Université de Montréal

Corrélats neuronaux de l'expertise auditive

par

Jean-Pierre Chartrand

Département de psychologie
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Doctorat en Psychologie
Option neuropsychologie clinique

8 Juillet, 2010

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée :

Corrélats neuronaux de l'expertise auditive

Présentée par :
Jean-Pierre Chartrand

a été évaluée par un jury composé des personnes suivantes :

Franco Lepore, président-rapporteur
et représentant du doyen de la FES
Pascal Belin, directeur de recherche
Isabelle Peretz, co-directrice
Sylvie Hébert, membre du jury
Claude Alain, examinateur externe

Résumé

La voix humaine constitue la partie dominante de notre environnement auditif. Non seulement les humains utilisent-ils la voix pour la parole, mais ils sont tout aussi habiles pour en extraire une multitude d'informations pertinentes sur le locuteur. Cette expertise universelle pour la voix humaine se reflète dans la présence d'aires préférentielles à celle-ci le long des sillons temporaux supérieurs. À ce jour, peu de données nous informent sur la nature et le développement de cette réponse sélective à la voix. Dans le domaine visuel, une vaste littérature aborde une problématique semblable en ce qui a trait à la perception des visages. L'étude d'experts visuels a permis de dégager les processus et régions impliqués dans leur expertise et a démontré une forte ressemblance avec ceux utilisés pour les visages.

Dans le domaine auditif, très peu d'études se sont penchées sur la comparaison entre l'expertise pour la voix et d'autres catégories auditives, alors que ces comparaisons pourraient contribuer à une meilleure compréhension de la perception vocale et auditive. La présente thèse a pour dessein de préciser la spécificité des processus et régions impliqués dans le traitement de la voix. Pour ce faire, le recrutement de différents types d'experts ainsi que l'utilisation de différentes méthodes expérimentales ont été préconisés.

La première étude a évalué l'influence d'une expertise musicale sur le traitement de la voix humaine, à l'aide de tâches comportementales de discrimination de voix et d'instruments de musique. Les résultats ont démontré que les musiciens amateurs étaient meilleurs que les non-musiciens pour discriminer des timbres d'instruments de musique mais aussi les voix humaines, suggérant une généralisation des apprentissages perceptifs causés par la pratique musicale. La seconde étude avait pour but de comparer les potentiels évoqués auditifs liés aux chants d'oiseaux entre des ornithologues amateurs et des participants novices. L'observation d'une distribution topographique différente chez les ornithologues à la présentation des trois catégories sonores (voix, chants d'oiseaux, sons de l'environnement) a rendu les résultats difficiles à interpréter. Dans la troisième étude, il était question de préciser le rôle des aires temporales de la voix dans le traitement de catégories d'expertise chez deux groupes d'experts auditifs, soit des ornithologues amateurs et des luthiers. Les données comportementales ont démontré une interaction entre les deux groupes d'experts et leur catégorie d'expertise respective pour des tâches de discrimination et de mémorisation. Les résultats obtenus en imagerie par résonance magnétique fonctionnelle ont démontré une interaction du même type dans le sillon temporal supérieur gauche et le gyrus cingulaire postérieur gauche. Ainsi, les aires de la voix sont impliquées dans le traitement de stimuli d'expertise dans deux groupes d'experts auditifs différents. Ce résultat suggère que la sélectivité à la voix humaine, telle que retrouvée dans les sillons temporaux supérieurs, pourrait être expliquée par une exposition prolongée à ces stimuli.

Les données présentées démontrent plusieurs similitudes comportementales et anatomo-fonctionnelles entre le traitement de la voix et d'autres catégories d'expertise. Ces aspects communs sont explicables par une organisation à la fois fonctionnelle et économique du cerveau. Par conséquent, le traitement de la voix et d'autres catégories sonores se baserait sur les mêmes réseaux neuronaux, sauf en cas de traitement plus poussé. Cette interprétation s'avère particulièrement importante pour proposer une approche intégrative quant à la spécificité du traitement de la voix.

Mots-clés : voix, expertise, audition, musiciens, timbre, imagerie cérébrale.

Abstract

The human voice is the most meaningful sound category of our auditory environment. Not only is the human voice the carrier of speech, but it is also used to extract a wealth of relevant information on the speaker. Voice-sensitive areas have been identified along the superior temporal sulci of normal adult listeners. Yet little data is available on the nature and development of this selective response to voice. In the visual domain, a vast literature focuses on a similar problem regarding face perception. Several studies have identified processes and regions involved in visual expertise, demonstrating a strong resemblance to those used for faces.

In the auditory domain, very few studies have compared voice expertise to expertise for other sound categories. Such comparisons could contribute to a better understanding of voice perception and hearing. This thesis aims to clarify the nature of the processes and regions involved in voice perception. Different types of experts and different experimental methods were used in three separate studies.

The first study assessed the influence of musical expertise on timbre voice processing, by using using behavioral voice and musical instrument discrimination tasks. The results showed that amateur musicians performed better than non-musicians in both tasks, suggesting a generalization of auditory abilities associated with musical practice. The second study compared event related potentials evoked by birdsongs in bird experts and non-expert participants. Because a different topographical distribution was observed among bird experts in all sound categories, a definitive interpretation was difficult to make.

In the third study, we asked whether the voice-sensitive areas would be recruited by different categories of sounds of expertise in guitar makers, bird experts and non-experts. The behavioral data showed an interaction between the two groups of experts and their respective category of expertise for memory and discrimination tasks. The functional magnetic resonance imaging results showed an interaction of the same type in the left superior temporal sulcus and the left posterior cingulate gyrus. The results show that the voice selective areas do not exclusively process voice stimuli but could also contribute to expert-level processing of other sound categories. Therefore, cortical selectivity to human voice could be due to a prolonged exposure to voice.

The data presented demonstrate several behavioral and anatomo-functional similarities between cerebral voice processing and other types of auditory expertise. These common aspects can be explained by a functional and economical brain organization. Consequently, sound processing would rely on shared neural networks unless necessary. This interpretation is particularly important to suggest an integrative approach for studying voice processing specificity.

Keywords : voice, expertise, audition, musicians, timbre, neuroimaging.

Table des matières

Liste des tableaux

Article 4 : Auditory expertise recruits voice-selective auditory cortex

Liste des figures

Article 1 : Auditory recognition expertise and domain specificity

Article 2 : Superior voice timbre processing in musicians

Article 3 : Brain response to birdsongs in bird experts

Article 4 : Auditory expertise recruits voice-selective auditory cortex

Liste des sigles et des abréviations

ANOVA	*Analysis of variance*
EEG	*Electroencephalography*
IRMf	Imagerie par résonance magnétique fonctionnelle
fMRI	*Functional magnetic resonance imaging*
FFA	*Fusiform face area*
GCPg	Gyrus cingulaire postérieur gauche
GCP	Gyrus cingulaire postérieur
MEG	Magnétoencéphalographie / *Magnetoencephalography*
MS	Millisecondes / *Milliseconds*
S	Secondes / *Seconds*
SE	*Standard error*
STG	*Superior temporal gyrus*
STS	Sillon temporal supérieur / *Superior temporal sulcus*
TEP	Tomographie par émission de positons
TVA	*Temporal voice area*

À mes parents, sans qui rien de tout ceci
n’aurait été possible.

Remerciements

Au cours de la réalisation de ma thèse, les mots d'encouragements furent très nombreux. Cependant, je retiens surtout les appuis authentiques que j'ai reçus. Cette page est donc dédiée aux personnes qui ont su, au fil des ans, m'appuyer de la manière la plus franche qui soit.

Ce document se voulant être une thèse, je ne peux commencer qu'en remerciant la personne qui m'a à la fois permis de la commencer...et de la terminer! J'offre mes plus sincères remerciements à Pascal, qui m'a accueilli dans son jeune laboratoire alors que j'étais moi-même débutant. Malgré la distance physique qui nous a séparés pendant une bonne partie de la réalisation de cette thèse, j'ai toujours senti qu'il était derrière moi et que mon projet faisait partie d'un travail d'équipe. Parmi ses qualités qui m'auront marqué tout au long du doctorat, je retiens surtout sa confiance, son enthousiasme et son empressement fou à essayer de nouvelles idées. J'espère bien un jour intégrer ces belles qualités (mais à plus faible dose, sans offense bien sûr!).

Je dois aussi remercier Isabelle, ma co-directrice, qui s'est chargée de m'intégrer à son équipe au moment où je me suis retrouvé orphelin de mon laboratoire vocal. D'un ton calme et rassurant, elle s'est toujours montrée disponible pour m'orienter et me rassurer dans plusieurs étapes cruciales de mon projet de thèse. En tant que co-directrice, elle possédait la balance du pouvoir, et ce fut très apprécié.

Mes remerciements s'adressent ensuite à mes parents. Pour des parents, il est peut-être difficile de comprendre adéquatement le parcours de son enfant aux études graduées. Tout ce que je peux dire, c'est qu'ils m'ont toujours offert un appui inconditionnel. Par ailleurs, ces remerciements ne pourraient être complets sans la mention des valeurs familiales qui m'ont influencé de près ou de loin dans la réalisation de ce projet. Provenant de ma mère, il s'agit de son sens de l'organisation et de la prévision sans égal alors que chez mon père, il s'agit de son optimisme candide et puissant. À eux deux, ils ont réussi à me transmettre un désir de vaincre l'adversité, peu importe la forme qu'elle peut prendre.

L'immensité de la reconnaissance et de la redevance que ces influences m'inspirent ne pourrait être traduite par de simples mots.

Puisqu'il est question de souligner les contributions familiales, je dois inévitablement mentionner mon chère frère Jean-Luc. Il s'agit de la personne qui me permet continuellement de conserver une attitude ludique, jeune et rebelle, malgré l'apparition récente de quelques cheveux blancs que je peux compter sur les doigts de ma main. La quantité de spectacles musicaux que nous sommes allés voir au cours des dernières années est un bon exemple des activités « frèresques » qui m'auront donné l'occasion de décrocher à plusieurs moments critiques!

Je dois aussi beaucoup à ma copine Julie, avec laquelle je partage un parcours très similaire depuis le début du baccalauréat. Dès nos premières rencontres, je me doutais que nos attitudes semblables nous mèneraient loin. Un doctorat plus tard, j'ai toujours la même conviction. Un gros merci à celle qui croit en mes idées et qui me supporte dans celles qui lui plaisent moins! Du même côté de la famille, je lance un petit clin d'oeil à mes beaux-parents, qui m'ont toujours offert un appui empathique, ayant eux aussi traversé la même épreuve académique.

Je voudrais remercier particulièrement certains de mes amis proches, dont Karl, René et Guylaine, avec qui j'ai pu partager mes bonnes et moins bonnes expériences. Je remercie aussi tous les autres amis qui estiment que leur contribution devrait être souligée ici. Je garde de très bons souvenirs des étudiants et collègues du BRAMS et du CERNEC que j'ai eu la chance de côtoyer de près ou de loin. Je suis persuadé qu'il sera très difficile pour moi de retrouver un milieu de travail aussi riche et stimulant. J'en profite pour souhaiter bon courage aux personnes qui lisent ce texte en s'imaginant l'impossible : la fin de leur propre thèse!

Finalement, j'aimerais aussi remercier un objet inanimé qui m'a suivi tout au long de mon doctorat et qui ne m'a jamais laissé tomber : mon ordinateur portable. Ayant accusé de durs coups physiques et subi une variété de climats, ce dernier s'est toujours porté garant de mes requêtes. Merci Thinkpad T42!

Introduction générale

La voix humaine. Cette composante prédominante de notre environnement auditif est principalement associée à la production du langage. Bien que porteuse de la parole, la voix humaine comporte plusieurs informations importantes qui permettent d'identifier notre locuteur, son âge, son sexe ou son état affectif. La présence d'une multitude d'informations contenues dans la voix fait de celle-ci un véritable « visage auditif » (Belin, Bédard, & Fecteau, 2004). D'autre part, le cerveau humain possède une capacité exemplaire pour extraire ces informations. À titre d'exemple, c'est avec aisance que l'on peut rapidement reconnaître la voix d'un pair ou détecter l'efficacité d'une blague dans le rire d'une amie. Considérant ces habiletés du cerveau qui sont propres à la voix humaine, on pourrait qualifier les humains d' « experts » en voix. Une question pourtant non-résolue demeure dans la nature et le développement de ces habiletés pour la voix humaine. Une meilleure compréhension de ce phénomène aurait des répercussions importantes sur notre conception du système auditif et de ses mécanismes, tant sur le plan comportemental que cérébral. C'est en se basant sur la notion que tous les humains sont experts pour traiter la voix humaine que la présente thèse tentera d'éclaircir cette question.

La voix humaine prend une place considérable dans notre environnement sonore, et ce depuis le tout jeune âge. En effet, la voix de la mère est un signal acoustique prépondérant dans l'environnement amniotique (Fifer & Moon, 1994). Très rapidement, les nouveau-nés démontrent des signes physiques de préférence à la voix de leur mère, tels qu'un ralentissement du battement de coeur ou une augmentation du débit de la tétée (DeCasper & Fifer, 1980; Fifer & Moon, 1994). Des données plus récentes indiqueraient même une préférence pour la voix de la mère avant la naissance, chez le foetus à terme (Kisilevski et al., 2003). Cette prise de contact très précoce avec la voix humaine, de même qu'une exposition répétée à celle-ci, devraient se traduire par la présence de marqueurs dans les aires auditives cérébrales.

En effet, depuis une dizaine d'années, on reconnaît la présence d'aires préférentielles à la voix dans les lobes temporaux chez l'humain (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin, Zatorre, & Ahad, 2002). Plus précisément, on observe une plus grande activité cérébrale au niveau des sillons temporaux supérieurs bilatéraux lorsqu'on présente des extraits de voix humaine, comparativement à des sons non-vocaux naturels ou à des sons de voix humaine qui ont été embrouillés. La nature et le développement de cette réponse sélective à la voix le long des sillons temporaux supérieurs demeurent cependant peu explorés. À ce jour, peu de données nous informent sur la cause de cette spécialisation du cerveau pour la voix (Belin, Bédard, & Fecteau, 2004).

Une version analogue de cette problématique a déjà été abordée dans le domaine visuel, en référence à la spécificité du traitement des visages par le gyrus fusiforme (Kanwisher, McDermott, & Chun, 1997). La présence de données divergentes a favorisé le maintient de deux camps très polarisés (Gauthier & Bukach, 2007). D'un côté, un groupe estime que le gyrus fusiforme répond sélectivement aux visages et qu'il contient des processus qui sont spécifiques à ceux-ci (Kanwisher, 2000; Kanwisher & Yovel, 2006). À l'autre pôle, on croit que le gyrus fusiforme serait impliqué dans tout traitement de catégorie visuelle homogène pour laquelle un individu aurait été suffisamment exposé pour développer une expertise visuelle (Gauthier, Skudlarski, Gore, & Anderson, 2000; Gauthier & Bukach, 2007).

Ce débat concernant la spécificité des visages, couramment surnommé « problème de l'expertise », est nourri par plusieurs études comportementales et d'imagerie cérébrale comparant l'expertise visuelle acquise dans d'autres domaines à celle des visages, laquelle serait universelle chez la plupart des humains (voir Kanwisher & Yovel (2006) pour une revue de la littérature). Dans le cas de la perception de la voix, on pourrait également considérer que la grande majorité des individus seraient des experts pour en extraire différentes informations. Contrairement au domaine visuel, il semblerait y avoir très peu d'études comparant directement l'expertise auditive et l'expertise pour la voix humaine. Certaines études ont par contre examiné l'influence de l'expertise musicale sur la perception de différents aspects langagiers (Besson, Schön, Moreno, Santos, & Magne, 2007; Lee & Hung, 2008; Nikjeh, Lister, & Frisch, 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007). Les données suggèreraient un transfert des habiletés acquises à travers la pratique musicale dans le traitement du langage, démontrant que la spécificité du langage peut être comprise à partir d'une expertise auditive telle que l'expertise musicale. Selon nous, l’expertise auditive serait tout aussi pertinente pour mieux comprendre la spécificité de la perception de la voix.

Présentation des objectifs de recherche et des études

Le but de la présente thèse est de préciser la spécificité des processus et régions impliqués dans le traitement de la voix. Comme il a été fait en vision, nous estimons qu’il serait pertinent de recruter des experts qui ont été exposés à une catégorie homogène de stimuli auditifs pendant une longue période, et dont l’expérience serait davantage comparable à celle que l’on retrouve pour la voix humaine plutôt qu’à d’autres catégories de sons de l’environnement. Par exemple, la capacité de certains ornithologues amateurs d’identifier plus de 200 chants d’oiseaux différents est similaire à la capacité des êtres humains à identifier leurs pairs par le timbre de leur voix. La comparaison des processus et régions impliqués dans l’expertise auditive avec ceux impliqués dans le traitement de la voix permettrait de raffiner nos connaissances sur le fonctionnement des aires auditives et d’évaluer l’effet de l’expérience sur le développement de celles-ci. Chacune des sections qui suivent sera présentée sous forme d’article.

Premier article : perception de la voix et spécificité

Dans ce premier article de recension de la littérature, nous aborderons tout d’abord la problématique de l’expertise en faisant un résumé de la littérature dans le domaine visuel. Il sera ensuite question de la perception de la voix et les différentes expertises auditives. Finalement, nous proposerons qu’une question telle que « La voix est-elle spéciale? » peut être investiguée en partie avec l’étude de l’expertise auditive.

Deuxième article : discrimination du timbre chez les musiciens

Bien que la majeure partie de cette thèse soit orientée sur les méthodes d’imagerie cérébrale, le deuxième article abordera la question de la spécificité de la voix à l’aide de données comportementales. Dans l’étude qui y est décrite, nous évaluerons en quoi une expertise musicale peut influencer le traitement de la voix humaine, à l’aide de tâches de discrimination de voix humaines et d’instruments de musique.

Hypothèses de la première étude: Une expérience en pratique musicale devrait avoir un effet positif sur le traitement non-langagier de la voix humaine. En d'autres termes, les musiciens amateurs recrutés devraient être meilleurs que les non-musiciens pour traiter des timbres d'instruments de musique et de voix humaine. Cette hypothèse est formulée à partir des données qui ont démontré l'avantage des musiciens dans le traitement de matériel non-musical (e.g., Münzer, Berti, & Pechmann, 2002).

Troisième article : potentiels évoqués chez les ornithologues amateurs

S'inspirant des données en électroencéphalographie (ÉEG) et en magnétoencéphalographie (MEG) démontrant de plus grandes ondes auditives reliées à l'instrument de pratique chez les musiciens (Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Shahin, Bosnyak, Trainor, & Roberts, 2003), le troisième article aura comme objectif d'identifier et de comparer les potentiels évoqués liés aux chants d'oiseaux chez des ornithologues amateurs et des participants novices. Les données récoltées proviendront de la présentation de chants d'oiseaux, de voix humaines et de sons de l'environnement lors d'une tâche de détection de cible de type *oddball*.

Hypothèses de la deuxième étude : La perception de chants d'oiseaux chez les ornithologues amateurs devrait provoquer de plus grandes ondes auditives P2, comparativement aux participants novices.

Quatrième article : rôle des aires de la voix dans l'expertise auditive

Finalement, l'étude présentée dans le quatrième article s'inspire fortement d'un des articles qui a lancé le débat de l'expertise dans le domaine visuel (i.e., Gauthier, Skudlarski, Core, & Anderson, 2000). Dans cette étude, il sera question de préciser le rôle des aires temporales de la voix dans le traitement de catégories d'expertise chez deux groupes d'experts auditifs, notamment des ornithologues amateurs et des luthiers. Ces deux populations sont choisies en fonction de pratiques auditives qui selon nous ressemblent à celles couramment utilisées pour la voix humaine. De plus, le recrutement de deux groupes d'experts permet d'avoir un plan symétrique : chaque groupe d'expert sera comparé à l’autre groupe qui sera alors considéré novice. Cette étude comportera une partie comportementale et une partie en imagerie par résonance magnétique fonctionnelle (IRMf).

Hypothèses de la troisième étude: Une interaction entre le groupe d’expertise et la catégorie d’expertise devrait être décelée, tant au niveau des tâches comportementales qu’au niveau des activations cérébrales qui répondent habituellement à la voix (STS). Plus précisément, les ornithologues amateurs devraient être meilleurs que les luthiers pour la mémorisation et la discrimination des chants d’oiseaux. Cette supériorité comportementale devrait se traduire en une plus grande réponse corticale aux chants d’oiseaux qu’aux sons de guitares. Les résultats inverses sont attendus pour les luthiers.

Articles théoriques et empiriques

Article 1 : Auditory recognition expertise and domain specificity

Jean-Pierre Chartrand, Isabelle Peretz et Pascal Belin

Publié dans Brain Research (2008), volume 1220, pages 191 à 198.

Auditory recognition expertise and domain specificity

Jean-Pierre Chartrand[1], Isabelle Peretz[1], and Pascal Belin[1 2].

[1]BRAMS, Université de Montréal, Montréal, Canada

[2]Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

Correspondence and requests for reprints to

Jean-Pierre Chartrand,

Département de Psychologie, C.P. 6128, succursale Centre-Ville,

Montréal, Québec, Canada, H3C 3J7

Tel : + 1 1 (514) 343-6111 ext. 2995; fax: +1 1 (514) 343-2175;

4363 words, 18 text pages, 2 figures

Abstract

Auditory recognition expertise refers to one's ability to accurately and rapidly identify individual sound sources within a homogeneous class of stimuli. Compared to the study of visual expertise, the field of expertise in sound source recognition has been neglected. Different types of visual experts have been studied extensively both in behavioral and neuroimaging studies, leading to a vigorous debate about the domain specificity of face perception. In the present paper, we briefly review what is known about visual expertise and propose that the same framework can be used in the auditory domain to ask the question of domain specificity for the processing and neural correlates of the human voice. We suggest that questions like “are voices special ?” can be partially answered with neuroimaging studies of “auditory experts”, such as musicians and bird experts, who rely on subtle acoustical parameters to identify auditory exemplars at a subordinate level. Future studies of auditory experts can not only serve to answer questions related to the neural correlates of voice perception, but also broaden the understanding of the auditory system.

Section: 7. Cognitive and Behavioral Neuroscience
Keywords: auditory expertise, bird experts, musicians, music, voice

1. Introduction

Every once in a while, we happen to meet people who excel in specific activities, fields of knowledge or cognitive tasks. These people, called experts, have been trained for years in order to be faster or be more proficient in their domain of expertise. This fascination for people who have developed specific skills over time is witnessed by the numerous studies conducted in cognitive, motor and especially vision research. Radiologists, chick sexers, bird, dog, and car experts all seem to have behavioral similarities with human face experts (i.e., virtually all healthy adults). They all seem to have both qualitative and quantitative advantages over novices in recognizing individual exemplars of the visual category of their expertise. Processing these different classes of stimuli of expertise might rely on similar processes, and consequently, on similar cortical regions. This possibility raises questions regarding the modularity of face processing because it would entail that faces are not special (Nachson, 1995). This has led to hotly debated and polarized opinions about the uniqueness of face recognition and its neural correlates. Proponents of the domain-specificity hypothesis argue that face perception is carried out by highly specialized processes and dedicated brain regions (Kanwisher, 2000). Opponents to this hypothesis adhere instead to the expertise hypothesis by which faces recruit general mechanisms that are fine-tuned by experience (Gauthier and Bukach, 2007). No comparable research effort has taken place in the auditory domain, even though it has been suggested that the human voice is an ‘auditory face’ (Belin et al., 2004), and that specific cortical regions are involved in voice processing (Belin et al., 2002). Questions about cognitive and anatomical modularity of voice processing have only been raised recently (Belin et al., 2004). Questions such as ‘in what way are voices special?’ and ‘is voice expertise a simple exposure and learning effect?’ can shed light on the development and functioning of the auditory system. Indeed, studying different types of auditory experts can open new avenues for understanding the modularity of auditory and voice perception. However, we will first examine the visual expertise literature.

2. The visual expertise debate

2.1 Functional modularity for visual objects

The most studied form of visual expertise is face processing. Face recognition is a relatively universal domain of expertise in humans ((Carey, 1992) and (Diamond and Carey, 1986)). Different behavioral studies show that face recognition is somewhat different from object recognition. For instance, when face recognition performance is compared to recognition performance for other categories of objects, participants usually identify faces more frequently and as quickly at the subordinate level as at the basic level (e.g., Bill Clinton vs. a human) (Tanaka, 2001). Subordinate-level categories are more informative about the objects they represent, but different exemplars categorized at a subordinate level are more similar to one another and require more effort to discriminate between them (Tanaka and Gauthier, 1997). For example, two human face exemplars offer more information if they are categorized at a subordinate level, but are more difficult to discriminate (e.g. Stephen Harper vs George W. Bush). Basic-level categories, inversely, offer less information but are much more discernible. For example, the “human” category is easily distinguished from another living being category, such as “dog”, but contains less information about facial features of its exemplars. This preference for the use of subordinate-level labels for human faces can be linked to the daily need to identify people by their unique identity rather than by their basic category (e.g. faces or humans). Non-facial categories can also be preferentially processed at the subordinate level by visual experts, hence supporting the expertise hypothesis. Tanaka and Taylor (1991) have shown that bird and dog experts use subordinate-level categories and names to describe birds and dogs, whereas novices tend to use basic levels (e.g., dog for novices vs. Bloodhound for dog experts). Moreover, experts were as fast when they used subordinate-level as when they used basic-level categorizations, contrarily to the novices who were faster for the basic-level names.

Another difference between face and object recognition is demonstrated by the face-inversion effect (Yin, 1969). This effect suggests that face-specific mechanisms, as opposed to other visual mechanisms involved in object recognition, are more disrupted when visual stimuli are turned upside-down. Inversion effects have also been observed in dog experts (Diamond and Carey, 1986), suggesting that mechanisms considered being face-specific could also be involved in expertise object recognition.

Another important question in the expertise debate is whether novices can be trained to become experts. This question is of primary interest, because it addresses the effect of training and exposure on behavioral markers of expertise. In fact, novices can be trained in the laboratory to recognize a category of novel and non-natural objects, such as “greebles” ((Gauthier and Tarr, 1997) and (Gauthier et al., 1998)). “Greebles” are 3D-rendered objects that share a common configuration. They all have a vertically-oriented body with four protruding parts, and can be categorized into different classes that are referred to as “gender” and “family”. Gauthier and Tarr (1997) trained participants to recognize 30 different greebles at three levels of categorization: family, gender, and individual levels. These participants were then tested to identify upright and inverted Greeble parts in three conditions: studied-configuration, transformed-configuration or isolated part. The results demonstrated that experts who had extensive practice with a homogeneous object category such as Greebles showed recognition effects usually associated with faces, such as being more sensitive to configuration changes.

In sum, expertise effects have been observed with visual objects such as cars, birds, faces and greebles. These data seem to support that face and object processing expertise involve the same mechanisms. However, it may be that different classes of objects involve distinct mechanisms that produce similar configuration effects. Brain imaging studies provide some additional information in this regard.

2.2. Anatomical modularity for object perception

Several functional neuroimaging studies have identified a cortical region, called the Fusiform Face Area (FFA) that responds preferentially and consistently to faces (e.g., Kanwisher et al., 1997). In subsequent work, this region has been shown to be also involved in processing objects of expertise other than faces. For example, Gauthier et al. (2000) reported that objects of visual expertise (birds and cars) activate the right FFA more strongly than non-expertise stimuli in bird and car experts. There was also a high correlation between a behavioral test of object expertise and the relative activation of the FFA for birds and cars. This led the authors to conclude that the FFA would be involved in visual expertise, rather than in face processing per se. While Xu (2005) has replicated this finding, other studies have obtained non-significant trends for FFA activations with Lepidoptera experts (Rhodes et al., 2004) or negative results with car and 3D artificial object experts ((Grill-Spector et al., 2004) and (Yue et al., 2006)). Moreover, it can be argued that face-specific processes can be recruited for face-like stimuli after training, since birds and cars could be visually interpreted with face-like configurations (Kanwisher, 2000).

It is worth mentioning that conventional fMRI procedures have a limited spatial resolution. This has a great impact when comparing regions as small as the FFA, which is represented by a small number of voxels. Using high resolution fMRI, a recent study conducted by Grill-Spector et al. (2006) (corrigendum: Grill-Spector et al., 2007) showed that the FFA is a heterogeneous region of higher and lower face-selective patches. Thus, previous studies asserting the presence of a highly selective face module were in fact describing averages of high and lower face-selective subregions.

Since the current data do not support either the domain-specificity or the expertise hypothesis, we may provisionally assume that face and objects of expertise rely on similar cortical regions and processes. The FFA controversy primarily questions the existence of a face-dedicated brain module. Yet, the visual expertise framework addresses broader questions about the development and plasticity of the visual system (Bukach et al., 2006). By studying greeble laboratory training, Gauthier et al. (1999) observed that training-induced expertise with greebles led to increased right FFA activations. These results suggest that expertise can in part explain the specialization of the FFA for faces. The surprising part is that only seven hours of training were sufficient to lead to functional cortical changes. This is an example of how researchers in the field of visual expertise have developed interesting methods and paradigms that can be used to further study the learning visual brain. These methods can also be borrowed and adapted to research in the auditory domain to study expertise. Although research in auditory recognition expertise is much less advanced, let us summarize what is currently known about auditory experts.

3. Voice perception

It has been proposed that we are all experts at face recognition. The same can be suggested about our ability to recognize voices, although forensic studies on earwitness identification show that voices are poorly recognized ((Hollien et al., 1983) and (Olsson et al., 1998)). Nevertheless, the human voice is probably the most meaningful sound category of our auditory environment. Contained in the human voice timbre are information that enable us to identify a speaker's gender (Mullennix et al., 1995), age (Linville, 1996), and emotional state (Scherer, 1995). These types of information are also present in human faces, but on some occasions they may not be available, such as when speaking to someone on the phone. Thus, the voice can be considered as an “auditory face” (Belin et al., 2004), from which rich linguistic and non-linguistic information can be extracted.

In addition to these considerations, cortical regions along the right superior temporal sulcus (rSTS) have been shown to respond preferentially to human voices over non-vocal sounds and scrambled non-speech vocal sounds ((Belin et al., 2000) and (Belin et al., 2002)). These bilateral regions, also called temporal voice areas (TVA), are spatially limited across participants, suggesting a clear anatomo-functional regionalization (Fig. 1). Clinical studies have also confirmed that voice discrimination or recognition can be impaired mostly after right-hemisphere lesions ((Peretz et al., 1994) and (Van Lancker and Kreiman, 1987)). These studies lead us to believe that there exists a form of voice expertise comparable to face recognition expertise. Could there be, as in the visual domain, other types of “voice” or sound experts?

Little is known about auditory experts, other than musicians, but there exist people who are highly trained to identify certain irregularities in the production of voice. Speech therapists and voice pathologists need to develop special skills to assess and diagnose voice pathologies. Kreiman et al. (1993) suggest that these professionals assess pathological voices using mental representations that they develop through their careers. Consequently these representations might vary across them, being influenced by internal and external factors such as memory, attention, and the acoustical context of the task ((Kreiman et al., 1992) and (Gerratt et al., 1993)). These professionals also differ in their use of strategies (Kreiman et al., 1992). Bhuta et al. (2004) assessed whether there was a correlation between subjective voice assessment (GRBAS scale) and objective acoustical voice assessment. Out of 19 acoustical parameters, only three noise parameters correlated with the perceptual voice analysis scale. The authors suggested that perhaps noise is the information voice professionals perceive in a dysphonic voice. No studies have assessed the link between non-pathological voice characteristics and individuals' strategies and performance in different conditions (Kreiman, 1997). Thus, the study of voice experts does not provide much insight for understanding voice or auditory recognition expertise.

4. Auditory experts

Expertise in object recognition has been studied mainly in the visual domain, but some visual objects, such as birds and musical instruments, can also be recognized through audition. Bird experts and musicians can be considered as “auditory experts”, relying on unique sound information to identify individual sound sources at a subordinate level. Bird experts have developed excellent skills at recognizing birds visually, but sometimes they might just hear a bird and not even see it. In order to identify these non-visible birds, they need to identify them by their songs, or calls in some instances. Just as the visual objects they represent, birdsongs are different across species, and even sometimes inside the same species. Bird experts have thus developed the ability to identify birdsongs in an auditory environment. The same can be said about musicians, when they need to identify a specific musical instrument in an orchestra for example. Much information is available about these two groups, because bird experts have been studied in vision research ((Tanaka and Taylor, 1991) and (Gauthier et al., 2000)), and musicians have been studied extensively for their musical and auditory skills.

No sample of auditory experts has been studied as extensively as musicians. Several reasons can explain this. First, they are easily available, and differ in levels of training. Some of them are trained at an academic level, which translates into a more uniform motor, cognitive, perceptual and conceptual knowledge across professionals. Other musicians, who have been trained at an amateur level, can differ from non-musicians due to years of practice (e.g. Tervaniemi et al., 2006). Also, questions about the development of expertise can be raised, since children can start learning a musical instrument as young as four years old. Musicians can be easily compared to novices, that is, non-musicians, who are capable of sophisticated listening strategies although they are not professionally trained (for a review, see Bigand and Poulin-Charronnat, 2006). In sum, both the homogeneity and heterogeneity of abilities and practices found in musicians are important reasons that make them an interesting sample of auditory experts. Musicians have been largely studied with musical stimuli, comparing their musical skills to non-musicians. These skills include superior pitch, temporal and also spatial processing in music conductors (Münte et al., 2001).

Musicians are known to have better pitch discrimination than non-musicians. Pitch discrimination studies generally demonstrate that trained musicians have a frequency discrimination threshold about half the size of non-musicians' ((Spiegel and Watson, 1981) and (Kishon-Rabin et al., 2001)). Moreover, if the participants are classical instrumentists with over ten years of practice, the observed threshold difference between musicians and non-musicians is multiplied by a factor of two (Micheyl et al., 2006). Another pitch-related ability influenced by musical training is absolute pitch. Absolute pitch is the ability to easily identify the pitch of a tone without any prior reference. The development of this ability seems to require both genetic factors and early musical training (Baharloo et al., 1998). These data demonstrate that musical background and practice are intimately associated with pitch discrimination and recognition in musicians. Musicians are also more accurate at processing temporal information. More precisely, they are superior at processing immediate temporal information, as opposed to temporal tasks that involve encoding of an interval stimulus in long-term memory (Rammsayer and Altenmuller, 2006). Finally, music conductors show improved auditory localization in the peripheral space, compared to musicians and non-musicians ((Münte et al., 2001) and (Nager et al., 2003)). Interestingly, congenitally and late-onset blind people also present such improved auditory spatial abilities ((Röder et al., 1999) and (Fieger et al., 2006)). Moreover, people who were blinded in infancy show better pitch direction judgement than late-onset blind people and control participants (Gougoux et al., 2004). Usually, factors such as the need to orient with sounds and reorganization of the visual cortex are used to explain such observations in blind people. These people might have developed such abilities with different types of stimuli in their everyday lives. On the other hand, musicians are of particular interest to the present topic because they gained their training with a restricted class of auditory stimuli. In this case, a pertinent question to ask is whether the skills acquired in the context of musical training can transfer to categories of stimuli other than musical instruments.

4.1 Transfer of auditory abilities in musicians

A few studies have assessed musicians' performance in perceptual tasks not exclusive to music, such as pitch and timbre discrimination. These studies focus on musicians' transfer of auditory skills to nonmusical stimuli. Such studies have investigated musicians' ability to decode speech prosody ((Magne et al., 2006), (Schön et al., 2004) and (Thompson et al., 2004)) and detect pitch violations in foreign languages (Marques et al., 2007). These studies show evidence for musicians' advantage in processing pitch and contour both in music and language. The results suggest the existence of shared processes in music and language.

Central to the auditory recognition expertise problem is the perception of timbre, because it is the most useful acoustical parameter that musicians and non-musicians alike use to recognize musical instruments. Timbre can be defined as an “attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different” (ANSI, 1973). For instance, a violin tone can be easily discriminated from a guitar tone of the same pitch and loudness, by listening to its usually slower attack and particular spectral envelope. Essentially, violin and guitar notes are easily distinguishable because they have different timbres.

Musicians seem to outperform non-musicians in processing musical instruments' timbre. McAdams et al. (1995) evaluated the perceptual structure of musical timbre in musicians, amateur musicians and non-musicians. Using a three-dimensional spatial model, they identified the logarithm of the attack time, the spectral centroid (the proportion of low and high frequency energies) and the spectral flux (the variation of spectral energies across time) to be the acoustic correlates of the dimensions used by the participants to discriminate timbres in a dissimilarity-rating task. They also observed that musicians, compared to amateur musicians and non-musicians, appeared to be more precise and coherent in their judgments. Thus, musicians do not seem to differ in the way they process musical instruments. They are just more consistent in their answers.

Münzer et al. (2002) compared musicians to non-musicians in timbre recognition tasks with different categories of sounds: musical timbre (chords), speech (phoneme combinations such as fa, li…) and pure tones. The authors found that musicians had better recognition performance than non-musicians. Another study confirms this timbre processing advantage in musicians. Chartrand and Belin (2006) compared the performance of musicians and non-musicians in two timbre discrimination tasks: one using instrumental timbres, the other using voices. In the voice task, the participants were instructed to indicate if pairs of syllables were pronounced from the same speaker. In the instrumental discrimination task, the participants had to indicate if two melodies of three notes were played on the same instrument. The musicians performed better at both tasks, but they tended to be slower. This latter effect can be interpreted in different ways. The first possibility is that the tested musicians had better auditory skills before their musical training even started, explaining why they also have better voice discrimination performance. Another possibility is that the musicians' training with musical timbres may have caused improved timbre processing irrespective of the sound category. The latter would support the hypothesis that the discrimination of voices and musical instruments share similar resources, supporting a domain-generality hypothesis.

4.2 Neuroimaging studies of the perception of musical instruments in musicians and non-musicians

Musicians' cortical response to musical tones is well documented due to a growing number of EEG and MEG studies showing enhanced brain responses to N1 and P2 components of the evoked potentials. Pantev et al. (1998) identified the auditory evoked field component N1m to be larger in musicians than in non-musicians, when passively exposed to piano tones. The augmented N1m dipole moment in musicians was found to be specific to the instruments of practice (Pantev et al., 2001; but see Lütkenhöner et al., 2006, for negative results). The P2 component seems to show robust effects of musical expertise. In an EEG study, Shahin et al. (2003) had violinists, pianists and non-musicians listen passively to violin, piano and pure tones. Piano and violin tones evoked larger P2 amplitudes in musicians, compared to control participants. Contrary to the results of Pantev et al. (2001), both piano and violin tones evoked similar amplitudes in violinists and pianists. Moreover, the P2 and P2m component has been found to increase with increasing spectral complexity of stimuli in musicians ((Kuriki et al., 2006) and (Shahin et al., 2005)). The effect is also less marked in non-musicians ((Shahin et al., 2005) and (Kuriki et al., 2007)). In summary, the results suggest both use-dependency and spectral complexity effects in P2(m) enhancements. However, these P2 effects are still poorly understood. A study of Sheehan et al. (2005) evaluated the impact of speech discrimination training on P2 amplitudes. The experimental group received speech discrimination training, while the control group was only exposed to the stimuli. While only the trained group gained better speech discrimination performance, P2 amplitudes were enhanced in both groups. Thus, the hypothesis that P2 enhancement automatically reflects an increase in behavioral discrimination performance is not supported. Higher P2 amplitudes seem to be unspecific effects of exposure to complex sounds and speech stimuli.

In an attempt to test the hypothesis that experience with a specific class of sound stimuli could evoke larger P2 components, Chartrand et al. (2007) investigated the auditory evoked potentials of birdsongs in bird experts. Bird experts and novices were exposed to three sound categories: birdsongs, environmental sounds and human voices. Similar amplitudes were found for the N1 components across groups. On the contrary, P2 amplitudes seemed to be more frontally distributed in bird experts, but this was observed not only for birdsongs but also for the other two sound categories. These results reflect a difference in sound processing in bird experts, which may be explained by the use of a different strategy during the task, or an attentional effect elicited by the perception of birdsongs. Furthermore, birdsongs do not contain as many harmonics as those found in musical instruments or voices. This difference in harmonic complexity was reflected in lower P2 amplitudes for birdsongs, both in musicians and non-musicians. Based on these results, it is quite difficult to ascertain that the auditory P2 component is related to timbre processing, which is thought to be the main mechanism involved in auditory recognition expertise.

Nevertheless, the generators of the P2 component appear located in the secondary auditory cortex (Shahin et al., 2003) in agreement with fMRI studies of timbre processing. In an fMRI study, Menon et al. (2002) investigated functional brain activations when participants were presented with series of six-note melodies played with two different synthesized timbres: Timbre A had a low spectral centroid, no spectral flux and a fast attack (15 ms), while Timbre B had a higher spectral centroid, greater spectral flux and a slow attack (80 ms). Two significant clusters of brain activations were found bilaterally only when Timbre B was compared to Timbre A. Equal levels of activations were found in the right and left temporal lobes, surrounding the primary auditory cortex, as well as the belt regions of the superior temporal gyri (STG) and the superior temporal sulci (STS). However, the left temporal activations were posterior to the right temporal activations suggesting an hemispheric asymmetry in neural processing of timbre. Converging fMRI results have been obtained by Warren et al. (2005). Alternating between noise and harmonic sounds while changing the spectral envelope resulted in a lateralized activation of the middle right STS. It is interesting to note that the two studies, which manipulated different physical parameters of timbre, found anterior STS activations. These brain activations are similar to the ones observed with voice perception studies.

In principle, brain damage to the STS should result in processing difficulties of timbre. We are aware of only one study in which a deficit in timbre recognition has been noted (Kohlmetz et al., 2003). In that study, the patient selectively lost musical timbre perception for percussion and piano tones after a right temporal lesion. It was interpreted as the result of a difficulty to perceive rapid changes of the auditory spectrum, necessary for the identification of piano and percussion timbres but not for wind and string instruments. Unfortunately, the patient was not tested in voice recognition. Thus, the domain specificity of the disorder cannot be assessed.

5. Conclusions: is the expertise controversy transposable to the auditory domain?

The goal of the present article was to suggest the use of an expertise framework in the auditory domain in order to assess domain specificity in voice recognition. We have identified musicians and bird experts as potential auditory experts, who are able to recognize highly familiar sound sources. This ability should be compared with their ability to recognize human voices. To illustrate this idea, we first described vision studies with expert object recognition. In the auditory domain, however, there is as yet no clear consensus about what is an “auditory object” ((Kubovy and Van Valkenburg, 2001) and (Griffiths and Warren, 2004)). Here, we considered musical timbres as a distinct category. In the case of bird experts, we considered birdsongs, which involve both spectral and temporal changes. These differences between birdsongs and musical instruments are shown in Fig. 2. Although the spectral envelope is important in the identification of musical instruments, it is not the sole basis for timbre recognition. The analysis of complex sounds also includes the analysis of temporal envelope changes (Warren et al., 2005). In future studies, it will be interesting to see if two classes of experts, like musicians and bird experts, rely on the same acoustical information and cortical regions to identify their objects of expertise at a subordinate level. To this aim, we plan to use an fMRI procedure highly similar to that of Gauthier et al. (2000), and present different auditory exemplars of birds, musical instruments, voices and environmental objects to musicians and bird experts. We predict that bird experts and musicians would show similar cortical activations in the rSTS to those of voice.

Finally, it is important to note that the study of auditory recognition expertise could also benefit from promising experimental designs and procedures currently used in auditory research fields like voice, speech and music processing. For instance, the use of sound morphing can enable us to create novel exemplars from combining two natural categories of sounds like voices and musical instruments (Bélizaire et al., 2007). This technique could permit us to create entire classes of hybrid or novel auditory objects.

References

ANSI, 1973. Psychoacoustical Terminology. S3.20. American National Standards Institute, New York.

Baharloo, S., Johnston, P.A., Service, S.K., Gitschier, J., Freimer, N.B., 1998. Absolute pitch: an approach for identification of genetic and nongenetic components. Am. J. Hum. Genet. 62, 224–231.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403,309–312.

Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Belin, P., Zatorre, R., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.

Bélizaire, G., Fillion-Bilodeau, S., Chartrand, J.P., Bertrand-Gauvin, C., Belin, P., 2007. Cerebral response to ‘voiceness’: a functional magnetic resonance imaging study. NeuroReport 18, 29–33.

Bhuta, T., Patrick, L., Garnett, J.D., 2004. Perceptual evaluation of voice quality and its correlation with acoustic measurements. J. Voice 18, 299–304.

Bigand, E., Poulin-Charronnat, B., 2006. Are we ‘experienced listeners’? A review of the musical capacities that do not depend on formal musical training. Cognition 100, 100–130.

Bukach, C.M., Gauthier, I., Tarr, M.J., 2006. Beyond faces and modularity: the power of an expertise framework. Trends Cogn. Sci. 10, 159–166.

Carey, S., 1992. Becoming a face expert. Philos. Trans. R. Soc. Lond., B Biol. Sci. 335, 95–103.

Chartrand, J.P., Belin, P., 2006. Superior voice timbre processing in musicians. Neurosci. Lett. 405, 164–167.

Chartrand, J.P., Fillion-Bilodeau, S., Belin, P., 2007. Brain response to birdsongs in bird experts. NeuroReport 18, 335–340.

Diamond, R., Carey, S., 1986. Why faces are and are not special: an effect of expertise. J. Exp. Psychol. Gen. 115, 107–117.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. NeuroImage. 23, 840-848.

Fieger, A., Röder, B., Teder-Sälejärvi, W., Hillyard, S.A., Neville, H.J., 2006. Auditory spatial tuning in late-onset blindness in humans. J. Cogn. Neurosci. 18, 149–157.

Gauthier, I., Bukach, C., 2007. Should we reject the expertise hypothesis? Cognition 103, 322–330.

Gauthier, I., Tarr, M.J., 1997. Becoming a ‘greeble’ expert: exploring mechanisms for face recognition. Vision Res. 37, 1673–1682.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191–197.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., Gore, J.C., 1999. Activation of the middle fusiform ‘face area’ increase with expertise in recognizing novel objects. Nat. Neurosci. 2, 568–573.

Gauthier, I., Williams, P., Tarr, M.J., Tanaka, J., 1998. Training ‘greeble’ experts: a framework for studying expert object recognition processes. Vision Res. 38, 2401–2428.

Gerratt, B.R., Kreiman, J., Antonanzas-Barroso, N., Berke, G.S., 1993. Comparing internal and external standards in voice quality judgments. J. Speech Hear. Res. 36, 14–20.

Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R., Belin, P., 2004. Neuropsychology: pitch discrimination in the early blind. Nature 430, 309.

Griffiths, T.D., Warren, J.D., 2004. What is an auditory object? Nat. Rev., Neurosci. 5, 887–892.

Grill-Spector, K., Sayres, R., Ress, D., 2007. Corrigendum: high-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat. Neurosci. 10, 133.

Grill-Spector, K., Sayres, R., Ress, D., 2006. High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat. Neurosci. 9, 1177–1185.

Grill-Spector, K., Knouf, N., Kanwisher, N., 2004. The fusiform face area subserves face perception, not generic within-category identification. Nat. Neurosci. 7, 555–562.

Hollien, H., Bennett, G., Gelfer, M.P., 1983. Criminal identification comparison: aural versus visual identification resulting from a simulated crime. J. Forensic Sci. 28, 208–221.

Kanwisher, N., 2000. Domain specificity in face perception Nat. Neurosci. 3, 759–763.

Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

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, 125–143.

Kohlmetz, C., Müller, S.V., Nager, W., Münte, T.F., Altenmüller, E., 2003. Selective loss of timbre perception for keyboard and percussion instruments following a right temporal lesion. Neurocase 9, 86–93.

Kreiman, J., 1997. Listening to voice: theory and practice in voice perception research. In: Johnson, K., Mullenix, J.W. (Eds.), Talker Variability in Speech Processing. Morgan Kaufmann Publishers Inc., San Francisco, pp. 88–108.

Kreiman, J., Gerratt, B.R., Kempster, G.B., Erman, A., Berke, G.S., 1993. Perceptual evaluation of voice quality: review, tutorial, and a framework for further research. J. Speech Hear. Res. 36, 21–40.

Kreiman, J., Gerratt, B.R., Precoda, K., Berke, G.S., 1992. Individual differences in voice quality perception. J. Speech Hear. Res. 35, 512–520.

Kubovy, M., Van Valkenburg, D., 2001. Auditory and visual objects. Cognition 80, 97–126.

Kuriki, S., Ohta, K., Koyama, S., 2007. Persistent responsiveness of long-latency auditory cortical activities in response to repeated stimuli of musical timbre and vowel sounds. Cereb. Cortex 17, 2725–2732.

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, 4046–4053.

Linville, S.E., 1996. The sound of senescence. J. Voice 10, 190–200.

Lütkenhöner, B., Seither-Preisler, A., Seither, S., 2006. Piano tones evoke stronger magnetic fields than pure tones or noise, both in musicians and non-musicians. NeuroImage 30, 927–937.

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, 199–211.

Marques, C., Moreno, S., Castro, S.L., Besson, M., 2007. Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. J. Cogn. Neurosci. 19, 1453–1463.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., Krimphoff, J., 1995. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. Psychol. Res. 58, 177–192.

Menon, V., Levitin, D.J., Smith, B.K., Lembke, A., Krasnow, B.D., Glazer, D., Glover, G.H., McAdams, S., 2002. Neural correlates of timbre change in harmonic sounds. NeuroImage 17, 1742–1754.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219, 36–47.

Mullennix, J.W., Johnson, K.A., Topcu-Durgun, M., Farnsworth, L.M., 1995. The perceptual representation of voice gender. J. Acoust. Soc. Am. 98, 3080–3095.

Münte, T.F., Kohlmetz, C., Nager, W., Altenmüller, E., 2001. Neuroperception. Superior auditory spatial tuning in conductors. Nature 409, 580.

Münzer, S., Berti, S., Pechmann, T., 2002. Encoding of timbre, speech and tones: musicians vs. non-musicians. Psychol. Beitr. 44, 187–202.

Nachson, I., 1995. On the modularity of face recognition: the riddle of domain specificity. J. Clin. Exp. Neuropsychol. 17, 256–275.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., Münte, T.F., 2003. The fate of sounds in conductors' brains: an ERP study. Brain Res. Cogn. Brain Res. 17, 83–93.

Olsson, N., Juslin, P., Winman, A., 1998. Realism of confidence inearwitness versus eyewitness identification. J. Exp. Psychol., Appl. 4, 101–118.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport 12, 169–174.

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

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., Belleville, S., 1994. Functional dissociations following bilateral lesions of auditory cortex. Brain 117, 1283–1301.

Pernet, C., Charest, I., Bélizaire, G., Zatorre, R.J., Belin, P., 2007. The Temporal Voice Areas: spatial characterization and variability. 13th Annual Meeting of the Organization for Human Brain Mapping. Neuroimage 36, Suppl 1, S109.

Rammsayer, T., Altenmuller, E., 2006. Temporal information processing in musicians and nonmusicians. Music Percept. 24, 37–48.

Rhodes, G., Byatt, G., Michie, P.T., Puce, A., 2004. Is the fusiform face area specialized for faces, individuation, or expert individuation? J. Cogn. Neurosci. 16, 189–203.

Röder, B., Teder-Sälejärvi, W., Sterr, A., Rösler, F., Hillyard, S.A., Neville, H.J., 1999. Improved auditory spatial tuning in blind humans. Nature 400, 162–166.

Scherer, K.R., 1995. Expression of emotion in voice and music. J. Voice 9, 235–248.

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

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, 1781–1785.

Shahin, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J. Neurosci. 23, 5545–5552.

Sheehan, K.A., McArthur, G.M., Bishop, D.V.M., 2005. Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? Cogn. Brain Res. 25, 547–553.

Spiegel, M.F., Watson, C.S., 1981. Factors in the discrimination of the tonal patterns. III. Frequency discrimination with components of well-learned patterns. J. Acoust. Soc. Am. 69, 223–230.

Tanaka, J.W., 2001. The entry point of face recognition: evidence for face expertise. J. Exp. Psychol. Gen. 130, 534–543.

Tanaka, J.W., Gauthier, I., 1997. Expertise in object and face recognition. In: Medin, D.L., Schyns, P.G. (Eds.), Mechanisms of Perceptual Learning. Advances in Research and Theory, Vol. 36. Academic Press Inc., London, pp. 83–125.

Tanaka, J.W., Taylor, M., 1991. Object categories and expertise: is the basic level in the eye of the beholder? Cognit. Psychol. 23, 457–482.

Tervaniemi, M., Castaneda, A., Knoll, M., Uther, M., 2006. Sound processing in amateur musicians and nonmusicians:event-related potential and behavioral indices. NeuroReport 17, 1225–1228.

Thompson, W.F., Schellenberg, E.G., Husain, G., 2004. Decoding speech prosody: do music lessons help? Emotion 4, 46–64.

Van Lancker, D., Kreiman, J., 1987. Voice discrimination and recognition are separate abilities. Neuropsychologia 25, 829–834.

Warren, J.D., Jennings, A.R., Griffiths, T.D., 2005. Analysis of the spectral envelope of sounds by the human brain. NeuroImage 24, 1052–1057.

Xu, Y., 2005. Revisiting the role of the fusiform face area in visual expertise. Cereb. Cortex 15, 1234–1242.

Yin, R.K., 1969. Looking at upside-down faces. J. Exp. Psychol. 81, 141–145.

Yue, X., Tjan, B.S., Biederman, I., 2006. What makes faces special? Vision Res. 46, 3802–3811.

Figure 1. Representation of spatial variability across the Temporal Voice Area (TVA). Participants were passively listening to human voices, musical instruments and environmental sounds (Pernet, Charest, Bélizaire, Zatorre & Belin, 2007; unpublished results). (a) About 50% of the voxels show stronger selectivity. (c) A cluster analysis based on individual peaks of activation revealed the existence of two main clusters in each hemisphere.

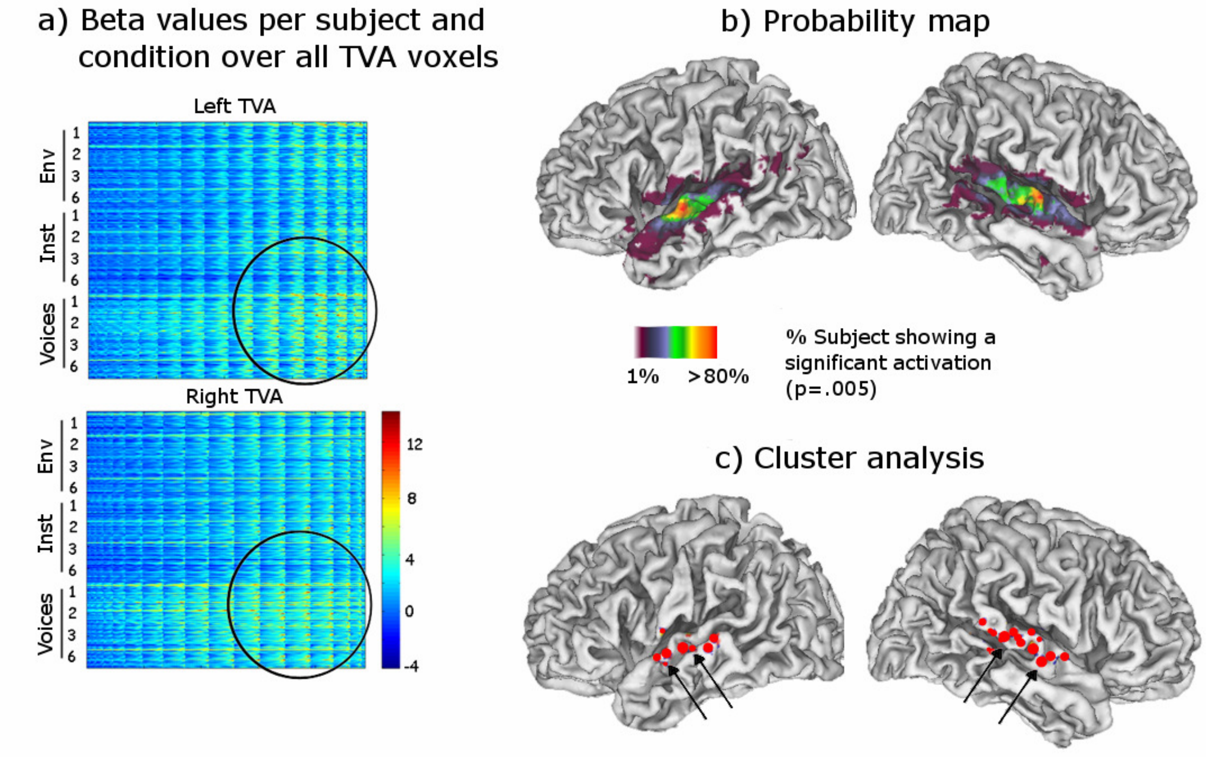
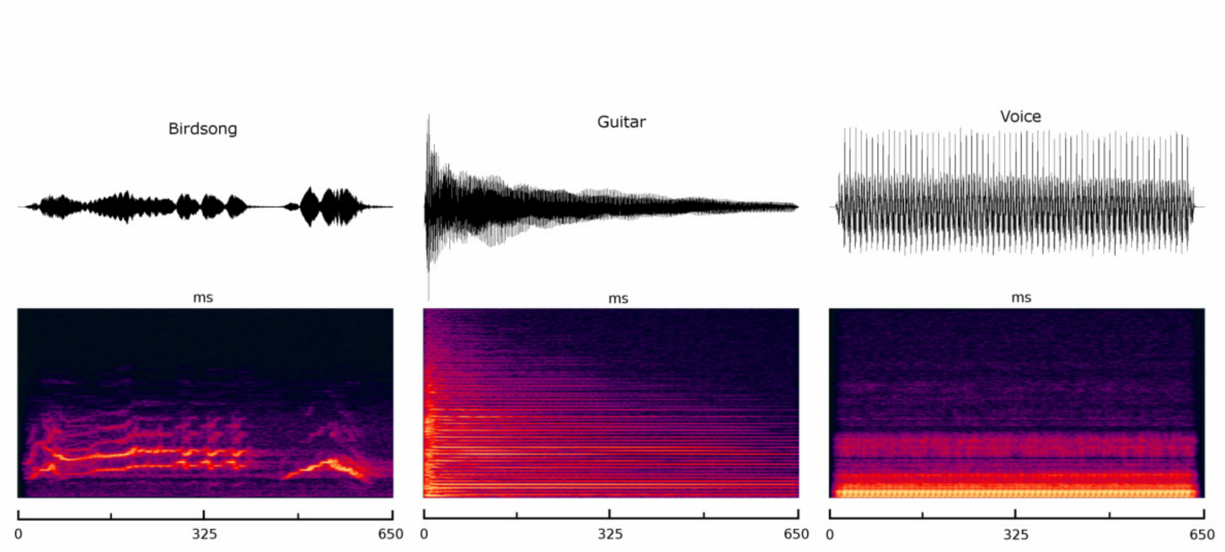


Figure 2. Spectral analysis of complex sound stimuli. Spectral analysis and waveforms of a birdsong, a guitar note played at C3, and a male human pronouncing the french vowel |a|. Birdsongs usually contain pitch and envelope variations that are necessary to their identification, as opposed to musical instruments and human voices.



Article 2 : Superior voice timbre processing in musicians

Jean-Pierre Chartrand et Pascal Belin

Publié dans Neuroscience Letters (2006), volume 405, pages 164 à 167.

Running head: Superior timbre processing in musicians

Superior voice timbre processing in musicians

Jean-Pierre Chartrand[1] and Pascal Belin[12]
[1]Université de Montréal, Montréal, Canada
BRAMS, Département de Psychologie, Université de Montréal
[2]Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

Correspondence and requests for reprints to

Jean-Pierre Chartrand
Laboratoire de Neurocognition vocale
Centre de Recherche en Neuropsychologie et Cognition
Département de Psychologie
C.P. 6128 succursale Centre-Ville
Montréal, Québec, Canada, H3C 3J7

Tel : + 1 1 (514) 343-6111 ext. 1495;
fax: +1 1 (514) 343-5787;

2894 words; 2 figures

Abstract

After several years of exposure to musical instrument practice, musicians acquire a great expertise in processing auditory features like tonal pitch or timbre. Here we compared the performance of musicians and non-musicians in two timbre discrimination tasks: one using instrumental timbres, the other using voices. Both accuracy (d-prime) and reaction time measures were obtained. The results indicate that the musicians performed better than the non-musicians at both tasks. The musicians also took more time to respond at both tasks. One interpretation of this result is that the expertise musicians acquired with instrumental timbres during their training transferred to timbres of voice. The musician participants may also have used different cognitive strategies during the experiment. Higher response times found in musicians can be explained by a longer verbal-auditory memory and the use of a strategy to further process auditory features.

Keywords: voice, musicians, expertise, musical training

Introduction

In auditory perception, one of the most studied forms of expertise has been that of musicians. Practicing a musical instrument usually starts during childhood or adolescence and after several years of exposure these people acquire great expertise with processing of auditory features like tonal pitch or timbre [13,10]. Musical training requires and involves different kinds of auditory processes such as recognition and discrimination of instrument tones, and musicians improve all those processes over time. Musicians' expertise with processing of timbre of musical instruments has also been demonstrated [13].

Electrophysiological data show that musical training is accompanied by specific cortical modifications. The P2 and N1c (sub-component of N1 occuring at electrode T8) components of the auditory evoked potential in response to music sounds are found to be enhanced in musicians [14]. The neuromagnetic N1m component is likewise enhanced in musicians [11], especially for the musical timbres of their own instrument [12]. Moreover, when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones, the P2 component is enhanced bilaterally and the N1c is enhanced in the right hemisphere [3]. The P2 is a particularly sensitive indicator of neural plasticity since it can be enhanced early in childhood, for example when 4- to 5-year-old children are exposed to musical training [16]. Both behavioral and electrophysiological data demonstrate that changes occur in musicians and that these changes are closely associated with the processing of musical features. But to what extent can musical expertise with musical timbres be generalized to other types of timbre from other sound categories?

The American National Standards Institute [1] defined timbre as "that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different". However, another definition of timbre, which is "an invariant quality based on perceivable transformations across pitch and/or loudness that is assumed to underlie the ability to identify one instrument or voice" [7] has been used often in the psychoacoustic literature. We find this definition of timbre being more operational and useful for the present research, because each musical instrument can have its own timbre across different pitches and it extends the notion of timbre to other categories such as voice.

In order to find out if musicians have a superiority over non-musicians in processing timbre, Pitt [13] observed how the two groups perceived dimensions of timbre and pitch. He asked participants to tell if two consecutive musical tones were different in timbre and/or pitch. Participants had a four-choice categorization task: *no change, pitch change, instrument change, and both changes*. Results showed that when timbre changed, the musicians were far more accurate than the non-musicians. Because the participants had a four-choice task, they had to focus on both timbre and pitch variations on each trial. It has been shown that timbre variations can affect the judgement of pitch [9,15], especially in non-musicians [13]. It is possible that the non-musicians who participated in Pitt's study were more impaired than musicians when they had to judge timbre variations because they needed to concentrate on both dimensions simultaneously. That study only partially demonstrated that musicians do have a superiority over non-musicians in processing musical timbres, because it could have been a consequence of their superiority in pitch processing. Münzer et al. [10] conducted three experiments to determine if musician's familiarity with tonal processing can be generalised to other kinds of superiority when processing other auditory features. In different recognition tasks, they found that musicians were better than non-musicians with musical instrument timbres, speech, and tones.

Speaker discrimination can be viewed as a particular case of timbre processing within a homogeneous category of sounds, and it is not known whether discriminating human voices and musical instruments involve different or similar processes. The present experiment asked the question of whether musical expertise with instrument timbres transfers to vocal timbres. In other words, will musicians be better than non-musicians at a vocal timbre discrimination task? We compared groups of musicians and non-musicians at two timbre discrimination tasks, one intrumental and one vocal. We predicted that musicians would perform better than non-musicians at both tasks, based on the results of Münzer et al. [10]. However, it is possible that musicians do not perform better than non-musicians since voice processing is a common and frequent task in both groups.

Methods

Thirty-six participants were recruited at the University of Montréal. Most of them were undergraduate students. The sample was composed of 17 musicians (9 women, 8 men) and 19 non-musicians (11 women, 8 men). A one-way ANOVA was performed to compare age differences between musicians (mean = 24.23, S.D. = 5.3) and non-musicians (mean = 23.94, S.D. = 5.93). The difference between groups was not significant ($F(1, 34) = 0.023$, $p > 0.05$). Musicians had at least 3 years of regular practice with an instrument or singing when included in this group (the years of training varied from 3 to 25 years). Out of the 17 musicians, there were 3 participants who were singers. When they were questioned about their musical training history, they reported that they also had courses with musical instruments, such as piano. Considering their specific training with voice production, they might be better than the other musicians at discriminating voices. However, the number of singers is so small that it would not make a large difference in the results. None of the musicians reported having absolute pitch. Both musicians and non-musicians reported having no auditory impairment. They participated on a voluntary basis and were not paid.

Two classes of stimuli were used: sounds of musical instruments and human voices. Thirty-two samples coming from four categories of musical instruments (piano, strings, guitar, brass) were created on a Roland JV-80 keyboard synthesizer. Each category included four instruments which played two sequences of notes: C-D-G and C-E-G. Those stimuli were recorded with Cooledit software (Syntrillium, 2000) in stereo and converted to mono with a sampling frequency of 22.05 kHz and a 16 bit resolution. The musical instrument samples had a mean length of 857 ms (S.D. = 27).

The human voices were taken from recordings of American vowels kindly provided by Hillenbrand [8]. The samples were also arranged in four categories: voices from women, men, boys, and girls. Each category contained four speakers pronouncing the syllables "had" and "heed". The samples were recorded in mono with a sampling frequency of 16 kHz and a 16 bit resolution. The mean length was 591 ms (S.D. = 87). Both classes of stimuli were normalized on mean energy (RMS) using Matlab (Mathworks).

Tasks were designed to be as similar as possible for the instrument and voice sounds. Sounds were presented in pairs with a 1s inter-onset-interval. Half of the pairs came from the same source (same instrument, or same voice); but all the pairs differed on the spectro-temporal pattern, whether the source was the same or not: the two melodies of a pair were always different for the instruments task, and the two syllables of a pair were always different for the speakers task. The experiment was divided into 4 blocks of 96 pairs, 2 for the instruments and 2 for the voices. Response time was recorded from the onset of the sound. The next trial was initiated 2 ± 0.5 s after the response to the preceding pair. Accuracy was measured with the d-prime (d') measure of sensitivity [6].

Participants were set in a sound proof cabin in front of a computer keyboard. On each trial, a pair of stimuli was presented via Beyerdynamic DT 770 headphones. Instructions were exactly the same for the voice and instrument tasks: “Indicate with the keyboard if the two sounds presented are produced from the same sound source or not, while responding as fast as possible and maintaining the lowest error-rate possible”; i.e., same or different instrument, or same/different speaker.

The design of the experiment was 2 × 2 mixed factorial, with musical training (musician versus non-musician) as the between-participants factor and task (voice discrimination versus instrument discrimination) as the within-participants factor. Two ANOVAs were performed, in which the dependent variables were accuracy (d') and reaction time. Since there was no main effect of participant gender on reaction times ($F(1, 34) = 0.624$, $p > 0.05$) and discrimination performance ($F(1, 34) = 0.562$, $p > 0.05$), data were pooled across male and female participants. Mean response times of the participants were calculated for each task after removing responses that were two standard deviations above the mean for that participant were removed. Thus, corrected response time means were computed and used in the analyses. All the required ANOVA assumptions were checked during the analysis. Homogeneity of variance was equal across all categories (i.e., musicians versus non-musicians and voices versus musical instruments) for between-participant ANOVAs. In addition, Greenhouse–Geisser correction was used for repeated measures analysis to correct for potential violations of the sphericity assumption.

Results

There was a main effect of task on d' scores, ($F(1, 34) = 31.482$, $p < 0.05$). The voice discimination task was more difficult than the musical instruments discrimination task for all the participants. There was also a main effect of group: the musicians were found to perform significantly better than non-musicians at both the voice discrimination and the instrument discrimination tasks, ($F(1, 34) = 10.834$, $p < 0.05$). The interaction between musical training and task was almost significant, ($F(1, 34) = 2.950$, $p = 0.095$), suggesting that musicians had a greater advantage over the non-musicians in the instrument discrimination task. The results are shown in Fig. 1. A weak but significant correlation was found between voice and instrument discrimination performance ($r = 0.359$, $p < 0.05$). In order to eliminate outliers, individual ratios were computed and then transformed into z-scores. No participants had z-score above or below 3.29 standard deviation from the mean ratio, so no participants were removed from the correlation. As shown in Fig. 2, the singing musicians had particularly good voice and instrument discrimination performance.

Response times tended to be longer for musicians, ($F(1, 34) = 3.497$, $p = 0.07$). The musical training × task interaction was not significant, ($F(1, 34) = 0.315$, $p > 0.05$). The simple effect of musical training on the voice discrimination task was significant, ($F(1, 34) = 4.16$, $p < 0.05$), the musicians having significantly longer response times than the non-musicians (cf. Fig. 1).

Discussion

As predicted, the musicians were found to perform better than the non-musicians at both voice and musical instrument discrimination tasks. The musicians took more time to respond in the voice discrimination task. All participants performed better at the instrument discrimination task than at the vocal discrimination task.

One interpretation of the results is that expertise with musical instrument timbres generalizes to other timbre tasks, such as voice discrimination. Human voice is special in the sense that it is the carrier of speech and contains rich paralinguistic information about the speaker's identity, gender, emotional state [2]. But when it comes to auditory processing of acoustical features, it is unknown whether voice discrimination involves processes that are different from the ones implicated in instrument discrimination or not. Since there is a relation between the participants' musical education and timbre discrimination performance, we suppose that during their musical training, the musicians have learned to better discriminate timbres of various sound sources, but this remains to be further assessed. However, since the singing musicians seemed to be among the best performers in both discrimination tasks, experiments exploring the specific link between singing and timbre processing should be conducted.

Also, to accept this interpretation, we need to suppose that voice timbre and instrument timbre discrimination involve similar processes. If voice and instrument timbre processing were mediated by different processes, one would have expected the musicians to have only performed better at the instrument discrimination task. Musicians and non-musicians would have been expected to perform at the same level with the voice timbre discrimination task, since voice processing is a common domain of expertise across the whole population [2]. Not only did the musicians perform better than non-musicians at both tasks, but they showed a tendency to have a greater advantage in the instruments discrimination task. This result is not in direct contradiction with the proposal that the processes involved in the two tasks are similar. Considering the training that musicians have received with musical instruments, they were much more familiar than the non-musicians with the task involving instrument timbres. While the present study was limited to the question of timbre processing across different sound categories, it raises the question of functional modularity of timbre processing. Other experiments should be performed to examine whether the two types of timbre discrimination can interfere with each other (see Gauthier and Kurby [5], for an example of an interference task in visual perception).

An alternative explanation of the musicians' better performance is that they could have a greater non-verbal I.Q. This characteristic was not measured in our experiment. A study conducted by Brandler and Rammsayer [4] tested different aspects of mental abilities across musicians and non-musicians. It demonstrated that musicians were significantly superior only on auditory verbal memory and reasoning scales. It is true that a larger short-term auditory memory might explain why musicians outperformed non-musicians. Actually, auditory short-term memory was not measured in this study. However, if this explanation was true, then other studies would need to be conducted in order to assess whether this difference between musicians and non-musicians is caused by musical practice. Since little is known about musical instrument and voice timbre processing, for the moment it is difficult to attribute direct links between specific processes and good auditory discrimination performance. These results may partially explain our finding that the musicians had longer response times in both tasks.

The musicians' longer response times can be explained by arguments proposed by Münzer et al. [10]. In conclusion to their experiments, these authors said that musicians could process sound at a deeper level when they were given more time to encode the stimuli (168 and 200 ms versus 68 ms stimulus duration). The authors estimated that given more time for encoding, performance increases and an advantage for the musicians shows up. They proposed that musicians could have proceeded even further with the analysis of the auditory features. Although the participants were all instructed to respond as fast as possible while maintaining the lowest error-rate possible, the musicians took more time to respond overall and may have processed the sounds at a deeper level than the non-musicians, or used different cognitive strategies.

Finally, participants performed less well at the voice discrimination task than at the instrument discrimination task, but this piece of result is less relevant, since the objective of this study was to assess the link between musical training and timbre processing and thus to compare the two groups. The task demands were exactly identical in the two discrimination tasks, except for the timbre source. There were three differences though. While voices contained approximately one pitch during the whole syllable, musical instruments constituted of three different pitches. This may have lowered the difficulty of the instrument discrimination task. A difference in stimuli length can be noted too. The voice sounds had a mean total duration of 650 ms while the instrument sounds had a mean duration of 850 ms. The participants had 200 ms more of auditory information to process. Those two differences between voice and musical instrument sounds by themselves can explain why the instrument discrimination task led to a better performance.

Conclusion

The main goal of the present study was to assess if musical timbre expertise can be generalized to timbres of other auditory domains, like voices. Our results suggest that it can, but this statement is not without limitations. First of all, it is not known yet if musicians are better than non-musicians at other voice processing tasks, like voice recognition for example. Second, it is also not known yet if musicians would perform better with other kinds of auditory stimuli, like environmental sounds. Future studies need to focus on multiple types of sound categories and other types of sound processing to further evaluate differences between muscians and non-musicians.

References

[1] ANSI (1973). *Psychoacoustical terminology. S3.20.* New York: American National Standards Institute.

[2] Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice : neural correlates of voice perception. *Trends in Cognitive Sciences,* 8, 129-135.

[3] Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex, 14*, 1088-1099.

[4] Brandler, S., & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music,* 31(2), 123-138.

[5] Gauthier, I. & Curby, K. M. (2005). A perceptual traffic jam on highway N170. *Current Directions in Psychological Science*, 14(1), 30-33.

[6] Green, D. M. & Swets, J. A. (1966). *Signal detection and psychophysics.* New York: Wiley.

[7] Handel, S., & Erickson, M. L. (2001). A rule of thumb: the bandwith for timbre invariance is one octave. *Music Perception,* 19(1), 121-126.

[8] Hillenbrand, J., Getty, L. A., Clark, M. J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America,* 97, 3099-3111.

[9] Moore, B. C. J. & Glasberg, B. R. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *Journal of the Acoustical Society of America, 87,* 2163-2177.

[10] Münzer, S., Berti, S., & Pechmann, T. (2002). Encoding of timbre, speech, and tones: Musicians vs. non-musicians. *Psychologische Beitrage,* 44(2), 187-202.

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

[12] Pantev, C., Roberts, L. E., Schulz, M., Engelien, A., Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport, 12*, 1-6.

[13] Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance, 20*, 976-986.

[14] Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of Neuroplastic P2 and N1c Auditory Evoked Potentials in Musicians. *The Journal of Neuroscience, 23(12)*, 5545-5552.

[15] Singh, P. G. & Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *Journal of the Acoustical Society of America, 95*, 2650-2661.

[16] Trainor, L. J., Shahin, A., & Roberts, L. E. (2003). Effects of musical training on the auditory cortex in children. *Annals New York Academy of Science, 999*, 506-513.

Figure 1. Mean peformance (+SE) for non-musicians and musicians at both voice and instrument discrimination tasks. Musicians are better than non-musicians at both tasks. The musical instruments discrimination task was easier for both groups.

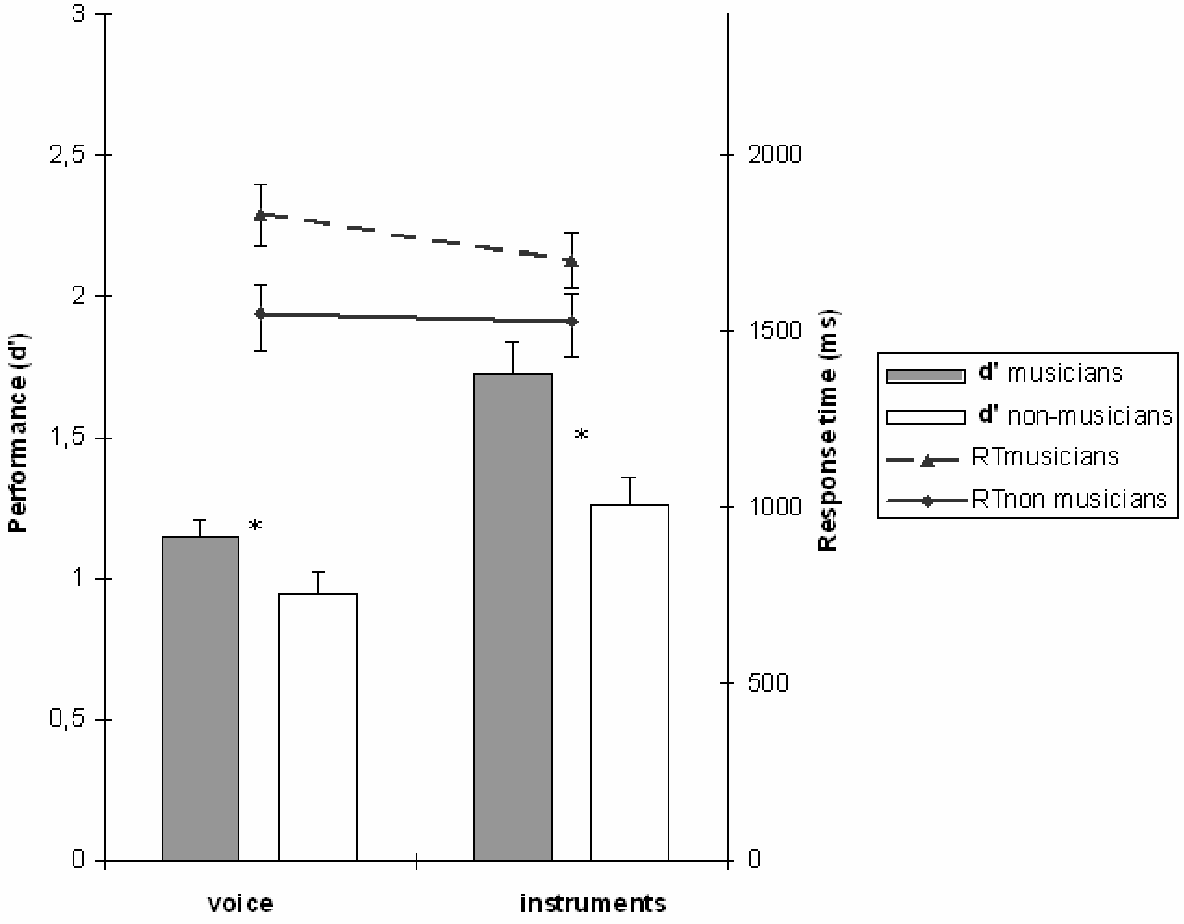
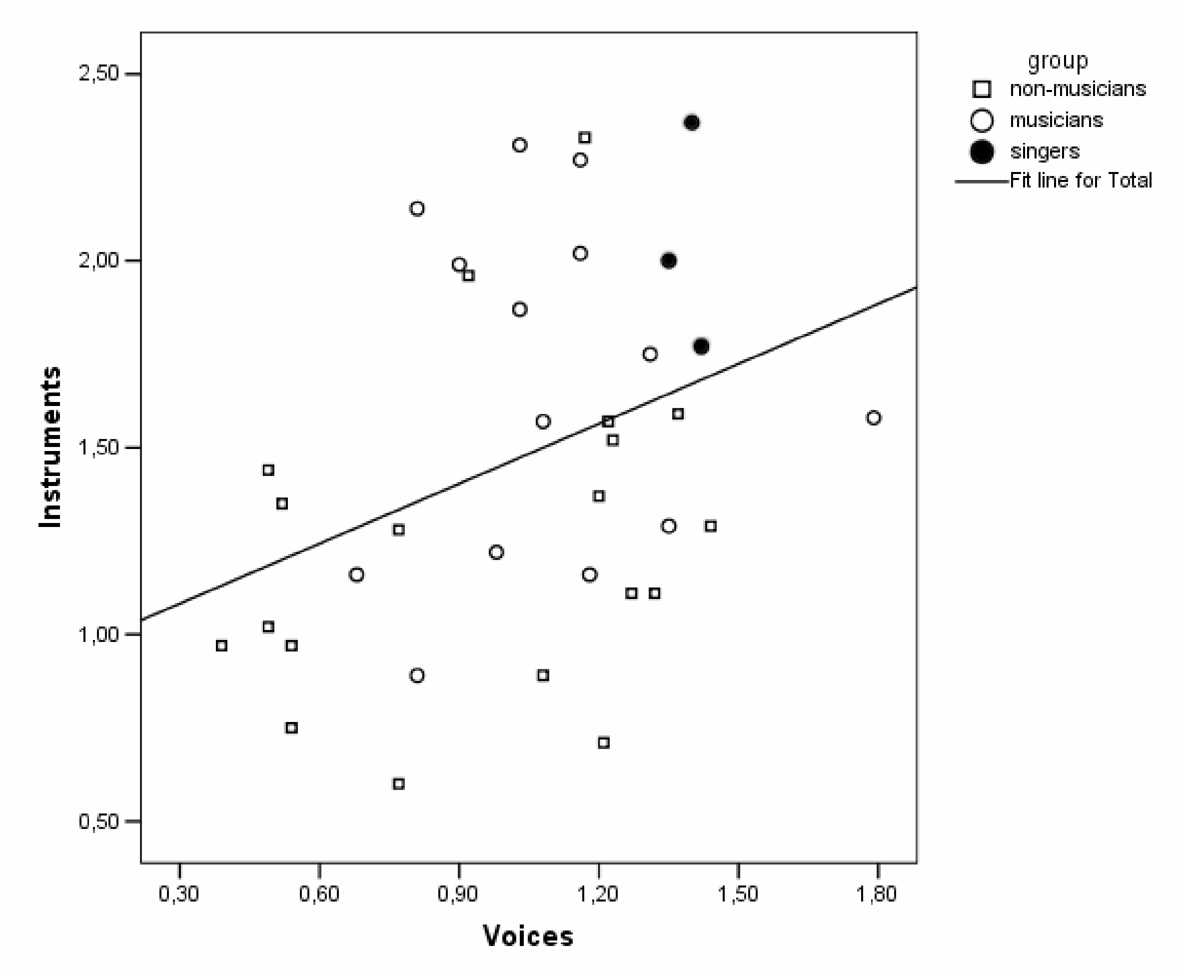


Figure 2. Correlation between instrument and voice discrimination tasks performance for both musicians and non-musicians.



Article 3 : Brain response to birdsongs in bird experts

Jean-Pierre Chartrand, Sarah Fillion-Bilodeau et Pascal Belin

Publié dans Neuroreport (2007), volume 18 no 4, pages 335 à 340.

Running head: Response to birdsongs in bird experts

Brain response to birdsongs in bird experts

Jean-Pierre Chartrand, Sarah Filion-Bilodeau and Pascal Belin
Université de Montréal, Montréal, Canada
BRAMS, Département de Psychologie, Université de Montréal
Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

Correspondence and requests for reprints to

Jean-Pierre Chartrand
Laboratoire de Neurocognition vocale
Centre de Recherche en Neuropsychologie et Cognition
Département de Psychologie
C.P. 6128 succursale Centre-Ville
Montréal, Québec, Canada, H3C 3J7

Tel : + 1 1 (514) 343-6111 ext. 1495;
fax: +1 1 (514) 343-5787;

2761 words; 4 figures

Abstract

Auditory expertise has mostly been studied in relation to musical processing, but expert auditory processing can also involve non-musical auditory stimuli, such as birdsongs in bird experts. In the present study, the neural correlates of bird expertise were investigated by using electroencephalography to measure auditory evoked potentials in bird experts and novices. Auditory stimuli of three categories (birdsongs, environmental sounds, voices) were presented in a pseudo-random order while subjects performed a simple target detection task (pure tone). We observed similar amplitude and distribution of the N100 component in bird experts and novices. In contrast, the amplitude of the P200 component was significantly smaller in bird experts at the Pz and Cz electrodes, reflecting a more frontal topography of this positivity. Notably, this group difference was observed for the birdsongs but also for the voices and environmental sounds, suggesting a general processing difference in bird experts not restricted to the category of expertise.

Keywords: bird experts; auditory expertise; EEG; auditory-evoked potentials; audition; voice

Introduction

Our perceptual environment is composed of a multitude of visual and auditory stimuli. Some of them are socially relevant, such as faces and voices, which allow both discrimination and recognition of peers as well as of their emotional states. Some other categories of stimuli are only of interest to a restricted section of the population. One example of such a stimulus category is birdsong, which is studied by people who enjoy watching and listening to birds as a hobby or as a professional activity. These individuals, because of their interest in, and repeated exposure to, bird stimuli, have developed an expertise with birds; that is, they display a better ability to identify bird pictures and songs, and to differentiate between them, than the general population.

Expertise has been defined as the appropriate acquisition and use of knowledge in a particular domain [1]. But the main focus of this study is on perceptual expertise, that is, expertise on the basis of perceptual stimulus recognition and processing. Cognitive processes and cortical representations related to visual expertise are being increasingly explored and compared with face expertise (see [2,3] for different points of view on the subject), whereas studies of expertise in the auditory domain are comparatively rare. Some studies have concentrated on music perception and others on voice perception, but there is a need to integrate these different auditory domains into a common framework.

In auditory perception, musicians have been studied because of the importance of music in human activities. A large amount of information has been gathered on behavioural and cortical modifications that follow auditory training, showing that musicians are characterized by several auditory processing differences when compared with nonmusicians. Musicians are better at perceiving dimensions of pitch and timbre [4], have a good tonal memory [5], and they perform better than nonmusicians at discriminating human voice and musical instrument timbres, although their response times are longer [6]. These behavioural differences are accompanied by neurophysiological differences. In particular, the P2 and N1c (right hemisphere, latency about 150 ms) electrophysiological components, evoked by pure and musical tones, have been found to have a greater amplitude in musicians than in nonmusicians [7]. Pantev et al. [8] also found a similar effect for the N1m, the neuromagnetic equivalent of the N1 component, particularly for sounds of the instrument of practice [8].

Sounds of human voice are a good example of the category of highly familiar sounds. All normal humans can to some extent be considered voice experts because of extensive experience with voice sounds. Thus, the study of voice processing can have important implications for the study of auditory expertise. Little is yet known about the functional anatomy of voice processing. Cortical regions along the bilateral superior temporal sulci show an enhanced response to voice stimuli [9,10], but the exact reason for this region's preferential response to voice stimuli is still unknown. Does the enhanced response along the superior temporal sulci to vocal sounds reflect cerebral mechanisms that are tuned to the acoustic structures of voices? Or does it reflect the expertise that normal listeners have necessarily developed within this particular sound category?

Here, the study of auditory-evoked potentials in bird experts and nonexperts was designed to address the question of auditory expertise. The hypotheses, on the basis of results obtained from experiments with musicians [11], were that bird experts would show larger P2 amplitudes for bird stimuli, when compared with novices.

Methods

Participants

Fourteen amateur bird experts and 14 novices were recruited for the experiment. Half of the participants were men and half were women in both groups. The participants reported having no auditory or intellectual deficits. The ages of the amateur bird experts (mean±SD: 43±3.6 years) and the novices (32±4.2 years) were not significantly different (T=-1.976, P=0.059). The amateur bird experts were contacted through advertisements placed on the websites of local amateur ornithologist associations and at an amateur ornithologist congress, and they were recruited on the basis of their bird identification activities. They had a mean of 11.5±6.987 SD years of experience. All participants gave written informed consent and received C$40 compensation for their participation.

Stimuli

Participants heard 450 different sound stimuli, 150 in each of three categories: birdsongs, human vocalizations and environmental sounds. The birdsongs were selected from the Chants d'oiseaux du Québec et de l'Amérique du Nord (2004) audio CD. Other sound stimuli came from commercially available sources and from recordings in the laboratory. Sounds were edited using Cool Edit Pro (Syntrillium corporation, Phoenix, Arizona, USA) to a sampling rate of 22 050 Hz, a 16-bit resolution, and a duration of 200 ms with a 10-ms linear attack and decay envelope filter. They were normalized by root mean square using Matlab (The MathWorks Inc., Natick, Massachusetts, USA). The average spectrum of the three sound categories is shown in Fig. 1.

Procedure

Participants were first assessed for their manual laterality using the Edinburgh inventory [12]. Three out of the 14 novices were left-handed and four out of the 14 amateur bird experts were left-handed. They also filled a questionnaire about their ornithology activities and their general health.

The participants were installed in a sound-proof cabin and were presented with different sounds in a pseudo-random order, split into 20 blocks of 3 min. The sounds were presented via Beyerdynamic DT 770 headphones at a self-adjusted comfort level of about 65 dB sound pressure level. The participants were instructed to detect a target stimulus that consisted of a 1000 Hz sinusoidal pure sound with a 10% probability of occurrence. They had to press a button each time they heard the target stimulus.

Electroencephalography recording and analysis

Electroencephalography (EEG) data were continuously recorded using a Brainamp amplifier (Brainproduct-MR 64 channel-Standard; 62 EEG electrodes, one EOG, one ECG, Brain Products, Munich, Germany). The 64 Ag/AgCl electrodes were installed using a BrainCap, 10-20 array. The reference electrode was FCz. The electrode impedances were kept below 10 kΩ throughout the recording. Continuous EEG data were recorded at a sampling frequency of 250 Hz using a 0.5–70 Hz band-pass filter. The epochs corresponding to each condition were extracted from the continuous data with a window of 1500 ms beginning 504 ms before stimulus onset.

EEG files of each participant were analysed using EEGlab [13] in Matlab (The Mathworks, Inc.). Epochs were further reduced before the analysis (-200 to 400 ms) and were visually inspected and cleaned for artefacts. For each waveform being measured, the mean voltage was calculated in a predefined window: the N1 wave was measured with a time window of 80–120 ms and the P2 wave was measured with a time window of 180–240 ms. Peak latencies for N1 and P2 were determined in their respective windows using peak values.

Statistical analyses

The design of the experiment was a 2×3 mixed factorial, with bird expertise (experts vs. novices) as the between-subjects factor and sound category (birdsongs vs. environmental sounds vs. voices) as the within-subjects factor. To assess group differences in evoked-response-potential (ERP) waveforms, a bootstrap analysis was performed at the three sites on EEG data as implemented in EEGLab [13] for each sound condition. Repeated measures analyses of variance (ANOVAs) were used to assess the latent differences in components. Additionally, the amplitudes of N1 and P2 at the three sites were analysed in six ANOVAs with sound condition (birdsongs vs. voices vs. environmental sounds) as a within-subjects factor. Effect sizes were computed as partial $\eta 2$ values. The mean between the Greenhouse–Geisser and Huynh–Feldt correction was used, as suggested by Stevens [14]. A Bonferroni correction was used, with $P < 0.008$ considered statistically significant.

Results

Component amplitudes

Bootstrap analyses revealed that the novices had greater P2 component amplitude than bird experts at Pz and Cz for each of the three sound categories (Fig. 2). Moreover, bird experts showed greater amplitude on the Fz electrode at about 300 ms but only for the voice condition (Fig. 2).

ANOVAs were also performed on individually measured peak amplitude and latency of the N1 and P2 components. The main effect of sound category was significant at Pz for the N1 component [F(1.95,50.71)=37.48; P<0.008; η_p^2=0.59]. Birdsongs evoked a smaller N1 component (m=-0.99) (P<0.008), whereas N1 amplitudes of voice (m=-1.83) and environmental sounds (m=-1.93) were not significantly different (P>0.008). The interaction Sound x Bird expertise was not significant [F(1.95,50.71)=1.53; P>0.008], showing that both novices and bird experts showed this reduced parietal N1 (Figs 2 and 3). The main effect of the sound category was also significant at Fz for the P2 component [F(1.31,34.17)=14.68; P<0.008; η_p^2=0.36], reflecting smaller P2 amplitudes at Fz for the birdsongs (m=1.32) than for the voices (m=2.33) and environmental sounds (m=2.40) (P>0.008). Again the interaction Sound x Bird expertise was not significant [F(1.31,34.17)=0.79; P>0.008], showing similar P2 reduction for both novices and bird experts (Figs 2 and 3).

Component latencies

ANOVAs performed on individually measured peak latencies revealed a main effect of sound condition for the P2 component at Pz [F(1.29,33.51)=7.52; P<0.008; η_p^2=0.22]. The P2 peak amplitudes evoked by voices (191.8±36) and birdsongs (189.4±34) at Pz occurred before those evoked by environmental sounds (233.1±61).

Scalp topography

Figure 4 shows scalp topography maps at 100 and 200 ms after onset. Novices and bird experts showed a similar central negative component at 100 ms for all sound conditions. The central positivity observed at 200 ms (Fig. 4, right panel) was much more frontal in the novices and was slightly right hemispheric-dominant in the bird experts.

Discussion

The goal of this study was to explore the neural correlates of expert auditory processing in bird experts. Two main findings emerged: first, the main difference related to expertise was that the bird experts had a smaller P2 response than the novices at Pz and Cz, for all three sound categories. Second, birdsongs were found to elicit smaller N1 amplitudes at Pz, and smaller P2 amplitudes at Fz, than the other two sound categories, irrespective of the bird expertise of the participants.

Smaller P2 in bird experts

The results are not in concordance with the prediction on the basis of results obtained with musicians [7], that the bird experts would show greater P2 amplitudes. The N1 and P2 components are the two principal auditory components to appear after the onset of a sound stimulus. The N1 component is usually modulated by attentional processes [15], whereas some studies have demonstrated that the P2 component is modulated by different auditory discrimination learning tasks [16,17].

Sheehan et al. [18] investigated the impact of speech discrimination training on the P2 component, while controlling for exposure to training stimuli during pretest and posttest phases of the experiment. Both groups were exposed to instances of speech stimuli during pretraining and posttraining testing, but only the experimental group received explicit speech discrimination training. Their results showed that P2 amplitudes, in response to speech stimuli, increased after the training task, but the increase could be observed in both the trained and the untrained groups. The authors concluded by warning against the assumption that an increase in P2 amplitude automatically reflects an increase in discrimination ability. Unfortunately, the bird experts who participated in this experiment were not assessed for birdsong discrimination but were rather recruited based on years of bird listening activities. This prevented us from investigating possible correlations between behavioural performance and component amplitudes.

Importantly, the P2 amplitudes were also much more frontally distributed in the bird experts as shown in Fig. 3. This suggests an alternative interpretation: bird experts may have processed all sound categories differently from the novices, perhaps because of greater experience with sound source recognition and identification. Effects of learning, which were initially specific to the trained category, may have generalized to other categories of sounds. Chartrand and Belin [6] observed a similar pattern in musicians:

superior musical instrument discrimination performance was correlated with superior voice timbre discrimination performance, suggesting that training with one category may have influenced the processing of unrelated sound categories.

Goldstone [19] described four mechanisms involved in perceptual learning: attention weighting, imprinting, differentiation, and unitization. Attention weighting suggests that in some stages of perceptual information processing, attention can be shifted towards features and dimensions of the stimuli that are important to the participant for the proper completion of the task. It is plausible that when the bird experts heard birdsongs, they shifted their attention towards what they believed to be relevant acoustical features, as they may do during their bird identification activities, and thus performed the required task with a different strategy to that used by the novices. Amateur musicians, for example, have sound processing advantages with acoustical features that are important for their musical practice [20]. An essential aspect of future investigations of auditory expertise will be the precise measurement of discrimination and recognition performance for the expert and nonexpert stimulus categories.

One possible factor having influenced the present pattern of results is that the bird experts were older than the control group (43 vs. 32 years old on average), although the difference did not reach significance. In previous research, Anderer et al. [21] quantified the effects of ageing on auditory ERP latencies and amplitudes. Although they found that the P2 amplitudes increased with advancing age up to 60 years, their distribution was more frontal with increase in age. Although these data are interesting and provide some explanation to support the frontal distribution in the bird experts, they cannot account for their smaller P2 amplitudes at the Cz and Pz electrode sites.

Smaller responses to birdsongs in novices and bird experts

The birdsongs elicited smaller N1 and P2 amplitudes in the Cz and Pz sites, respectively. This may be explained partly by the fact that the three sound categories (i.e. voices, birdsongs and environmental sounds) do not have the same long-term average spectral distribution (Fig. 1). In a study conducted by Shahin et al. [11], P2 auditory-evoked responses were modulated by the spectral complexity of musical sounds in pianists and nonmusicians. In that study, participants were presented with three piano tones that differed in the number of harmonics they contained and a pure tone matched to the fundamental frequency. The P2-evoked responses were larger in both musicians and nonmusicians when they were presented with the more complex piano tones. According to the average power spectrum graph and the spectral analysis examples of the three sound categories used in this experiment (Fig. 1), the birdsongs have much more acoustic energy in the high-frequencies range than do the other two sound categories. Although the effect described in the Shahin et al. [11] study is not the same as that described in this study, both results present some evidence that the auditory-evoked P2 component seems to be modulated by the spectral information contained in sound stimuli.

Conclusion

This study has shown that ERP-amplitude differences between groups of auditory experts and novices may not always translate into larger P2 amplitudes in the expert group. Instead, the expert group may have used a different strategy, reflected in a more frontal positivity.

References

1. Kolodner JL. Towards an understanding of the role of experience in the evolution from novice to expert. Int J Man Mach Stud 1983; 19:497–518.

2. Gauthier I, Skudlarski P, Gore JC, Anderson AW. Expertise for cars and birds recruits brain areas involved in face recognition. Nat Neurosci 2000; 3:191–197.

3. Gauthier I, Curby KM. A perceptual traffic jam on highway N170. Curr Dir Psychol Sci 2005; 14:30–33.

4. Pitt MA. Perception of pitch and timbre by musically trained and untrained listeners. J Exp Psychol Human 1994; 20:976–986.

5. Minsel TR. Memory for tones in musicians and non-musicians: an expert/novice study. Dis Abs Int: Sci Eng 1996; 57:1479.

6. Chartrand JP, Belin P. Superior voice timbre processing in musicians. Neurosci Lett 2006; 405:164–167.

7. Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 2003; 23:5545–5552.

8. Pantev C, Roberts LE, Schulz M, Engelien A, Ross B. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 2001; 12:1–6.

9. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. Nature 2000; 403:309–312.

10. Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. Cogn Brain Res 2002; 13:17–26.

11. Shahin A, Roberts LE, Pantev C, Trainor LJ, Ross B. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. Neuroreport 2005; 16:1781–1785.

12. Oldfield RC. The assessment and analysis of handedness: the Edinburgh Inventory. Neuropsychologia 1971; 9:97–113.

13. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. J Neurosci Meth 2004; 134:9–21.

14. Stevens J. Applied multivariate statistics for the social sciences. 3rd ed. Mahwah, New Jersey: Lawrence Erlbaum; 1996.

15. Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, et al. Modulation of early sensory processing in human auditory cortex during auditory selective attention. Proc Natl Acad Sci USA 1993; 90: 8722–8726.

16. Reinke K, He Y, Wang C, Alain C. Perceptual learning modulates sensory evoked response during vowel segregation. Cogn Brain Res 2003; 17: 781–791.

17. Trainor LJ, Shahin A, Roberts LE. Effects of musical training on the auditory cortex in children. Ann NY Acad Sci 2003; 999:506–513.

18. Sheehan KA, McArthur GM, Bishop DVM. Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? Cogn Brain Res 2005; 25:547–553.

19. Goldstone RL. Perceptual learning. Annu Rev Psychol 1998; 49:585–612.

20. Tervaniemi M, Castaneda A, Knoll M, Uther M. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. Neuroreport 2006; 17:1225–1228.

21. Anderer P, Semlitsch HV, Saletu B. Multichannel auditory event-related brain potentials: effects of normal aging on the scalp distribution of N1, P2, N2 and P300 latencies and amplitudes. Electroencephaloge Clin Neurophysiol 1996; 99:458–472.

Figure 1. Auditory stimuli. (a) Long-term average spectrum of the three sound categories. X-axis, frequency (Hz); Y-axis, amplitude (arbitrary units). (b) Spectrograms of examples of stimuli from each sound category. X-axis, time;Y-axis, frequency (0-11 025 Hz).

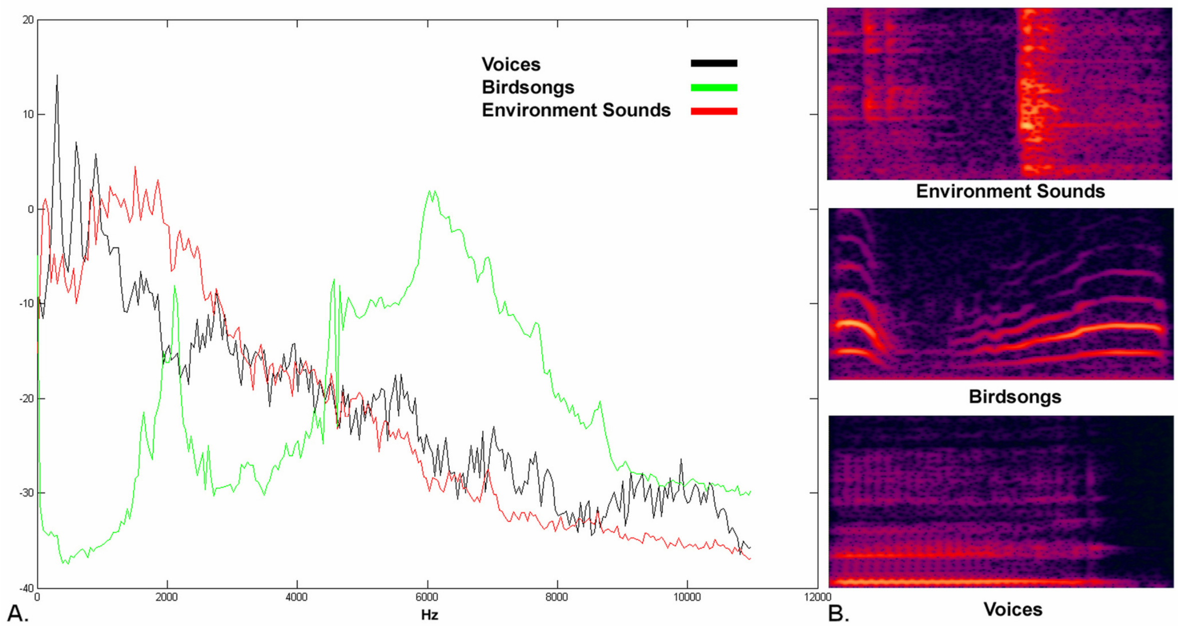


Figure 2. Auditory-evoked potentials. Auditory-evoked potentials elicited by voices (left), birdsongs (middle), and environmental sounds (right). X-axis, time (ms);Y-axis, scalp potential (μV). Rows correspond to three main electrodes: Fz, Cz, and Pz. Significant differences in the bootstrap test are highlighted in yellow.

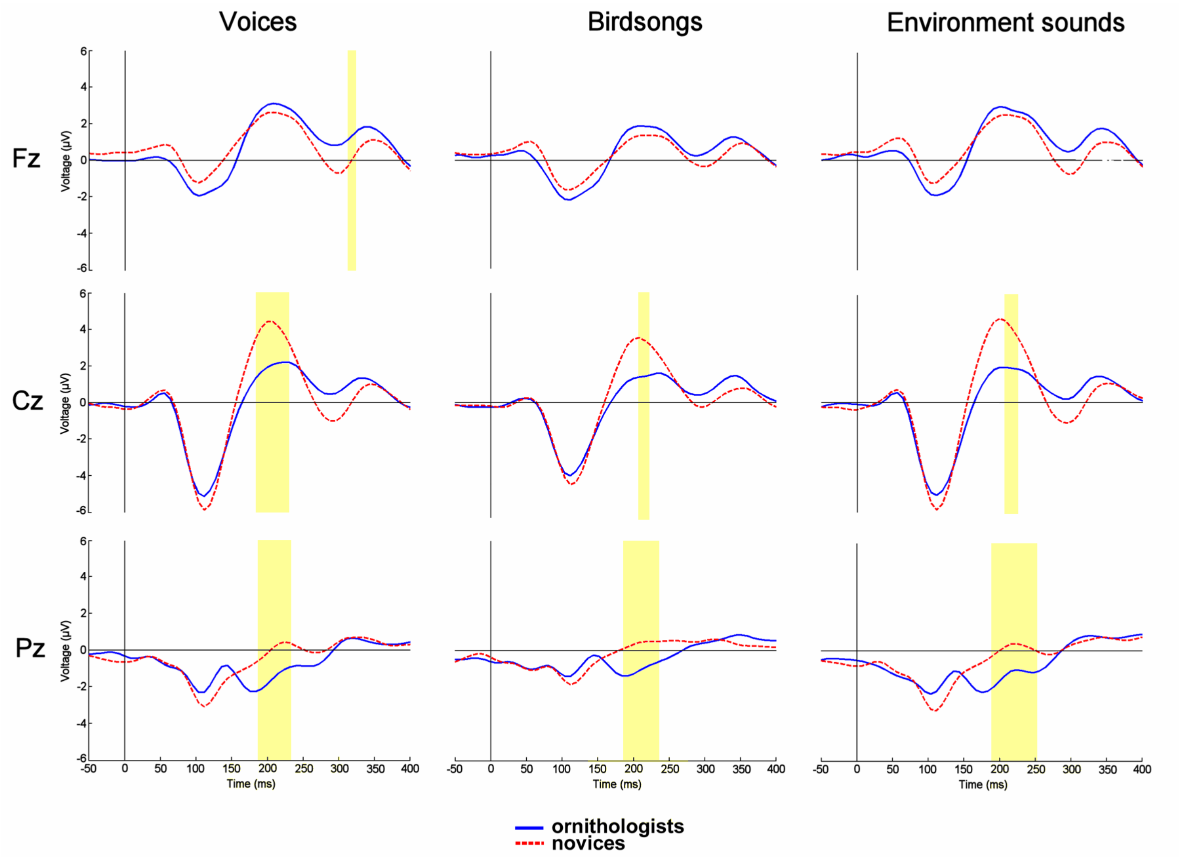


Figure 3. N1and P2 amplitudes. Group-average amplitude of the N1and P2 components at Fz (upper panel), Cz (middle panel), and Pz (lower panel) sites. Columns correspond to the three stimulus categories. Y-axis, peak amplitude (μV). Bars indicate standard deviation.

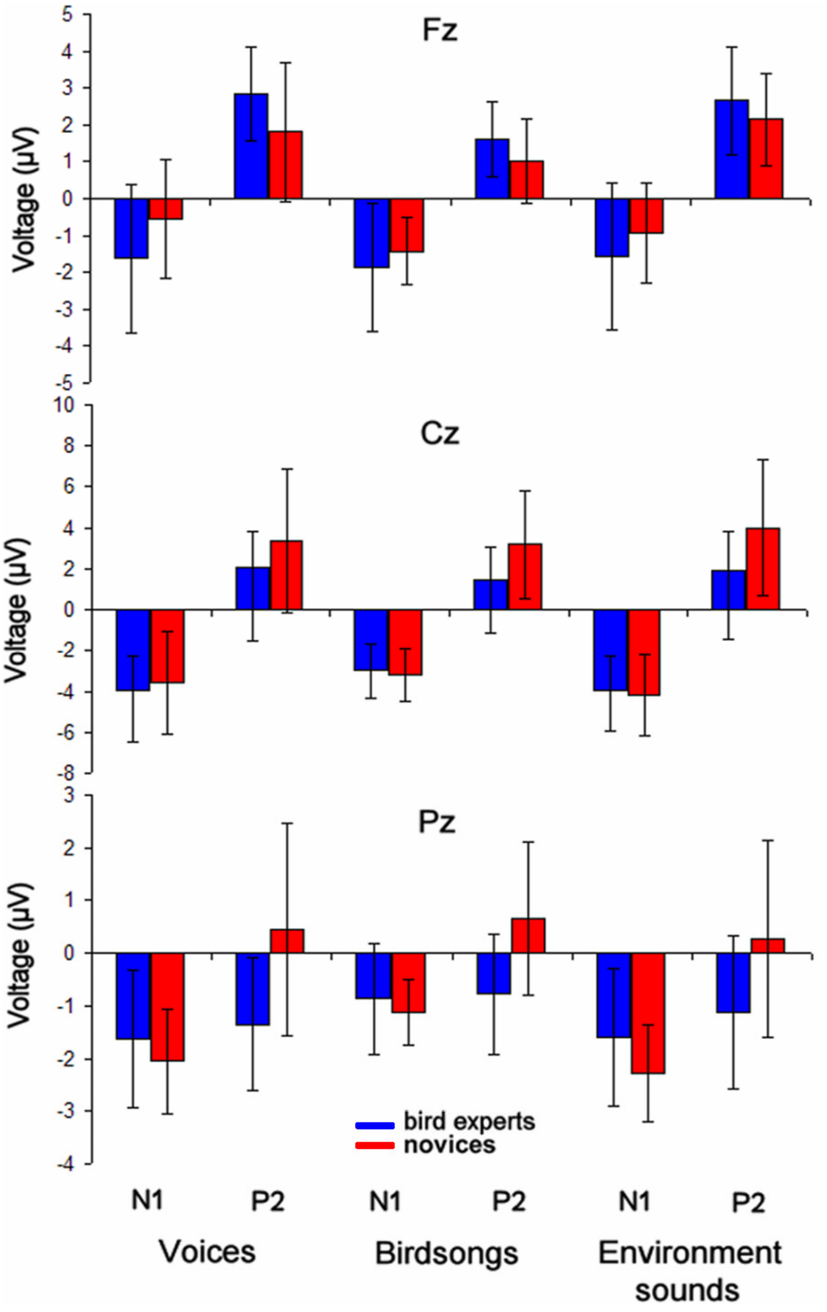
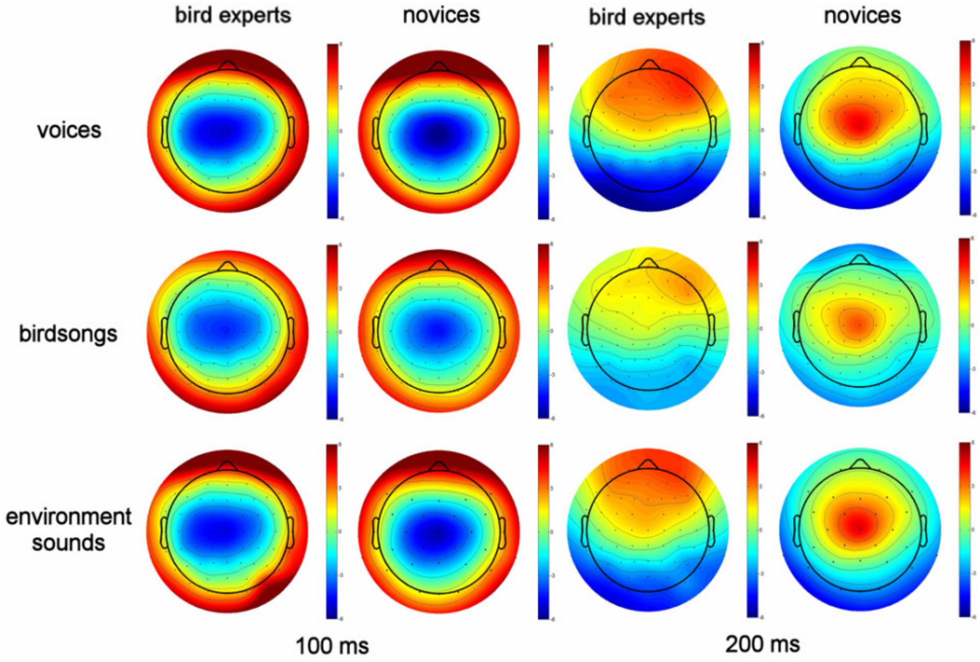


Figure 4. Electroencephalography scalp topography. Scalp potential maps are shown at 100 and 200 ms for bird experts and novices, for all three sound categories.Colourscales indicate scalp potential (μV).The right frontal positivity in bird experts is clearly seen in the right panel (200 ms).



Article 4 : Auditory expertise recruits voice-selective auditory cortex

Jean-Pierre Chartrand, Isabelle Peretz et Pascal Belin

Soumission en préparation

Auditory expertise recruits voice-selective auditory cortex

Jean-Pierre Chartrand[1], Isabelle Peretz[1], and Pascal Belin[1 2].

[1]BRAMS, Université de Montréal, Montréal, Canada

[2]Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

Correspondence and requests for reprints to:
Jean-Pierre Chartrand,
Département de Psychologie, C.P. 6128, succursale Centre-Ville,
Montréal, Québec, Canada, H3C 3J7
Tel : + 1 1 (514) 343-6111 ext. 2997; fax: +1 1 (514) 343-2175;

2769 words, 8 text pages, 4 figures, 1 table

Abstract

Whether the temporal voice areas (TVA) of auditory cortex are engaged by the acoustical structure of voices or by the expertise with voices of normal listeners is unknown. Here we measured cerebral activity in auditory experts — ornithologists and guitar makers — and found a stronger response to expert than non-expert sound categories in the TVA, demonstrating their involvement in expert-level cerebral sound processing.

There is a long-standing debate in visual neuroscience on the functional significance of activations in the face-selective areas of the fusiform gyrus (FFA). While some researchers believe they reflect selectivity for faces as an ecologically-relevant category of visual stimuli[1], others interpret these activations as reflecting expertise-dependent processing of visual stimuli — faces or not[2]. Although it has not been asked yet in auditory neuroscience, the analogous expertise problem is particularly relevant for the understanding of auditory cortex functional organization. The temporal voice areas (TVA) of auditory cortex are known to respond preferentially to sounds of human voice in normal subjects[3,4]. Yet it remains unknown whether a greater response to voices reflects pure selectivity to voices or the prolonged experience with this specific sound category.

Brain response to specific classes of stimuli has been studied extensively in musicians using various imaging methods such as electroencephalography[5], magnetoencephalography[6], and functional magnetic resonance imaging (fMRI)[7]. The general findings show that long-term training with a musical instrument is associated with an enhanced response of the auditory cortex or neighboring regions, suggesting the role of specialized cortical sites in the perception of a homogeneous sound category other than the human voice. The neural basis of auditory expertise has been largely studied with musicians, who acquire a great deal of aptitudes such as pitch and rhythm perception through musical practice[8]. However, a unique aspect of voice perception lies in one's ability to identify or discriminate a single individual among a great number of different people, by recognizing the timbre of this individual's voice. Very few studies have looked at other types of auditory expertise which, in our opinion, would be more comparable to the expertise that everyone has with the human voice. In order to address this gap, we recruited two groups of auditory experts who present with similar auditory abilities to voice perception. We asked whether the TVA would be recruited by different categories of sounds of expertise in guitar makers (n=6), bird experts (n=8) and non-experts (n=8). More precisely, we hypothesised that the TVAs would be involed in expertise processing in both groups of experts, since the bilateral superior temporal sulcus (STS) has been showed to respond to timbre variations[9].

To assess behavioral expertise, the participants performed memory and discrimination auditory tasks with birdsongs and guitar sounds (see supplementary Methods). As expected, experts performed better with their respective sound of expertise, for both memory ($F(1,12) = 25.414, p < 0.001$) and discrimination ($F(1,12) = 34.377, p < 0.001$) tasks as shown by the significant interactions in Figure 1. In contrast, their performance with other sound categories (human voices and coins) was not different from the other group. There was no difference between the two groups of experts for the memory ($F(1,12) = 0.41, p = 0.843$); and discrimination ($F(1,12) = 0.87, p = 0.369$) tasks.

In the second part of the study, the participants underwent an fMRI scanning session. First, a voice localizer run contrasting vocal and nonvocal sounds was performed for individual localization of the TVA (see supplementary Methods). In both experts and non-experts, the TVA showed the typical pattern of voice-selective activity centered on middle and anterior parts of the superior temporal sulcus bilaterally[3,4]. Comparing the TVA between experts vs. non-experts showed no significant difference ($P < 0.01$ FDR). Consequently, the TVA averaged from the three groups (left max.: x,y,z = -60, -15, 0; 13230 mm³; $t = 12.79$; right max.: x,y,z = 60, 3, -9; 17172 mm³; $t = 10.24$; $P < 0.01$ FDR, $k = 20$) were used as the region of interest in which effects of expertise were investigated. Participants were then scanned in an event-related protocol in which they listened to brief samples of human voices, birdsongs and guitar sounds presented in a pseudo random order with an 5s SOA while performing a one-back task (Cf. Supplementary methods). Results from the one-back task revealed a significant main effect of sound ($F(2,24) = 4.405, p = 0.023$). Both groups of experts have responded more slowly to the human voice samples (886.78ms) than the birdsongs (832.85ms) or guitar sounds (827.8ms). The group x sound interaction was not significant ($F(2,24) = 3.223, p = 0.058$).

The fMRI effects of interest were investigated in the above-defined TVA regions of interest only. In these bilateral clusters, the response to sounds of human voice was strongest for all three groups. Including the non-experts in the analysis, there was no main effect of group ($F(2,54)= 5.854$; $P < 0.005$ uncor., $k = 5$). The main goal of this study was to assess the presence of a common region involved when experts listened to their sounds of expertise. While excluding the non-experts from the planned contrast, we observed a significant group x sound of expertise interaction in a 513mm³ cluster in the left TVA only (max: x, y, z = -60, -12, -12; $t = 3.33$, $p < 0.005$ uncorrected). In this cluster, the same pattern of interaction observed at the behavioural level was found: activity in response to bird songs was greater than in response to guitar sounds in the bird expert group, while the opposite pattern was found for the guitar experts. No significant expertise interaction was found in the right TVA.

By limiting the analysis to the TVAs as a region of interest, we have found that the central part of the left TVA responds more strongly to sounds of expertise. These findings provide a neural correlate of auditory expertise, an issue which so far has mainly been studied in the particular population of musicians. Musicians' expertise has both multi-modal and motor components, recruiting many different brain areas[8,10]. By comparing brain activations associated with two different classes of expertise stimuli – one learned in part through musical practice and the other one learned in a non-musical context – this study demonstrates, for the first time, the presence of cortical areas involved in auditory expertise in musical or non-musical practice. Since all humans are considered to be voice experts, other types of auditory expertise such as the ones assessed in this study are easily comparable to voice perception. Thus, the presented results redefine the role of the TVAs, which are thought to be highly selective to human voice[3,4]. The results show that they do not exclusively process voice stimuli but might also contribute to expert-level processing of other sound categories. One premise in the study of perceptual expertise is that experts have acquired their expertise through exposure and stimulus processing. Gradually, they become more knowledgeable, faster, more accurate and less distracted[11]. While the litterature is sparse regarding this subject, it is conceivable to think that the voice selectivity of the TVAs found in human adults has emerged, or is at least enhanced from prolonged exposure to vocal stimuli.

Moreover, the TVAs are located along the bilateral mid to anterior STS. This anatomical region is known to be part of the proposed auditory "what" pathway[12,13], responsible for extracting the acoustic signatures of auditory objects. The involvement of the STS in voice expertise, but also in bird and guitar expertise, reinforces the proposal that the bilateral STS might have a generic role in higher-level auditory processing and recognition. For example, the right STS is suggested to be the last and third anatomo-functional step for the analysis of the spectral envelope of auditory objects[14].

Finally, these data inform the debate on expertise vs. category specificity in visual neuroscience, by showing that similar issues are valid in other sensory domains, potentially re-igniting and widening this debate. Comparing other auditory expertise domains with voice perception is an essential procedure to better understand the development of voice selectivity in the auditory cortex. The next steps in understanding the nature of voice areas and other expertise domains lie in both developmental and laboratory-training studies. It has been recently shown that 7-month-olds but not 4-month-olds have increased responses in left and right superior temporal cortex to the human voice when compared to nonvocal sounds[15]. However, little is known about the normal developmental course of voice processing, as compared to speech processing. Also, training adult participants with novel auditory stimuli would enable the assessment of pre and post-training neural correlates. Future studies should unravel how both development and exposure interact with each other to better understand “how voices are special”.

References

1. Kanwisher, N., & Yovel, G. Phil Trans R Soc B. **361**, 2109-2128 (2006).
2. Gauthier, I., & Bukach, C. Cognition. **103**, 322-330 (2007).
3. Belin, P. et al. Nature. **403**, 309-312 (2000).
4. Belin, P. et al. Brain Res Cogn Brain Res. **13**, 17-26 (2002).
5. Shahin, A. et al. J. Neurosci. **23**, 5545-5552 (2003).
6. Pantev, C. et al. Neuroreport. **12**, 169-174 (2001).
7. Margulis, E.H. et al. Hum Brain Mapp. **30**, 267-275 (2009).
8. Stewart, L. Clin Med. **8,** 304-308 (2008).
9. Menon, V. et al. Neuroimage. **17**, 1742-1754 (2002).
10. Peretz, I. et al. Annu Rev Psychol. **56**, 89-114 (2005).
11. Palmeri, T. J., Wong, A. C. N., & Gauthier, I. Trends Cogn Sci. **8**, 378-86 (2004).
12. Belin, P., & Zatorre, R. J. Nat Neurosci. **3**, 965-966 (2000).
13. Wang, W. J., Wu, X. H., & Li, L. Neurosci Bull. **24**, 173-182 (2008).
14. Griffiths, T. D. et al. Hear Res. **229**, 46-53 (2007).
15. Grossman, T., Oberecker, R., Koch, S. P., & Friederici, A. D. Neuron. **65**, 852-858 (2010).

Figure 1: Experts' and non-experts' behavioral performance with auditory memory and discrimination tasks. a) Birdsongs and guitar sounds: the interactions between sound category and group of expertise were significant for both the memory tasks ($F(1,12) = 25.291$, $p < 0.025$) and the discrimination tasks ($F(1,12) = 7.033$, $p < 0.025$). b) Voices and environmental sounds (coins): in contrast, there were no significant interaction for the memory ($F(1,12) = 9^{e-6}$, $p = 0.99$) and discrimination ($F(1,12) = 0.87$, $p = 0.37$) tasks, suggesting that expertise did not generalize to other sound categories.

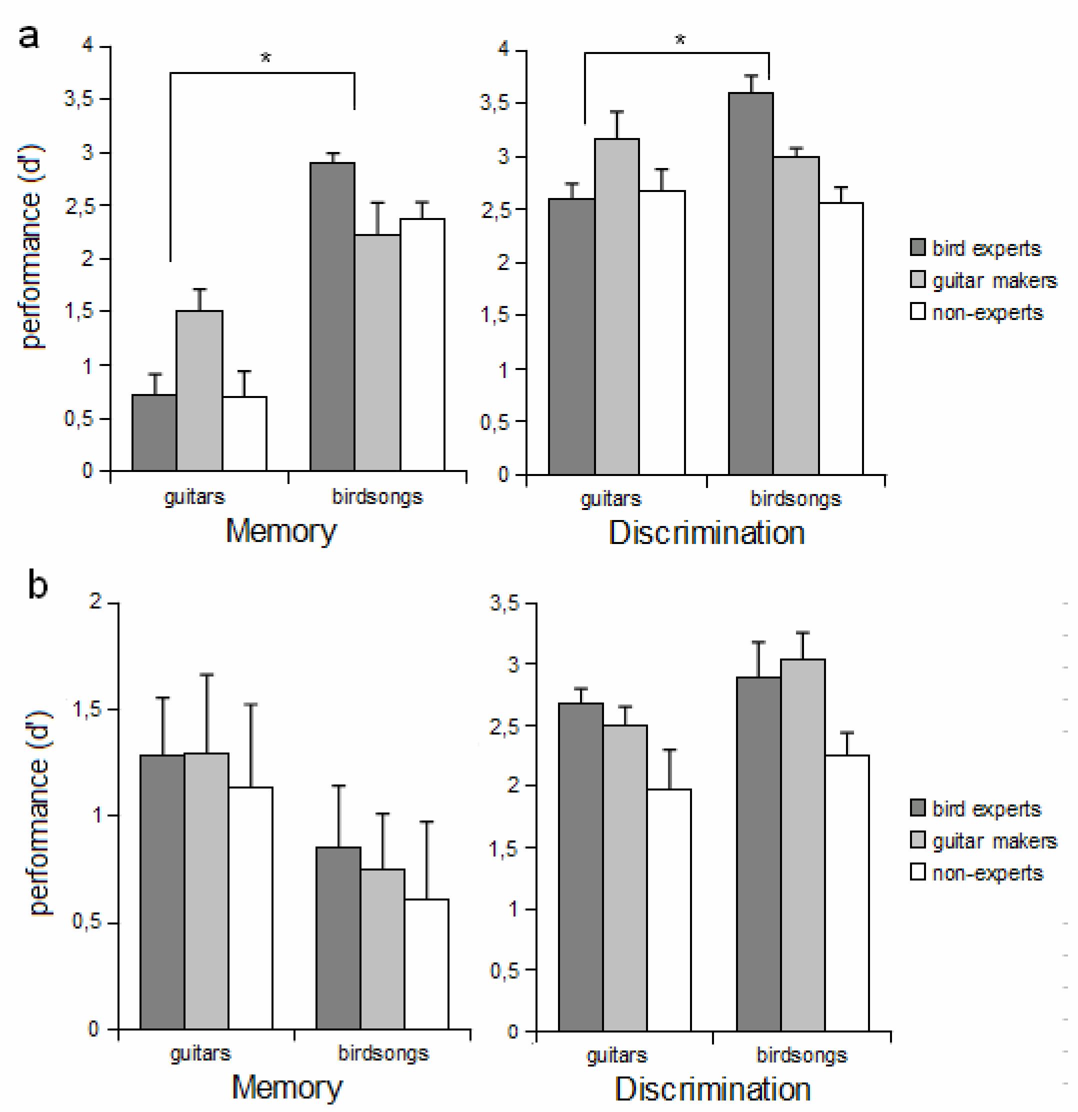
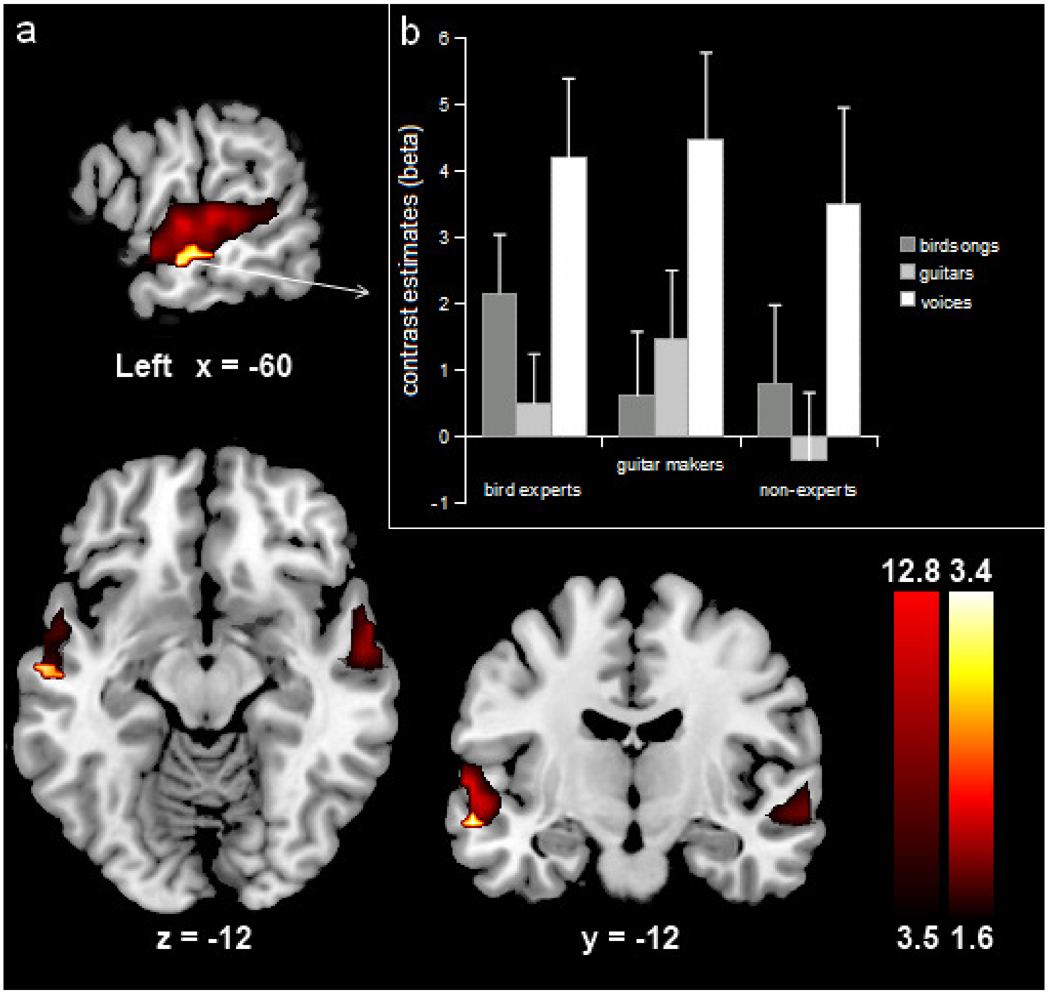


Figure 2: a) Group-average ROIs (red) corresponding to the vocal vs non-vocal contrast from the localizer run (left max.: x,y,z = -60, -15, 0; 490 voxels; $t = 12.79$; right max.: x,y,z = 60, 3, -9; 636 voxels; $t = 10.24$; $P < 0.01$ FDR, $k = 20$). Brain activation (yellow) corresponding to the positive interaction group x sound of expertise (max.: x,y,z = -60, -12, -12; 513 mm^3; $t = 3.33$; $P < 0.005$ uncor.). The Significant signal changes are showed in colorscale (t statistics) overlaid on an anatomical MR image based on the ICBM Single Subject MRI template. The interaction map (yellow) is showed at a lower threshold for display purposes ($P < 0.05$ uncor.). b) Bar graphs present activations for each sound condition for each group at the maximum beta value for the group x sound interaction. Error bars indicate standard error of the mean.



Supplementary Methods

Participants

Sixteen healthy participants were recruited based on their auditory expertise. Eight bird experts (age = 39.5 ± 15.78; 7 males) were recruited in Montreal local bird clubs and mailing lists in the province of Québec (Canada). They all reported regular practice of watching and listening to birdsongs. Eight guitar makers (age = 36.5 ± 10.88; 8 males) were recruited from guitar shops in the Greater Montreal Area (Quebec, Canada). A third group of 8 novice participants was added and recruited in the Greater Montreal Area as well (age = 33 ± 13,09; 8 males). The inclusion criteria for all the expert participants were: at least five years of experience in the respective field of expertise (bird experts: 20.75 ± 14.57; guitar makers: 19.67 ± 8.07). No participant reported neurological or auditory defects. All the guitar makers reported playing guitar on a regular basis and four bird experts reported playing a musical instrument, though not at a professionnal level (piano, oboe, and clarinet). Experience with guitar and bird sounds for the experts participants are summarized in supplementary table 1. None of the non-experts reported playing a musical instrument, as playing a musical instrument was an exclusion factor for the non-experts. The participants were paid for their participation to the experiment and provided written informed consent. The protocol was approved by the research ethics board of *Regroupement Neuroimagerie Québec* (CMER-RNQ).

Behavioral tasks and analysis

The memory tasks first involved a learning part, in which the participants were asked to listen carefully to sound samples. They were told that they would have to recognize the targets among distractors in the second part. During the learning part, each target was presented twice, with an ISI of 1500 ms and a 2000 ms interval between the pairs. The recognition part was administered immediately after the learning part. For each sound presented, the participants were asked to rate the sound familiarity on a 10-point scale (10 being completely sure they had previously heard it). The same procedure was repeated with all the memory tasks. The discrimination tasks followed the memory tasks to avoid interference effects and consisted of pairs of sounds for which the participants had to tell if they were produced from the same sound source (i.e.: same bird species, guitar model, speaker, coin) on a 10-point scale (10 being sure the sounds are from different sources). The sounds were presented with an ISI of 1000 ms. The order in which the tasks were administered was rotated across participants for each group in order to reduce practice or fatigue effects. Prior to the experiment, the difficulty levels of each task had been validated with a non-expert sample (n=28). Items being too easy or too difficult were removed in order to get an average 75% performance rate in the non-expert group. Additional stimuli and task details are shown in supplementary table 2. The 10-point scale values were converted into binary scale values, then converted to the d' statistic for both memory and discrimination tasks. The performance was analyzed using 2 (group) x 2 (sound category) mixed factorial designs with repeated-measures on the second factor. The non-experts' data were not included in the analysis but where nonetheless included in the graphs (Figure 1) to show their performance.

MRI data acquisition and analysis

The functional and anatomical images were acquired on a 3 Tesla Magnetom TRIO Siemens scanner (Siemens Vision Imager, Siemens, Munich, Germany), at the Unité de Neuroimagerie Fonctionnelle (UNF) du Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM) (Montréal, Canada). The functional images were acquired using an echo-planar pulse sequence (FOV 192 mm, matrix size: 64 x 64, slice thickness 3 mm, TE = 30 ms, flip angle 90º, axial slice). The localizer run had 32 interleaved ascending slices, covering the whole brain (TA = 2s, TR = 10s; 62 volumes). The one-back runs consisted of 17 interleaved ascending slices oriented with the participant's temporal lobes (TA = 1040 ms, TR = 2500 ms; 250 volumes). These slices did not cover the whole brain and where aligned with each participant's temporal lobes (supplementary figure 1). T1 weighted 3-d images were acquired using the following parameters: 176 slices, voxel size: 1 x 1 x 1.2 mm, slice thickness = 1 mm TR = 2300 ms, TE = 2.98 ms, inversion recovery preparation time = 900 ms, flip angle 9º, FOV 256 x 240 mm, matrix size: 256x256, slice thickness = 1.2 mm.

Localizer scan

This part contained 40 8s blocks of sounds (16 bit, mono, 22050 Hz sampling rate): 20 blocks of vocal sounds (speech and non-speech) and 20 blocks of non-vocal sounds (i.e. industrial sounds, environmental sounds, as well as some animal vocalizations). All the samples were normalized for mean amplitude using equal loudness contour. The participants were instructed to listen passively to these blocs of sounds while keeping their eyes closed. The auditory stimuli were presented at a comfortable sound level using MRI-compatible electrodynamic headphones (MR confon GmbH, Magdeburg, Germany).

Event-related scanning

In an event-related design, the participants were presented with 30 samples (1.25 s, 16 bit, mono, 44 100 Hz sampling rate) per sound category (human voices, birdsongs, guitar samples, silence), with a 3.75s ISI. Different sounds used for each of the two one-back runs. The human voices included speech (french words) as well as nonspeech vocalizations, while the birdsongs were selected from the « Chants d'oiseaux du Québec et de l'Amérique du Nord » audio CD (2004; Peterson Guides coll., Broquet/Cornell laboratory of ornithology) . The guitars sounds consisted of chords, single tones, and melody excerpts obtained from various online sources. For each sound category, there was a 20% probability that a sound would be played twice in succession occuring in a pseudo-random order across the whole run. The participants were instructed to listen carefully to each sound and to press a button with their index when the exact same sound was consecutively presented twice (1-back task). Two guitar makers and one non-expert were excluded from the entire study because they did not answer for more than half of the trials in these runs.

Data preprocessing and analysis

Image processing and statistical analysis were performed using SPM5 (Wellcome Trust Centre for Neuroimaging; www.fil.ion.ucl.ac.uk/spm). The functional images were spatially realigned, coregistered with the anatomical image, corrected for slice-timing, normalized (MNI T1 template, ICBM, NIH P-20 project) and spatially smoothed using a 6 mm isotropic gaussian kernel. Participant-specific GLMs were estimated with the four sound conditions (3 sound categories + silence). The first level contrasts consisted of each sound condition vs the silence condition. These contrasts were then included and estimated in the 3 (group) x 3 (sound condition) mixed factorial second-level analysis. Once again, the non-experts' data were included in the model estimation in order to plot the contrast estimates but were not part of the interaction statistical test. All the voxels from the functional one-back scans were used for the full brain analysis showed in supplementary figure 2. For the ROI analysis, the bilateral TVAs averaged from the three groups (left max.: x,y,z = -60, -15, 0; 490 voxels; $t = 12.79$; right max.: x,y,z = 60, 3, -9; 636 voxels; $t = 10.24$; $P < 0.01$ FDR, $k = 20$) were specified as an explicit mask when setting up the second level statistical model in SPM. The final maps were displayed on a high definition structural brain template based on the ICBM single subject template brain atlas (http://www.loni.ucla.edu/ICBM/).

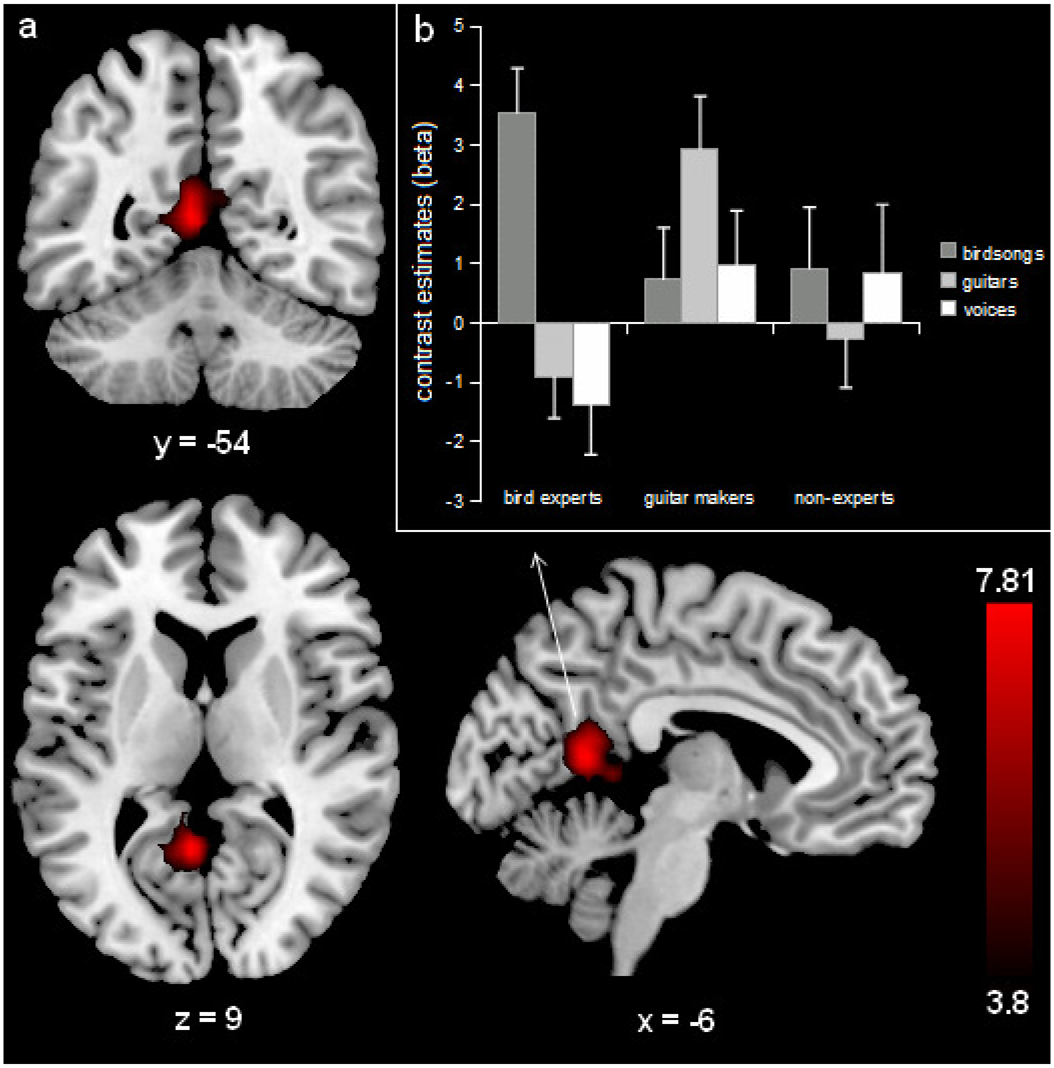
Supplementary Results

A second method was used to assess a group x sound interaction. For this method, no functional ROI was used. Testing for the positive interaction resulted in a 6372 mm^3 cluster in the left posterior cingulate gyrus (supplementary figure 1) (max: $x, y, z = -6, -54, 9$; $t = 7.81$, $p < 0.01$ FDR). This full-brain analysis showed a stronger response to sounds of expertise in the left posterior cingulate cortex, which has been associated with face and voice familiarity in a previous study[1]. This could mean that both groups of experts have developed a familiarity with their own class of stimuli, like non-experts would do with faces or voices.

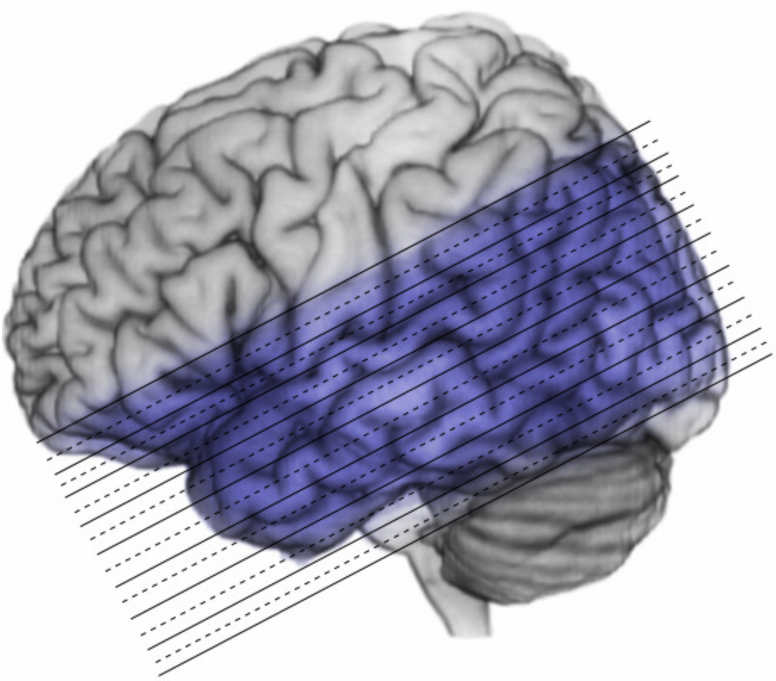
Supplementary references

1. Shah, N. J. et al. Brain. **124**, 804-815 (2001).

Supplementary Figure 1: a) Brain activation (red) corresponding to the positive group x sound interaction with a full brain analysis (max.: *x,y,z* = -6, -54, 9; 236 voxels; $t = 7.81$; $P < 0.01$ FDR). Significant signal changes are showed in colorscale (t statistics) overlaid on an anatomical MR image based on the ICBM Single Subject MRI template. b) Bar graphs present activations for each sound condition for each group at the maximum beta value for the group x sound interaction. Error bars indicate standard error of the mean.



Supplementary Figure 2: In blue, the brain regions covered by the 17 interleaved ascending slices during the one-back runs. The slices were oriented with each participant's temporal lobes.



Supplementary table 1: auditory training in both groups of experts

bird experts	years of practice	age of training onset
1	10	20.0
2	16	13.0
3	12	18.0
4	30	12.0
5	45	18.0
6	6	22.0
7	10	18.0
8	37	25.0
guitar makers	**years of practice**	**age of training onset**
1	18.0	8.0
2	16.0	19.0
3	22.0	12.0
4	14.0	14.0
5	35.0	17.0
6	17.0	15.0

Supplementary table 2: Stimuli length and task trials for the behavioral tasks. The initial number of trials (memory) or pairs (discrimination) are showed in parentheses, before correcting for a performance rate of 75% in control participants (n = 28).

memory	sample length (ms)	targets	trials (initially)
birdsongs	1325	10	19 (20)
guitars	625	5	10
voices	600	10	19 (20)
pennies	1000	6	9 (12)
discrimination	sample length (ms)	samples	pairs (initially)
birdsongs	1150	36	188 (194)
guitars	625	16	56
voices	600	38	133 (192)
pennies	1000	12	75 (120)

Discussion générale

Le but de la présente thèse était de préciser la spécificité des processus et des régions impliqués dans le traitement de la voix. L'étude de divers experts auditifs, à l'aide de différentes méthodes d'expérimentation, semble être une solution pertinente et novatrice pour comparer la perception de la voix humaine à d'autres catégories auditives. Dans les prochaines pages, il sera question de discuter de l'apport scientifique des résultats présentés et des limites inhérentes à ceux-ci. D'autre part, de nouvelles hypothèses de recherche seront proposées afin de mieux répondre à la question conductrice de cette thèse : « La voix est-elle spéciale? ». Finalement, nous proposerons une intégration des données entourant les différents types d'expertises auditives à la perception de la voix.

Entraînement musical et traitement de la voix

L'hypothèse à l'origine de la première étude était qu'un entraînement musical aurait un impact positif sur le traitement de la voix humaine, puisqu'il avait déjà été démontré que les musiciens possédaient de bonnes habiletés auditives pour le traitement du timbre et de la hauteur (Münzer, Berti, & Peckman, 2002; Pitt, 1994; Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Micheyl, Delhommeau, Perrot, & Oxenham, 2006; Spiegel & Watson, 1984). Nos résultats ont montré que les musiciens amateurs étaient meilleurs que les non-musiciens pour discriminer des timbres de voix humaine, contribuant ainsi à la littérature qui s'intéresse aux effets d'un entraînement musical sur la perception auditive et plus particulièrement sur la perception de la voix.

Les résultats de notre étude ont récemment été corroborés par une étude qui a examiné d'autres paramètres que le timbre vocal (Eadie, Van Boven, Stubbs, & Giannini, 2010). Cette étude s'est penchée sur l'influence d'un entraînement musical sur la discrimination des hauteurs et du jugement de voix dysphoniques. Les résultats indiquent que comparativement aux non-musiciens, les musiciens montrent une meilleure fidélité intra et inter-sujets pour des jugements de voix dysphoniques. De plus, les auteurs ne remarquent qu'une corrélation faible à modérée entre la capacité de discrimination de la hauteur et les jugements de la voix. Cette tendance suggère que la meilleure capacité de discrimination des hauteurs chez les musiciens n'est pas suffisante pour prédire leur fidélité dans le jugement de voix dysphoniques. Les auteurs évoquent l'implication d'autres habiletés auditives, qui ne sont cependant pas explicitées par ceux-ci et qui demeurent à être identifiées. En somme, cette étude réplique l'avantage des musiciens pour traiter la voix humaine et nous renseigne que cette différence ne peut être expliquée uniquement par une meilleure performance de discrimination des hauteurs.

À la lumière de ces données, il nous paraît difficile de déterminer de quelle façon l'entraînement musical améliore le traitement de la voix. Plusieurs études de Kraus et collaborateurs effectuées à l'aide des potentiels évoqués au niveau du tronc cérébral donnent certains indices. Ces études suggèrent que l'expérience musicale aurait une influence générale sur le système auditif et plus particulièrement dans les structures sous-corticales. Plus précisément, les résultats indiquent entre autres un encodage plus fidèle des informations linguistiques tonales dans le tronc cérébral des musiciens, comparativement aux non-musiciens (Musacchia, Sams, Skoe, & Kraus, 2007). Par ailleurs, la fidélité de cet encodage est corrélée avec les performances comportementales, de même que la durée et le début de l'entraînement musical (Musacchia et al., 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007). Outre le meilleur encodage de la hauteur, des données similaires sont observées avec le timbre d'un stimulus non-linguistique, tel qu'un pleur de bébé (Strait, Kraus, Skoe, & Ashley, 2009).

Finalement, une dernière étude a tenté d'établir un lien entre les réponses corticales et sous-corticales. Musacchia, Strait, et Kraus (2008) ont présenté la syllabe « *da* » à des musiciens et des non-musiciens. Les musiciens se distinguaient tout d'abord par un meilleur encodage de la périodicité et de la fondamentale au niveau du tronc cérébral. Chez le même groupe, il y avait une association entre l'encodage de la fondamentale, la pente P1-N1 et la latence des ondes N2-P2. Les auteurs ont aussi vérifié le lien entre les mesures physiologiques et l'étendue de l'entraînement. Ils ont observé que le nombre d'années d'entraînement était corrélé avec une plus faible latence de la réponse d'apparition au tronc cérébral, et que l'âge du début de l'entraînement était lié aux amplitudes des harmoniques dans le tronc cérébral. Par ailleurs, les scores d'aptitude musicale, mesurés à l'aide de deux échelles comportementales (Seashore et MAT-3) étaient associés à la pente des ondes N1-P2, de sorte qu'un meilleur résultat était corrélé à une pente plus forte. L'ensemble de ces données démontre que l'entraînement musical présente une association complexe entre les habiletés musicales et les réponses auditives sous-corticales et corticales. La présence d'une multitude de corrélations entre plusieurs mesures rend tout de même laborieuse une interprétation qui, selon nous, se voudrait intégrative. Il est difficile d'indiquer, par exemple, l'ordre ou la direction dans laquelle ces adaptations se sont produites chez les musiciens.

Pour expliquer ces résultats, Musacchia et collaborateurs (2008) font référence à la théorie de la hiérarchie inversée, selon laquelle l'apprentissage serait un processus descendant guidé, qui commencerait à des niveaux de traitement plus complexes (corticaux) et descenderait vers des niveaux de traitement de bas niveau (sous-corticaux) lorsque nécessaire (*Reverse hierarchy theory*; Ahissar & Hochstein, 2004). Dans le cas des musiciens, les apprentissages auditifs initiaux se formeraient au niveau cortical, puis se propageraient aux réseaux sous-corticaux, suite à des demandes plus grandes. Ce processus impliquerait par exemple une rétroaction du cortex auditif vers le collicule inférieur par la voie corticofugale (Suga, Gao, Zhang, Ma, & Olsen, 2000). Bien que ce modèle ait initialement été suggéré pour les apprentissages visuels, des résultats récents valident l'utilisation d'un tel modèle en audition (Nahum, Nelken, & Ahissar, 2010). Cette procédure permettrait une meilleure sensibilité neuronale en augmentant le ratio signal-bruit sur les informations acoustiques complexes perçues par le tronc cérébral, dont celles qui seraient entre autres impliquées dans la musique ou la communication vocale.

Les résultats comportementaux obtenus dans notre troisième étude sont également compatibles avec la théorie de la hiérarchie inversée. La partie comportementale de cette étude a révélé une interaction entre le groupe d'expertise et la performance aux tâches de mémoire et de discrimination pour les sons d'expertise (chants d'oiseaux et guitares). Ces résultats ont permis d'objectiver une différence de performance entre les deux groupes d'experts dans leur propre catégorie d'expertise. Or, compte tenu que les luthiers possédaient une expérience musicale, on aurait pu s'attendre à de meilleures performances dans les autres catégories auditives, comme les voix humaines par exemple (cf. deuxième article). Les luthiers recrutés nous ont tous mentionné que leur principale activité avec les guitares était dédiée à la fabrication et à l'entretien de celles-ci, plutôt qu'à la pratique musicale. Il se peut que leurs habiletés musicales aient diminué avec une faible pratique, entraînant du même coup une baisse de leur avantage auditif général. Il est possible aussi qu'ils n'aient jamais atteint un niveau d'entraînement musical suffisamment élevé pour qu'un effet soit perceptible. La meilleure performance des luthiers et des ornithologues amateurs, observée uniquement dans leur propre catégorie d'expertise, pourrait être expliquée par un recours à des habiletés de haut niveau, comme l'utilisation d'étiquettes verbales et de représentations en mémoire à long-terme. Selon la théorie de la hiérarchie inversée, il n'y aurait pas de modifications apportées aux réseaux neuronaux de bas niveau lorsque les processus de haut-niveau sont suffisants pour accomplir la tâche. Il est plausible que les experts de la troisième étude n'aient pas bénéficié du développement d'une meilleure sensibilité des structures auditives sous-corticales pendant le développement de leur expertise.

Les meilleures performances comportementales des musiciens dans le traitement vocal pourraient être le reflet d'une influence du réseau sous-cortical dans l'évaluation de paramètres acoustiques similaires aux instruments de musique et à la voix, comme la hauteur et le timbre. Il est donc très difficile de s'appuyer uniquement sur ces résultats pour se prononcer en faveur des similarités entre le traitement de la voix et des instruments de musique. Par ailleurs, davantage d'études seraient nécessaires afin de déterminer s'il existe des effets comportementaux propres à la voix ou des tâches vocales de haut niveau pour lesquelles les musiciens ne seraient pas meilleurs que les non-musiciens. Cela pourrait être le cas par exemple dans des tâches potentiellement peu pratiquées par les musiciens, telles que l'évaluation de l'âge ou du sexe, ou la reconnaissance d'individus encodés en mémoire à long-terme.

Réponse corticale aux chants d'oiseaux

La seconde étude présentée avait pour but de comparer les potentiels évoqués auditifs liés aux chants d'oiseaux entre des ornithologues amateurs et des participants novices. Dans un deuxième temps, l'identification d'une composante répondant plus fortement aux chants d'oiseaux aurait permis de comparer celle-ci à la réponse de la voix humaine. Plusieurs données de la littérature justifiaient l'intérêt porté à l'onde auditive P2 comme corrélat électrophysiologique de l'expertise auditive, que ce soit avec du matériel verbal ou non-verbal (Sheehan, McArthur, & Bishop, 2005; Shahin, Bosnyak, Trainor, & Roberts, 2003).

Non seulement la deuxième étude n'a pas été en mesure d'identifier des potentiels de plus grande amplitude pour les chants d'oiseaux chez les experts, mais ceux-ci ont démontré une distribution topographique nettement plus frontale pour l'amplitude P2. Par conséquent, les résultats ont été difficiles à comparer entre les deux groupes, puisque les ornithologues démontraient un comportement différent des novices pour les trois catégories de stimuli. Plusieurs explications possibles de cette différence ont déjà été discutées dans l'article et ne seront pas reprises dans la présente section. Par contre, de nouvelles données concernant l'onde P2 ont émergé depuis la réalisation de cette étude et auraient avantage à être considérées dans la planification d'une nouvelle étude. Nous savons désormais que cette composante représente aussi un indice de familiarité pour des sons de l'environnement, tels que des vocalisations animales (Kirmse, Jacobsen, & Schröger, 2009). De plus, Charest et collaborateurs (2009) ont observé une positivité fronto-temporale plus forte pour la voix, comparativement à des sons non-vocaux. Les auteurs indiquent que cette positivité émerge aussi tôt que 162 ms et plafonne à une latence d'environ 200 ms, ce qui ressemble fortement à l'onde P2. Dans une prochaine étude, il serait capital de vérifier l'effet de la présentation de voix et d'instruments de musique variant selon le degré de familiarité chez des musiciens et des non-musiciens. Le recours à des musiciens, plutôt qu'à des ornithologues amateurs, offrirait l'opportunité de comparer des effets pour des catégories sonores bien documentées, tout en conservant des caractéristiques acoustiques très similaires entre les deux catégories.

Malgré la grande importance accordée à l'onde auditive P2 par la littérature, peu d'informations existent quant à sa localisation et à son étendue cérébrale. Bien que Charest et collaborateurs (2009) n'aient pas inclus d'analyse de source dans leur étude, ils suggèrent que la positivité frontale à la voix correspondrait aux régions corticales préférentielles à la voix, soit le long des sillons temporaux supérieur bilatéraux. De leur côté, les études qui se sont penchées sur la pratique musicale estiment que les générateurs de cette onde sont situés dans les aires auditives primaires (e.g., Shahin et al., 2003; Baumann, Meyer, & Jäncke, 2008). Les méthodes comme l'ÉEG ou la MEG offrent la meilleure résolution temporelle parmi les méthodes d'imagerie non-invasives (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasamaa, 1993). En contrepartie, la localisation de sources et leur étendue demeure imparfaite dès que plusieurs sources sont actives en même temps (Sharon, Hamalainen, Tootell, Halgren, & Belliveau, 2007). Puisqu'une partie importante du problème de l'expertise est basée sur l'implication d'une même région corticale dans le traitement de la voix et d'autres catégories auditives, l'utilisation de potentiels évoqués n'est pas la méthode optimale pour répondre à cette question. C'est la raison pour laquelle nous nous sommes tournés vers l'IRMf dans la troisième étude.

Localisation anatomo-fonctionnelle de l'expertise auditive

La troisième étude avait pour but d'identifier une région commune à la perception de sons d'expertise chez deux groupes d'experts. Notre hypothèse principale était que les aires de la voix montreraient un effet d'expertise. En limitant nos analyses aux aires temporales de la voix comme région d'intérêt, nous avons identifié que la partie centrale du STS gauche répond plus fortement aux stimuli d'expertise. À notre connaissance, nous démontrons pour la première fois un locus commun à la perception de sons d'expertise chez deux groups d'experts qui ne sont pas tous musiciens.

Ce résultat corrobore des études précédentes qui ont démontré une réponse préférentielle pour l'instrument de pratique dans les aires temporales (Pantev, Roberts, Schulz, Engelien, & Ross, 2001; Shahin, Roberts, Pantev, Aziz, & Picton, 2007; Margulis, Mlsna, Uppunda, Parrish, & Wong, 2009). Dans une étude similaire à la nôtre, Margulis et collaborateurs (2009) ont évalué le même type d'interaction chez des flutistes et des violonistes. Les résultats ont montré que la seule région temporale significative pour l'instrument de pratique était la partie postérieure du gyrus temporal supérieur gauche. Les paramètres utilisés lors de cette étude peuvent expliquer la différence des résultats obtenus avec notre étude. Margulis et al. ont présenté de longs extraits musicaux (12s) lors d'une écoute passive, ce qui laissait la liberté aux participants de porter leur attention sur un ou plusieurs aspects de chaque séquence présentée. Compte tenu de cette variabilité, il est probable que la seule interaction significative soit obtenue dans une région généralement impliquée dans le traitement de la musique (e.g., Ohnishi et al., 2001). De plus, cette étude n'a pas explicitement interrogé les aires de la voix. Une implication des aires de la voix dans le traitement des extraits musicaux aurait pu être diluée à cause des corrections pour les comparaisons multiples. Dans notre expérience, les participants devaient porter leur attention sur l'ensemble des caractéristiques acoustiques parce qu'ils devaient répondre seulement si le son suivant était exactement le même. Ce type de tâche a sans doute permis aux experts de traiter les stimuli de leur catégorie à un plus haut niveau d'abstraction, puisqu'ils étaient familiers avec ces stimuli.

Liebenthal et collaborateurs (2010) rapportent l'implication du STS central gauche dans la catégorisation de sons hautement familiers. À l'aide de l'ÉEG et de l'IRMf en simultané, ils ont mesuré la réponse de participants avant et après qu'ils aient appris à catégoriser 1) des sons non-phonémiques non-familiers ressemblant à de la parole et 2) des patrons phonémiques familiers. Ils ont remarqué une plus grande activation du STS postérieur gauche pour l'apprentissage des stimuli non-familiers, alors que les sons phonémiques familiers activaient la partie centrale du STS gauche, indépendamment de l'entraînement. L'onde P2 révélait le même patron : une plus forte amplitude après l'entraînement pour les sons non-familiers, alors qu'elle était invariable pour les sons phonémiques familiers. Les auteurs concluent que le facteur principal qui différencie les parties centrale et postérieure du STS gauche réside dans le degré de familiarité avec les catégories sonores. Cette familiarité correspondrait à un plus grand degré d'abstraction et une meilleure représentation à long-terme dans le STS central gauche pour les stimuli familiers, alors que le STS postérieur s'appuierait davantage sur des propriétés acoustiques. Cette organisation hiérarchique pour le traitement de stimuli familiers est cohérente avec ce qui a été observé pour le traitement de la voix. Plusieurs études ont démontré, à l'aide de l'IRMf, que la reconnaissance de la voix en tant que source sonore implique le STS, principalement du côté droit, à différents niveaux : la partie antérieure est influencée par des processus descendants tels que l'analyse détaillée de l'identité, alors que la partie postérieure reflète davantage la complexité spectro-temporale (Belin & Zatorre, 2003; Von

Kriegstein & Giraud, 2004; Warren, Scott, Price, & Griffiths, 2006). D'après ces données, les experts de notre étude auraient peut-être traité les sons de leur domaine d'expertise de la même façon que des extraits de voix humaines sont traités lorsque ces dernières doivent être comparées ou identifiés.

Le fait que l'interaction entre chaque groupe d'expertise et sa catégorie respective ait eu lieu dans l'hémisphère gauche peut trouver une explication dans la nature des stimuli utilisés. Peu de contrôles acoustiques ont été apportés aux échantillons sonores, mises à part la durée et la normalisation de l'amplitude moyenne. Deux raisons expliquent cette décision. D'une part, nous voulions avoir la plus grande variété de sons naturels possibles, afin de maximiser la familiarité avec les sons perçus dans chaque catégorie. D'autre part, l'interaction entre les deux catégories d'expertise permettait de minimiser les chances d'obtenir une région commune qui répondrait en fonction des caractéristiques acoustiques. Malgré cette absence de contrôles, la présence de transitions de fréquences très rapides, à la fois retrouvées pour les guitares et les chants d'oiseaux, peut contribuer à une préférence de l'hémisphère gauche pour ces stimuli. Dans une étude de morphométrie, les aires sensibles à la hauteur du gyrus de Heschl bilatéral ont été comparées chez des musiciens et des non-musiciens (Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005). Schneider et collaborateurs ont découvert que les personnes qui écoutaient plus souvent des instruments de musique percussifs ou de faible richesse spectrale (*fundamental pitch listeners)* présentaient un plus grand volume de matière grise dans l'hémisphère gauche. Cette donnée était valide autant pour les musiciens que les non-musiciens, démontrant une relation entre les pratiques d'écoute sur le cortex auditif sensible aux changements temporels rapides. Ceci est en accord avec la proposition qu'il y aurait une asymétrie hémisphérique pour le traitement de changements sonores rapides et la richesse spectrale pour le traitement des stimuli linguistiques ou non (Belin et al., 1998; Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002).

Le lobe temporal gauche aurait aussi une préférence pour les vocalisations animales. Des études ont montré une adaptation aux vocalisations animales dans le gyrus temporal supérieur gauche (Altmann, Doerhmann, & Kaiser, 2007; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008). De plus, une autre étude a obtenu une activation du STS central gauche aux vocalisations de chats, dont la coordonnée maximale est presqu'identique à celle obtenue dans notre étude (Fecteau, Armony, Joanette, & Belin, 2004). La plus forte activation chez les ornithologues peut être expliquée à la fois par une sensibilité aux changements rapides et par une représentation accrue de cette catégorie animale.

Les résultats obtenus dans le STS gauche démontrent un traitement plus poussé des stimuli présentés, suivant le principe d'organisation fonctionnelle du cortex auditif (Rauschecker, 1998), ainsi que le recours à la voie ventrale pour identifier les stimuli et leurs caractéristiques (Zatorre & Belin, 2000). Cette donnée n'est pourtant pas complète pour expliquer l'expertise auditive des experts recrutés, puisqu'une autre interaction entre chaque groupe d'expert et sa catégorie d'expertise a été retrouvée lorsque l'analyse n'était pas limitée aux aires sensibles à la voix.

L'évaluation de l'interaction entre le groupe d'expertise et sa catégorie d'expertise a révélé l'implication du gyrus cingulaire postérieur gauche (GCPg) dans le traitement des stimuli d'expertise. Une première explication de ce résultat tient de la nature de la tâche utilisée. Dans une tâche très similaire à la nôtre, Stevens (2004) a évalué les régions cérébrales impliquées dans la mémorisation à court-terme de voix humaines. Dans deux conditions d'une tâche *two-back*, les participants devaient reconnaître la voix présentée ou se souvenir du mot prononcé. Le résultat le plus constant entre les participants était une réponse plus forte du GCP médian pour la mémorisation du locuteur, comparativement à la mémorisation du mot. Ce résultat est appuyé par d'autres études démontrant une plus grande activation du CGP pour détecter une cible auditive, alors que les stimuli nouveaux ne l'activent pas (Stevens, Skudlarski, Gatenby, & Gore, 2000; Kiehl, Laurens, Duty, Forster, & Liddle, 2001). Un premier rôle du GCP impliquerait la mémorisation d'informations acoustiques pour effectuer une action dans l'immédiat. Ce lien entre le CGP et le cortex auditif est validé par les études de traçage rétrograde et antérograde effectuées chez le macaque, qui démontrent des connections réciproques entre la partie postérieure du gyrus temporal supérieur et le gyrus cingulaire postérieur (Yukie, 1995; Seltzer & Pandya, 2009). Ces données dénotent le rôle du GCP durant la tâche *one-back*, mais l'activation de celui-ci de manière préférentielle aux stimuli d'expertise réside dans le rôle de ce dernier pour la reconnaissance de stimuli familiers. Plusieurs études d'imagerie ont démontré une activation du GCP pour la reconnaissance de voix et de visages familiers (Shah et al., 2001), de même que pour la vue de noms de personnes familières (Maddock, Garrett, & Buonocore, 2001; Sugiura et al., 2008). Nous suggérons que le fait d'entendre un son d'une catégorie que l'on connaît très bien peut évoquer des souvenirs pour cette catégorie, comme c'est le cas pour la mémoire auto-biographique ou des connaissances personnelles.

En somme, les résultats obtenus en IRMf révèlent un rôle combiné des aires temporales et limbiques postérieures dans le traitement de stimuli auditifs d'expertise. Comme il a été observé, ce réseau est semblable à ce qui a été retrouvé pour les voix humaines ou d'autres catégories sonores, bien qu'il ne soit pas aussi étendu. Alors que l'implication du STS gauche est le résultat d'un traitement hiérarchique des stimuli et une abstraction de la catégorie d'appartenance, le gyrus cingulaire postérieur gauche est quant à lui responsable des représentations mnésiques et sémantiques liées aux sons perçus. Le développement des expertises que nous avons observées sont aussi dépendantes des caractéristiques de base des stimuli, considérant la préférence de l'hémisphère gauche pour ces stimuli.

Quelques limites de la troisième étude sont également à considérer. La première de ces limites est le faible nombre de participants recrutés. Compte tenu du faible nombre de participants dans les deux groupes, il est plus difficile de généraliser les résultats sur la population étudiée, puisqu'il est habituellement conseillé d'avoir entre 8 et 16 participants dans chaque groupe pour une analyse de groupe à effets aléatoires (*random effects analysis*; Friston, Holmes, & Worsley, 1999). De plus, il aurait été intéressant d'évaluer l'activité corticale dans d'autres régions, telles que les aires visuelles ou frontales. Ces régions ne constituaient pas l'intérêt principal de cette étude, parce qu'elles ne sont pas considérées comme préférentielles à la voix. Par contre, l'implication de régions frontales dans la reconnaissance de la voix a déjà été montrée (von Kriegstein, Kleinschmidt, Sterzer, & Giraud, 2005). Bien que non-spécifiques, certaines régions frontales auraient peut-être été utilisées chez les groupes d'experts. Finalement, l'ajout d'une condition auditive supplémentaire comprenant des sons de l'environnement non-vocaux aurait permis d'avoir un niveau de base commun aux trois catégories auditives, comme il est habituellement fait en vision (e.g., Gauthier et al., 2000). Qui plus est, la présence d'une quatrième catégorie sonore aurait pu servir de substitut à la session *localizer* (Friston, Rotshtein, Geng, Sterzer, & Henson, 2006). L'utilisation d'une session *localizer* fonctionnelle pour déterminer les régions préférentielles à la voix est une pratique courante qui procure différents avantages. Habituellement de courte durée, elle maximise les régions qui répondent à la voix pour chaque individu, et ce indépendamment des autres tâches effectuées. Elle limite aussi la quantité de voxels inclus dans les comparaisons, ce qui augmente la puissance statistique. Pourtant, certains de ces avantages peuvent aussi être des inconvénients. Par exemple, étant donné que le *localizer* que nous avons utilisé et les sessions suivantes ne sont pas de la même nature (écoute passive vs tâche attentionnelle; longueur des événements), il est difficile d'indiquer avec certitude le rôle des neurones qui ont répondu à la fois à la voix lors d'une tâche passive et à l'interaction obtenue pendant une tâche active (Friston et al., 2006). Il est à souligner que l'élimination d'une session *localizer* aurait nécessité des analyses indépendantes pour les deux groupes d'experts, afin de conserver la qualité orthogonale des contrastes. À titre d'exemple, l'analyse effectuée chez les ornithologues serait un contraste de conjonction telle que « voix vs environnement ∩ oiseaux vs environnement ». Les questions reflétées par de tels contrastes sont évidemment légèrement différentes de ce qui a été fait dans la troisième étude, mais elles nous semblent tout aussi pertinentes et surtout complémentaires. Plutôt que de démontrer les régions sensibles à deux catégories d'expertise, ce type d'analyse aurait affiché la région sensible à la fois aux voix et à une autre catégorie d'expertise, lors d'une même tâche.

Afin de vérifier les critiques faites par Friston et collaborateurs (2006), Berman et collaborateurs (2010) ont vérifié si la distribution neuronale de la réponse spécifique aux visages était affectée par le type de tâche. Bien que la distribution n'était pas affectée par l'utilisation d'une tâche passive ou de type *one-back*, l'utilisation d'une tâche attentionnelle semblait être plus fiable dans les résultats. Il y avait aussi une influence notable de la catégorie qui sert de contraste (visages vs maisons ou visages vs visages embrouillés). Alors que les visages contrastés avec des visages embrouillés produisaient de plus grandes activations, ils produisaient aussi plus de variance. Ces conclusions sont applicables au gyrus fusiforme uniquement, donc il est difficile de se prononcer sur ce qui adviendrait des aires de la voix pour de telles comparaisons.

La voix est-elle spéciale?

Il nous semblerait inopportun et irrespectueux envers le lecteur de ne pas tenter de répondre à cette question initiale, qui se voulait être la fondation même de cette thèse. À la lumière des données qui ont été présentées et discutées, nous croyons que cette question est encore trop jeune pour être parfaitement résolue. Bien que les musiciens professionnels, les ornithologues amateurs ou les luthiers soient exposés à leur catégorie d'expertise depuis une ou plusieurs dizaines d'années, il existera toujours un biais inévitable envers la voix humaine, puisqu'elle est présente depuis la naissance et demeure le véhicule du langage.

De par sa nature, la voix comporte des aspects uniques, peu importe la comparaison qu'on puisse en faire. Cet argument est d'ailleurs soulevé par Gauthier et Bukach (2007) en ce qui concerne la comparaison faite entre les visages et d'autres types de stimuli visuels. Par ailleurs, certaines leçons peuvent être tirées du débat de la spécificité des visages dans le domaine visuel. Une de celles-ci est que lorsqu'un débat devient polarisé, les auteurs sont portés à classifier les données ou les articles selon un mode binaire. En contrepartie, un avantage certain de la présence d'un débat actif est que les données négatives ont un poids tout aussi important que les données positives, ce qui offre la possibilité de planifier et d'explorer des hypothèses de recherche nouvelles. Considérant ces points, nous suggérons d'éliminer la question « La voix est-elle spéciale? » et de la remplacer par « De quelle façon la voix est-elle spéciale? ». Poser cette nouvelle question consisterait à se demander quels aspects de la voix humaine sont uniques à celle-ci, en tenant compte des nombreuses similarités qu'elle peut afficher avec d'autres stimuli.

Si on se fie à l'évolution des débats de la spécificité du cerveau pour le traitement des visages ou du langage, il y aura nécessairement émergence de données contradictoires dans le domaine de la voix. Comme il est suggéré par Zatorre et Gandour (2008) en ce qui a trait au débat de la spécificité du traitement pour le langage et la musique, il est primordial de ne pas diviser, mais plutôt d'unir et de mieux comprendre les données contradictoires.

Un exemple d'intégration se retrouve dans la méta-analyse de Hein et Knight (2008). Ils ont relevé les résultats de plus de 100 études dans le but de caractériser la distribution des régions recrutées par différentes tâches le long du STS. En effet, le STS est impliqué dans une multitude de fonctions dont la théorie de l'esprit, l'intégration audiovisuelle, le traitement du mouvement, des visages, de la parole et de la voix. Grâce à une analyse de regroupement (*cluster analysis*), les auteurs ont noté que des fonctions cognitives différentes recrutent des régions très similaires du STS. Il est possible que cette situation s'applique également au domaine de l'audition : la participation d'une même région au traitement de différentes catégories auditives ne signifierait pas nécessairement l'implication des mêmes fonctions. Les auteurs suggèrent que la clé pour comprendre le fonctionnement du STS réside dans l'exploration des réseaux de coactivation plutôt que dans l'unique observation de régions circonscrites. Ainsi, ils encouragent une plus grande utilisation de la connectivité fonctionnelle afin de mieux comprendre les fonctions associées au STS. En audition, l'utilisation plus courante de cette pratique rendrait mieux compte de la complexité de l'expertise auditive et du traitement de la voix. En effet, c'est à l'aide de techniques novatrices telles que la connectivité fonctionnelle que l'on pourra comprendre les réelles différences entre ces deux domaines.

Conclusion

Le but de la présente thèse était d'explorer l'expertise auditive à l'aide de différentes méthodes, afin de la comparer à celle pour la voix, qui est considérée comme universelle chez les humains. Les données obtenues, de même que la discussion qui a suivi, ont montré que certaines régions corticales et processus sont communs au traitement de la voix et des stimuli auditifs d'expertise. Ces aspects communs sont explicables par une organisation à la fois fonctionnelle et économique du cerveau, qui semblerait traiter des stimuli auditifs de manière spécifique seulement lorsque c'est nécessaire. Dans les études à venir, il sera important de continuer à caractériser les différences et les ressemblances entre la voix et d'autres catégories auditives, non pas pour isoler ces catégories entre elles, mais plutôt pour en arriver à une compréhension commune et intégrée du cerveau auditif.

Bibliographie

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences, 8*, 457–464.

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 364*, 285-299.

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex, 17*, 2601–8.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., et al. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience, 10*, 536-540.

Belin, P., Bédard, C., & Fecteau, S. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences, 8*, 129-135.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P. & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-12.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research. Cognitive Brain Research, 13*, 17-26.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *NeuroReport, 14*, 2105–2109.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience, 25*, 399-410.

Berman, M. G., Park, J., Gonzalez, R., Polk, T. A., Gehrke, A., Knaffla, S., et al. (2010) Evaluating functional localizers : The face of the FFA. *NeuroImage, 50*, 56-71.

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., et al. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*, 127.

Coez, A., Zilbovicius, M., Ferrary, E., Bouccara, D., Mosnier, I., Ambert-Dahan, E., Bizaguet, E., Syrota, A., Samson, Y., & Sterkers, O. (2008). Cochlear Implant Benefits in Deafness Rehabilitation : PET Study of Temporal Voice Activations. *The Journal of Nuclear Medicine, 49*, 60-67.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*, 1174–1176.

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia, 46*, 2776-2786.

Eadie, T. L., Van Boven, L., Stubbs, K., & Giannini, E. (2010). The Effect of Musical Background on Judgments of Dysphonia. *Journal of Voice, 24*, 93-101.

Fifer, W. P., & Moon, C. M. (1994). The role of mother's voice in the organization of brain function in the newborn. *Acta Paediatrica Supplement, 397*, 86-93.

Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *NeuroImage, 10*, 1-5.

Friston, K. J., Rotshtein, P., Geng, J. J., Sterzer, P., & Henson, R. N. (2005). A critique of functional localisers. *NeuroImage, 30*, 1077-1087.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face perception. *Nature Neuroscience, 3*, 191-197.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*, 322-330.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.

Hämäläinen, M. S., Hari, R., Ilmoniemi, R. J., Knuutila, J., Lounasamaa, O. V., 1993. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics, 65*, 413–497.

Hein, G., & Knight, R. T. (2008). Superior Temporal Sulcus – It's my Area : or is it? *Journal of Cognitive Neuroscience, 20*, 2125-2136.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area : a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*, 4302-4311.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759-763.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area : a cortical region specialized for the perception of faces. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 361*, 2109-28.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: an event related fMRI study, *Psychophysiology, 38*, 133–142.

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*, 125-143.

Kisilevski, B. S., Hains, S. M. J., Lee, K., Xie, X., Huang, H., Ye, H. H., Zhang, Y., & Wang, Z. (2003). Effects of experience on fetal voice recognition. *Psychological Science, 14*, 220-224.

Lee, C. Y., & Hung, Y. H. (2008). Identification of Mandarin tones by English-speaking musicians and non-musicians. *The Journal of the Acoustical Society of America, 124*, 3235-3248.

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport, 12*, 2653-2657.

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40*, 291-305.

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,* Apr. 9. (Epub ahead of print).

Maddock, R. J., Garrett, A. S., & Buonocore, M. H. (2001). Remembering familiar people : the posterior cingulate cortex and autobiographical memory retrieval. *Neuroscience, 104*, 667-676.

Margulis, E. H., Mlsna, L. M., Uppunda, A. K., Parrish, T. B., & Wong, P. C. M. (2009). Selective Neurophysiologic Responses to Music in Instrumentalists with Different Listening Biographies. *Human Brain Mapping, 30*, 267-275.

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

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*, 15894-15898.

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

Nahum, M., Nelken, I., & Ahissar, M. (2010). Stimulus uncertainty and perceptual learning : Similar principles govern auditory and visual learning.

Nikjeh, D. A., Lister, J. J., & Frisch, S. A. (2009). Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech : influence of music training. *Ear and hearing, 30*, 432-446.

Ohnishi, T., Matsuda, H., Asada, T, Aruga, M., Hirakata, M., Nishikawa, M., & al. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex,* 11, 754–760.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., & Ross, B. (2001). Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport, 12*, 169–174.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology: Human Perception & Performance, 20*, 976-986.

Rauschecker, L. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8*, 516-521.

Seltzer, B., & Pandya, D. N. (2009). Posterior cingulate and retrosplenial cortex connections of the caudal superior temporal region in the rhesus monkey. *Experimental Brain Research, 195*, 325-334.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005). Structural, Functional, and Perceptual Differences in Heschl's Gyrus and Musical Instrument Preference. *Annals of the New York Academy of Sciences, 1060*, 387-394.

Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., Markowitsch, H. J., et al. (2001). The neural correlates of person familiarity. A functional magnetic resonance imaging study with clinical implications. *Brain, 124*, 804-815.

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*, 5545–5552.

Shahin, A. J., Roberts, L. E., Pantev, C., Aziz, M., & Picton, T. W. (2007). Enhanced anterior-temporal processing for complex tones in musicians. *Clinical Neurophysiology, 118*, 209-220.

Sheehan, K. A., McArthur, G. M., Bishop, D. V. M., 2005. Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Cognitive Brain Research, 25*, 547–553.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and non-musicians. *Journal of the Acoustical Society of America, 76*, 1690-1695.

Stevens, A. A., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging, 18*, 495– 502.

Stevens, A. A. (2004). Dissociating the cortical basis of memory for voices, words and tones. *Cognitive Brain Research, 18*, 162–171.

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency – effect of training on subcortical processing of vocal expressions of emotion. *The European Journal of Neuroscience, 29*, 661-668.

Suga, N., Gao, E., Zhang, Y., Ma, X., & Olsen, F. J. (2000). The corticofugal system for hearing : Recent progress. *Proceedings of the National Academy of Sciences, 97*, 11807-11814.

Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y., Maeda, Y., Matsue, Y., et al. (2008). Anatomical segregation of representations of personally familiar and famous people in the temporal and parietal cortices. *Journal of Cognitive Neuroscience, 21*, 1855-1868.

von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage, 22,* 948–955.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *NeuroImage, 31*, 1389-1397.

Wong, P. C., 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.

Yukie, M. (1995). Neural connections of auditory association cortex with the posterior cingulate cortex in the monkey. *Neuroscience Research, 22*, 179-187.

Zatorre, R. J., & Belin, P. (2000). 'What', 'where' and 'how' in auditory cortex. *Nature Neuroscience, 3*, 965-966.

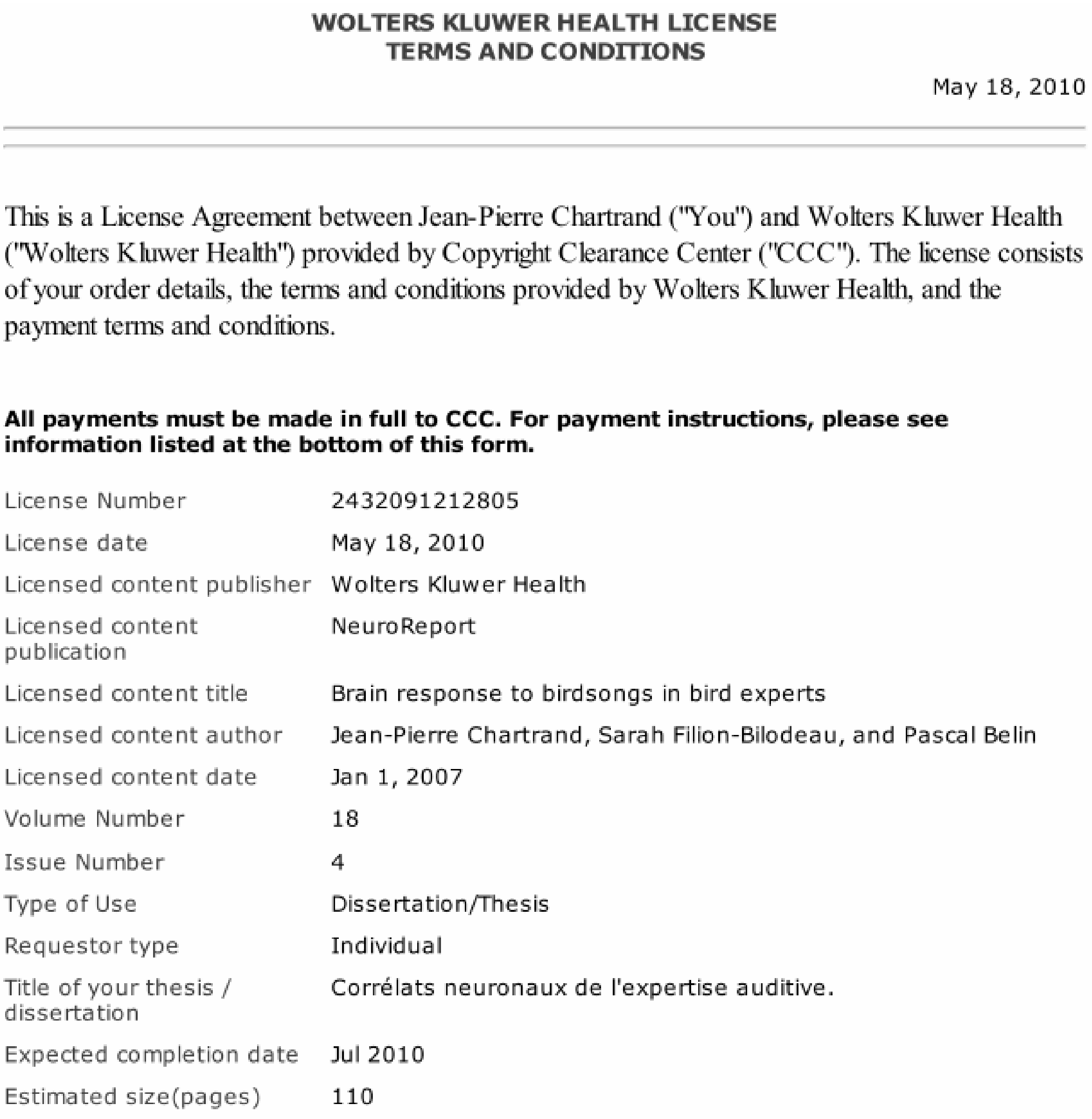
Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946-953.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex : music and speech. *Trends in Cognitive Sciences, 6*, 37-46.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch : moving beyond the dichotomies. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363*, 1087-1104.

Annexes

Accord des éditeurs pour l’inclusion de l’article « Brain response to birdsongs in bird experts » dans la thèse.



WOLTERS KLUWER HEALTH LICENSE
TERMS AND CONDITIONS

May 18, 2010

This is a License Agreement between Jean-Pierre Chartrand ("You") and Wolters Kluwer Health ("Wolters Kluwer Health") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Wolters Kluwer Health, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

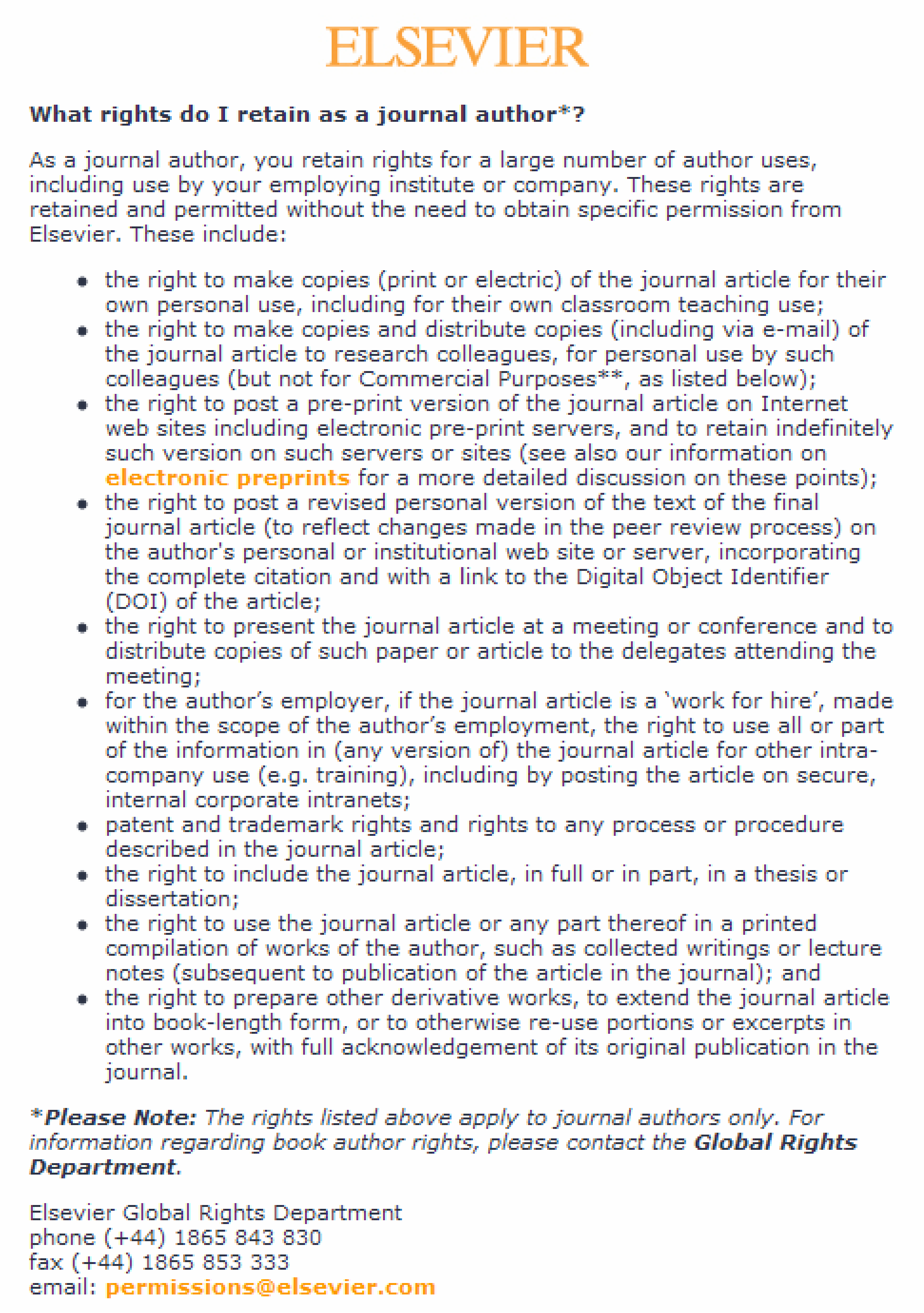
License Number	2432091212805
License date	May 18, 2010
Licensed content publisher	Wolters Kluwer Health
Licensed content publication	NeuroReport
Licensed content title	Brain response to birdsongs in bird experts
Licensed content author	Jean-Pierre Chartrand, Sarah Filion-Bilodeau, and Pascal Belin
Licensed content date	Jan 1, 2007
Volume Number	18
Issue Number	4
Type of Use	Dissertation/Thesis
Requestor type	Individual
Title of your thesis / dissertation	Corrélats neuronaux de l'expertise auditive.
Expected completion date	Jul 2010
Estimated size(pages)	110

Terms and Conditions

1. A credit line will be prominently placed and include: for books - the author(s), title of book, editor, copyright holder, year of publication; For journals - the author(s), title of article, title of journal, volume number, issue number and inclusive pages.
2. The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, or authors of the material, or to Wolters Kluwer/Lippincott, Williams & Wilkins.
3. Permission is granted for one time use only as specified in your correspondence. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. Once term has expired, permission to renew must be made in writing.
4. Permission granted is non-exclusive, and is valid throughout the world in the English language and the languages specified in your original request.
5. Wolters Kluwer Health/ Lippincott, Williams & Wilkins, cannot supply the requestor with the original artwork or a "clean copy."
6. The requestor agrees to secure written permission from the author (for book material only).
7. Permission is valid if the borrowed material is original to a LWW imprint (Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Igaku-Shoin, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language).
8. If you opt not to use the material requested above, please notify Rightslink within 90 days of the original invoice date.
9. Other Terms and Conditions:

v1.0

Permission des éditeurs pour l'inclusion des articles « Auditory recognition expertise and domain specificity » et « Superior voice timbre processing in musicians » dans la thèse.



ELSEVIER

What rights do I retain as a journal author*?

As a journal author, you retain rights for a large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier. These include:

- the right to make copies (print or electric) of the journal article for their own personal use, including for their own classroom teaching use;
- the right to make copies and distribute copies (including via e-mail) of the journal article to research colleagues, for personal use by such colleagues (but not for Commercial Purposes**, as listed below);
- the right to post a pre-print version of the journal article on Internet web sites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites (see also our information on **electronic preprints** for a more detailed discussion on these points);
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on the author's personal or institutional web site or server, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article;
- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for the author's employer, if the journal article is a 'work for hire', made within the scope of the author's employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g. training), including by posting the article on secure, internal corporate intranets;
- patent and trademark rights and rights to any process or procedure described in the journal article;
- the right to include the journal article, in full or in part, in a thesis or dissertation;
- the right to use the journal article or any part thereof in a printed compilation of works of the author, such as collected writings or lecture notes (subsequent to publication of the article in the journal); and
- the right to prepare other derivative works, to extend the journal article into book-length form, or to otherwise re-use portions or excerpts in other works, with full acknowledgement of its original publication in the journal.

****Please Note:*** *The rights listed above apply to journal authors only. For information regarding book author rights, please contact the* ***Global Rights Department****.*

Elsevier Global Rights Department
phone (+44) 1865 843 830
fax (+44) 1865 853 333
email: **permissions@elsevier.com**