté un test musical dans le protocole de Schön et collaborateurs et avons testé un échantillon aléatoire dans **l'étude 1** puis avons comparé un groupe de musiciens à un nouveau groupe de non musiciens dans **l'étude 2**. Dans les deux études, les différentes mesures récoltées révélaient que la dimension linguistique était mieux apprise que la dimension musicale (comme le montrent les effets principaux de la dimension dans les analyses des données comportementales, ainsi

que les différences de tailles d'effet de familiarité dans l'analyse des potentiels évoqués dans les deux dimensions). Plusieurs hypothèses explicatives sont à prendre en considération.

8.1 Hypothèse d'une plus grande familiarité vis à vis du langage que vis à vis de la musique.

Tout d'abord, la première hypothèse, relativement générale, est que la dimension linguistique est inconditionnellement plus familière pour les participants que ne l'est la dimension musicale. Nous avons vu dans la partie introductive que les deux domaines linguistique et musicaux partagent certaines caractéristiques communes et notamment au niveau des connaissances implicites. Dans le domaine linguistique, les connaissances phonotactiques implicites définissent l'ordre des phonèmes qui sont le plus probables d'apparaître les uns après les autres au sein des mots (Friederici & Wessels, 1993). Dans le domaine musical, les auditeurs possèdent des connaissances tonales implicites qui définissent l'importance perceptive des notes les unes par rapport aux autres dans un contexte tonal précis (Tillmann et al., 2000). Cependant, pour arriver à maîtriser parfaitement ces deux activités de communication, ces connaissances ne sont surement pas suffisantes et de nombreuses années d'apprentissage sont nécessaires pour acquérir un ensemble de connaissances beaucoup plus complexes. Ainsi, tous les participants ont suivi un apprentissage linguistique formel incluant l'apprentissage de l'écriture, de la lecture et de manière cruciale de la production du langage. Ils sont donc experts de la langue française à de multiples niveaux et ce n'est vraisemblablement pas le cas pour le domaine musical, même pour les participants musiciens. Cette expertise vis à vis de la langue française va impliquer une connaissance poussée des règles phonotactiques qui pourraient être à l'origine des meilleures performances dans la dimension linguistique. En effet, les connaissances des fréquences de positionnement des syllabes dans la langue française sont importantes pour la segmentation d'un flux de paroles. Une étude récente s'est intéressée au rôle des connaissances

phonotactiques au niveau syllabique dans l'apprentissage d'un langage artificiel (Mersad & Nazzi, 2011). Pour ce faire, les auteurs ont créé deux langages, un premier langage qui était congruent par rapport aux règles de fréquence de positionnement phonotactique des syllabes ainsi qu'un deuxième langage qui était incongru vis à vis de ces règles. Seul le groupe ayant été familiarisé avec le langage congru au niveau phonotactique apprenait les mots du langage artificiel. Ces connaissances phonotactiques reposent inévitablement sur la capacité à discriminer les phonèmes entre eux alors que les connaissances tonales reposent sur la capacité à discriminer deux notes entre elles. Cependant, des différences importantes au niveau acoustique existent entre domaine linguistique et domaine musical. Ceci me mène à formuler une deuxième hypothèse explicative d'ordre acoustico-perceptive.

8.2 Hypothèse acoustico-perceptive.

En effet, la détection fine d'évènements acoustiques se déroulant très rapidement dans le temps est nécessaire afin d'arriver à faire la différence entre les syllabes /py/ et /gy/ par exemple. Par opposition, la discrimination entre deux notes de piano va plutôt impliquer la capacité à discriminer une information relativement stable dans le temps, la hauteur (fréquence fondamentale). Dans les tests réalisés dans les **études 1 et 2**, alors que les consonnes étaient relativement différentes les unes des autres (/s/, /p/, /g/, /m/) toutes les notes des mélodies du test musical possédaient le même timbre de piano. De plus, ces dernières étaient toutes séparées par des intervalles relativement faibles. Il est possible que la distance perceptive entre les différentes consonnes du test linguistique soit plus grande que la distance perceptive entre les différents intervalles inter-notes du test musical. Toutes ces différences laissent penser que la discrimination des phonèmes était certainement plus facile que la discrimination des petits intervalles se traduisant par le patron de résultats comportementaux.

L'hypothèse acoustique semble inhérente aux expériences s'intéressant à comparer le traitement du

langage et de la musique. En effet, une étude menée par Koelsch et collaborateurs (2009) montre des résultats comparables aux nôtres. Cette étude s'intéressait à la comparaison entre le traitement du langage et le traitement de la musique dans une tâche de mémoire de travail. Les participants devaient mémoriser soit l'information musicale (hauteur) soit l'information verbale (la syllabe) de séquences de quatre syllabes chantées. Les résultats montraient que les performances des participants dans la tâche linguistique étaient significativement supérieures aux performances dans la tâche musicale. Deux études réalisées au sein de l'équipe ont reporté chez l'adulte et chez l'enfant des résultats qui vont aussi dans le même sens (Schön et al., 2004; Magne et al., 2006). Dans ces expériences, les participants devaient réaliser des tâches similaires dans les deux dimensions, linguistiques et musicales. Dans une première session les participants devaient détecter des incongruités de hauteur créées par une augmentation de 35 ou de 120% de la fréquence fondamentale du dernier mot de phrases simples et dans une seconde session, ils devaient détecter des incongruités de hauteur créées par une augmentation d' $1/5$ ou d' $1/2$ ton au niveau de la dernière note de phrases musicales simples. Si l'on convertit la modification d' $1/5$ de ton en un pourcentage, on aboutit à une modification de 3%. Alors que la modification sensée être équivalente dans le langage était de 35%. Cette simple comparaison révèle une différence d'un facteur 10 entre les deux modifications. Cet exemple révèle la difficulté ainsi que la nécessité de bien choisir et de contrôler les stimuli utilisés dans les tests musicaux surtout dans l'objectif de comparer le traitement du langage et de celui de la musique.

Finalement, il est également important de garder en mémoire que les stimuli utilisés lors de nos tests musicaux consistaient en des mélodies de trois notes jouées avec un timbre de piano issu de la base General MIDI alors que les participants étaient familiarisés avec des syllabes artificielles chantées. Les spectres parlent d'eux même et révèlent des différences acoustiques claires entre les stimuli musicaux et ceux utilisés pendant l'apprentissage. Ces différences acoustiques ne doivent certainement pas faciliter les premiers stades d'encodage de l'information sensorielle pour la reconnaissance des items. En outre, ce n'est pas uniquement au niveau de l'encodage des informations présentées lors des tests que des

différences entre langage et musique peuvent exister mais également au niveau de la consolidation en mémoire, du rappel de l'information en mémoire et au niveau de la résistance aux interférences.

8.3 Hypothèse de résistance aux interférences.

Une troisième hypothèse alternative est que la dimension linguistique est moins sensible aux interférences que ne l'est la dimension musicale. Il me semble que différents effets d'interférence peuvent être observés à différents moments du protocole expérimental utilisé au cours de cette thèse. Ainsi, des problèmes d'interférences peuvent survenir au niveau des tests et ceci nécessite de faire un bref rappel des stimuli utilisés lors de ces tests. Dans chaque essai, les participants devaient décider quel item (premier ou deuxième) ressemblait le plus à ce qu'ils avaient entendu pendant la phase de familiarisation. A chaque essai, un des deux items était un mot du langage alors que l'autre était un part-mot. Les part-mots sont construits avec les mêmes syllabes que les mots et ressemblent donc énormément aux mots au niveau phonologique. De plus, ils ont déjà été entendus pendant la phase d'apprentissage donc possèdent déjà un certain niveau de familiarité. Ainsi, les performances sont globalement plus faibles lorsque le test implique des part-mots que lorsqu'il implique des non-mots. Cette différence fondamentale laisse penser que les représentations des part-mots peuvent interférer plus fortement sur les représentations des mots que ne pourraient le faire les non-mots. Cette idée rappelle l'effet de proximité phonologique observé dans des tâches de rappel en mémoire et de reconnaissance d'items. L'effet de proximité consiste à observer de meilleures performances de rappel pour une séquence de lettres qui ne se ressemblent pas sur le plan acoustique que pour une séquence composée de lettres similaires au niveau acoustique (e.g. B, V, G par rapport à F, K, R). Cet effet de proximité phonologique se transpose dans la dimension musicale en un effet de proximité de hauteur (pitch proximity). Dans une tâche d'apprentissage implicite comme celle que j'ai utilisé lors de ma thèse, les effets d'interférences et de proximité de hauteur sont susceptibles d'avoir plus affecté les

performances dans le test musical que les effets d'interférence et de proximité phonologique présentes dans le test linguistique. De plus, une autre étude s'est intéressée aux effets d'interférences pour des stimuli musicaux (Schendel & Palmer, 2007). Pour ce faire, les auteurs ont testé les effets de suppression articulatoire exercés par des syllabes chantées ou par le mot « the » sur le rappel en mémoire de séquences musicales et verbales. Les résultats montraient d'une part que les séquences linguistiques étaient mieux rappelées que les séquences musicales. D'autre part, les résultats montraient que les effets de suppression articulatoire étaient plus importants pour les séquences musicales que pour les séquences linguistiques suggérant que ces dernières sont moins sensibles aux interférences que les séquences musicales. Finalement, des interférences entre les deux dimensions ont pu aussi avoir eu lieu le long de la phase de familiarisation. En effet, des études ont montré que le niveau d'attention au cours de l'apprentissage influence les performances dans le test (Toro et al., 2005b ; Perruchet & Pacton, 2008). Dans la mesure où je ne donnais aucune instruction quant au focus attentionnel à réaliser, il est possible que les participants aient pu effectuer des changements de focus attentionnels entre les deux dimensions pendant la phase d'apprentissage. Cette dernière possibilité pourrait directement impliquer que les participants aient besoin d'une période d'apprentissage plus longue pour présenter un niveau de performances équivalent dans les deux dimensions.

8.4 Perspectives et interprétations.

Jusqu'à présent, je me suis attelé à présenter les principales hypothèses explicatives quant à la supériorité des performances dans la dimension linguistique par rapport à la dimension musicale. Certaines d'entre elles ouvrent des pistes pour des recherches futures même si le problème inhérent des stimuli semble difficile à résoudre. Toujours est-il que la première chose à faire serait à mon sens de remplacer le timbre de piano par une syllabe qui n'aurait pas été entendue pendant la phase de familiarisation mais présentant des caractéristiques phonologiques les plus proches de celles entendues

pendant la familiarisation. L'hypothèse de plus grande familiarité du langage par rapport à la musique pourrait être abordée en réalisant des stimuli construits de manière à minimiser les connaissances phonotactiques des participants, par exemple des phonèmes d'une langue étrangère. L'hypothèse de durée insuffisante d'apprentissage pourrait simplement être étudiée en manipulant les durées d'apprentissage.

Enfin, au niveau théorique, les meilleures performances linguistiques que musicales constituent un résultat intéressant dans l'avancement de la compréhension des domaines linguistiques et musicaux ainsi que pour l'exploration d'éventuels liens entre le traitement du langage et le traitement de la musique. En effet, depuis de nombreuses années, un débat scientifique existe autour de la conception modulariste du langage amenée par Pinker. Selon Pinker, les processus computationnels impliqués dans le traitement du langage sont spécifiques au langage. Dans le cadre de l'étude des apprentissages implicites, les données expérimentales vont clairement dans le sens d'un mécanisme d'apprentissage indépendant de la modalité, du stade développemental, et de la dimension. Comme je le montrais plus haut, la situation dans le cas présent est pour le moins subtile. Le patron de résultat observé dans les **études 1 et 2** est consistant et montre une supériorité du langage face à la musique. Je serais donc amené à suggérer que le processus d'apprentissage implicite dans la dimension linguistique est indépendant du processus d'apprentissage de la dimension musicale lorsque le langage à apprendre est chanté. Néanmoins, je me suis permis de grouper les données des **études 1 et 2** à l'occasion de l'écriture de l'article soumis à *Frontiers in Auditory Cognitive Neurosciences* (**étude 4**). Dans ce cas, on observe une corrélation significative entre les performances dans le test linguistique et celles dans le test musical ($r = .3$; $p = .05$) : les sujets présentant des performances élevées dans le test linguistique présentent également des performances élevées dans le test musical. Ce dernier résultat est important pour plusieurs raisons. D'une part, il montre que lorsque j'augmente le nombre d'observations ($N = 50$ participants) j'augmente la puissance des tests statistiques. De plus, il vient nuancer la suggestion d'indépendance des deux processus d'apprentissage en s'appuyant sur des études ayant reporté chez

l'adulte et chez l'enfant une corrélation significative entre les compétences linguistiques et les compétences musicales suggérant là aussi des processus communs aux deux domaines. Par exemple, Anvari et collaborateurs (2002) montraient chez l'enfant que les aptitudes musicales (reflétées par les performances dans des tâches de perception et de production de mélodies, de rythmes et de progressions d'accords) prédisaient les performances dans différentes tâches de conscience phonologique. De même, chez l'adulte, Slevc & Miyake (2006) montraient que les aptitudes musicales prédisent les capacités de production et de perception dans une seconde langue. Ces deux études suggèrent l'existence d'un lien étroit entre la capacité à distinguer les représentations acoustiques des événements auditifs linguistiques et non linguistiques. Ce point est une différence cruciale qui pourrait contribuer à expliquer les différences entre musiciens et non musiciens dans les études réalisées au cours de cette thèse. Je m'attarderai plus longuement sur ce point dans la troisième partie de la discussion et passe désormais à la deuxième partie de la discussion.

Chapitre 9 : Comparaison des approches comportementales et électrophysiologiques.

L'objectif de cette discussion n'est pas de tenir un plaidoyer en faveur de l'approche électrophysiologique et donc indirectement de diminuer l'importance des informations comportementales mais plutôt de tenir un plaidoyer en faveur des approches multi méthodologiques qui permettent, à mon sens, d'obtenir plusieurs points de vue complémentaires en terme de qualité de l'information fournie sur le fonctionnement du système cérébral au niveau de processus cognitifs complexes comme celui de l'apprentissage des structures linguistiques et musicales. Pour introduire cette discussion je commence par aborder la question du statut des représentations des items linguistiques puis celui des items musicaux utilisés lors des **études 1, 2 et 3**. S'en suit une discussion autour des approches comportementales et électrophysiologiques dans le cadre de l'apprentissage implicite des structures linguistiques et musicales.

9.1 Quel(s) statut(s) pour les items des tests ?

Tout d'abord, concentrons-nous sur la question du statut des items linguistiques dans les protocoles d'apprentissages implicites. Il est intéressant de remarquer que cette question était déjà abordée dans le papier de Saffran et collaborateurs (1996b) : « If learners can use this type of information, then the distribution of subwords units in the input could serve as an initial bootstrapping device for generating candidate lexical hypotheses. ». Par la suite, Saffran précisait cette hypothèse en proposant que les représentations émergeant d'un apprentissage statistique pourraient servir aux enfants de candidats lexicaux disponibles pour intégrer leur langue maternelle (Saffran et al., 2001). En effet, plusieurs caractéristiques du protocole d'apprentissage implicite de langages artificiels permettent de penser que les représentations des items appris implicitement sont d'ordre lexical ou de manière plus appropriée, sont d'ordre proto-lexical. D'une part, le fait que les participants arrivent à discriminer un mot à forte

PT d'un mot à faible PT suggère que ces deux items possèdent des représentations différentes pour les participants. Indépendamment de leur statut, lexical ou non, ces représentations semblent suffisamment différentes pour que les participants arrivent à discriminer les mots des part-mots au niveau comportemental. Avant cette phase de test, les participants commencent toujours par écouter pendant plusieurs minutes le flux continu de mots. Cette période d'écoute peut être considérée comme une phase durant laquelle les participants sont familiarisés avec les formes acoustiques des différents mots. Selon Perruchet et Vinter (2002, 2008), à chaque présentation du mot « pogysi », par exemple, la trace mnésique de sa représentation se consolide petit à petit tout en diminuant la force de la représentation de « sipygy » déjà consolidée auparavant. De plus, dans le contexte de la littérature s'intéressant aux représentations lexicales dans le langage réel, les stimuli que nous utilisons pour aborder l'apprentissage de langages artificiels correspondent à des pseudo-mots. Les pseudo-mots sont des items linguistiques qui respectent les règles orthographiques de la langue considérée et qui sont prononçables (par opposition aux non mots qui ne sont pas prononçables : jgkdoe). Sur cet ensemble de considérations s'ajoutent plusieurs études dont l'objectif principal était précisément de tester cette hypothèse de (proto) lexicalité des items (Graf et al., 2007 ; Kolinsky et al., 2009 ; Voloumanos et al., 2009). En effet, dans le cadre général de l'apprentissage du langage, une autre question d'intérêt est celle de l'acquisition du sens des mots ou plus généralement l'appariement des mots avec les concepts (word to world mapping problem). Graf et collaborateurs ont eu l'idée originale de faire suivre la phase classique de familiarisation par une phase d'appariement sensée mimer l'appariement de la forme acoustique du mot avec un objet présenté visuellement. Cette étude était une des premières à révéler un lien entre la segmentation des mots dans un langage artificiel et l'apprentissage de ces mots en montrant que les enfants arrivent à associer un objet et un mot à forte PT mais pas à associer un objet et un part mot ni un non mot. Cet exemple semble avoir clarifié la situation concernant les items linguistiques. Cependant qu'en est-il pour les items musicaux issus des expériences réalisées lors de cette thèse ? De manière générale, existe-t-il un lexique musical qui serait le pendant du lexique

linguistique ?

Pour Peretz et Coltheart (2003) la réponse à la question est oui. En effet, ces auteurs se sont basés sur un ensemble d'études de patients déficitaires pour le traitement de la musique pour développer un modèle fonctionnel du traitement de la musique. Dans ce modèle, les auteurs font apparaître un lexique musical qui est supposé contenir l'ensemble des représentations des phrases musicales ayant été entendues tout au long de la vie des participants. Dans ce modèle, les auteurs considèrent que le lexique musical est le siège d'une procédure de sélection menant à la reconnaissance de la mélodie. Une étude récente précise ce modèle pour les mélodies non familières et les auteurs font l'hypothèse que le lexique musical ne conserverait pas que l'information mélodique mais plutôt qu'il contiendrait différents paramètres musicaux permettant la reconnaissance de nouvelles mélodies (comme le timbre par exemple ; Daltrozzo & Schön, 2009). Un autre pan de recherche apparaît intéressant dans le cas présent et concerne les représentations en mémoire dans le chant. Ces études se sont intéressées à étudier le rappel en mémoire des mélodies et des paroles contenues dans des chansons (Crowder et al., 1990; Hébert & Peretz, 2001; Morrongiello & Roes, 1990; Peretz et al., 2004; Serafine et al., 1984). Les résultats de ces études montrent que la dimension linguistique est mieux retenue en mémoire que la dimension musicale, suggérant que les représentations des structures musicales contenues dans le chant sont moins stables, moins robustes que les représentations linguistiques.

Ces notions de stabilité et de force des représentations issues d'un apprentissage implicite sont proposées par Cleeremans et Jimenez (2002) comme étant un point important pour la compréhension du lien entre le phénomène d'apprentissage et les modifications conscientes ou non associées. Pour ces auteurs, les représentations apprises implicitement correspondent à des patrons d'activation de réseaux neuronaux distribués dans le cerveau. Ces patrons d'activation peuvent varier au niveau de plusieurs dimensions: leur stabilité dans le temps, leur force et leur distinctivité. La stabilité dans le temps implique une composante de mémoire de travail pour que la représentation soit maintenue active pendant le traitement. La force des représentations se base sur la quantité et le niveau d'activation des

réseaux neuronaux impliqués. Finalement la distinctivité implique le degré de recouvrement entre deux représentations différentes issues de la même situation. De manière importante, pour ces auteurs les connaissances explicites impliquent des représentations en mémoire de meilleure qualité que les connaissances implicites. Dans la figure 2 (issue de l'article de Cleeremans & Jimenez, 2002), la force des représentations en abscisse est reportée en fonction des différentes situations d'apprentissage depuis la situation implicite, explicite puis la situation automatisée. On remarque ainsi que les représentations implicites (à gauche) sont bien plus faibles que les représentations explicites (au milieu) sur plusieurs paramètres que je ne commenterai pas.

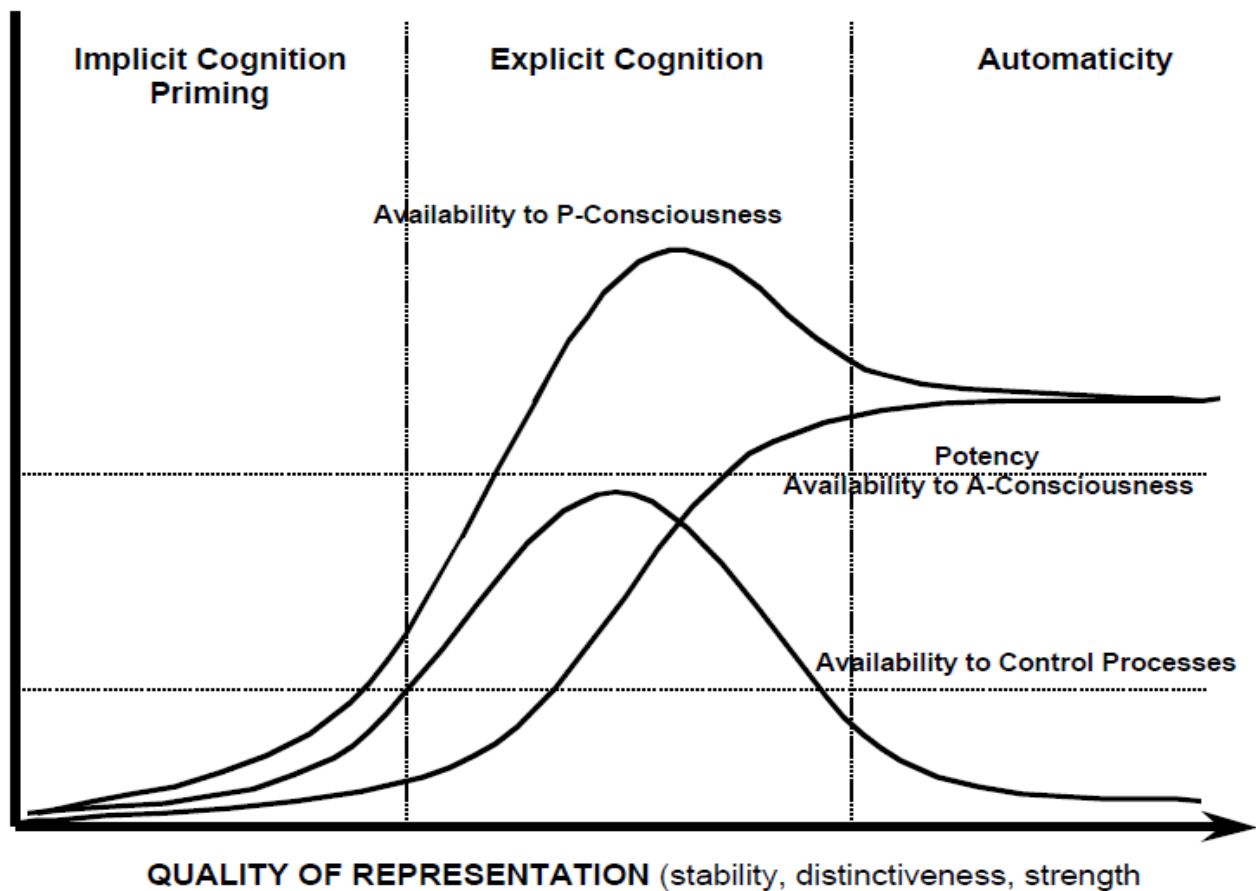


Figure 2: Représentation graphique des relations entre la qualité des représentations (implicites, explicites et automatiques, en abscisses) et plusieurs paramètres de conscience: puissance, disponibilité vis à vis du contrôle, disponibilité vis à vis de l'expérience subjective des participants. (Cleeremans & Jimenez, 2002)

Dans un article plus récent, Seth et collaborateurs (2008) mettaient l'accent sur l'intérêt d'utiliser des

mesures implicites et indirectes comme les méthodes de neuroimageries que sont l'IRMf et l'EEG. Ces méthodes pourraient apporter des informations importantes et complémentaires aux mesures comportementales dans la caractérisation des représentations conscientes et les modifications qualitatives des représentations implicites et explicites. Ainsi, des études dans différents domaines ont reporté des effets électrophysiologiques sans pour autant reporter d'effet au niveau comportemental. L'étude de Peretz et collaborateurs (2009) a révélé des effets électrophysiologiques en réponse à des changements de hauteur chez des patients amusiques sans qu'ils n'arrivent à détecter ce changement au niveau comportemental. Une autre étude a révélé des différences électrophysiologiques sous tendant un apprentissage sans modification des performances comportementales (Tremblay et al., 1998). Ainsi, si des dissociations entre mesures comportementales et électrophysiologiques existent, l'intérêt des mesures indirectes semblent décisives dans l'étude des processus sous tendant un apprentissage implicite.

9.2 Les points forts de l'électro-encéphalographie.

En effet, dans le cas des protocoles d'apprentissage implicites, l'utilisation de l'approche électrophysiologique peut avoir un intérêt pour plusieurs raisons et à plusieurs moments du protocole expérimental. Commençons par les tests. La première raison en faveur de l'EEG est plutôt quantitative. Au cours de ces tests, les participants devaient choisir l'item, premier ou deuxième, qui ressemblait le plus à ce qu'ils avaient entendu juste avant. De manière importante, les participants n'effectuaient ainsi qu'un seul appui par paire d'items sur le bouton correspondant à l'item choisi. Parallèlement à ces réponses comportementales, l'EEG était enregistré en continu permettant d'obtenir un potentiel évoqué pour chaque item des paires présentées lors de ces tests. L'EEG permet d'être plus précis que la mesure comportementale au niveau quantitatif en fournissant une mesure électrophysiologique pour les différents niveaux de familiarité.

La deuxième raison est plutôt qualitative. En effet, comme je le disais précédemment, les

représentations des items linguistiques et musicaux formées lors de la phase d'apprentissage sont implicites et relativement faibles. D'une part, leur rappel en mémoire est certainement facilement contaminé par les processus de prise de décision impliqués dans la réponse comportementale et d'autre part, le caractère explicite de la tâche est susceptible d'altérer la qualité des représentations implicites des items. L'EEG permettrait ainsi, à terme, de se passer de la phase de test ou tout du moins de la rendre la plus implicite possible en modifiant sa conception (e.g. présentation des items de manière isolée plutôt que par paires, utilisation d'une échelle de familiarité plutôt qu'un choix dichotomique).

La troisième raison renvoi à la sensibilité statistique de l'EEG vis à vis de la mesure comportementale. En effet, alors que le seuil individuel pour rejeter l'hypothèse nulle dépend directement de la distribution binomiale et est assez élevé (car N , c'est à dire le nombre d'essais, est petit), le seuil individuel avec les potentiels évoqués est calculé à partir des distributions t de Student des mots et des part mots. De plus, alors que le point de mesure comportementale est unique pour chaque paire, le point de mesure des PEs est multiple, car on peut tester la différence entre deux conditions à plusieurs points temporels (time-points). En effet, la majorité des participants présentait un effet électrophysiologique dans les deux dimensions. Cela permet d'une part de raffiner les interprétations des données comportementales et offre la possibilité d'étudier le processus d'apprentissage implicite au niveau individuel. De plus, cela permet de suivre précisément les dynamiques cérébrales du traitement des items linguistiques et musicaux avant même que la réponse comportementale n'ait eu lieu. Ce paramètre est ainsi décisif lorsque les participants sont des nouveau-nés ou même lorsqu'ils présentent des troubles moteurs ou de communication.

Ainsi l'EEG présente plusieurs points forts au niveau des tests mais aussi au niveau de la phase d'apprentissage. En effet, il serait difficile de suivre l'apprentissage en temps réel avec la seule mesure comportementale récoltée dans un test en choix forcé. Il pourrait être envisagé d'alterner de courtes sessions d'apprentissage avec plusieurs tests comportementaux. Cependant, comme je le disais précédemment, le caractère explicite du test va inévitablement changer le statut des représentations des

items appris dès la première passation du test. L'EEG permet quant à lui de suivre la dynamique de la phase d'apprentissage et permet de révéler l'émergence du corrélat neuronal du processus de segmentation. Ainsi, l'utilisation systématique de l'EEG pendant la phase d'apprentissage permettrait peut être de se passer de la phase de tests ce qui a un intérêt important pour répondre ou préciser les questions sur les effets d'interférences. Finalement, l'utilisation de l'EEG permet d'une certaine manière d'obtenir une information reproductible, formatée et peut être moins sensible au stade de développement, à l'espèce étudiée et à certains déficits cognitifs présents dans certaines pathologies que ne l'est la mesure comportementale.

9.3 Réconciliation des deux approches.

Seth et collaborateurs mettaient un accent sur l'utilisation de mesures indirectes dans le cadre de l'apprentissage implicite. Cependant, ces auteurs avançaient également que des progrès pouvaient être effectués aussi bien au niveau des mesures comportementales. Ainsi, j'ai pu essayer d'aller dans ce sens en conduisant deux nouveaux types d'analyses qui jusque là ne semblent pas avoir été utilisées ou reportées dans la littérature. La première analyse prenait en compte le contraste de PT entre l'item familier et l'item non familier pour chaque essai. Dans un essai impliquant le mot gimysy et le part-mot pygyimi, le contraste de PT entre les deux items était de 0.68 ($1 - 0.32$). Si la distinctivité des représentations des deux items se base essentiellement sur les PT, le premier type d'essai devrait être plus facile que le deuxième. Cela devrait se traduire par de meilleures performances pour le premier que pour le deuxième type d'essai. Ainsi la figure 3 illustre cette analyse pour le test musical et semble suggérer que les musiciens sont sensibles au contraste alors que les non musiciens ne semblent pas présenter une aussi grande sensibilité.

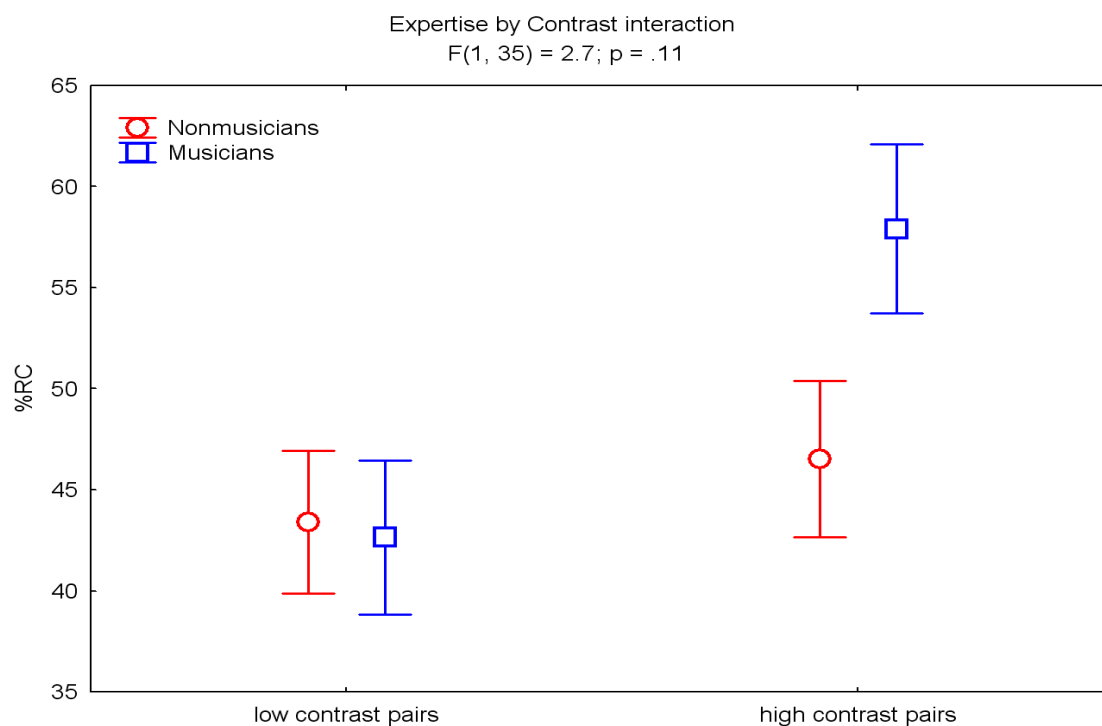


Figure 3: Graphique illustrant les performances moyennes de chaque groupe (en ordonnées) pour les paires de tests à fort contraste (gauche) et pour les paires de tests à faible contraste (droite) dans le test musical.

Pour tenter de caractériser les effets d'interférences que je soupçonnais se produire au fur et à mesure que les tests se déroulent, j'ai également comparé la moyenne des performances pour les 12 premiers essais à la moyenne des performances pour les 13 derniers essais. Comme l'illustre la figure 4, cette analyse permet encore une fois d'apporter des précisions et même de valider nos interprétations vis à vis des interférences présentes au cours du test linguistique.

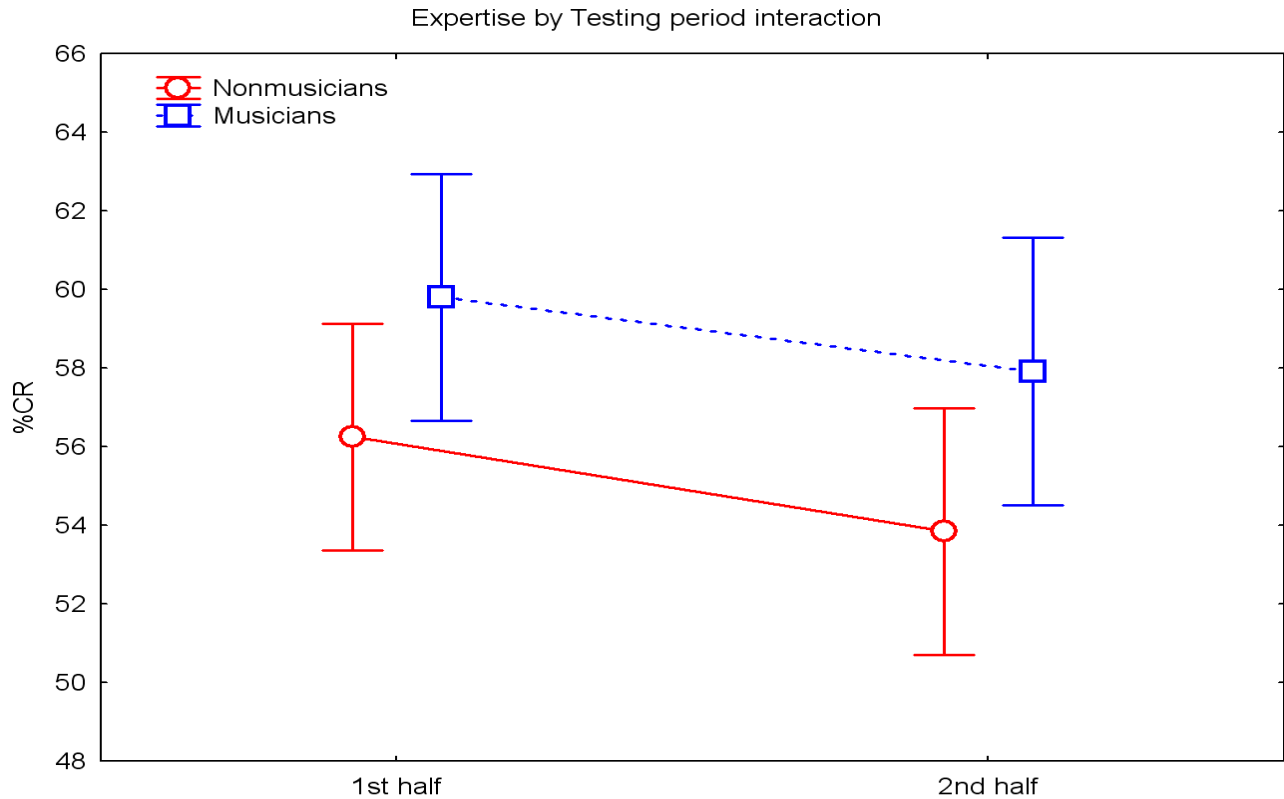


Figure 4: Graphique représentant les performances moyennes (% de réponses correctes) des deux groupes (en ordonnées) pour la première et la deuxième moitié du test linguistique (en abscisses).

Sur la base de cette discussion, il semble clair que chacune des approches fournit des points de vue complémentaires l'une vis à vis de l'autre. Ce point est essentiel car il permet de raffiner les interprétations des résultats vis à vis du processus étudié. En effet, l'utilisation des deux approches permet évidemment de tester la corrélation des deux variables mesurées. C'est la façon la plus communément utilisée dans la littérature. Plusieurs études s'intéressant à l'apprentissage de langages artificiels ont révélé une corrélation significative entre l'amplitude de la N400 dans la phase d'apprentissage et les performances dans le test consécutif. Ces résultats sont intéressants mais sont à considérer avec précaution car ces analyses perdent tout leur intérêt lorsque le groupe présente des performances en moyenne non différentes du hasard. Pour ma part, j'ai montré que la taille d'effet de familiarité présent au niveau de la composante N400 des PE était plus grande pour l'analyse effectuée

sur les essais corrects que pour l'analyse effectuée sur tous les essais (corrects et incorrects). Ce résultat est intéressant dans la mesure où il complète bien les résultats comportementaux. Un autre résultat intéressant dans ce contexte concerne le test musical pour les musiciens. Dans ce test, les deux groupes présentaient une sensibilité aux PT avec de meilleures performances pour les mélodies à forte PT que pour les mélodies à faible PT. Les données EEG révélaient une composante de type MMN plus ample pour les mélodies non familières chez les musiciens seulement. Ce résultat déjà intéressant en soit prend un intérêt encore plus important au regard de l'analyse de cette MMN en réponse aux deux sous types de mélodies non familières. Je discuterai ce point de manière plus précise dans la dernière partie de la Discussion mais je peux d'ores et déjà avancer que la complémentarité des approches est intéressante quant à la précision des interprétations fonctionnelles des données électrophysiologiques. En effet, une corrélation entre des performances comportementales et l'amplitude d'une composante plutôt perceptive comme la N1 ou la MMN sera bien différente de la corrélation trouvée entre ces performances et l'amplitude d'une composante plutôt cognitive.

Ainsi, une fois mis bout à bout, les indices comportementaux et électrophysiologiques sont complémentaires et permettent de mieux comprendre les facteurs décisifs en jeu dans des processus cognitifs complexes et, comme nous le verrons dans la dernière partie de la discussion, de compléter, raffiner les modèles de segmentation du langage.

Chapitre 10 : L'expertise musicale chez l'adulte et l'apprentissage de la musique chez l'enfant facilitent la segmentation de la parole.

Dans cette dernière partie je tente d'intégrer les résultats issus des **études 2, 3 et 4** réalisées au cours de ma thèse dans le cadre précis du modèle d'apprentissage du langage récemment proposé par Rodriguez-Fornells et collaborateurs (2009). Je commence par faire un bref rappel sur le cerveau musicien en présentant les caractéristiques principales qui justifient son étude dans le cadre général du traitement du langage et de la musique. Je présenterai ensuite le modèle d'apprentissage du langage en mettant l'accent sur les structures cérébrales impliquées et susceptibles d'être les cibles d'une plasticité induite par la pratique musicale chez l'adulte, puis, proposerai des hypothèses sur la spécificité de cette plasticité musicale en regard des résultats de **l'étude 3** chez l'enfant.

10.1 Le cerveau musicien, rappel.

Je reportais en détail dans la partie introductive traitant de l'expertise musicale un ensemble de résultats montrant que les musiciens présentent de nombreuses différences vis à vis des non musiciens. Dans la partie traitant des études longitudinales, j'apportais des précisions en révélant que ces différences en faveur des musiciens trouvent spécifiquement leur origine dans la pratique intensive de la musique. Ainsi, nous commençons à obtenir de nombreux indices montrant que la pratique et l'apprentissage de la musique provoquent des changements fonctionnels et structurels notamment au niveau auditif et sensorimoteur. En effet, les musiciens présentent des indices comportementaux et électrophysiologiques, corticaux et sous corticaux, suggérant qu'ils détectent mieux, au niveau attentif et préattentif, les changements de plusieurs paramètres acoustiques des sons non linguistiques et linguistiques. Les musiciens sont également plus sensibles au niveau comportemental (attentif) et électrophysiologique (attentif et préattentif) dans la détection d'incongruités contextuelles et structurelles dans la musique et le langage. Ainsi, l'hypothèse de transfert d'apprentissage de la musique

vers le langage est de plus en plus étudiée, éprouvée et semble être bien acceptée dans la communauté scientifique. Les deux études (**études 2 et 3**) que j'ai réalisées lors de ma thèse ont révélé des résultats intéressants pour la comparaison des musiciens et des non musiciens et la compréhension du traitement du langage et de la musique. Cependant, il me semble que ce n'est qu'une fois inclus et intégrés dans le contexte théorique et expérimental du modèle d'apprentissage du langage proposé par Rodriguez-Fornells et collaborateurs, que les résultats des expériences réalisées révèlent toute leur valeur.

10.2 Le modèle d'apprentissage du langage proposé par Rodriguez-Fornells et collaborateurs.

Depuis Wernicke, les études s'intéressant aux bases cérébrales sous tendant les processus d'intégration, de perception et de compréhension du langage sont fleurissantes, amenant chacune des mesures complémentaires et souvent des résultats différents voir contradictoires, probablement dus à leur dépendance avec le type de tâche utilisé. Cette caractéristique est un des points principaux sur lesquels se basent Hickok et Poeppel (2007) pour proposer un nouveau modèle inspiré par celui de Wernike et de celui développé pour la modalité visuelle. Les auteurs soulignent la nuance entre les tâches impliquant une compréhension du langage et celles impliquant une simple perception du langage. Pour ces auteurs, les tâches perceptives n'impliquent pas d'accès au lexique mais impliquent préférentiellement une composante de mémoire de travail. Par opposition, les tâches de compréhension nécessitent inévitablement un accès au lexique. Ainsi, pour tenter de comprendre cette dissociation, Hickok et Poeppel proposent qu'il existerait différents niveaux de représentations impliqués dans le processus de mise en correspondance du son avec son sens. Nous verrons un peu plus loin que cette proposition est intéressante notamment au sujet des connaissances implicites phonotactiques. Hickok et Poeppel proposent ainsi que le traitement de la parole soit organisé autour de deux voies distinctes, la voie dorsale et la voie ventrale qui présentent chacune des spécificités de traitement et de fonction. Dans ce

modèle d'organisation corticale du traitement de la parole, la voix dorsale est considérée, dans son ensemble, comme une interface sensorimotrice impliquant un large réseau fronto- pariéto-temporal. La fonction essentielle des structures impliquées serait de mettre en correspondance les représentations d'ordre phonologique issues d'une analyse spectro-temporelle dans le Gyrus Temporal Supérieur (GTS) avec des représentations d'ordre articulatoires plutôt au niveau des aires pré-motrices et de l'aire de Broca. La voix ventrale est quant à elle considérée comme une interface lexico conceptuelle permettant de mettre en correspondance les représentations issues du GTS avec des représentations d'ordre lexical et conceptuel principalement au niveau du Gyrus Temporal Moyen.

Alors que Hickok et Poeppel évoquaient un élément développemental dans la justification de la voie dorsale, leur modèle ne prenait évidemment pas en compte toutes les études récemment conduites au sujet de l'apprentissage implicite de langages artificiels. Dans ce contexte, le modèle proposé par Rodriguez-Fornells et collaborateurs est spécifiquement destiné à combler ce manque et tente ainsi de mieux cerner les processus neurophysiologiques impliqués dans l'apprentissage du langage chez l'adulte. Ce modèle propose de faire une synthèse des données anatomo-fonctionnelles, électrophysiologiques et comportementales récoltées dans une série d'études s'intéressant aux premiers stades de l'acquisition d'un langage révélées dans des tâches de segmentation plutôt implicites. Ce modèle intègre également des données anatomo-fonctionnelles, électrophysiologiques et comportementales des premiers stades de l'acquisition du sens. Dans ce modèle, la voix dorsale est également présente et implique le GTS postérieur supposé contenir les représentations de séquences phonologiques transmises via le faisceau arqué au Cortex Pré Moteur censé mettre en correspondance ces séquences phonologiques avec des séquences articulatoires ordonnées.

La voix ventrale est considérée comme une interface d'intégration sémantico-conceptuelle connectant le GTS au Gyrus Temporal Moyen et au Gyrus Frontal Inférieur ventral via plusieurs faisceaux: le uncinate fasciculus, le inferior longitudinal fasciculus et le inferior fronto occipital fasciculus. Finalement, les structures hippocampiques, para-hippocampiques et plus généralement le Lobe

Temporal Médian connectées via le faisceau longitudinal supérieur sont intégrées dans le modèle en tant qu'interface lexico épisodique. Si ce bref résumé du modèle ne se veut pas exhaustif, il n'en demeure pas moins qu'il permet de mettre en évidence les liens intimes qui peuvent exister entre le système de traitement de l'information auditive et le système moteur. Cette caractéristique est particulièrement intéressante pour ce qui concerne les résultats de cette thèse. En effet, comme je le disais plus haut, les musiciens présentent des caractéristiques neurophysiologiques au niveau auditif et moteur différentes des non musiciens. De plus, une étude récemment publiée a montré un fort recouvrement des aires cérébrales impliquées dans le traitement lexico-phonologique et mélodique en comparant des stimuli musicaux, linguistiques parlés et chantés (Schön et al., 2010). Finalement, une méta-analyse de Vigneau et collaborateurs (2006) a reporté les aires corticales activées pendant différents types de tâches linguistiques. Ces deux études révèlent un fort recouvrement des aires corticales impliquées dans le traitement linguistique et musical. Mises bout à bout, ces données expliquent l'intérêt de l'étude des musiciens professionnels ainsi que la comparaison des deux dimensions linguistiques et musicales contenues dans le chant. Dans la section suivante, je discuterai tout d'abord des résultats obtenus lors des tests et des différences observées entre musiciens et non musiciens. Pour se faire, je centre la discussion autour de la voix dorsale décrite dans le modèle de Rodriguez-Fornells et collaborateurs et propose trois hypothèses susceptibles de rendre compte des différences entre musiciens et non musiciens. Dans une première partie, je fais l'hypothèse d'un meilleur fonctionnement du système impliqué dans le traitement de l'information auditive chez les musiciens. Dans une seconde partie, je fais l'hypothèse d'un meilleur fonctionnement du système moteur des musiciens. Dans une troisième partie, je fais l'hypothèse d'une meilleure communication entre ces deux systèmes. Dans une seconde section, je discuterai brièvement les résultats préliminaires des phases d'apprentissages et je terminerai par une discussion des résultats de l'étude longitudinale chez l'enfant et leurs implications (**étude 3**).

10.3 Vers une hypothèse de meilleur système perceptif des musiciens.

Au niveau des tests, les PEs révèlent un effet principal de l'expertise musicale au niveau de la N1 dans les deux dimensions, linguistiques et musicales. Ce premier résultat est intéressant car il confirme les résultats de nombreuses études ayant montré des effets d'expertise similaires dans des tâches perceptives préattentives et attentives utilisant des stimuli linguistiques et musicaux (Pantev et al., 1998 ; Shahin et al., 2003, 2005 ; Musacchia et al., 2008). Même si l'amplitude de la N1 est sensible au niveau d'attention des participants (Hillyard et al., 1973, 1987), une étude a montré que cet effet est vraisemblablement le reflet d'une meilleure qualité des représentations neuronales des attributs acoustiques des sons plutôt que le reflet d'un effet d'attention sélective (Baumann et al., 2008). De plus, le générateur de la N1 semble être localisé au niveau du cortex auditif (gyrus de Heschl latérale et Planum Temporale). Cette structure participerait à la formation des représentations acoustique et serait la porte d'entrée de l'information auditive dans le réseau complexe impliqué dans les deux modèles décrits plus haut. Ainsi, sur la base de nos résultats, outre les avantages au niveau sous cortical, les musiciens en comparaison avec les non musiciens, présenteraient des représentations des sons plus distinguables les unes des autres. Cet avantage doit probablement provenir d'un meilleur encodage des attributs acoustiques des sons grâce à une meilleure analyse spectro-temporelle, plus fine et plus efficace que les non musiciens (Tallal & Gaab, 2006).

Dans les tests musicaux, les PEs révélaient deux résultats surprenants mais amenant des précisions sur les caractéristiques fonctionnelles du système auditif des musiciens. En effet, seuls les musiciens présentaient un effet de familiarité des items au niveau de la P2 ainsi qu'au niveau d'une composante de type MMN en réponse à la deuxième note des mélodies peu familières. Ces deux composantes précoces trouvent leurs générateurs probablement au niveau du Planum Temporale (Godey et al., 2001 ; Pulvermüller & Shtyrov, 2006 ; Molholm et al., 2005). De manière intéressante, cette structure est connue pour être modifiée par l'expertise musicale. En effet, comparés aux non musiciens, les musiciens présentent un PT plus développé avec une plus grande densité de matière grise (Keenan et

al., 2001, Luders et al., 2004, Schlaug et al., 1995b). Considérés dans leur ensemble, les données de la littérature ainsi que les effets de familiarité retrouvés dans les tests musicaux sur la P2 et la MMN suggèrent que les musiciens catégorisent mieux l'information tonale et l'information d'intervalle que les non musiciens. Ainsi, le Planum Temporale qui présente un rôle important dans la distribution des informations dans la voie dorsale est sans doute une structure dont le fonctionnement pourrait expliquer les différences d'efficacité d'apprentissage des structures linguistiques et musicales chez les musiciens et les non musiciens. Si un meilleur fonctionnement global du système de traitement des informations auditives est susceptible de contribuer aux différences observées entre musiciens et non musiciens, l'implication du système moteur au niveau du Gyrus Frontal Inférieur et du Cortex Pré-Moteur est à prendre en considération.

10.4 Vers une hypothèse de meilleur fonctionnement du système moteur des musiciens.

Des composantes négatives sensibles au degré de familiarité des items linguistiques et musicaux se développaient aux alentours de 600 ms pour le langage et 700 ms pour la musique. Les items linguistiques et musicaux peu familiers évoquaient une négativité plus ample que les items familiers. Malgré des latences différentes entre les deux dimensions, les distributions topographiques frontocentrales étaient relativement similaires. De manière intéressante, les études sur lesquelles se base le modèle de Rodriguez Fornells et collaborateurs reportent également une composante de topographie similaire qui se développe pendant la période d'apprentissage (Cunillera et al., 2006 ; 2009 ; De Diego Balaguer et al., 2007). Cette composante semble avoir été localisée au niveau d'une large région incluant le Cortex Pré Moteur, le Gyrus Frontal supérieur et le Gyrus Pariétal Inférieur (Cunillera et al., 2009).

Il me semble important de rappeler que les items à différencier par les participants ont été entendus pendant quelques minutes seulement. Au fur et à mesure de l'apprentissage, les items sont présentés de

manière répétitive devenant de plus en plus familiers pour les participants. Ainsi, lors d'un essai de test, les participants se retrouvent à choisir entre un item familier (le mot, présenté environ 100 fois) et un autre moins familier (le part-mot, présenté environ 25 fois). Dans la littérature N400 classique, les mots naturels présentant une faible fréquence d'occurrence (donc peu familiers) évoquent une N400 plus ample que des mots à forte fréquence d'occurrence (donc familiers) (Young & Rugg, 1992 ; Van Petten & Kutas, 1990). De plus, ce que nous considérons comme des mots dans nos tests ne sont finalement que des pseudo-mots vis à vis du langage naturel. O'Rourke et Holcomb ont étudié l'effet N400 dans une tâche de décision lexicale impliquant des mots et des pseudo-mots. Alors que la N400 en réponse aux mots réels était maximale à 400 ms, la N400 en réponse aux pseudo-mots était justement plutôt maximale à 600 ms. Les auteurs proposaient que la plus grande latence aux pseudo-mots résulterait d'une étape supplémentaire de contrôle attentionnel d'origine top-down. De manière intéressante, dans le modèle de Rodriguez-Fornells et collaborateurs, une étape de contrôle d'origine top-down semble être assurée par le CPM. Dans notre cas, nous disposons d'un système EEG à 32 capteurs, il nous est ainsi difficile d'estimer correctement la ou les sources impliquées dans la génération de cette composante. Néanmoins, sur la base de la littérature, il est envisageable de faire l'hypothèse d'une plus forte contribution du CPM chez les musiciens que chez les non musiciens. En effet, les musiciens présentent des aires motrices plus développées que celles des non musiciens suggérant de meilleures représentations motrices que chez les non musiciens (Hutchinson et al., 2003). De plus, le Gyrus Frontal Inférieur (aire de Broca) est une structure adjacente au CPM qui est impliqué aussi bien dans le traitement des structures linguistique que dans le traitement de structures musicales (Vigneau et al., 2006 ; Koelsch et al., 2005). Les musiciens présentent une épaisseur corticale au niveau du GFI plus importante que les non musiciens (Slumming et al., 2002). Une étude à récemment montré que l'intégrité des fibres de matière blanche dans cette région prédit les performances d'apprentissage d'une grammaire artificielle (Flöel et al., 2009). Le GFI est activé aussi bien dans une tâche de complétion de mélodies que dans une tâche de complétion de phrases (Brown et al., 2006) Ces données pourraient

suggérer une plus forte implication du GFI chez les musiciens que chez les non musiciens.

10.5 Vers une hypothèse de meilleure communication entre les structures auditives et motrices.

Si un meilleur fonctionnement indépendant du PT et du CPM peuvent expliquer les différences observées entre musiciens et non musiciens, une plus forte connectivité entre ces deux structures pourrait également rendre compte des différences observées. Ces deux structures sont reliées par le faisceau arqué et ce dernier est modifié par l'apprentissage de la musique. En effet, il a été montré que l'apprentissage de la musique induit chez l'enfant une augmentation de la quantité de fibres nerveuses dans le faisceau arqué (Wan & Schlaug, 2010). Ainsi, une plus grande connectivité entre ces deux régions pourrait mener à une plus forte activation coordonnée des deux structures qui pourrait être reflétée par l'effet N400 plus grand chez les musiciens que chez les non musiciens.

10.6 Dynamique d'apprentissage : différences entre musiciens et non musiciens (étude 5).

Au niveau des périodes d'apprentissage, nos résultats amènent des précisions sur les différences entre musiciens et non musiciens. D'une part, nos résultats sont en accord avec les résultats d'études précédentes ayant reporté des patrons de modulations de l'amplitude de la N400 différents en fonction des performances comportementales (Abla et al., 2008 ; Cunillera et al., 2009). Généralement, les auteurs reportent un patron en U inversé spécifiquement pour le groupe de participants ayant de bonnes performances comportementales. Ce patron de modulation de l'amplitude de la N400 est interprété comme reflétant une alternance entre le processus de segmentation sous-tendu par l'augmentation de la N400 au début de l'apprentissage, et un processus de reconnaissance des mots appris implicitement sous-tendu par la diminution de l'amplitude de la N400 (Cunillera et al., 2006, 2009 ; Rodriguez-

Fornells et al., 2009). Dans nos analyses, alors que les non musiciens présentent une augmentation de l'amplitude de cette composante le long de l'apprentissage, les musiciens présentent quant à eux un patron en U inversé caractérisé par une augmentation de l'amplitude N400 pendant les 2 premières minutes de l'apprentissage suivie d'une diminution de la N400 jusqu'à la fin de l'apprentissage. L'analyse frequency-tagging semble conforter les différences observées au niveau des PEs (même si les résultats ne sont que préliminaires). En effet, nos résultats révèlent une dynamique de la puissance spectrale à la fréquence d'occurrence des mots différente chez les musiciens et les non musiciens. Ce résultat préliminaire est déjà intéressant en soit car il montre que, même dans une condition clairement plus implicite que celle utilisée par Buiatti et collaborateurs (2009), le frequency-tagging semble être encore un outil intéressant pour étudier l'apprentissage de langages artificiels. Sur la base des dynamiques temporelles des modulations de la N400 et de la fréquence du mot, ainsi que les propositions de Rodriguez-Fornells sur l'interprétation de la dynamique d'évolution de la N400, j'aurai tendance à suggérer que l'augmentation de la puissance spectrale à la fréquence des mots serait reliée au processus de reconnaissance des mots du langage.

10.6 Intérêt et précisions amenées par l'étude longitudinale.

Les résultats précédents chez le musiciens adultes, bien que très informatifs, restent sujets aux inhérentes critiques de la présence de facteurs explicatifs non pris en compte tels que des facteurs génétiques susceptibles d'offrir des prédispositions de bases pour la pratique musicale (Pulli et al., 2008). Dans ce contexte, **l'étude 3** permet de s'affranchir de ces critiques. En effet, les deux groupes d'enfants ont été constitués de manière pseudo-aléatoire et n'étaient pas différents au niveau de tests neuropsychologiques classiquement utilisés dans ce type d'étude. Les résultats de cette étude sont clairement spectaculaires tant au niveau des données comportementales qu'au niveau des données électrophysiologiques des tests. Avant le début de l'apprentissage de la peinture ou de la musique, les

deux groupes d'enfants n'arrivaient pas à reconnaître les mots du langage. Les PE révélaient quant à eux une composante négative aux alentours de 600 ms, sensible au degré de familiarité des items et rappelant clairement la morphologie de la négativité observée chez l'adulte. Deux années scolaires après le début de l'étude, les résultats révélaient une forte augmentation des performances pour le groupe musique seulement. De plus, cette augmentation de performances était accompagnée d'une augmentation de l'effet de familiarité au niveau de la composante négative chez les enfants du groupe musique uniquement. Ces résultats inattendus (bien qu'espérés) me permettent de préciser l'origine des effets observés chez l'adulte musicien. D'une part, les effets observés chez les enfants musiciens ne semblent pas être liés à une plus forte augmentation générale du niveau d'attention ni même du niveau intellectuel que chez les enfants du groupe peinture. En effet, les deux groupes ont améliorés leurs performances de manière équivalente dans plusieurs tests neuropsychologiques. D'autre part, il ne s'agit pas d'un niveau de motivation ou d'un niveau d'implication dans les activités apprises différents entre les deux groupes. En effet, les données préliminaires d'expériences incluses dans le projet MUSAPDYS et s'intéressant aux caractéristiques du geste d'écriture révèlent des effets bénéfiques et spécifiques de l'apprentissage de la peinture sur le geste d'écriture. Dans le domaine de l'apprentissage implicite chez l'enfant, Conway et collaborateurs (2009) ont montré que des enfants sourds puis équipés d'un implant présentaient un déficit d'apprentissage implicite au niveau visuel (Conway et al., 2011). Ces auteurs proposaient que l'absence de stimulations auditives entraînerait un déficit cognitif général, non spécifique à la modalité auditive suggérant leur auditory scaffolding hypothesis. Ainsi, au vu des résultats de l'étude menée chez l'enfant, de ceux reportés chez l'adulte et des structures corticales mises en évidence dans le modèle de Rodriguez-Fornells, il pourrait y exister plusieurs structures cibles telles que le PT, le CPM et le GFI avec une implication forte des faisceaux arqué et longitudinal supérieur.

De manière importante, ces résultats élargissent le champ d'action de la pratique musicale et suggèrent que celle-ci peut avoir des effets positifs sur des capacités cognitives et linguistiques de haut niveau.

Ainsi, il semble important de promouvoir l'apprentissage de la musique dès le plus jeune âge, ce qui n'est malheureusement pas le cas dans la politique éducative développée par le gouvernement actuel.

10.7 Conclusions et Perspectives.

L'ensemble des travaux réalisés au cours de cette thèse ont été menés avec la volonté de faire coexister plusieurs approches, plusieurs visions et plusieurs échelles. La synthèse de ces travaux aura été un passage particulièrement difficile à réaliser mais apportera je l'espère des précisions sur plusieurs sujets d'intérêt tant scientifique, qu'éducatif et plus globalement sociétal. En effet, le cerveau musicien est un modèle dont l'étude permettra certainement de préciser les modèles cognitifs impliqués dans différents domaines ainsi que le lien entre structures et fonctions. Cependant, après avoir réalisé l'ensemble de ces travaux en utilisant principalement une méthode électrophysiologique de surface, j'aimerais pouvoir me rapprocher des sources ou tout du moins pouvoir accéder à une information spatiale plus précise. Ainsi, la réalisation de travaux utilisant des approches neurophysiologiques chez le singe (LFPs et spikes) et le patient épileptique (sEEG) permettrait de mieux caractériser les dynamiques d'activités ainsi que les connexions entre les différentes structures impliquées dans le réseau neuronal complexe impliqué dans le processus de segmentation de flux auditifs. De plus, après avoir réalisé cette étude longitudinale chez des enfants de 8 ans, j'aimerais réaliser des études similaires chez le bébé. Finalement, les premiers résultats très encourageants des expériences réalisées chez l'enfant ouvrent des perspectives de recherche excitantes et offrent un bel espoir quant à l'utilisation de la musique pour le développement cognitif des enfants et pourquoi pas pour la remédiation à certaines pathologies.

Cinquième Partie

RÉFÉRENCES

BIBLIOGRAPHIQUES

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press.

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line Assessment of Statistical Learning by Event related Potentials. *Journal of Cognitive Neuroscience*, 20 (6), 952-964.

Altmann, G. T., Dienes, Z., Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 21, 899–912.

Anvari, S., Trainor, L., Woodside, J., & Levy, B. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, 83(2), 111-130.

Aslin, R., Saffran, J., & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.

Aslin, R. N. (2000, July). Interpretation of infant listening times using the headturn preference technique. Paper presented at the International Conference on Infant Studies, Brighton, England.

Bermudez, P., Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*. 1060, 395-399.

Besson, M., Macar, F. (1987). An event-related-potentials analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*. 24 (1), 14-25.

Besson, M., Faïta, F., Requin, J. (1994). Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neuroscience Letters*. 168 (1-2), 11-105.

Besson, M., Faïta, F. (1995). An Event-Related Potential (ERP) Study of Musical Expectancy: Comparison of Musicians With Nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*. 21 (6), 1278-1296.

Besson, M., Schön, D. (2001). Comparison between language and music. *Annals of the New*

York Academy of Sciences. 930, 232-258.

Besson, M., Schön, D. (2003). Comparison between music and language. Dans I. Peretz & R. Zatorre (Éd.), *The Cognitive Neuroscience of Music* (p. 269–293). Oxford: Oxford University Press.

Bidelman, G. M., Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*. 1355, 112-125.

Bigand, E., Perruchet, P., Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Current Psychology of Cognition*. 17 (3), 577-600.

Bigand, E. (2004). Can an expert musical ear develop from passive listening to music? *Revue de Neuropsychologie*, 14(1-2), 191-221.

Brattico, E., Näätänen, R., Tervaniemi, M. (2001). Context effects on pitch perception in musicians and non-musicians: evidence from ERP recordings. *Music Perception*. 19, 1-24.

Brown, S., Martinez, M. J., Parsons, L. M. (2006). Music and language side by side in the brain : a PET study of the generation of melodies and sentences. *European Journal of Neuroscience*. 23, 2791-2803.

Buiatti, M., Peña, M., Dehaene-Lambertz, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *Neuroimage*. 44 (2), 509-519.

Butzlaff, R. (2000). Can music be used to teach reading? *J Aesthet Educ*. 34, 167.

Chobert, J., Marie, C., François, C., Schön, D., Besson, M. (soumis à JOCN). Enhanced passive and active processing of syllables in musician children.

Cleeremans, A., Destrebecqz, A., Boyer, M. (1998). Implicit learning: news from the front. *TICS*. 2 (10), 406-416.

Cleeremans, A., Jiménez, L. (2002). Implicit learning and consciousness: A graded, dynamic perspective. In R.M. French & A. Cleeremans (Eds.), *Implicit Learning and Consciousness*, Hove, UK: Psychology Press, pp. 1-40.

Colombo, J., Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*. 4, 219-223.

Conway, C. M., Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 31(1), 24.

Conway, C. M., Pisoni, D., & Kronenberger, W. (2009). The Importance of Sound for Cognitive Sequencing Abilities: The Auditory Scaffolding Hypothesis. *Current Directions in Psychological Science*, 18(5), 275-279.

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science*. 14 (1), 69-82.

Cunillera, T., Toro, J. M., Sebastián-Gallés, N., Rodríguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: an event-related brain potential study. *Brain Research*. 1123 (1), 168.

Cunillera, T., Gomila, A., Rodríguez-Fornells, A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC Neuroscience*. 18, 9.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Gallés, N., Ortiz, H., Pujol, J., Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage*. 48 (3), 541.

Cutler, A., Carter, D. M. (1987). The predominance of strong initial syllables in the english vocabulary. *Computer Speech and Language*. 2, 133-142.

Daltrozzo, J., Schön, D. (2009a). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*. 21 (10), 1882.

Daltrozzo, J., Schön, D. (2009b). Is conceptual processing in music automatic? An electrophysiological approach. *Brain Research*. 1270, 88-94.

Daltrozzo, J., Tillmann, B., Platel, H., Schön, D. (2010). Temporal aspects of the feeling of

familiarity for music and the emergence of conceptual processing. *Journal of Cognitive Neuroscience*. 22 (8), 1754.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., Bachoud-Lévi, A. C. (2007). Different Neurophysiological Mechanisms Underlying Word and Rule Extraction from Speech. *PLoS ONE*. 2 (11), e1175.

De Diego Balaguer, R., Fuentemilla, L., Rodriguez-Fornells, A. (2011). Brain Dynamics sustaining rapid rule extraction from speech. *Journal of Cognitive Neuroscience*. doi:10.1162/jocn.2011.21636

Dienes, Z., Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cogn Sci* 28, 531-558.

Dowling, W. J., Kwak, S., Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics*, 57, 136-149.

Ehrle, N., Samson, S. (2005). Auditory discrimination of anisochrony: influence of the tempo and musical backgrounds of listeners. *Brain and Cognition*. 58, 133-147.

Elger, C. E., Grunwald, T., Lehnertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., Van Roost, D., Heinze, H. J. (1997). Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia*. 35 (5), 657- 667.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In Barkow J., Cosmides, L., Tooby, J. (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, England: Oxford University Press.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458-467.

Fiser, J., Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual

Feature Hierarchies. *Journal of Experimental Psychology: General*. 134 (4), 521.

Forgeard, M., Winner, E., Norton, A., Schlaug, G. (2008). Practicing a Musical Instrument in Childhood is Associated with Enhanced Verbal Ability and Nonverbal Reasoning. *PlosONE*. 3(10), e3566.

Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C., Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *Neuroimage*. 47 (4), 1974-1981.

Foxton, J. M., Talcott, J. B., Witton, C., Brace, H., McIntyre, F., Griffiths, T. D. (2003). Reading skills are related to global, but not local, acoustic pattern perception. *Nature Neuroscience*. 6, 343-344.

Friederici, A. D., Pfeifer, E., Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*. 1(3), 183-192.

Friederici, A. D., Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech recognition. *Perception & Psychophysics*. 54 (3), 287-295.

Friederici, A. D., Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: first-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*. 25 (1), 157-176.

Fujioka, T., Trainor L. J., Ross, B., Kakigi R., Pantev, C. (2004). "Musical training enhances automatic encoding of melodic contour and interval structure." *Journal of Cognitive Neuroscience*. 16, 1010-1021.

Gaser, C., Schlaug, G. (2003). Brain Structures Differ between Musicians and Non-Musicians. *Journal of Neuroscience*. 23, 9240-9245.

Gervain, J., Macagno, F., Cogoï, S., Peña, M., Mehler, J. (2008). The neonate brain detects speech structure. *PNAS*. 105, 14222.

Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a

comparison of data in the same patients. *Clinical Neurophysiology*. 112 (10), 1850-1859.

Habermeyer, B., Herdener, M., Esposito, F., Hilti, C. C., Klarhöfer, M., di Salle, F., Wetzel, S., Scheffler, K., Cattapan-Ludewig, K., Seifritz, E. (2009). Neural correlates of pre-attentive processing of pattern deviance in professional musicians. *Human Brain Mapping*. 30 (11), 3736-3747.

Hackett, T. A., Preuss, T. M., Kaas, J. H. (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *Journal of Comparative neurology*. 441 (3), 197-222.

Hahne, A., Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*. 11 (2), 194-205.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., Varpula, T. (1982). Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: implications for their neural generation. *Electroencephalography and clinical neurophysiology*. 54 (5), 561-569.

Hayes, J. R., Clark, H. H. (1970). Experiments in the segmentation of an artificial speech analog. *Cognition and the development of language*, 221-234.

Hébert, S., Peretz, I., & Gagnon, L. (1995). Perceiving the tonal ending of tune excerpts: The roles of pre-existing representation and musical expertise. *Canadian Journal of Experimental Psychology*, 49, 193-209.

Hickok, G., Poeppel, D. (2007). The Cortical organization of speech processing. *Nature Reviews Neuroscience*. 8, 393-402.

Hillyard, S. A., Hink, R. F., Schwent, V. L., Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*. 182 (108), 177-180.

Hillyard, S. A., Woldorff, M., Mangun, G. R., Hansen, J. C. (1987). Mechanisms of early selective attention in auditory and visual modalities. *Electroencephalogr Clin Neurophysiol Suppl*. 39: 317-324.

Hutchinson, S., Lee L. H. L., Gaab N. Schlaug, G (2003). Cerebellar volume: Gender and musicianship effects. *Cerebral Cortex*. 13, 943-949.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., Schlaug, G (2009). Musical training shapes structural brain development. *Journal of Neuroscience*. 29 (10), 3019-3025.

Jackendoff, R. (2002). *Foundations of language*. Oxford University Press.

Jäncke, L. (2002). What is special about the brains of musicians? *Neuroreport*. 13 (6), 741-742.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371-375.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological review*. 83, 323-355.

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance*. 21, 293-307.

Keenan, J. P., Thangaraj, V., Halpern, A. R., Schlaug, G (2001). Absolute pitch and planum temporale. *Neuroimage*. 14 (6), 1402-1408.

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 (2 Suppl), 125-143.

Koelsch, S., Schröger, E., Tervaniemi, M. (1999). Superior pre-attentive auditory processing in musicians. *Neuroreport*. 10 (6), 1309-1313.

Koelsch, S., Gunter, T., Friederici, A. D., Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience*. 12 (3), 520-541.

Koelsch, S., Mulder, J. (2002a). Electric brain responses to inappropriate harmonies during listening to expressive music. *Clinical Neurophysiology*. 113(6), 862-869.

Koelsch, S., Schmidt, B. H., Kansok, J. (2002b). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*. 39, 657-663.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*. 7 (3), 302-307.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., Schlaug, G. (2005). Adults and children processing music: an fMRI study. *Neuroimage*. 25, 1068-1076

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an fMRI study. *Human Brain Mapping*. 30 (8), 859-873.

Krampe, R. T., Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J Exp Psychol*. 125, 331-359.

Kraus, N., Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Review Neuroscience*. 11, 599.

Krumhansl, C. L., Shepard, R. N. (1979). Quantification of the Hierarchy of Tonal Functions Within a Diatonic Context. *Journal of Experimental Psychology: Human Perception and Performance*. 5(4), 579-594.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Review Neuroscience*. 5, 831.

Kujala, T., Tervaniemi, M., Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biological Psychology*. 74, 1-19.

Kutas, M., Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*. 207 (4427), 203-205.

Kutas, M., Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*. 307 (5947), 161-163.

Kutas, M., Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *TICS*. 4 (12), 463-470.

Lappe, C., Herholz, S. C., Trainor, L. J., Pantev, C. (2008). Cortical plasticity induced by short-

term unimodal and multimodal musical training. *Journal of Neuroscience*. 28 (39), 9632-9639.

Lieberman, M., Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In: M. Aronoff and R. Oerhle (Eds), *Language Sound Structure* (157-223). Cambridge, MA: MIT Press.

Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography and clinical neurophysiology*. 92 (3), 204-214.

Loui, P., & Wessel, D. (2008). Learning and linking an artificial musical system: Effects of set size and repeated exposure. *Musicae Scientiae*, 12(2), 207–230.

Loui, P., Wessel, D., Hudson Kam, C. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music perception*. 27 (5), 377-388.

Loveless, N.E., Brunia, C.H.M., (1990). Effects of rise time on late components of the auditory evoked potential. *Journal of Psychophysiology*. 4, 369-380.

Luders, E., Gaser, C., Jancke, L., Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage*. 22 (2), 656-664.

Magne, C., Schön D., Besson M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *J Cogn Neurosci*. 18(2), 199-11.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S., Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *PNAS*. 97 (8), 4398-4403.

Makeig, S., Bell, A. J., Jung, T. P., Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Adv Neural Inform Proc Syst*. 8, 145-151.

Marie, C., Kujala, T., Besson, M. (2010). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex*. In press.

Marie, C., Magne, C., Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*. 23 (2), 294-305.

May, P. J. C., Tiitinen, H. (2009). Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. *Psychophysiology*. 46, 1-57.

McLaughlin, J., Osterhout, L., Kim, A. (2004). Neural correlates of second -language word learning: minimal instruction produces rapid change. *Nature Neuroscience*. 7 (7), 703.

McNealy, K., Mazziota, J. C., Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *J Neurosci*. 26, 7629-7639.

McNealy, K., Mazziota, J. C., Dapretto, M. (2010). The neural basis of speech parsing in children and adults. *Developmental Science*. 13 (2), 385-406.

McMullen, E., Saffran, J. (2004). Music and Language: A Developmental Comparison. *Music Perception*. 21 (3), 289-311.

Mersad, K., Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory and Cognition*. Epub ahead of print.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*. 219 (1-2), 36-47.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., Tervaniemi M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research*. 1194, 81-89.

Milovanov, R., Huotilainen, M., Esquef, P. A., Alku, P., Välimäki, V., Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters*. 460(2), 161-165.

Molholm, S., Martinez, S., Ritter, W., Javitt, D. C., Foxe, J. J. (2005). The neural circuitry of pre-attentive auditory change detection an fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex*. 15, 545-551.

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, *Cerebral Cortex*. 19, 712-713.

Münte, T. F., Altenmüller, E., Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Review Neuroscience*. 3 (6), 473-478.

Musacchia, G, Sams, M, Skoe, E, Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *PNAS*. 104(40), 15894-15898.

Musacchia, G, Strait, D., Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. *Hearing Research*. 241(1-2), 34-42.

Näätänen, R., Gaillard, A. W. K., Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*. 42, 313-329.

Näätänen, R., Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*. 24 (4), 375-425.

Näätänen, R., Lehtokoksi, A., Lenne, M. (1997). Language-specific phoneme representations revealed by electrical and magnetic brain responses. *Nature*. 385, 432-434.

Nespor, M., Vogel, I. (1983). Prosodic structure above the word. In: A. Cutler & D. R. Ladd (Eds), *Prosody: Models and Measurements*. Berlin, Germany: Springer-Verlag.

Neville, H. J., Kutas, M., Chesney, G., Schmidt, A. L. (1986). Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. *Journal of Memory and Language*. 25 (1), 75-92.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., Garrett, M. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*. 3, 151-165.

O'Rourke, T. B., Holcomb, P. J. (2002). Electrophysiological evidence for the efficiency of

spoken word processing. *Biological Psychology*. 60 (2-3), 121-150.

Osterhout, L., Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*. 31 (6), 785-806.

Osterhout, L., Holcomb, P. J. (1993). Event-related potential and syntactic anomaly: evidence of anomaly detection during the perception of continuous speech. *Language and Cognitive processes*. 8, 413-437.

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*. 999, 497.

Pacton, S., Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *JEP, Learning Memory and Cognition*. 34 (1), 80-96.

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology*. 118, 177-185.

Pakarinen, S., Huotilainen, M., Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clinical Neurophysiology*. 121 (7), 1043-1050.

Pantev, C., Hoke, M., Lehnertz, K., Lütkenhöner, B., Anogianakis, G., Wittkowski, W. (1988). Tonotopic organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and clinical neurophysiology*. 69 (2), 160-170

Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and clinical neurophysiology*. 94, 26-40.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*. 392 (6678), 811-814.

Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., Ross, B. (2001). Timbre-specific

enhancement of auditory cortical representations in musicians. *Neuroreport*. 12 (1), 169-174.

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(45), 14100-14107.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*. 10 (6), 717-733.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*. 6 (7), 674-681.

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press US.

Peña, M., Bonatti, L., Nespore, M., Mehler J. (2002). Signal-Driven Computations in Speech Processing. *Science*. 298, 604.

Peretz, I., Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*. 6 (7), 688-691.

Peretz, I., Brattico, E., Järvenpää, M., Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain*. 132, 1277-1286

Perruchet, P. (2008). Implicit learning. In *Cognitive psychology of memory*. Vol.2 of *Learning and memory: A comprehensive reference* (J. Byrne, Ed.) (pp. 597-621). Oxford: Elsevier.

Picton, T. W., Goodman, W. S., Bryce, D. P. (1970). Amplitude of evoked responses to tones of high intensity. *Acta oto-laryngologica*. 70 (2), 77-82.

Picton, T. W., Sassa, J. M., Dimitrijevic, A., Purcell, D. (2003). Human auditory-state responses. *International Journal of Audiology*. 42 (4), 177-219.

Pulli, K., Karma, K., Norio, R., Sistonen, P., Göring, H. H. H., Järvelä, I. (2008). Genome-wide linkage scan for loci of musical aptitude in Finnish Families : evidence for a major locus at 4q22. *Journal of Medical Genetics*. 45, 451-456.

Pulvermüller, F., Shtyrov, Y. (2006). Language outside the focus of attention : The Mismatch

Negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology*. 79, 49-71.

Randel, D. M. (1978). *Harvard concise dictionary of music*. Cambridge, MA: Harvard University Press.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*. 6, 855-863.

Rodriguez-Fornells, A., Cunillera, T., Mestress-Misse, A., De Diego Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B*. 364, 3711-3735.

Rohrmeier, M., Rebuschat, P., Cross, I. (2010). Incidental and online learning of melodic structure. *Consciousness and Cognition*. Epub ahead of print.

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. *Neuroscience Letters*. 308, 33-36.

Saffran, J. R., Aslin, R. N., Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*. 274, 1926.

Saffran, J. R., Newport, E. L. Aslin, R. N. (1996a). Word segmentation: the role of distributional cues. *J Mem Lang*. 35, 606-621.

Saffran, J., Johnson, E., Aslin, R., & Newport, E. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70(1), 27-52.

Saffran, J. R., Senghas, A., Trueswell, J. C. (2001). The acquisition of language by children. *PNAS*. 98, 12874.

Saffran, J.R. (2003). Statistical language learning: Mechanisms and Constraints. *Current Directions in Psychological Science*. 12 (4), 110.

Saffran, J. R., Hauser, M., Seibel, R. L., Kapfhamer, J., Tsao, F., Cushman, F. (2008). Grammatical pattern learning by infants and cotton-top tamarin monkeys. *Cognition*. 107, 479.

Sanders, L. D., Newport, E. L., Neville, H. J. (2002). Segmenting nonsense: an event-related-

potential index of perceived onsets in continuous speech. *Nature Neuroscience*. 5(7), 700-703.

Sanders, L. D., Neville, H. J. (2003). An ERP study of continuous speech processing. I. Segmentation, semantics, and syntax in native speakers. *Brain Res. Cogn. Brain Res.* 15 (3), 228-240.

Sanders, L. D., Neville, H. J. (2003a). An ERP study of continuous speech processing. II. Segmentation, semantics, and syntax in non-native speakers. *Brain Res. Cogn. Brain Res.* 15 (3), 214-227.

Schellenberg, E. G (2004). Music lessons enhance IQ. *Psychol Sci* 15, 511-514.

Schellenberg, E. G (2006). Long-term positive associations between music lessons and IQ. *J Educ Psychol*. 98, 457-468.

Schendel, Z. A., Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory and Cognition*. 35 (4), 640-650.

Schlaug, G, Jäncke, L., Huang, Y., Staiger, J. F., Steinmetz, H. (1995a). Increased corpus callosum size in musicians. *Neuropsychologia*. 33(8), 1047-1055.

Schlaug, G, Jäncke, L., Huang, Y., Steinmetz, H. (1995b). In vivo evidence of structural brain asymmetry in musicians. *Science*. 267 (5198), 699-701.

Schlaug, G (2001). The brain of musicians. A model for functional and structural adaptation. *Annals of the New York Academy of Sciences*. 930, 281-299.

Schmithorst, V. J., Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience letters*. 321 (1-2), 57-60.

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. *Nature Neuroscience*. 5(7), 688-694.

Schön, D., Magne, C., Besson, M.(2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*. 41(3), 341-349.

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*. 15 (1), 450-461.

Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *TICS*. 12 (8), 314.

Shahin, A., Bosnyak, D. J., Trainor, L. J., Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*. 23 (13), 5545-5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*. 16 (16), 1781-1785.

Slevc, L. R., Miyake, A. (2006). Individual differences in second- language proficiency: does musical ability matter? *Psychological Science*. 17 (8), 675-681.

Sloboda, J. (2005). *Exploring the musical mind: cognition, emotion, ability, function*. Oxford University Press, p 269.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *Neuroimage*. 17 (3), 1613-1622.

Snyder, J. S., Alain, C., Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*. 18 (1), 1-13.

Tallal, P., Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends In Neuroscience*. 29 (7), 382-390.

Tallon-Baudry, C., Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *TICS*. 3 (4), 151-162.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*. 13, 10.

Tervaniemi, M., Rytkönen, M., Schröger, E., Ilmoniemi, R. J., Näätänen, R. (2001). Superior

formation of cortical memory traces for melodic patterns in musicians. *Learning and Memory*. 8, 295-300.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental Brain Research*. 171, 1-10.

Tervaniemi, M., Castaneda, A., Knoll, M., Uther, M. (2006). Sound processing in amateur musicians and non-musicians: Event-related potential and behavioral indices. *Neuroreport*. 17, 1225-1228.

Thiessen, E. D., Hill, E., Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*. 7, 53.

Tillmann, B., McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)-similarities. *J Exp Psychol Learn Mem Cogn*. 30, 1131-1142.

Tillmann, B., Poulin-Charronnat, B. (2009). Auditory expectations for newly acquired structures. *The Quarterly Journal of Experimental Psychology*. 63 (8), 1646-1664.

Toro, J. M., Trobalón, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception and Psychophysics*. 67 (5), 867.

Toro, J. M., Trobalón, J. B. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*. 97 (2), B25-34.

Trainor, L. J., Clark, E. D., Huntley, A., Adams, B. A. (1997). The acoustic basis of preferences for infant-directed singing. *Infant behavior and development*. 20, 383-396.

Trainor, L. J., Austin, C. M., Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological science*. 11, 188-195.

Trainor, L. J., McDonald, K. L., Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive*

Neuroscience. 14, 1-13.

Tremblay, K., Kraus, N., McGee, T. (1998) The time course of auditory perceptual learning: neurophysiological changes during speech-sound training. *Neuroreport*. 9 (16), 3557.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., Otis, B. (2001). Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear and hearing*. 22(2), 79-90.

Van Petten, C., Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition*. 18 (4), 380-393.

Van Zuijen, T., Sussman, E., Winkler, I., Näätänen, R., Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity – a mismatch negativity study sounds comparing musicians and non-musicians. *Cognitive Brain Research*. 23, 270-276.

Vaughn, K. (2000) Music and mathematics: Modest support for the oft-claimed relationship. *J Aesthet Educ*. 34, 149-166.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, 30(4), 1414-32.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstor, V. A.(2005). To musicians, the message is in the meter - preattentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage*. 24, 560-564.

Wan, C. Y., Schlaug, G (2010). Music making as a toll for promoting brain plasticity across the life span. *Neuroscientist*. 16 (5), 566-577.

Woldorff, M. G, Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., Bloom, F. E. (1993). *PNAS*. 90 (18), 8722-8726.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*. 10, 420-422.

Young, M. P., Rugg, M. D. (1992). Word frequency and multiple repetition as determinants of

the modulation of event-related-potentials in a semantic classification task. *Psychophysiology*. 29 (6), 664-676.

Yvert, B., Fischer, C., Bertrand, O., Pernier, J. (2005). Localization of human supratemporal auditory areas from intracerebral auditory evoked potentials using distributed source models. *Neuroimage*. 28 (1), 140-153.