

Université de Montréal

**Perception des vocalisations: études comportementales et  
d'imagerie par résonance magnétique fonctionnelle  
événementielle**

par

Shirley Fecteau

Faculté de médecine

Thèse présentée à la Faculté des études supérieures  
en vue de l'obtention du grade de Ph.D.  
en Sciences biomédicales  
option neuropsychologie

Juin 2004



© Shirley Fecteau, 2004

W

4

U58

2004

v. 112

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Cette thèse intitulée :

Perception des vocalisations: études comportementales et d'imagerie par résonance  
magnétique fonctionnelle événementielle

présentée par :  
Shirley Fecteau

a été évaluée par un jury composé des personnes suivantes :

Jacques Montplaisir, président-rapporteur  
Yves Joanette, directeur de recherche  
Pascal Belin, co-directeur  
Mario Beauregard, membre du jury  
Patrik Vuilleumier, examinateur externe  
Roger Godbout, représentant du doyen de la FES



## Résumé

Peu d'études en neurosciences ont exploré la voix humaine. Ce travail s'intéresse à la perception des vocalisations (p. ex., rire, toux) chez l'humain. Les études constituant le corps de cette thèse visent à examiner les performances comportementales et les substrats neuronaux du traitement perceptuel de vocalisations. Le premier chapitre présente les principales études abordant l'émotion vocale et la problématique du travail de recherche. La recherche effectuée est ensuite rapportée sous forme de chapitres distincts. Enfin, les résultats des expérimentations font l'objet d'une discussion générale.

Le premier article explore si le jugement des dimensions émotionnelles de valence, d'intensité et d'authenticité de vocalisations est modulé selon l'âge et le sexe des participants. Les résultats montrent que le jugement de valence est constant entre les participants et l'âge est un facteur significatif dans le jugement des trois dimensions, alors que le sexe ne l'est pas.

Le second article examine l'effet d'amorçage comportemental de vocalisations émotionnelles et non-émotionnelles chez un groupe d'hommes. Les résultats indiquent un effet significatif d'amorçage pour les vocalisations, lequel est modulé selon le sexe du locuteur, mais pas selon la charge émotionnelle.

Le troisième article étudie la spécificité des substrats neuronaux sous-jacents à la voix humaine. Les résultats montrent une réponse sélective des sillons temporaux supérieurs pour les vocalisations humaines, comparativement à des stimuli auditifs nonhumains, suggérant une spécificité de la réponse neuronale pour la voix humaine.

Le quatrième article explore les structures neuronales associées à la perception d'émotion exprimée via des vocalisations. Des réponses des régions temporales et frontales pour les vocalisations émotionnelles, comparativement aux non-émotionnelles sont

observées. Aussi, des activités cérébrales sont associées à des catégories spécifiques de vocalisations. Par exemple, une augmentation des réponses de l'amygdale est associée aux rires et aux vocalisations sexuelles et celles de l'hypothalamus aux pleurs.

Enfin, le dernier chapitre constitue la discussion générale de la thèse. L'ensemble des résultats est résumé et les modèles de la perception de la voix de Belin, Fecteau et al. (2004) et du traitement des émotions d'Adolphs (2002a,b) sont présentés. Une discussion sur l'évolution du langage conclut la thèse.

**Mots-clés :** voix, émotion, perception, spécificité de l'espèce, amorçage, imagerie par résonance magnétique fonctionnelle événementielle

## Abstract

Only a few studies in neuroscience have explored the processing of human voice. The goal of this thesis is to study the perception of nonlinguistic vocalizations in humans (e.g., laughs, coughs). The papers that compose this thesis explore behavioral performances and neural basis of perception of vocalizations. The first chapter presents the most relevant studies on perception of emotions expressed vocally and the rationale of this thesis. The following chapters present experiments conducted to explore the processing of voice perception. Finally, results from these experiments are discussed in a broader perspective.

The first paper explores whether judgment of the emotional dimensions of valence, intensity, and authenticity of vocalizations are modulated by age and sex of participants. Results reveal that judgment of valence was consistent across participants and that age significantly influenced judgment of these emotional dimensions, whereas sex did not.

The second paper examines the behavioral priming effect of emotional and non-emotional vocalizations in healthy male participants. A priming effect for the vocalizations was observed, which was influenced by the speaker's sex, but not by the emotional load.

The third paper studies the specificity of the neural substrates underlying perception of human voice. Stronger responses in the superior temporal sulci for the human vocalizations, relative to the nonhuman auditory stimuli were found, suggesting a species-specificity of the neural basis for human voice.

The fourth paper explores the neural structures of perception of emotional vocalizations. Stronger activation for the emotional vocalizations, relative to the neutral ones, were observed in the temporal and frontal regions. Also, specific emotional vocal categories elicited stronger responses in cerebral regions: laughs and sexual vocalizations induced greater activity in the amygdala, and cries in the hypothalamus.

Finally, the last chapter is the general discussion of the thesis, which summarizes the main results reported in the experiments. Theoretical models of voice perception proposed by Belin, Fecteau et al. (2004) and of emotional perception proposed by Adolphs (2002a,b) are presented. A brief discussion on the evolution of language concludes the thesis.

**Keywords** : voice, emotion, perception, species-specificity, priming, event-related functional magnetic resonance imaging

## Table des matières

Identification du jury .....	ii
Résumé en français .....	iii
Résumé en anglais .....	v
Table des matières .....	vii
Liste des tableaux .....	ix
Liste des figures .....	x
Liste des abréviations .....	xi
Dédicace .....	xii
Remerciements .....	xiii
 INTRODUCTION GÉNÉRALE .....	 1
 ARTICLE 1:	
Jugement de dimensions émotionnelles des vocalisations nonlinguistiques: différences selon l'âge .....	40
Résumé .....	43
Introduction .....	44
Méthode .....	45
Résultats .....	47
Discussion .....	48
Conclusion .....	54
Remerciements .....	55
Légende des figures.....	56
Références .....	61
 ARTICLE 2:	
Amorçage comportemental dans la perception des vocalisations émotionnelles: l'influence du sexe du locuteur .....	89
Résumé .....	92
Introduction .....	93
Méthode .....	93
Résultats .....	94
Discussion .....	95
Remerciements .....	97

Légende des figures .....	98
Références .....	100

## ARTICLE 3:

Spécificité des réponses cérébrales à la voix humaine? Une étude d'IMRf événementielle .....	101
Résumé .....	104
Introduction .....	105
Méthode .....	106
Résultats .....	109
Discussion .....	110
Remerciements .....	116
Légende des figures .....	117
Légende du tableau .....	119
Références .....	127

## ARTICLE 4:

Réponses neuronales associées à la perception de vocalisations émotionnelles: une étude d'IRMf événementielle .....	132
Résumé .....	135
Introduction .....	136
Méthode .....	137
Résultats .....	140
Discussion .....	141
Remerciements .....	147
Légende des figures .....	148
Légende du tableau .....	149
Références .....	154

DISCUSSION GÉNÉRALE .....	163
---------------------------	-----

Références générales .....	196
----------------------------	-----

## Liste des tableaux

### INTRODUCTION

Tableau 1.	Études rapportant des caractéristiques du locuteur qui sont marquées par les vocalisations chez les animaux nonhumains .....	2
Tableau 2.	Principales études de neuroimagerie fonctionnelle explorant le traitement de la prosodie émotionnelle chez l'individu sain .....	18
Tableau 3.	Principales études de neuroimagerie fonctionnelle explorant le traitement des vocalisations nonlinguistiques émotionnelles chez l'individu sain .....	22
Tableau 4.	Exemples de paramètres acoustiques des patrons prosodiques associés à l'expression d'émotions (Scherer et Oshinsky, 1977; Scherer, 1986) .....	31

### ARTICLE 3

Table 1.	Coordonnées des réponses significatives des STS pour les différents contrastes d'intérêt .....	125
----------	------------------------------------------------------------------------------------------------	-----

### ARTICLE 4

Table 1.	Coordonnées des contrastes d'intérêt .....	152
----------	--------------------------------------------	-----

## Liste des figures

### INTRODUCTION

Figure 1.	Régions activées: stimuli vocaux émotionnels versus non-émotionnels .....	24
-----------	------------------------------------------------------------------------------	----

### ARTICLE 1

Figure 1.	Scores du jugement de valence .....	57
Figure 2.	Scores du jugement d'intensité .....	58
Figure 3.	Scores du jugement d'authenticité .....	59
Figure 4.	Interaction du âge x sexe du jugement d'authenticité .....	60

### ARTICLE 2

Figure 1.	Effet d'amorçage, scores de valence, d'intensité et d'authenticité .....	99
-----------	--------------------------------------------------------------------------	----

### ARTICLE 3

Figure 1.	Exemples de stimuli .....	120
Figure 2.	Contraste stimuli auditif > silence .....	121
Figure 3.	Contrastes vocalisations humaines > sons nonhumains .....	122
Figure 4.	Contrastes a) vocalisations animales > sons nonvocaux, b) vocalisations humaines > sons nonvocaux .....	123
Figure 5.	Superposition des trois contrastes d'intérêt .....	124

### ARTICLE 4

Figure 1.	Contraste vocalisations émotionnelles > neutres .....	150
Figure 2.	Contrastes a) rires > vocalisations neutres, b) vocalisations sexuelles > vocalisations neutres, c) pleurs > vocalisations neutres .....	151

### DISCUSSION GÉNÉRALE

Figure 1.	Échelles utilisées dans cette étude pour juger les vocalisations humaines sur les dimensions de valence, d'intensité et d'authenticité .....	167
Figure 2.	Modèle du noyau affectif proposé par Russell (2003) .....	168
Figure 3.	Modèle de perception de la voix proposé par Belin, Fecteau et al. (2004) .....	183
Figure 4.	Modèle de perception d'un stimulus émotionnel proposé par Adolphs (2002a,b) .....	189



## Liste des abréviations

D	droit
G	gauche
FFA	«fusiform face area»
IRMf / fMRI	imagerie par résonance magnétique fonctionnelle / «functional magnetic resonance imaging»
GTM / MTG	gyrus temporal médian / «middle temporal gyrus»
GTS / STG	gyrus temporal supérieur / «superior temporal gyrus»
STS	sillon temporal supérieur / «superior temporal sulcus»
TEP / PET	tomographie par émission de positrons / «positron emission tomography»

*À Christian*

## Remerciements

Je tiens tout d'abord à remercier du fond du coeur Yves Joannette et Pascal Belin. Je suis débarquée dans vos bureaux complètement essoufflée et perdue, et vous m'avez fait confiance sans poser de question sur mon parcours précédent. Sans votre soutien, je n'aurais pas poursuivi mes études doctorales. Merci d'avoir cru en moi tout au long de ces deux dernières années. Merci.

Plus particulièrement, je remercie Yves parce ce j'ai toujours sentie que vous étiez là pour moi, que je pouvais me réfugier sous votre aile si besoin. Merci aussi pour votre calme et votre écoute. Dès notre première rencontre, vous m'aviez demandé ce que je voulais faire lorsque je serai grande et vous avez toujours respecté mes choix.

Je remercie Pascal pour sa générosité, tu as toujours été extrêmement accueillant et chaleureux. Tu étais à tout moment très ouvert et disponible pour discuter et répondre à mes questions. Enfin, un grand merci pour nos discussions enrichissantes sur la voix et surtout pour ta passion pour la recherche que j'avais perdue mais que tu m'as fait retrouver. Merci de m'avoir donné la chance de poursuivre mon doctorat.

I also want to thank Jorge Armony for his availability, his calm, his patience, and for having believed in me. Not only are you a great collaborator, but also a friend. I really enjoyed working with you. You taught me a lot about imaging, but more importantly about science. You helped me believe in myself again.

Je tiens absolument à remercier les gens qui m'ont permis de faire un changement de trajet de vie sans trop faire de vague. Plus particulièrement, je veux remercier Marie-José Rivest qui est venue me chercher. Merci pour votre écoute, votre disponibilité et surtout parce que vous m'avez fait souffler, vous avez momentanément pris en charge ma vie pour ensuite me repositionner. Je tiens également à remercier Dr Fernand Roberge, Dr Patrik

Vinay, Dr Gilles Bleau et Solange Cantin qui m'ont permis de me repositionner tout en douceur.

C'est aussi très important pour moi de remercier François Bherer, Patricia Jelenic et Frédéric Gougoux qui m'ont grandement soutenue. Merci pour votre écoute, votre patience et surtout pour votre amitié. Merci aussi à Dr Marie Labelle pour ses encouragements et qui m'a aussi permis de souffler durant un moment.

Je tiens aussi à dire merci aux gens de l'Hôpital Notre-Dame: Jean-Maxime Leroux, Boualem Mensour, Gilles Beaudoin ainsi qu'aux technologues de la résonance magnétique qui ont été patientes et souriantes malgré les nombreux problèmes techniques que nous avons eus. Un gros merci à Pierre Ahad pour ses judicieux commentaires et qui m'a fait bénéficier de son expertise en acoustique, merci aussi pour ta disponibilité.

Je tiens aussi à remercier les gens du BIG (*Brain Imaging Group*) qui m'ont accueillie dans leur laboratoire comme étant une des leurs. Merci à François, j'ai eu un réel plaisir à te côtoyer.

Merci aussi à Catherine et Mathieu qui ont expérimenté pour moi, je vous remercie surtout pour votre amitié et les fou rires qu'on a eus. Catherine, tu as été une colocataire de bureau tout simplement géniale. Merci aussi à Karyne pour ton amitié. Merci aussi aux filles du labo de Yves, plus particulièrement Nathalie, Tania et Evelyne, avec qui j'ai toujours eu du plaisir à discuter.

I want to thank Frederick who always believed in me; you gave me wings to change my trajectory. A big thanks also to Roger for his generosity and his humor. You both worked very hard to find the vocal stimuli I needed. Without you, I couldn't have finished

that quickly. Merci également à Jean-Sébastien pour le temps qu'il m'a donné pour ma collecte de stimuli et surtout pour son amitié.

Je n'aurais pu terminer ma thèse sans mes bons vieux amis de Québec: Annie, François, Isabelle, Robert, Brigitte, Éliane et Yvan. Vraiment, je vous remercie d'avoir été toujours là pour moi, même si par moments je disparaissais derrière mon écran pendant de longues périodes. Un merci spécial à Annie et Steve qui m'ont soutenue de façon plutôt intensive (!) pendant mon changement de trajectoire. Enfin, un gros merci à Hugo pour son soutien, son humour et surtout pour sa patience lors du dernier droit de mes études.

Enfin, je termine mes remerciements par ceux qui sont mes plus chers amis, mes parents. Vous avez toujours cru en moi, vous m'avez toujours encouragée et vous m'avez montré à croire en qui je suis, à m'écouter. Il est impossible de décrire à quel point vous êtes importants pour moi. Tout simplement merci d'être là.

Je remercie finalement les Instituts de Recherche en Santé du Canada et la Faculté de médecine pour leur soutien financier.

## Introduction

*Il y a autant de mouvements dans la voix qu'il y a de mouvements dans l'esprit, et l'esprit est profondément affecté par la voix.*

Cicérone, De Oratore

De même que les extraits langagiers, les vocalisations nonlinguistiques, tels le rire, le soupir et le raclement de la gorge représentent une source importante de renseignements sur le locuteur (voir revue de Kreiman, 1997). Ces indices marquent l'identité individuelle (Tosi, 1979; Imaizumi, Mori et al, 1997), l'âge (Hartman et Danahuer, 1976; Helfrich, 1979; Linville, 1996), le sexe (Lass, Hughes et al., 1976; Smith, 1979; Mullenix, Johnson et al., 1995; Whiteside, 1998; Bachorowski et Owren, 1999), la personnalité (Scherer, 1972; Siegman, 1978; Scherer, 1979), la classe sociale (Robinson, 1979), l'appartenance à une communauté raciale et ethnique (Giles, 1979; Walton et Orlikoff, 1994) et des indices sur la situation sociale dans laquelle l'individu se trouve (Brown et Fraser, 1979; Brown et Levinson, 1979; Giles, Scherer et al., 1979). Une autre information fondamentale contenue dans les vocalisations nonlinguistiques est l'état émotionnel du locuteur. L'expression et la perception des émotions exprimées par ces vocalisations constituent un moyen de communication essentiel pour l'adéquation des interactions sociales. Une utilisation et/ou une perception inappropriée des vocalisations est rapidement attribuée à un diagnostic clinique (Scherer et Zei, 1988) et ce, même chez le nouveau-né (Corwin et Golub, 1996).

Le traitement des vocalisations nonlinguistiques constitue un des premiers processus socio-communicatifs et émotionnels dans l'ontogenèse (Barr, Hopkins et al., 2000). Très tôt, l'humain s'exprime par des vocalisations nonlinguistiques, par exemple, le rire apparaît déjà vers l'âge de 4 mois (Rush et Ekman, 2001). Ces vocalisations constituent un des outils les plus efficaces de communication entre le nouveau-né et ses parents, lesquelles



sont inhérentes (Murray, 1979; Cosmides, 1983) et ont même parfois une valeur adaptative et de survie pour le nouveau-né (Zeskind et Marshall, 1988). Le traitement des vocalisations n'est pas uniquement une composante essentielle dans l'ontogenèse chez l'humain, mais également chez la plupart des espèces animales nonhumaines.

Dans la phylogenèse, les vocalisations nonlinguistiques constituent un élément important. En effet, les vocalisations que nous utilisons actuellement semblent être essentiellement issues d'une continuité phylogénique (Robinson, 1976; Lamendella, 1977; Scherer, 1985). Par exemple, on peut attribuer un bénéfice adaptatif aux cris de peur, lesquels représentent un signal exprimant clairement l'émotion et dont son identification est relativement universelle: il signale un danger au reste de la cohorte, même à de longues distances. Les vocalisations humaines et animales nonhumaines partagent plusieurs attributs fondamentaux tant sur le plan fonctionnel qu'acoustique (p. ex., van Lawick-Goodall, 1968 et voir revue de Ghazanfar et Hauser, 2001). Similaires à l'humain, les vocalisations des animaux marquent plusieurs caractéristiques du locuteur (voir tableau 1).

Tableau 1. Études rapportant des caractéristiques du locuteur qui sont marquées par les vocalisations chez les animaux nonhumains.

identité	Rendall, Rodman et al. (1996); Rendall, Owren et al. (1998); Miller, Miller et al. (2001); Weiss, Garibaldi et al. (2001); Rendall (2003)
âge	Smith (1977)
sexe	Rendall, Rodman et al. (1996); Rendall, Owren et al. (1998); Miller, Miller et al. (2001); Weiss, Garibaldi et al. (2001); Rendall (2003)
état de reproduction	Semple et McComb (2000)
gabarit	Fitch (1997)
appartenance à un groupe	Rendall, Rodman et al. (1996); Rendall, Owren et al. (1998); Miller, Miller et al. (2001); Weiss, Garibaldi et al. (2001); Rendall (2003)
statut	Sebeok (1968, 1977); Smith (1977)
état motivationnel	Goodall (1986); Masataka et Khoda (1988); Hauser (1991); Gouzoules et Gouzoules (2000)

Aussi, la plupart des animaux communiquent leurs états affectifs par la voix (Braun et Poeggel, 2001), plus particulièrement les primates (van Hoof, 1972; Green et Cliff, 1975; Seyfarth, Cheney et al., 1980; Gouzoules, Gouzoules, et al., 1984; Goodall, 1986; Jürgens, 1986,1998; Masataka et Khoda, 1988; Fichtel et Hammerschmidt, 2003; Rendall, 2003). Enfin, les mécanismes physiologiques qui sous-tendent les vocalisations nonlinguistiques humaines et animales nonhumaines sont similaires (Scherer, 1982) et plusieurs caractéristiques acoustiques sont comparables (Green et Cliff, 1975; Morton, 1977; Scherer et Kappas, 1988; Rendall, 2003). Par exemple, une des vocalisations les plus étudiées est le pleur chez le nouveau-né, en partie parce qu'il représente un élément commun des bébés de la majorité des espèces primates (Newman, 1985; Bard et Gardner, 1996). Plus précisément, la plupart des bébés primates produisent certains types spécifiques de vocalisations lorsqu'ils sont séparés de leur mère, des pleurs de douleur et des pleurs d'inconfort, similaires entre les espèces de primates. Enfin, sur le plan acoustique, certains patrons intonatifs des vocalisations sont similaires à ceux produits par l'humain (Jürgens, 1998).

Les vocalisations nonlinguistiques ont été peu étudiées par les neurosciences. Comme il vient d'être mentionné, certains chercheurs (Green et Cliff, 1975; Scherer, 1981a, 1985) suggèrent une universalité de plusieurs aspects des vocalisations entre les espèces animales, y incluant les humains. Toutefois, on ne connaît toujours pas la spécificité ou l'universalité du traitement perceptuel des vocalisations. L'implication de régions cérébrales dans la perception de la voix des conspécifiques a été rapportée chez l'humain et chez le singe rhésus. Cependant, on ne sait pas si ces régions répondent spécifiquement aux vocalisations des conspécifiques ou aux vocalisations en général. Pourtant la perception de la voix et l'étude des réseaux neuronaux impliqués dans ce traitement sont essentielles afin de mieux comprendre les processus socio-communicatifs et de cognition émotionnelle chez l'humain sain et pathologique, ainsi que de participer à des grands débats telles les similitudes et les



différences entre les primates, humains et nonhumains.

L'étude des vocalisations a été négligée par les chercheurs en neurosciences au profit de l'exploration de fonctions supérieures, c.-à-d., les vocalisations linguistiques. L'homologue linguistique des vocalisations émotionnelles nonlinguistiques constitue la prosodie émotionnelle. Celle-ci est la composante vocale émotionnelle de ces vocalisations. Toutefois, comme il sera discuté plus en détails à la section *1.1 Limitations de l'utilisation des stimuli prosodiques*, ces stimuli portent de l'information à la fois émotionnelle et linguistique. Cependant, bien que la littérature soit relativement riche en ce qui concerne la prosodie émotionnelle, les aspects fondamentaux des expressions nonlinguistiques produites par la voix humaine ne sont toujours pas clairement décrits. Un de ces aspects est la charge émotionnelle que transportent les vocalisations nonlinguistiques. On ne sait toujours pas, par exemple, si le jugement de la valence émotionnelle (positive / négative) est constant entre les individus. De plus, l'influence de l'âge et du sexe des locuteurs et des décodeurs sur le jugement de différentes dimensions émotionnelles (par exemple l'intensité, l'authenticité de l'émotion) et l'effet de ces dimensions sur la mémoire ne sont pas connus. Ce qui est surprenant puisque l'étude de l'émotion constitue un des plus importants corps de recherche en neurosciences. Il paraît donc essentiel d'explorer le traitement émotionnel via les vocalisations nonlinguistiques afin de faire les liens entre la littérature sur le langage et les expressions faciales et d'inclure les vocalisations nonlinguistiques dans les modèles de traitement de l'émotion.

Deux parties composent le chapitre de l'introduction de la thèse. La première aborde la perception de l'émotion vocale. Une brève description de la prosodie émotionnelle est présentée, suivie d'une discussion sur des propriétés confondantes des stimuli prosodiques: la culture et la langue. La culture et la langue rendent difficile l'interprétation des résultats rapportés dans la littérature sur l'émotion vocale. Ces limites motivent entre autres l'étude des vocalisations nonlinguistiques, laquelle fait l'objet central de cette thèse. Enfin, un bref

résumé des recherches antérieures sur la perception de l'émotion vocale exprimée par la prosodie et les vocalisations nonlinguistiques est présenté. Différents traitements sont explorés dans la littérature : perceptuel, hédonique, induction d'émotion, etc. Seules les études explorant le traitement perceptuel sont discutées, ce dernier constituant l'objet principal de l'ensemble de ce travail. La deuxième section de l'introduction décrit la problématique abordée dans la thèse.

### **1. Prosodie émotionnelle**

L'émotion vocale a deux principaux modes d'expression, les vocalisations nonlinguistiques et les patrons prosodiques, et ces deux modes ne couvrent pas les mêmes spectres émotionnels. L'étendue des paramètres acoustiques utilisés dans les patrons prosodiques ne permet pas d'exprimer un aussi grand nombre d'émotions comparativement aux vocalisations nonlinguistiques et limite aussi l'intensité émotionnelle et la spontanéité d'expression (Scherer, 1981a; Ross, Edmondson et al., 1986; Barr, Hopkins et al., 2000). Néanmoins, la recherche sur la perception de l'émotion vocale s'est principalement intéressée à l'étude de la prosodie.

La prosodie émotionnelle désigne l'information affective véhiculée par un locuteur à propos de son propre état émotif. Elle s'inscrit dans le développement des émotions, de la socialisation et du langage. Si les habiletés à percevoir les variations prosodiques sont déficitaires, les intentions de communication d'autrui ne seront pas correctement identifiées et interprétées. De tels déficits peuvent avoir des répercussions au niveau de la régulation des émotions. Par le fait même, le fonctionnement global de l'individu, plus particulièrement dans la réciprocité des interactions sociales, est touché. La prosodie émotionnelle constitue donc un outil essentiel dans le traitement de l'émotion (Joanette, Goulet et al., 1990).

### *1.1 Limitations de l'utilisation des stimuli prosodiques*

Comme il a été mentionné plus haut, les patrons prosodiques constituent les stimuli les plus utilisés dans les études antérieures pour explorer l'émotion vocale. Toutefois, l'utilisation même de stimuli langagiers limite l'étude de la perception des émotions. La production et la perception des émotions via des extraits langagiers sont influencées par l'appartenance culturelle et la langue des locuteurs et des décodeurs. La nature et l'importance de ces influences sont toutefois insuffisamment définies pour étudier l'émotion isolée de ces facteurs.

#### Culture

La production des patrons prosodiques émotionnels est en partie déterminée par des normes culturelles (McCluskey, Albas et al., 1975; Albas, McCluskey et al., 1976; McCluskey et Albas, 1981; Van Bezooijen, Otto et al., 1983; Mesquita, Frijda et al., 1997). La perception est donc susceptible de varier entre des décodeurs de culture différente. Malgré une reconnaissance des émotions via des stimuli prosodiques au-dessus du niveau chance, des différences entre cultures quant à la perception de ces émotions sont notables (Albas, McCluskey et al., 1976). Les similarités et les différences de perception des émotions ont été davantage étudiées en utilisant les expressions faciales (voir revue sur la perception des émotions faciales et vocales de Mesquita et Frijda, 1992 et voir Scherer et Wallbott (1994) pour une discussion sur l'universalité dans les réponses émotionnelles), mais le débat à savoir si la perception des émotions est universelle ou non est toujours ouvert.

#### Sémantique

L'influence de la langue, quant à elle, est principalement due au fait que la prosodie n'est pas la seule source langagière à pouvoir exprimer des émotions. La sémantique, l'élément linguistique se rapportant au sens des mots, peut porter une charge émotionnelle. Par exemple, la sémantique du mot «sérénité» traduit une émotion plus marquée comparativement à celle du mot «patin». L'information émotionnelle qu'exprime la

sémantique peut influencer celle que porte la prosodie. Par exemple, les charges émotionnelles des mots «sérénité» et «patin» prononcés avec le même patron prosodique ne sont pas comparables. Ces différences ont principalement été mesurées par des paradigmes de mémoire (Rubin et Friendly, 1986; Cahill et McGaugh, 1998). En acceptant l'hypothèse suggérant que l'importance de la charge émotionnelle est relative à la mémorisation du stimulus, on suppose que le traitement émotionnel des exemples donnés ci haut est différent. Ces différences doivent être connues lorsqu'on étudie la perception des émotions. Différentes stratégies méthodologiques ont été utilisées et développées dans le but de contrôler l'influence de la langue et de la culture.

### *1.2 Méthodes pour contrôler les limitations des stimuli prosodiques*

Afin d'étudier la perception de l'émotion vocale via le langage, l'information prosodique doit être isolée de la sémantique. Plusieurs stratégies méthodologiques ont donc été développées pour contrôler ou masquer le contenu sémantique des stimuli utilisés.

#### Stimuli asémantiques

La première stratégie constituait à contrôler la sémantique en présentant le même stimulus, dont le contenu sémantique est émotionnellement neutre, prononcé avec différents patrons prosodiques émotionnels. L'utilisation de lettres de l'alphabet (Dusenbury et Knower, 1938; Knower, 1941; Davitz et Davitz, 1959; Hornstein, 1967; Wolf, Gorski et al., 1972), de chiffres (Pfaff, 1954), de mots simples (p. ex., les jours de la semaine; Osser, 1964) et de phrases (Fairbanks et Pronovost, 1939; Pollack, Rubenstein et al., 1960; Beldoch, 1964; Dimitrovsky, 1964; Kramer, 1964; Levy, 1964; Burns et Beier, 1973; Schlanger, 1973; Ross, Duffy et al. 1973; Zuckerman, Lipets et al., 1975; Fenster, Blake et al., 1977; Safer et Leventhal, 1977; Brown, 1980; Wallbott et Scherer, 1986) a contribué à l'étude de la prosodie émotionnelle en contrôlant l'information sémantique. Toutefois, l'influence de la langue et de la culture sur la perception des émotions en utilisant ce type de stimuli n'est

pas comparable selon si les stimuli appartiennent ou non à la langue et/ou au groupe culturel de la population étudiée.

### Stimuli chuchotés

Une autre stratégie, visant à élider l'information sémantique, consiste à utiliser des voix chuchotées, desquelles on ne peut percevoir l'information sémantique (Knowler, 1941). Cette méthode a inspiré les chercheurs à manipuler acoustiquement les stimuli.

### Segmentation, filtres acoustiques et inversion

Des procédures acoustiques ont été développées afin de masquer l'intelligibilité du contenu sémantique. La segmentation, les filtres acoustiques et l'inversion temporelle constituent les principales procédures acoustiques. Avant d'aborder les procédures acoustiques, il est nécessaire de décrire brièvement les caractéristiques acoustiques de la prosodie. La prosodie s'exprime acoustiquement par des variations de fréquence, de timbre, d'amplitude et de durée. Les contreparties perceptives de ces paramètres sont, dans l'ordre, la hauteur, le timbre perçu (qualité identificatrice), l'intensité et la longueur. La fréquence correspond au nombre de vibrations des cordes vocales et est calculée en cycles par seconde (Hertz). La fréquence la plus basse, qu'on appelle la fréquence fondamentale (F0 ou fundamental ou harmonique zéro), donne au son sa hauteur. Les harmoniques correspondent aux autres ondes que la F0 et elles ont une amplitude décroissante à mesure que l'on s'éloigne de la F0. Les formants correspondent à une zone d'harmoniques amplifiées. Le timbre est cette qualité qui permet de discerner les voix. On parle de timbre de voix aigu pour les sons dont les composantes fréquentielles sont hautes et de timbre grave pour ceux dont les fréquences sont basses. L'amplitude est la fonction de la pression sonore et est calculée en décibels. Enfin, des composantes de la structure temporelle du patron prosodique (par exemple, le tempo, le rythme, les pauses et l'enveloppe) sont essentielles pour identifier l'émotion.



La F0 est l'information cruciale dans les patrons prosodiques. Les variations de la F0 représentent l'indice perceptuel qui caractérise le mieux le niveau d'engagement émotionnel du locuteur (Williams et Stevens, 1972; Ladd et Silverman, 1985; Scherer, 1986). Par exemple, la moyenne et les variations de la F0 sont, selon Pell (1999), les variables distinguant la gaieté, la tristesse, la colère et la condition neutre. Les relations entre les paramètres acoustiques et les catégories d'émotion sont discutées ultérieurement à la section 1.4 *Résultats sur la perception de la prosodie émotionnelle*.

La segmentation est parmi les procédures acoustiques visant à masquer le contenu sémantique de l'extrait langagier. Elle consiste à couper le stimulus en unités, lesquelles sont interverties pseudo-aléatoirement et re-séquencées afin d'obtenir un nouveau stimulus. Cette technique permet de garder intactes la F0, l'amplitude et le timbre, ainsi que de masquer le contenu sémantique du stimulus. Cependant, les pauses, le tempo et le rythme sont aussi modifiés, la prosodie est donc changée de façon importante. L'expression de l'émotion via la prosodie n'est pas représentative puisque des composantes mêmes de l'information prosodique sont modifiées. Pour étudier la perception des patrons prosodiques émotionnels, ceux-ci doivent être naturels et fiables, ce qui n'est pas le cas dans la segmentation. Une autre procédure, celle de l'inversion temporelle, consiste à faire jouer le stimulus langagier à l'envers. Dans cette condition, l'information sémantique est relativement masquée. Toutefois, comme dans le cas de la segmentation, la prosodie est changée de façon notable.

L'utilisation de certains filtres acoustiques permet de masquer le contenu sémantique alors que les indices prosodiques demeurent intacts et les caractéristiques acoustiques préservées ou modifiées diffèrent selon le filtre. Le filtre permet d'arrêter en tout ou en partie une zone de fréquence. La gamme de fréquences qui passe à travers le filtre s'appelle la bande passante. Le filtre passe-haut laisse passer les hautes fréquences (l'étendue des formants F2-F3) et supprime les autres, dont la F0. Le filtre passe-bas, quant à lui, laisse

passer les basses fréquences et élimine les autres. Selon le seuil fréquentiel du filtre, différentes informations sont altérées, éliminées ou conservées. Par exemple, les filtres avec bandes passe-bas de 80 à 300Hz préservent la F0 mais masquent certaines informations phonétiques (Lakshminarayanan, Ben Shalom et al., 2003). Le filtre avec une bande passe-bas d'un seuil de fréquence de 500Hz, quant à lui, masque le timbre et l'amplitude, mais préserve la F0, le tempo, le rythme et les pauses (Johnson, Emde et al., 1986). Parfois, à l'intérieur d'une même tâche, les seuils fréquentiels des bandes passe-bas choisis diffèrent selon le sexe des locuteurs (Scherer, 2000). L'application d'un seuil de 400Hz aux voix de femmes et d'un seuil de 300Hz aux voix d'hommes respecte les différences de la F0 des voix selon le sexe.

#### Rotation spectrale et langage rentrant

D'autres procédures acoustiques ont été récemment utilisées dans les études sur l'émotion vocale. Scott, Blank et al. (2000) se sont référés à la technique de rotation spectrale, laquelle détruit l'intelligibilité du langage, alors que la structure acoustique demeure intacte et la complexité de cette structure se compare à celle des stimuli langagiers originaux (voir Blesser, 1972 pour plus de détails). Lakshminarayanan, Ben Shalom et al. (2003), quant à eux, ont eu recours à la technique de langage rentrant, laquelle préserve l'enveloppe fréquentielle (conserve la F0, la durée, l'amplitude) alors que l'information phonémique est éliminée, le stimulus langagier est donc inintelligible.

L'utilisation des procédures acoustiques permet de masquer la sémantique des stimuli langagiers tout en respectant le patron prosodique. Néanmoins, la perception des émotions de ces stimuli se module selon le bagage langagier et culturel des locuteurs et des décodeurs puisque ces stimuli appartiennent quand même à une langue et à une culture donnée. L'effet de la langue et de la culture sur l'habileté à percevoir les patrons prosodiques émotionnels n'est donc pas comparable selon le bagage langagier et culturel des individus.

### Stimuli pseudo-langagiers asémantiques

Afin de diminuer l'effet de référence de la langue et de la culture, des phrases pseudo-langagières asémantiques («speechlike but semantically meaningless») ont été développées par Scherer, Banse et al. (2001). Ces phrases ont été construites à partir des systèmes phonologiques et phonétiques de langues indo-européennes. Elles n'appartiennent donc pas à une langue ou à une culture spécifique, mais représentent plutôt les références d'un large groupe d'individus. Dans cette étude, les neuf décodeurs, de culture et de langue différentes, ont reconnu certaines émotions véhiculées par les phrases et les matrices de confusion étaient similaires entre les décodeurs. Toutefois, un effet de la culture et/ou de la langue dans la reconnaissance des émotions était observé. Selon les auteurs, l'information segmentale et suprasegmentale utilisée par les décodeurs varie selon la langue et la culture, en dépit du fait que l'utilisation de ces stimuli diminue l'effet de référence.

Il n'y a pas que la sémantique de l'extrait langagier qui biaise le traitement de l'émotion vocale via la prosodie, la structure syntaxique semble aussi influencer le traitement de l'émotion vocale. L'existence d'une corrélation entre des frontières syntaxiques et des changements prosodiques est rapportée. Certains changements prosodiques, plus particulièrement les pauses, la longueur de la syllabe, le retour à la hauteur tonale initiale, ainsi que certaines variations de la F0, tendent à apparaître aux frontières syntaxiques (Klatt, 1975). Des études ont cependant montré que cette corrélation reste imparfaite (Beckman et Edwards, 1990; Ferreira, 1993; Gerken, 1996a,b), plus particulièrement en ce qui concerne la hiérarchie de la structure syntaxique (Chomsky et Halle, 1968; Morgan, Meier et al., 1987). Malgré le fait que la structure prosodique n'indique peut-être pas toujours précisément des marqueurs syntaxiques, elle se retrouve au premier plan dans l'input.

Pour toutes ces raisons, l'utilisation même d'extraits langagiers constitue un problème inhérent pour étudier l'émotion vocale. Chaque locuteur utilise sa culture et sa langue



comme références pour identifier les émotions exprimées par les stimuli langagiers, qu'ils soient asémantiques, manipulés acoustiquement ou pseudo-langagiers. Les différences de culture et de langue entre les locuteurs et les décodeurs sont au coeur même du traitement des émotions (Ekman, 1994; Russell, 1994; Mesquita, Frijda et al., 1997; Scherer, 1997a,b; Scherer, Banse et al., 2001). La nature et l'importance des inférences de ces stimuli dues à la langue et/ou à l'apprentissage ou à d'autres techniques de transmission culturelle sur le traitement des patrons prosodiques ne sont pas encore bien connues. Aussi, l'existence de caractéristiques universelles de patrons vocaux émotionnels et d'une habileté universelle à percevoir ces patrons (Kramer, 1964; Beier et Zautra, 1972; van Bezooijen, 1984), lesquelles sont basées sur la continuité phylogénique de l'émotion vocale, demeurent, quant à eux, non décrits.

Enfin, les stimuli ne sont pas seuls responsables de la difficulté à interpréter les résultats rapportés dans la littérature et de la divergence de ceux-ci. Les choix méthodologiques concernant les tâches utilisées sont également à considérer.

### *1.3 Types de tâches utilisées*

Différents paradigmes ont été utilisés pour étudier la perception des émotions (Bachorowski et Owren, 2003 pour une revue). Certains visent à explorer la perception de la catégorie de l'émotion (p. ex., la joie), d'autres à examiner la perception des dimensions émotionnelles des stimuli (p. ex., l'intensité).

#### Tâches de jugement catégoriel

La tâche la plus souvent utilisée dans la littérature est celle du choix forcé, dans laquelle les participants doivent sélectionner une catégorie d'émotion parmi plusieurs proposées (voir Scherer, 1986 pour une revue; Wallbott et Scherer, 1986; Pittam et Scherer, 1993), ou deux catégories pour un même stimulus donné (Scherer, Banse et al., 2001). Une autre tâche,

celle du choix libre («free-response labeling»), consiste à demander aux participants de décrire par des mots ou des phrases l'émotion perçue (Pakosz, 1983; Johnson, Emde et al., 1986; Sobin et Alpert, 1999). Les réponses des participants sont ensuite classées en catégories d'émotion selon un lexique de mots exprimant des émotions préalablement validé (Izard, 1972). Plusieurs auteurs (p. ex., Pittam et Scherer, 1993; Banse et Scherer, 1996) critiquent la tâche du choix forcé puisqu'elle limite la mesure de perception de l'émotion à un choix (ou à deux) parmi des catégories d'émotion fixes. La stratégie utilisée pour attribuer une catégorie d'émotion à un stimulus dans le cadre de cette tâche ne semble pas être toujours représentative de celle utilisée dans un contexte plus écologique. Par exemple, lorsque l'émotion exprimée par un stimulus est ambiguë et qu'il est relativement difficile de lui attribuer une seule catégorie, certains participants utilisent une stratégie d'exclusion des catégories non plausibles et choisissent la catégorie restante, même si elle ne correspond pas exactement à leur perception. Johnson, Emde et al. (1986) et Russell (1994) ont comparé les performances à la tâche de choix forcé et celle du choix libre et les résultats indiquent que cette dernière maximise l'opportunité à faire des jugements impliquant d'autres émotions. Par exemple, des participants ont caractérisé des stimuli de joie ou de peur comme exprimant de l'intérêt et de la surprise.

#### Tâches des dimensions émotionnelles

L'étude de la perception de dimensions émotionnelles permet, quant à elle, d'avoir plusieurs mesures émotionnelles pour un même stimulus donné. Différentes dimensions peuvent être mesurées. Scherer (1981a) a demandé aux participants de juger les dimensions de actif / passif, plaisant / déplaisant et fort / faible des stimuli sur des échelles continues. Plus récemment, Scherer, Banse et al. (2001) ont exploré les dimensions d'intensité et d'authenticité («naturelness») de l'émotion exprimée. De plus, l'utilisation d'une échelle continue permet d'obtenir un jugement plus fin, donc d'observer plus facilement des différences subtiles entre des groupes de participants.

Le choix de la tâche utilisée est généralement fait selon notre position quant aux différents modèles théoriques sur la perception de l'émotion (pour plus de détails sur cet important débat, voir Ortony et Turner, 1990; Levenson, Carstensen et al., 1991; Ekman, 1992a,b; Izard, 1992; Panksepp, 1992; Turner et Ortony, 1992; Russell et Barrett, 1999; Frank et Stennett, 2001). Certains chercheurs soutiennent que les émotions sont des unités discrètes (p. ex., Darwin, 1872/1965; Oatley et Johnson-Laird, 1987; Gray, 1990; Ekman, 1992a,b; Lang, Bradley et al., 1992), c'est-à-dire l'existence d'un nombre restreint d'émotions de base, incluant la tristesse, la joie, la colère, la surprise, la peur et le dégoût. Ils optent donc davantage pour les tâches de jugement catégoriel. Alternativement, d'autres chercheurs conceptualisent les émotions à partir de dimensions émotionnelles (Russell, 1980; Russell et Fehr, 1987). Cette conception, plus flexible que celle des émotions de base, met l'emphase sur des processus dynamiques incluant plusieurs composantes en lien avec la cognition en caractérisant les émotions par des dimensions tel le niveau d'activation («arousal») et de plaisant / déplaisant. Aussi, au lieu d'explorer si les décodeurs trouvent le stimulus plaisant ou non, d'autres chercheurs demandent aux décodeurs si l'émotion exprimée par l'encodeur est, par exemple, authentique ou non (Scherer, Banse et al., 2001).

#### *1.4 Résultats sur la perception de la prosodie émotionnelle*

Les études comportementales sur la perception de la prosodie émotionnelle examinent le plus souvent notre capacité à attribuer une émotion à un stimulus prosodique ou à le décrire en termes d'activation. Les études de neuroimagerie, quant à elles, explorent les bases neuronales associées à la perception de stimuli prosodiques. Cependant, il ne semble pas exister jusqu'à maintenant dans la littérature d'études qui font le lien entre les données comportementales et celles de neuroimagerie.

#### Résultats sur la perception de la prosodie émotionnelle : études comportementales

En général, les études antérieures suggèrent que notre habileté à percevoir l'état émotionnel exprimé par la prosodie est bonne (voir Russell, Bachorowski et al., 2003 pour une revue;

Siegman, 1978; Scherer, 1981a,b; Williams et Stevens, 1981; Scherer, 1995), quoique les performances sont généralement inférieures à celles d'identification des expressions faciales. Dans l'étude de Banse et Scherer (1996), les participants reconnaissent les quatorze émotions étudiées (colère chaude, colère froide, peur / panique, anxiété, désespoir, tristesse, allégresse, joie, intérêt, ennui, honte, fierté, dégoût et mépris). Il semble que les émotions négatives soient mieux reconnues que les émotions positives dans les tâches de catégorisation de l'émotion (Fenster, Blake et al., 1977) et dans les jugements des dimensions émotionnelles (Scherer, Banse et al., 2001), indépendamment des conditions de présentation des stimuli.

Les conditions dans lesquelles sont présentés les stimuli prosodiques ne semblent pas influencer de façon importante la perception catégorielle des émotions exprimées. Wallbott et Scherer (1986), utilisant une tâche de choix forcé de stimuli non filtrés et filtrés (avec une bande passe-bas d'un seuil fréquentiel de 400 Hz), ont observé que les performances à identifier les émotions négatives (colère et tristesse) étaient supérieures à celles des émotions positives (joie et surprise). Johnson, Emde et al. (1986) ont étudié la perception de stimuli de joie, de tristesse, de colère et de peur produits par des voix naturelles et synthétisées sujettes à trois conditions: filtrage, segmentation et inversion temporelle. Même si les performances diminuaient graduellement selon la difficulté des conditions, aucune d'elles n'altérait significativement les performances de perception de la catégorie émotionnelle. Une autre étude, celle de Lakshminarayanan, Ben Shalom et al. (2003), a mesuré l'effet de différentes manipulations spectrales sur la perception de stimuli prosodiques émotionnels de joie, de colère et de tristesse dans une tâche de choix forcé. Les stimuli étaient non filtrés, filtrés avec des bandes passe-bas et des bandes passe-haut, rentrants et re-synthétisés. Les performances étaient bonnes en dépit de la manipulation spectrale, même lorsque l'information spectrale était relativement minime. Cependant, les performances étaient significativement diminuées dans les conditions filtrées et rentrants. Il est surprenant que cette dernière condition altère les performances puisqu'elle préserve la



F0. Les auteurs suggèrent qu'elle compromet la syllabification, laquelle constitue un aspect intrinsèque de la structure temporelle et représente donc un indice essentiel aux patrons prosodiques.

Une littérature qui a grandement contribué à la recherche sur la perception des émotions est celle d'amorçage (p. ex., Vuilleumier, Armony et al., 2001), un processus de mémoire implicite (Schacter et Tulving, 1994). Malgré que certaines études ont exploré l'effet d'amorçage dans le domaine auditif en utilisant une tâche d'identification de mots ou de phrases (p. ex., Kempley et Morton, 1982), une tâche de complétion de mots (p. ex., McClelland et Pring, 1991), l'effet d'amorçage de stimuli émotionnels auditifs a néanmoins été très peu exploré. Quelques études ont rapporté un effet d'amorçage au niveau sémantique en utilisant des stimuli langagiers (Rubin et Friendly, 1986; Maratos, Allan et al., 2000) en observant un effet d'amorçage plus important pour les stimuli exprimant une émotion négative comparativement aux stimuli neutres. Toutefois, Church et Schacter (1994) n'ont pas observé d'effet d'amorçage plus important pour les stimuli avec un patron prosodique émotionnel comparativement à ceux avec un patron prosodique neutre.

#### Résultats sur la perception de la prosodie émotionnelle: études de neuroimagerie fonctionnelle

Plusieurs études d'imagerie fonctionnelle, particulièrement celles utilisant l'imagerie par résonance magnétique fonctionnelle (IRMf) et la tomographie par émission de positrons (TEP), ont exploré les substrats neuronaux du traitement perceptuel de l'émotion vocale. Tout comme pour les études comportementales, le principal intérêt de ces études de neuroimagerie constitue la perception de la prosodie émotionnelle.

Les études d'imagerie fonctionnelle rapportent la sollicitation de plusieurs régions cérébrales dans le traitement perceptuel de l'émotion vocale (pour une revue, voir Adolphs, 2002; Russell, Bachorowski et al., 2003). Ces études ne font toujours pas consensus quant

aux régions impliquées dans ce traitement, en partie en raison de leurs différences méthodologiques. Néanmoins, certains substrats neuronaux, de par leurs activations les plus souvent rapportées dans la littérature, semblent être plus impliqués que d'autres. Les régions les plus souvent observées dans le traitement perceptuel de l'émotion vocale sont les gyri temporaux, le cortex préfrontal, l'amygdale, l'insula et les ganglions de la base. Les tableaux 2 et 3, ainsi que la figure 1, présentent les principales études explorant les substrats neuronaux impliqués dans le traitement perceptuel de la prosodie émotionnelle et des vocalisations nonlinguistiques émotionnelles chez l'individu sain, respectivement.

Même si les études de lésion fournissent des indications importantes quant aux structures impliquées dans le traitement perceptuel de l'émotion vocale (p. ex., Monrad-Krohn, 1947; Heilman, Scholes et al., 1975; Ross, 1981; Gorelick et Ross, 1987; Ross, Anderson et al., 1989; Cancelliere et Kertesz, 1990; Darby, 1993; Habib, Daquin et al., 1995; Hornak, Rolls et al., 1996), elles ne sont pas présentées dans le tableau 2, le thème général de ce travail étant le traitement perceptuel de l'émotion vocale chez l'individu sain. Les différences entre ces deux types de littérature sont importantes tant au niveau méthodologique qu'à l'interprétation des résultats.

Tableau 2. Principales études de neuroimagerie fonctionnelle explorant le traitement de la prosodie émotionnelle chez l'individu sain

référence	région	x,y,z{mm}
<p><b>Kawashima, Itoh et al., 1993</b>  <u>technique</u>: TEP  <u>émotions</u>: ND (émotionnel, non émotionnel)  <u>stimuli</u>: histoires en Italien chez des participants japonais ne parlant pas Italien  <u>tâche</u>: écoute passive  <u>participants</u>: 8 hommes</p>	<p><u>régions d'intérêt</u>: partie operculum des gyri frontaux inférieurs, gyri temporaux transverses, partie postérieure des gyri temporaux supérieurs, gyri angulaires  <u>émotion &gt; condition basale</u>  partie operculum du gyrus frontal inférieur  gyrus temporal supérieur postérieur  gyrus temporal transverse  gyrus angulaire</p>	ND
<p><b>George, Parekh et al., 1996b</b>  <u>technique</u>: TEP  <u>émotions</u>: joie, tristesse, colère, neutre  <u>stimuli</u>: phrases  <u>tâche</u>: identification de l'émotion, réponse verbale  <u>tâche contrôle</u>: répétition du second mot de la phrase  <u>participants</u>: 8 hommes et 5 femmes</p>	<p><u>émotion &gt; tâche contrôle**</u>  D / G lobe préfrontal latéral  <u>émotion &gt; verbal **</u>  G tronc cérébral / thalamus</p>	<p>35 35 -6; -40 30 -8  -6 -30 2</p>
<p><b>Imaizumi, Mori et al., 1997</b>  <u>technique</u>: IRMf  <u>émotions</u>: joie, tristesse, surprise, neutre  <u>stimuli</u>: mots  <u>tâche</u>: identification de l'émotion et identification du locuteur  <u>participants</u>: 6 hommes</p>	<p><u>émotion &gt; condition basale **</u>  D gyrus frontal inférieur  G insula  G gyrus lingual  D / G cervelet  G gyrus frontal médian  D gyrus parahippocampal  D gyrus occipital latéral</p>	<p>-42 39 6  41 -2 23  6 -63 4  46 -62 -22*  34 42 20*  -24 -28 -16  -40 -78 12</p>

<p><b>Buchanan, Lutz et al., 2000</b>  <u>technique</u>: IRMf  <u>émotions</u>: joie, colère, tristesse, neutre  <u>stimuli</u>: mots  <u>tâche</u>: identification de l'émotion  <u>tâche verbale</u>: détection des mots «power» et «bower»  <u>participants</u>: 10 hommes</p>	<p><u>émotion &gt; verbal</u> **  D gyrus frontal inférieur 44 20 16  G gyrus cingulaire -12 4 20  D pariétal inférieur 40 -48 32  <u>tristesse &gt; joie</u>  D gyrus frontal antérieur moyen 36 44 24  <u>joie &gt; tristesse</u>  aucune réponse significative  <u>tristesse &gt; verbal</u> **  D gyrus frontal postérieur moyen 40 8 28  <u>joie &gt; verbal</u> **  D gyrus temporal supérieur / lobe pariétal inférieur 32 -28 28</p>	
<p><b>Wildgruber, Pihan et al., 2002</b>  <u>technique</u>: IRMf  <u>émotions</u>: joie, tristesse, neutre  <u>stimuli</u>: paires de phrases (première phrase, p1; seconde phrase, p2)  <u>tâche</u>: identification de l'émotion et discrimination de la phrase la plus <i>expressive</i>  <u>participants</u>: 6 hommes et 6 femmes</p>	<p><u>prosodie (joie, tristesse, neutre) &gt; silence (p1)</u>  D / G cortex frontal dorsolatéral 45 21 24; -45 33 15  D / G cortex frontal orbitobasal, insula antérieure 36 21 -3; -36 27 -3  D / G gyrus temporal supérieur 63 -24 3; -60 -18 0  D / G thalamus 9 -15 0; -12 -15 3  D pallidum 15 9 -6  D gyrus angulaire antérieur 9 -48 6  gyrus angulaire postérieur 0 -60 12  <u>prosodie (joie, tristesse, neutre) &gt; silence (p2)</u>  D / G cortex frontal dorsolatéral 45 6 24; -48 27 27  D / G cortex frontal orbitobasal, insula antérieure 54 24 -6; -36 27 -3  D / G gyrus temporal supérieur 66 -21 3; -63 -15 0  D lobule pariétal inférieur 57 -39 45  D pallidum 18 6 -3  G aire motrice supplémentaire -3 6 57  <u>discrimination de l'expressivité</u> **  D / G cortex frontal dorsolatéral 48 6 21; -51 21 36  D / G cortex frontal orbitobasal, insula antérieure 54 18 -12; -54 21 -6</p>	



	D lobule pariétal inférieur	57 -42 48
	G aire motrice supplémentaire	-3 12 51
<b>Kotz, Meyer et al., 2003</b>	<u>positive &gt; neutre</u> (stimuli non filtrés) **	
<u>technique</u> : IRMf	D / G gyrus fusiforme inférieur	52 13 4; -50 13 1
<u>émotions</u> : positive (ex. joie), négative (ex. colère), neutre	G gyrus fusiforme inférieur	-47 28 1
<u>stimuli</u> : phrases	G gyrus d'Heschl médian	-44 -22 9
sémantiquement neutres, non filtrées et filtrées (suppression du 3ième harmonique et les suivants)	G planum temporal	-53 -30 11
<u>tâche</u> : jugement de valence (positif, négatif ou neutre)	D / G sillon temporal supérieur / gyrus temporal médian	46 -52 12; -44 -49 9 46 -46 26
<u>participants</u> : 4 hommes et 8 femmes	D lobe pariétal inférieur	5 8 2; -7 2 4
	D / G noyau caudé	
	<u>négative &gt; neutre</u> (stimuli non filtrés) **	32 -5 16; -41 3 8
	D / G operculum rolandique / insula	-58 -13 12
	G gyrus d'Heschl latéral	
	<u>positive &gt; neutre</u> (stimuli filtrés)	-28 22 0
	G insula antérieure	-53 19 4; 44 21 13
	D / G gyrus fusiforme inférieur	-47 35 -2
	G gyrus fusiforme inférieur	28 38 6
	D gyrus fusiforme médian	-49 -14 7
	G gyrus de Heschl latéral	-55 -34 12
	G planum temporal	-32 -46 34
	G sillon intrapariétal	10 16 1; -9 19 3
	D / G noyau caudé	
	<u>négative &gt; neutre</u> (stimuli filtrés)	-29 22 0
	G insula antérieure	-44 21 13
	G gyrus fusiforme inférieur	-52 21 3
	G gyrus fusiforme inférieur	-47 35 -2
	D / G gyrus fusiforme	40 32 11; -44 32 7
	D / G gyrus fusiforme inférieur/ médian	40 19 20
	D gyrus frontal supérieur	-48 -14 7
	G gyrus de Heschl latéral	-54 -32 11
	G planum temporal	-31 -48 40
	G sillon intrapariétal	11 15 1; -9 11 5
	D / G noyau caudé	

<p><b>Mitchell, Elliott et al., 2003</b>  <u>technique</u>: IRMf  <u>émotions</u>: joie, tristesse, neutre  <u>stimuli</u>: phrases sémantiquement émotionnelles prononcées avec différents patrons prosodiques émotionnels, stimuli non filtrés et filtrés (bande passe-bas, seuil de 333Hz)  <u>tâche</u>: écoute passive  <u>participants</u>: 13 hommes</p>	<p><u>émotion &gt; silence</u> (stimuli non filtrés) **  D gyrus temporal médian  G gyrus temporal supérieur  D lobule pariétal inférieur  D gyrus frontal inférieur</p> <p><u>émotion &gt; neutre</u> (stimuli non filtrés) **  D gyrus temporal inférieur  D gyrus frontal médian  G gyrus frontal supérieur  D précuneus  gyrus cingulaire</p> <p><u>neutre &gt; émotion</u> (stimuli non filtrés)  G gyrus frontal inférieur  G gyrus frontal médian</p> <p><u>émotion &gt; silence</u> (stimuli filtrés)  D / G gyrus temporal médian  G gyrus frontal médian</p>	<p>60 -6 -9  -45 -21 6  45 -45 45  48 39 12  60 -54 -9  42 36 -3  -24 54 36  15 -54 33  0 -30 27  -51 36 12  -36 12 51  57 -9 -6; -54 -9 -6  -36 0 60</p>
<p><b>Sander, Roth et al., 2003b</b>  <u>technique</u>: IRMf, approche par bloc  <u>émotion</u>: joie, tristesse  <u>stimuli</u>: adjectifs  <u>tâche</u>: identification de l'émotion  <u>tâche contrôle</u>: identification de /a/  <u>participants</u>: 24 femmes</p>	<p><u>régions d'intérêt</u>: régions préfrontales, frontales, insulaires, temporopariétales et occipitales</p> <p><u>joie &gt; /a/</u>  D / G préfrontal  D / G frontal  D / G insula  D / G temporopariétal</p> <p><u>joie &gt; tristesse</u>  D / G frontal  D / G insula</p>	<p>ND</p>

\* pic maximal de cette région

\*\* contrastes rapportées dans la figure 1

Tableau 3. Principales études de neuroimagerie fonctionnelle explorant le traitement des vocalisations nonlinguistiques émotionnelles chez l'individu sain

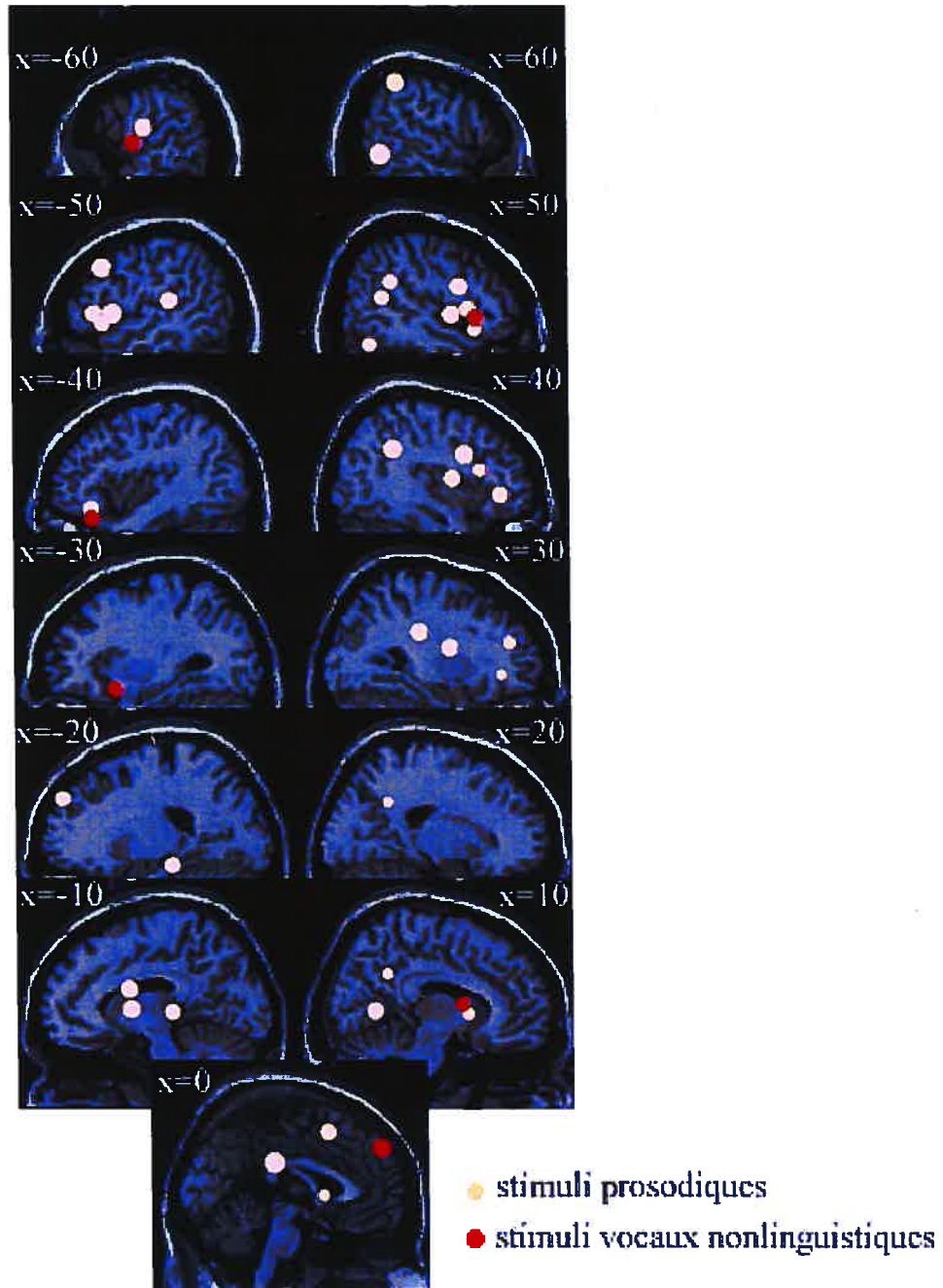
référence	région	x,y,z{mm}
<b>Morris, Scott et al, 1999</b> <u>technique</u> : TEP <u>émotions</u> : tristesse, joie, peur, neutre (nasales voisées) <u>tâche</u> : identification du sexe <u>participants</u> : 6 hommes	<u>émotion &gt; neutre</u> (nasales voisées) **	
	G gyrus temporal médian	-62 -6 -4
	gyrus frontal supérieur	0 54 42
	D noyau caudé	14 6 10
	D / G insula antérieure	-34 12 -16; 46 20 -2
	D / G cortex préfrontal ventral	-36 30 -14; 46 20 -2
	<u>peur &gt; joie + tristesse + neutre</u>	
	G insula	-34 18 -14
	D gyrus frontal supérieur	16 46 36
	G cortex préfrontal ventral	-72 -26 -10
G gyrus temporal médian	-34 18 -14	
D amygdale	18 -4 -22	
<b>Phillips, Young et al, 1998</b> <u>technique</u> : IRMf <u>émotions</u> : peur, dégoût, neutre (c.-à-d., joie moyenne ( <i>rire + soupir</i> )) <u>tâche</u> : identification du sexe <u>participants</u> : 6 hommes	<u>peur &gt; neutre</u> (joie moyenne)	
	D gyrus temporal inférieur postérieur	23 -53 -7
	noyau caudé	0 8 9
	D cortex frontal médian	3 44 4
	G cortex occipital médian	-17 -50 -2
	D / G gyrus temporal médian	38 -25 -7; -40 -17 -7
	D / G gyrus temporal supérieur	23 -31 15*; -52 -31 15
	D hippocampe	6 -44 4
	D gyrus cingulaire antérieur	6 25 -2
	D hippocampe-amygdale	12 -19 -13
	D thalamus	3 -3 4*
	<u>dégoût &gt; neutre</u> (joie moyenne)	
	D / G cortex occipital médian	9 -67 31; -29 -67 31
	D / G gyrus temporal supérieur	55 -33 15*; -55 -22 4*
	G gyrus cingulaire postérieur	-9 -64 15*
D cortex préfrontal dorsolatéral	40 28 4	
G thalamus	-3 -8 4	

	G cortex préfrontal rostral <u>peur &gt; dégoût</u> D hippocampe G gyrus cingulaire postérieur D cortex occipital médian D gyrus cingulaire antérieur D cortex frontal médian D noyau caudé D gyrus temporal médian <u>dégoût &gt; peur</u> G cortex occipital médian G gyrus cingulaire postérieur D cortex préfrontal dorsolatéral D / G gyrus temporal supérieur G thalamus	-26 36 15 38 -22 -7 -12 -58 9 26 -53 -7 3 28 -2 6 44 4 20 -28 15 55 -8 -2 -32 -67 31 -9 -61 15 40 31 4 52 -22 9*; -55 -19 4 -3 -11 4
<b>Sander et Scheich, 2001; voir aussi Sander, Brechmann et al., 2003</b> <u>technique</u> : IRMf <u>émotions</u> : joie, tristesse <u>tâche</u> : écoute passive <u>tâche contrôle</u> : soustraction de chiffres <u>participants</u> : 10 femmes et 1 homme	<u>région d'intérêt</u> : amygdale <u>rire &gt; silence</u> D / G amygdale <u>pleur &gt; silence</u> D / G amygdale <u>rire &gt; tâche</u> D / G amygdale <u>pleur &gt; tâche</u> D / G amygdale	ND

\* pic maximal de cette région

\*\* contrastes rapportés dans la figure 1

Figure 1. Régions activées: stimuli vocaux émotionnels versus non-émotionnels



Peu d'études ont pour objet d'explorer les réponses spécifiques à certains types d'émotions vocales (p. ex., joie, tristesse). L'expression de peur est l'émotion vocale la plus spécifiquement explorée et l'amygdale est reconnue pour son implication dans la perception des expressions faciales de peur (voir Adolphs, 2002; Haxby, Hoffman et al., 2002 pour une revue). L'équipe de recherche de Phillips a exploré si cette réponse de l'amygdale était spécifique à la modalité visuelle ou si cette réponse serait aussi observée dans la modalité auditive. Phillips, Young et al. (1998) ont rapporté l'implication de l'amygdale dans la reconnaissance des expressions vocales de peur. Ce résultat va dans le même sens que ceux obtenus par Scott, Young et al. (1997) qui ont montré une perception déficitaire des patrons prosodiques de peur chez des individus avec lésion bilatérale de l'amygdale. Toutefois, Anderson et Phelps (1998) ont observé une reconnaissance intacte des expressions vocales de peur chez des individus avec lésion bilatérale de l'amygdale.

#### Latéralisation

Le principal intérêt de la majorité des études d'imagerie cérébrale sur le traitement perceptuel de la prosodie émotionnelle a plutôt porté sur la latéralisation cérébrale des réponses aux stimuli émotionnels, c.-à-d. sur la contribution différentielle de l'hémisphère droit et gauche au traitement de l'information prosodique émotionnelle. Plusieurs hypothèses ont été proposées.

#### Hypothèse de la dominance de l'hémisphère droit

Traditionnellement, le traitement de la prosodie émotionnelle est rapporté comme impliquant davantage l'hémisphère droit, indépendamment de la valence de l'émotion (pour une revue, voir Joannette, Goulet et al., 1990; Roland, 1993). Cette hypothèse («the right hemisphere dominance hypothesis») provient essentiellement des études de lésion unilatérale droite (Heilman, Scholes et al., 1975; Tucker et Watson, 1977; Ross, 1981; Ley et Bryden, 1982; Heilman et Bowers, 1984; Borod, Koff et al., 1985; Bowers et Coslett,



1987; Ehlers et Dalby, 1987; Bryan, 1989; Blonder et Bowers, 1991; Lalande, Brawn et al., 1992; Borod, 1993; Darby, 1993; Ross, 1993; Starkstein, Federoff et al., 1994; Dysktra et al., 1995; Ross, Thompson et al., 1997; Schmitt, Hartje et al., 1997; Borod, Cicero et al., 1998; Breitenstein, Daum et al., 1998; Pell, 1999). La plupart de ces études ont rapporté une perception déficitaire des patrons prosodiques émotionnels chez les individus cérébrésés à l'hémisphère droit, alors que ce traitement semble intact chez les individus cérébrésés à l'hémisphère gauche (Starkstein, Federoff et al., 1994). Toutefois, comme nous le verrons plus loin dans cette même section, ce traitement perceptuel peut aussi être compromis chez cette dernière population.

Chez l'individu sain, des études comportementales d'écoute dichotique de stimuli prosodiques (Haggard et Parkinson, 1971; Safer et Leventhal, 1977; Ley et Bryden, 1982; Shipley-Brown et Dingwall, 1988; Herrero et Hillix, 1990; Schmitt, Hartje et al., 1997) et de vocalisations nonlinguistiques (Carmon et Nachshon, 1973) ont montré un avantage de l'oreille gauche pour les stimuli émotionnels, donc en faveur d'une dominance de l'hémisphère droit. Aussi, différentes études d'imagerie cérébrale chez l'individu sain rapportent des données allant aussi dans le sens d'une réponse latéralisée en faveur de l'hémisphère droit dans le traitement des patrons prosodiques émotionnels. Elles ont utilisé différentes techniques: l'IRMf (Buchanan, Lutz et al., 2000; Rama, Martinkauppi et al., 2001; Wildgruber, Pihan et al., 2002; Mitchell, Elliott et al., 2003), la TEP (Zatorre, Evans et al., 1992; Kawashima, Itoh et al., 1993; George, Parekh et al., 1996b), l'électroencéphalogramme (Pihan, Ackermann et al., 1997; Pihan, Altenmuller et al., 2000) et les potentiels évoqués (Kotz, Alter et al., 2000). Dans le cas de la perception des vocalisations nonlinguistiques, les résultats de Sander et Scheich (2001) sont en faveur de l'hypothèse de la dominance de l'hémisphère droit. Ils ont observé une réponse plus forte de l'amygdale droite que de l'amygdale gauche pour les vocalisations émotionnelles (rires et pleurs) comparativement au silence et comparativement à une tâche contrôle de

soustraction de chiffres. Il est à noter que l'amygdale constituait la seule région d'intérêt et sa réponse était bilatérale.

Toutefois, certaines études chez l'individu sain n'ont pas montré de résultats corroborant l'hypothèse de la dominance de l'hémisphère droit. Des études rapportent des réponses bilatérales, même accentuées à gauche, dans le traitement de la prosodie émotionnelle (p. ex., Kotz, Alter et al., 2003). L'hypothèse de l'hémisphère droit n'est pas non plus claire chez la personne cérébrolésée. Même s'il semble que les individus présentant des difficultés à identifier et à discriminer des patrons prosodiques émotionnels montrent plus souvent une lésion à l'hémisphère droit qu'à l'hémisphère gauche (Starkstein, Federoff et al., 1994), des déficits de perception de l'émotion via les patrons prosodiques suite à une lésion à l'hémisphère gauche sont aussi rapportés, lesquels sont généralement associés à une aphasie (Schlanger et Schlanger, 1976; Cancelliere et Kertesz, 1990; Van Lancker et Sidtis, 1992; Pell et Baum, 1997). Certaines études chez les individus cérébrolésés à l'hémisphère gauche proposent même une dominance de l'hémisphère gauche dans le traitement de stimuli prosodiques émotionnels (Seron, van der Kaa et al., 1982; Speedie, Coslett et al., 1984; Pell, 1998).

#### Hypothèse de la valence

En opposition à l'hypothèse de latéralisation, l'hypothèse de la valence propose une dominance de l'hémisphère gauche dans le traitement des émotions positives et une dominance de l'hémisphère droit pour les émotions négatives (Sackeim et Gur, 1979; Davidson et Fox, 1982; Sackeim, Greenberg et al., 1982; Davidson, 1992a,b, 1995; Sobotka, Davidson et al., 1992; Ross, Homan et al., 1994; Pell, 1998; Davidson et Irwin, 1999). Cependant, différentes études ne supportent pas l'hypothèse de la valence. Certaines études d'imagerie cérébrale chez l'individu sain n'ont pas observé de réponse latéralisée dans le traitement de la prosodie émotionnelle selon la valence émotionnelle (George, Parekh et al., 1996b; Buchanan, Lutz et al., 2000; Wildgruber, Pihan et al., 2002; Kotz,

Meyer et al., 2003) et des études chez des individus avec lésion bilatérale présentent des difficultés sévères dans la perception de la prosodie émotionnelle (Cancelliere et Kertesz, 1990; Ross, Thompson et al., 1997).

Des données peuvent supporter les deux hypothèses, celle de l'hémisphère droit et celle de la valence, selon les régions étudiées. L'étude de Sander, Roth et al. (2003) en est un exemple. Ils ont exploré les substrats neuronaux à l'aide d'IRMf des patrons prosodiques de joie et de tristesse (voir tableau 2 pour plus de détails). Dans cette même expérimentation, les auteurs rapportent une activation latéralisée en faveur de l'hypothèse de la dominance de l'hémisphère droit dans la région d'intérêt temporopariétale: la réponse, quoique bilatérale, était plus importante dans l'hémisphère droit, comparativement à celle de l'hémisphère gauche pour les stimuli émotionnels (versus la détection de /a/). Ils rapportent aussi des activations latéralisées en faveur de l'hypothèse de la valence dans la région d'intérêt occipitale: pour les stimuli de tristesse la région occipitale était plus fortement activée dans l'hémisphère droit que dans l'hémisphère gauche, alors que pour les stimuli de joie, la réponse de cette région était plus importante dans l'hémisphère gauche que dans l'hémisphère droit.

#### Hypothèse type d'émotion

Une autre hypothèse propose que les réponses cérébrales pour les émotions primaires (p. ex., la joie, la tristesse) sont latéralisées à droite, alors que celles pour les émotions sociales (p. ex., la culpabilité, la gêne, l'envie) sont plutôt latéralisées à l'hémisphère gauche (Ross, Homan et al., 1994). Toutefois, l'étude des émotions sociales est relativement récente et cette hypothèse ne semble pas avoir été spécifiquement testée (pour plus de détails sur les émotions sociales, voir Eisenberg, 2000; Greene, Sommerville et al., 2001; Moll, Oliveira-Souza et al., 2002).

### Latéralisation en fonction des paramètres acoustiques

Alternativement, il a été suggéré que la latéralisation du traitement prosodique varie en fonction des paramètres acoustiques prosodiques. La plupart des études conceptualisent la prosodie émotionnelle comme étant une entité distincte plutôt qu'une manipulation des paramètres prosodiques pour exprimer une émotion. Toutefois, l'importance de la contribution des paramètres acoustiques dans la perception des stimuli prosodiques émotionnels doit être prise en compte dans les études. Comme il a été mentionné à la section 1.3 *Méthode pour contrôler les limitations des stimuli prosodiques*, certains paramètres acoustiques jouent un rôle crucial dans la perception de la prosodie émotionnelle, particulièrement la F0, laquelle est l'information acoustique la plus utilisée pour identifier l'émotion d'un patron prosodique.

Les indices acoustiques utilisés pour reconnaître l'émotion via la prosodie sont eux-mêmes traités par des systèmes qui impliquent l'un ou l'autre hémisphère. Selon le modèle de Van Lancker et Sidtis (1992), les hémisphères gauche et droit sont responsables du traitement de différents aspects de la prosodie émotionnelle: l'hémisphère droit est supérieur au gauche dans l'extraction de l'information de la F0, alors que l'hémisphère gauche est plus important que le droit dans l'extraction de l'information temporelle. Par exemple, plusieurs chercheurs (Zatorre et Samson, 1991; Zatorre, Evans et al., 1992; Tzourio, Massiou et al., 1997; Zatorre et Belin, 2001; Meyer, Alter et al., 2002; Zatorre, Belin et al., 2002) ont rapporté que la F0 est traitée par l'hémisphère droit, alors que d'autres (Carmon et Nachshon, 1973; Robin, Tranel et al., 1990; Zatorre, 1988; Zatorre, Evans et al., 1992; Zatorre, Belin et al., 2002) ont observé que les indices temporeux sont traités dans l'hémisphère gauche. Il semble aussi que certains paramètres acoustiques ne soient pas impliqués dans un traitement latéralisé. Par exemple, dans le cas de l'étude de Wildgruber, Pihan et al. (2002), la latéralisation en faveur de l'hémisphère droit observée ne s'explique pas par la structure acoustique des stimuli (étendue de la F0 et la durée). On connaît encore insuffisamment les régions cérébrales impliquées dans ces traitements et si



ces paramètres sont traités différemment indépendamment du type d'information que ces paramètres apportent. Par exemple, il a été proposé que ces patrons de latéralisation sont indépendants du type de prosodie, émotionnelle ou linguistique (Zatorre, 1988; Ouellette et Baum, 1993). Il est aussi fort possible qu'un même paramètre acoustique soit traité par différentes régions d'un hémisphère ou de l'autre, selon l'aspect traité de ce paramètre. Par exemple, selon Ladd (1996), le lobe temporal latéral droit traite la direction du ton, mais peut ne pas traiter l'étendue et la hauteur du ton. Aussi l'oreille gauche semble montrer un avantage dans le traitement des basses fréquences, tel la F0, comparativement aux hautes fréquences (Ivry et Leiby, 1993, mais voir aussi Wolf, 1977).

L'hypothèse de la latéralisation en fonction des paramètres acoustiques peut s'inscrire dans les trois hypothèses précédemment présentées. Certains indices prosodiques utilisés pour percevoir les états émotionnels sont relativement marqués et stables (Williams et Stevens, 1972; Scherer, 1981a; Cosmides, 1983; Frick, 1985; Ladd, Silverman et al., 1985; Wallbott et Scherer, 1986; Murray et Arnott, 1993; Pittam et Scherer, 1993; Bachorowski et Owren, 1995; Wildgruber, Pihan et al., 2002). Il est possible de faire des liens entre les changements physiologiques qui sous-tendent certains états émotionnels et certains changements de paramètres acoustiques (Scherer et Oshinsky, 1977; Scherer, 1986). Le tableau 4 présente les paramètres acoustiques les plus souvent associés à certaines émotions. Donc, que l'hémisphère droit ou gauche soit davantage sollicité dans le traitement d'une émotion peut en partie être expliqué par le fait que certains indices acoustiques jouent un rôle plus important que d'autres dans la perception d'une émotion donnée et par le fait que le traitement de certains indices et/ou certains aspects de ces indices acoustiques soient latéralisés.

Tableau 4. Exemples de paramètres acoustiques des patrons prosodiques associés à l'expression d'émotions (tiré des études de Scherer et Oshinsky, 1977 et Scherer, 1986)

paramètres acoustiques	direction de l'effet	émotions						
		joie	peur	colère	tristesse	dégoût	ennui	surprise
F0: variation	faible		x	x	x	x	x	
	grande	x						x
F0: contour	montant	x	x			x		x
	descendant			x	x		x	
F0: niveau	haut	x	x	x				x
	bas				x	x	x	
tempo	rapide	x	x	x				x
	lent				x	x	x	
enveloppe	abrupte	x						x
	arrondie		x	x	x	x	x	
harmoni- ques	peu	x			x		x	
	plusieurs		x	x		x		x

En résumé, plusieurs résultats suggèrent que la perception de l'émotion vocale semble être une fonction dominante de l'hémisphère droit, particulièrement dans les régions frontale et temporale, cependant son degré de latéralisation et la contribution de l'hémisphère gauche ne sont pas encore établis.

La perception de l'émotion vocale nécessite le traitement de plusieurs composantes: acoustiques, sémantiques, attentionnelles, etc. Ces composantes impliquent elles-mêmes différentes structures et réseaux neuronaux distribués entre les deux hémisphères. En dépit de la nature distribuée du traitement, l'hémisphère droit, en particulier les régions frontales inférieures, semble être une composante critique du système, lequel travaille en



collaboration avec des régions plus postérieures de l'hémisphère droit, des régions frontales gauches et des structures sous-corticales.

Une réponse en faveur de l'hémisphère droit est déjà observée au niveau du traitement de la voix, indépendamment de la nature des stimuli vocaux. La réponse neuronale du cortex auditif du traitement perceptuel de la voix humaine est plus importante dans l'hémisphère droit, lorsqu'on la compare à celle du traitement de stimuli nonvocaux (Belin, Zatorre et al., 2000).

Les qualités émotionnelles des patrons prosodiques émotionnels exigent l'implication d'autres régions que celles uniquement associées au traitement des stimuli auditifs. Le traitement prosodique est composé de plusieurs tâches et fonctions, lesquelles sont elles-mêmes distribuées entre les deux hémisphères. Ainsi, le patron de latéralisation observé peut, jusqu'à un certain point, dépendre de la façon dont le traitement de la prosodie émotionnelle est évalué, c.-à-d. sur les composantes de la prosodie émotionnelle accentuées par ce qu'exige la tâche. Cependant, on connaît encore peu les différences de réponses cérébrales selon le type de traitement effectué. Par exemple, l'étude de Peper et Irle (1997) a montré que les régions frontales seraient impliquées dans le décodage de la valence de la prosodie émotionnelle, les régions dorsolatérales, pariétales et temporales dans l'étiquetage de la catégorie et dans l'appariement, alors que l'aspect d'activation («arousal») impliquerait l'ensemble de ces régions de l'hémisphère droit. Wildgruber, Pihan et al. (2002), quant à eux, proposent que l'induction des émotions implique une réponse latéralisée des stimuli prosodiques, alors qu'une réponse bilatérale serait observée dans le traitement perceptuel.

L'activation de certaines des régions impliquées dans le traitement de la prosodie émotionnelle est reliée à l'identification et à l'interprétation des expressions émotionnelles et/ou à l'engagement des processus de mémoire (Zatorre, Evans et al., 1992; Zatorre, Evans

et al., 1994; Imaizumi, Mori et al., 1997; Pihan, Ackermann et al., 1997; Pihan, Altenmüller et al., 2000). L'importance de ces composantes nécessaires pour effectuer la tâche est à considérer dans les asymétries des réponses cérébrales. Il a été suggéré que les effets de modulation de l'attention dans l'activation du cortex auditif impliquent une plus grande susceptibilité du cortex auditif droit comparativement au gauche (O'Leary, Johnson-Laird et al., 1996; Pugh, Offywitz et al., 1996; Grady, Van Meter et al., 1997).

L'information sur laquelle les participants portent attention diffère selon le type de tâche exécutée, donc le type de tâche utilisé lors de la mesure cérébrale peut biaiser les réponses neuronales et contribuer aux asymétries des réponses cérébrales dans le traitement de la prosodie émotionnelle. Plusieurs études ont utilisé une tâche d'identification du sexe pour explorer les réponses neuronales dans le traitement de stimuli émotionnels (p. ex., Phillips, Young et al., 1997; Phillips, Young et al., 1998.). Ces études suggèrent que les réponses neuronales aux stimuli émotionnels ne sont pas dépendantes des instructions explicites pour porter attention au contenu émotionnel des stimuli. Toutefois, d'autres études ont rapporté que la tâche d'identification de sexe (et même les tâches de classification) biaise les réponses cérébrales en faveur de l'hémisphère gauche. Des études comportementales (Jones, 1979; Jones et Anuza, 1982) et d'imagerie fonctionnelle (Morris, Frith et al., 1996) ont montré que classer des visages selon le sexe implique une forte dominance de l'hémisphère gauche, plus particulièrement chez les participants hommes, et ce même si l'hémisphère droit est davantage impliqué dans la discrimination de visages. En audition, aucune expérimentation jusqu'à maintenant semble avoir eu comme objet d'étude l'influence de la tâche d'identification du sexe sur les réponses neuronales dans la perception des émotions vocales.

La latéralisation de la prosodie émotionnelle semble également varier en fonction de la demande de la tâche chez l'individu sain (Behrens, 1985; Kotz, Alter et al., 2000; Pihan, Altenmüller et al., 2000) et chez l'individu cérébrolésé (Tompkins et Flowers, 1985). Il est

suggéré que l'activation de l'hémisphère gauche (frontal et temporal) corrèle avec l'augmentation des demandes cognitives lorsque les tâches exigent des traitements complexes.

#### Remarques méthodologiques

Il est difficile de comparer les réponses cérébrales et leur degré de latéralisation entre les différentes études. Les choix méthodologiques des études sont en partie responsables des différences entre les résultats. Plusieurs variables sont cruciales dans la mesure de la perception de l'émotion, par exemple le type de stimuli, la tâche utilisée, la population, l'approche (par bloc vs. événementiel). Ross, Thompson et al. (1997), Pell (1998) et Baum et Pell (1999) critiquent essentiellement l'absence de contrôle des variables cruciales dans le traitement de la prosodie émotionnelle, par exemple la prosodie linguistique. Les travaux de Weintraub, Mesulam et al. (1981) et de Brådvik, Dravins et al. (1991) ont même suggéré que seule la prosodie linguistique est traitée dans l'hémisphère droit (études de lésion unilatérale à l'hémisphère droit). Il est donc essentiel de bien isoler ces deux types de prosodie afin d'étudier les régions impliquées dans le traitement des patrons prosodiques émotionnels. Comme il a été mentionné à la section *1.1 Limitations de l'utilisation des stimuli prosodiques*, la sémantique peut constituer une variable confondante dans le traitement de la prosodie émotionnelle. L'information sémantique influence en effet la réponse cérébrale et elle impliquerait davantage l'hémisphère gauche (Vikingstad, George et al., 2000). Une étude récente, utilisant un filtre acoustique, a rapporté des activations différentes pour les mêmes stimuli prosodiques, selon qu'ils aient été filtrés ou non. Kotz, Meyer et al. (2003) ont observé une réponse du gyrus frontal inférieur pour les émotions positives et négatives exprimées par les stimuli filtrés (seule l'information prosodique était disponible), alors que pour les stimuli non filtrés (avec information sémantique), cette région répondait seulement pour les émotions positives. Des études de ce type, manipulant des filtres acoustiques, participeront certainement à une meilleure compréhension des réponses latéralisées du traitement perceptuel de la prosodie émotionnelle.

Les choix méthodologiques peuvent mener à l'exploration d'autres traitements que perceptuel, tel l'induction d'émotion. Les stimuli émotionnels peuvent facilement et rapidement induire des états émotionnels aux participants, particulièrement lorsqu'une approche par bloc est utilisée ou lorsque la durée des stimuli est importante. Dans ces cas, on observe une activation du système autonome. Les états émotionnels induits sont associés à des changements dans la gestuelle, l'expression faciale et d'autres indices physiologiques, tels la pression artérielle et la température corporelle. Les frontières entre la perception et l'induction d'une émotion demeurent à être étudiées. Ces frontières inconscientes (c.-à-d. l'individu n'exprime pas de changement d'état émotionnel mais une différence d'activité cérébrale est observée) ou conscientes (c.-à-d. l'individu observera un changement d'état émotionnel, lequel est associé à des changements physiologiques) sont d'autant plus différentes d'un individu à l'autre.

Comme mentionné au début de l'introduction de thèse, le traitement perceptuel constituant le principalement objet exploré dans le cadre de ce travail, seules les principales études de perception de l'émotion vocale sont présentées. Toutefois, il sera de plus en plus important d'étudier différents traitements cognitifs à l'intérieur d'une même étude. Il semble que dans certains cas, l'identification de l'émotion et le jugement hédonique sont en *opposition*. Dans Wallbott et Scherer (1986), même si les participants identifiaient correctement l'émotion de tristesse dans des phrases neutres sémantiquement, ils jugeaient ces stimuli comme étant plaisants. Ces dichotomies demeurent à être explorées par des expérimentations comportementales, électrophysiologiques et de neuroimagerie fonctionnelle.

En bref, l'émotion vocale a principalement été représentée par la prosodie émotionnelle. L'étude de la prosodie ne peut à elle seule caractériser le traitement de l'émotion vocale. Les vocalisations nonlinguistiques apportent aussi de l'information importante quant au



traitement de l'émotion vocale, tant au niveau des substrats neuronaux sous-jacents qu'au niveau comportemental. La seconde section de l'introduction constitue la problématique de la thèse. Plus spécifiquement, elle présente les questions auxquelles les études composant la thèse tentent de répondre.

## **2. Problématique**

Dans les interactions sociales chez l'humain et le primate nonhumain, un contrôle («monitoring») continu des expressions vocales est d'une importance fondamentale pour une adéquation des ajustements comportementaux (Bertrand, 1969; Chevalier-Sokolnikoff, 1974). L'expression et l'interprétation de l'affect vocal s'inscrivent dans les comportements interactifs et appartiennent aux caractéristiques fondamentales de la cognition sociale de notre espèce. Ainsi, une perception adéquate de l'émotion vocale, principalement exprimée par la prosodie et par les vocalisations nonlinguistiques, constitue un élément essentiel dans la régulation émotionnelle et dans les interactions sociales. Comme mentionné plus haut, alors que la perception de l'émotion via la prosodie a été grandement étudiée par la communauté scientifique des neurosciences, peu de chercheurs se sont penchés sur l'émotion exprimée vocalement qui ne porte pas d'information linguistique. La prosodie émotionnelle n'est pas la seule représentante de l'émotion vocale, les différences entre prosodie et vocalisations nonlinguistiques, discutées antérieurement, sont fondamentales. Les vocalisations nonlinguistiques sont moins susceptibles d'être influencées par le langage et la culture que les stimuli émotionnels porteurs d'information linguistiques ou pseudo-linguistiques. Les vocalisations nonlinguistiques représentent une façon universelle d'exprimer nos émotions et de percevoir les émotions qu'exprime autrui, allant bien au-delà des codes linguistiques.

Il est nécessaire d'étudier la perception des vocalisations nonlinguistiques. Les résultats des recherches rapportés dans la littérature ne sont pas concluants, principalement parce que peu d'études abordent cette question. La perception des vocalisations nonlinguistiques

représente un aspect nécessaire pour un développement psychologique sain. Par exemple, le pleur du bébé constitue la première vocalisation utilisée pour interagir avec ses pairs: la vocalisation du pleur est le premier outil vocal utilisé entre le bébé et ses parents. Le pleur constitue la plus puissante et précoce forme de communication qui représente un système de signaux sociaux, lequel guide le comportement des parents (Drummond, McBride et al., 1993). Au cours du développement, le pleur devient presque exclusivement un moyen d'expression de ses émotions, le langage étant devenu l'outil de communication prioritaire.

### *But de la thèse*

Cette thèse est composée de quatre études ayant le but commun d'explorer le traitement des vocalisations nonlinguistiques chez des volontaires adultes sains. Les performances comportementales et les substrats neuronaux des processus de perception et d'amorçage sont examinés. La discussion générale résume les quatre expérimentations qui compose la thèse et discute de leurs limites, ainsi que des modèles théoriques sur la perception de la voix et des émotions.

### Première étude

L'ensemble des vocalisations sans information linguistique est principalement composé de vocalisations porteuses d'information émotionnelle. Même si dans l'ontogenèse, ces vocalisations représentent un important moyen de communication, nous connaissons toujours très peu notre habileté à juger leur contenu émotionnel. Il ne semble pas exister d'études qui ont exploré l'habileté des individus sains à percevoir les vocalisations nonlinguistiques, possiblement parce que cette habileté nous paraît trop évidente. Toutefois, il est important d'explorer notre accord inter-juge de la perception de ces stimuli. Par exemple, jusqu'à quel point notre jugement de la valence de l'émotion (positive / négative) est-il comparable d'un individu à l'autre? Qu'en est-il des différentes dimensions émotionnelles, telles l'intensité et l'authenticité de l'émotion exprimée? Le jugement de ces dimensions (la valence, l'intensité et l'authenticité), sont-ils différents selon le groupe



d'âge ou selon le sexe des individus? Plusieurs études, examinant la perception des expressions faciales et de la prosodie émotionnelle, ont rapporté que la perception des émotions était modulée selon le sexe et l'âge des individus. Il est essentiel d'explorer si ces différences sont réservées à la modalité visuelle et linguistique, ou bien si elles s'étendent au domaine de l'émotion vocale sans information linguistique. Le but de cette première étude est donc d'explorer spécifiquement les jugements de valence, d'intensité et d'authenticité de l'émotion exprimée par des vocalisations nonlinguistiques chez des femmes et des hommes jeunes adultes et d'âge moyen.

### Seconde étude

Un autre traitement important d'étudier est celui de l'amorçage, un type de mémoire implicite. L'effet d'amorçage indique qu'une expérience passée (p. ex., présentation préalable de stimuli) influence la perception de stimuli étudiés. Cet effet permet donc entre autres d'identifier les caractéristiques d'un stimulus qui ont une influence sur la perception de stimuli étudiés. Dans le cadre de cette seconde expérimentation, nous nous sommes demandés s'il existait un effet d'amorçage auditif général pour les vocalisations nonlinguistiques. Si tel est le cas, cet amorçage, est-il modulé selon la valence de l'émotion et / ou, selon le sexe des locuteurs? Le but de cette seconde étude est d'examiner l'amorçage de vocalisations émotionnelles et neutres produites par des femmes et des hommes chez un groupe de participants d'hommes.

### Troisième étude

Jusqu'à maintenant, les expérimentations ont étudié différents traitements perceptuels des vocalisations nonlinguistiques au niveau comportemental. Il est aussi nécessaire d'examiner les substrats neuronaux de la perception des vocalisations. Une des questions qui est nécessaire d'explorer avant d'étudier des aspects spécifiques de la perception des vocalisations nonlinguistiques est celle de la voix humaine. Comment notre cerveau traite-t-il la voix humaine? Notre cerveau traite continuellement notre environnement sonore, mais sa

réponse cérébrale à la voix humaine est-elle spéciale comparativement à celle d'autres stimuli auditifs? Des études suggèrent que certaines régions sont impliquées dans le traitement de la voix, mais cette réponse cérébrale, est-elle sélective à la voix humaine? Il est tout à fait légitime et important de se poser la question si ces régions ne répondent pas tout simplement à la voix animale compte tenu de l'importance que prennent les vocalisations humaines dans la phylogenèse et des similitudes entre la voix humaine et celle animale. Le but de cette étude est d'identifier les substrats neuronaux sous-jacents à la perception de la voix humaine (vocalisations nonlinguistiques et langage) et d'examiner si ces mêmes régions répondent également aux vocalisations animales (p. ex., aboiements, miaulements).

#### Quatrième étude

L'expérimentation précédente a pour but d'étudier les substrats neuronaux sous-jacents à la perception de la voix humaine. Il paraît essentiel d'aller un peu plus loin dans l'exploration des régions cérébrales impliquées à la perception de la voix humaine et examiner si différentes régions répondent à des vocalisations spécifiques. Le but de la dernière étude est donc d'identifier les réponses cérébrales associées à la perception de vocalisations émotionnelles (rire, vocalisation sexuelle, pleur, cri de peur) et vocalisation non-émotionnelles (p. ex., toux).

**CHAPITRE 2**

**ARTICLE 1**

**JUGEMENT DE DIMENSIONS ÉMOTIONNELLES  
DES VOCALISATIONS NONLINGUISTIQUES:  
DIFFÉRENCE SELON L'ÂGE**

**ARTICLE 1**

**JUDGMENT OF EMOTIONAL NONLINGUISTIC  
VOCALIZATIONS: AGE-RELATED DIFFERENCES**

**Shirley FECTEAU, Jorge L. ARMONY,**

**Yves JOANETTE & Pascal BELIN**

*Applied Neuropsychology* (sous-presse)

Judgment of emotional nonlinguistic vocalizations: Age-related differences

Shirley Fecteau<sup>1,2,3</sup>, Jorge L. Armony<sup>3,4</sup>, Yves Joanette<sup>2,3</sup>, Pascal Belin<sup>2,3</sup>

<sup>1</sup> Département de psychiatrie, Université de Montréal, Montréal, Canada

<sup>2</sup> Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montréal, Canada

<sup>3</sup> Centre de recherche en neuropsychologie expérimentale et cognition (CERNEC), Université de Montréal, Montréal, Canada

<sup>4</sup> Douglas Hospital Research Centre, McGill University, Montréal, Canada

Correspondence should be addressed to Shirley Fecteau,  
Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville,  
Montréal, Qué., H3C 3J7, Canada

E-mail: shirley.fecteau@umontreal.ca; Telephone number: 514 343-2330

### Abstract

Humans make extensive use of vocal information to attribute emotional states to other individuals. To date, most studies exploring perception of vocal emotions have done so in the context of speech prosody, even though nonlinguistic emotional vocalizations represent an important, perhaps more universal means to express emotions. Here, we explored the perception of emotional nonlinguistic vocalizations in healthy individuals, with an emphasis on potential age- and sex-related differences. Sixty participants rated 563 positive (laughs, sexual vocalizations), negative (cries, screams of fear), and neutral vocalizations (e.g., coughs), according to the valence, intensity and authenticity of the emotion expressed. Ratings were consistent among individuals, suggesting that valence is an adequate measure of emotional categorization. An important effect of age emerged: age by vocalization category interactions were observed for both valence and intensity ratings: younger participants rated stimuli as more emotional than older individuals (i.e., higher valence for positive, lower for negative, and more intense for both positive and negative). We also found a sex effect in the authenticity ratings: older women rated the vocalizations as less authentic than younger women whereas authenticity judgments did not differ between the two age groups in men. Taken together, these findings suggest that, as previously observed for facial expressions and prosody, the judgments of emotional vocalizations may vary with age.



## Introduction

Processing nonlinguistic vocalizations plays an essential role in maintaining successful relationships and healthy psychological functioning. Nonlinguistic vocalizations are a primary means of communication, as they appear early in the ontogenesis: newborns express and understand emotions vocally long before learning language (e.g., Barr, 2000). From a phylogenetic point of view, they represent one of the most important means to communicate for most primates, especially to express their affective state (Rendall, 2003; Goodall, 1986; Masataka, 1988; Seyfarth, 1980; Gouzoules, 1984; Fichtel, 2003; Green, 1975). Moreover, emotional vocalizations of primates share fundamental attributes, such as the motivation of communicating an emotion. A striking example is the similar way primate infants, including human babies, use vocalizations of crying to communicate with their mother (Newman, 1985; Bard, 1996).

Although processing emotion through nonlinguistic vocalizations is crucial for adequate social interactions, it has received little attention. Most work on emotion used prosody and facial expressions. Among the findings of these studies of emotional processing are reports of age- and sex-related differences. For instance, older adults make more errors in indicating the emotion expressed through prosody (Kiss, 2001), as well as through facial expressions (Phillips, 2002). Also, women and men seem to perceive facial expressions differently: young women rate male faces more positively than young men (Van Strien, 2000), and men identify neutral faces as being angry more often than do women (Crews, 1994). However, no such sex-related differences have been reported for the perception of emotional prosody.

In this study, we tested whether these previous findings could be extended to the domain of nonlinguistic vocalizations. Therefore, we asked healthy individuals to judge three separate emotional dimensions, valence, intensity, and authenticity, conveyed by a large set of nonlinguistic vocalizations.

## Materials and method

### *Participants*

Sixty francophone adults divided into four equal-number groups according to age and sex (young women, mean age = 22.5 years, range of 19-29; young men, mean age = 23.7 years old, range of 19-30; middle-aged women, mean age = 54.6, range of 42-72; middle-aged men, mean age = 54.5, range of 44-67) took part in this study. They were exempt of neurological or psychiatric conditions and they were not taking anti-depressants. All participants had normal ( $n=59$ ) or corrected-to-normal audition ( $n=1$ ). The means of years of education were not statistically different between the four groups (15.3, 16.1, 16.5, 15.3 years, respectively).

### *Preselection of Stimuli*

An original set of 1007 nonlinguistic vocalizations from diverse sound libraries was gathered. Stimuli consisted of laughs, cries, vocalizations of sexual pleasure, and screams of fear, as well as neutral vocalizations such as coughs, throat clearing, etc. A group of twelve judges (four equal-number groups according to age and gender, mean age of 20 years old for young men and women; mean age of 56 years old for middle-aged women and men) were asked to decide, using a five-alternative forced-choice task, the type of vocalization (laugh, cry, sexual vocalization, scream of fear, or "other sounds"), and using a visual analog scale (1-100 points), the intensity of the emotion expressed.

This step allowed us to assign an emotional category to each vocalization (i.e., positive, neutral, or negative) independently from the ratings provided by the experimental group. Additionally, it permitted a preselection of stimuli in order to exclude ambiguous stimuli: inclusion of stimuli required at least seven judges to agree on one category while the reminder judges had to agree on a second category (i.e., stimuli that were assigned to more than two categories were discarded). The final set of stimuli consisted of 563 vocalizations

(165 laughs, 46 cries, 159 sexual vocalizations, 62 screams of fear, and 123 neutral vocalizations).

In order to study the perception of emotion in nonlinguistic vocalizations, stimuli were divided into three emotional categories: positive (laughs and sexual vocalizations), negative (cries and screams of fear) and neutral (e.g., coughs, clearing throat). These neutral sounds consisted of the sounds classified in the “other sounds” category. Statistical comparisons (t-tests) confirmed significant differences in intensity ratings between positive and neutral stimuli ( $p < 0.01$ ), and between negative and neutral stimuli ( $p < 0.01$ ).

#### *Experimental Task*

Sixty participants were instructed to rate the preselected 563 stimuli on three major dimensions of emotions using visual analog scales (1-100 points): valence (*very negative* to *very positive*), intensity (*not at all intense* to *extremely intense*), and authenticity (*not at all authentic* to *extremely authentic*) of the emotion expressed by the speaker. The instructions were: “What is your *judgment*? For the valence scale, is the expressed emotion negative or positive? For the intensity scale, is the expressed emotion very or not at all intense? And for the authenticity scale, is the expressed emotion very or not at all authentic?”. In the case of valence, if there was no emotion perceived (i.e., “neutral” sounds), participants were told to place their mark in the middle of the scale. Stimuli were equalized on energy (RMS) with Mitsyn (WLH) and CoolEdit Pro (Syntrillium Software Corporation). They were pseudo-randomly presented through headphones in a soundproof environment using Media Control Function (DigiVox). The experiment was divided in three blocks of around one hour in one single session.

#### *Data Processing*

The mean score of ratings for each emotional category was calculated for each participant, separately for the three judgments, and entered into statistical analyses.

## Results

A repeated measures ANOVA with one within-subjects factor, the emotional category (negative, positive, neutral), and two between-subjects factors: age (young and middle-aged) and sex was performed independently for each of the three judgments.

1. Valence ratings: we found a main effect of emotional category ( $p < 0.0001$ ; Greenhouse-Geisser correction for non-sphericity): post-hoc analysis (Tukey) showed that all groups rated stimuli consistently with their assigned category; that is, stimuli in the positive category (laughs, sexual vocalizations) were rated with the highest valence (mean value = 71.3), negative vocalization received the lowest ratings (mean value = 21.8), and valence ratings for neutral sounds were in the middle of the scale (mean value = 58.4). We also observed an emotional category by age interaction ( $p < 0.05$ ): post-hoc analysis revealed that the younger group rated negative stimuli more negative, and positive stimuli more positive than the older group, with no differences between groups for neutral vocalizations (see figure 1). No sex-related differences were observed.

Insert figure 1 about here

2. Intensity ratings: a main effect of emotional category ( $p < .0001$ ) was observed: post-hoc analysis (Tukey) showed that negative stimuli were rated more intense than the positive and neutral ones, and positive stimuli were rated more intense than the neutral ones. There was also a significant emotional category by age interaction ( $p < .0001$ ): the younger group rated negative and positive stimuli as more intense than the older group, while there was no difference between groups for neutral vocalizations (see figure 2). As with valence ratings, we did not find any sex-related differences in intensity ratings.

Insert figure 2 about here

3. Authenticity ratings: a main effect of emotional category ( $p < .0001$ ) was observed: post-hoc analysis (Tukey) revealed that neutral stimuli were rated more authentic than the positive ones, and positive stimuli were rated more authentic than the negative ones (see figure 3). We also found an age by sex interaction ( $p = .036$ ): older women rated all stimuli as less authentic than younger women, while there was no difference across male groups (see figure 4).

Insert figure 3 about here

Insert figure 4 about here

## Discussion

The main purpose of this study was to explore the influence of age and sex on three dimensions of perceived emotion in nonlinguistic vocalizations: valence, intensity, and authenticity. Overall, we found significant differences across groups in the emotional ratings for vocalizations expressing positive (happiness, sexual pleasure) and negative (fear, sadness) emotions, whereas neutral vocalizations were rated similarly.

### *Consistency of Valence Perception*

One important result of this study is that valence ratings of nonlinguistic vocalizations were found to be consistent across individuals. Regardless of group, individuals rated positive stimuli with a higher valence than neutral stimuli, and the neutral ones with a higher valence than the negative vocalizations (see figure 1). Thus, the assigned emotional category of our stimuli, captured by a continuous, bipolar dimension (positive-negative), was appropriate and consistent across all four groups of participants. This finding suggests that valence is an adequate measure of emotional categorization, consistent with previous studies using prosody (Bachorowski, 1999) and facial expressions (Russell, 1986; Russell,



1997). In addition, these findings confirmed that all groups understood the task and used the scales appropriately.

#### *Age differences in emotional perception*

A key finding of this study was that younger and older participants rated differently the perceived emotional valence and intensity of the nonlinguistic vocalizations. For these two ratings, younger participants rated both positive and negative stimuli by using a more extended range of the scale than the older participants (see figures 1 and 2).

Age groups seemed to rate emotions differently in terms of valence: younger adults judged positive stimuli as more positive and negative ones as more negative than older adults. That is, older individuals rated positive and negative vocalizations as less “emotional”. In terms of intensity perception, young adults rated positive and negative stimuli as more intense than older adults. Importantly, whereas differences of intensity ratings between emotional categories in the young adults’ judgments were observed, older participants rated all stimuli as equally intense: neutral stimuli were rated as intense as the negative and the positive ones.

These findings are consistent with previous studies suggesting age differences in processing emotions. Behavioral studies reported that older adults identify and discriminate emotions less accurately than younger adults through prosodic information (Kiss, 2001; Oscar-Berman, 1990; Allen, 1993), lexical emotional information (Grunwald, 1999), and facial expressions (Phillips, 2002; Oscar-Berman, 1990; Gunning-Dixon, 2003), especially for negative emotions (Moreno, 1993; McDowell, 1994; MacPherson, 2002; Phillips, 2002). An age decline in intensity rating in facial and prosodic information has also been reported (Thompson, 2001). Finally, older adults make more mistakes at perceiving emotional body cues (Montepare, 1999).



Neuroimaging studies suggest that the effect of aging on emotion processing may be due to an age-related reorganization of the cortical networks used for emotional perception. For example, old and young adults activate different neural networks during the perception of facial expressions (Gunning-Dixon, 2003). Other studies suggest an age-related reduction in the medial temporal activation in emotional perception using faces (Iidaka, 2002; Grady, 1995), and pictures (Grady, 1999), as well as in frontal lobe activation in response to facial expressions (Gunning-Dixon, 2003). These age-related differences in neural activity may not be emotion-specific, but rather due to general functional and structural alterations in the temporal and frontal regions. Activity in these regions appears to be negatively correlated with aging as shown by imaging (Martin, 1991; Murphy, 1996; Petit-Taboue, 1998; Esposito, 1999, for a review see Raz, 2000), and molecular studies (Bliss, 1973; Bliss, 1993; Morris, 1986; Meunier, 1993). Structural alterations in these regions have also been reported in neuroimaging (Albert, 1996; Convit, 1997; Fama, 1997; Golomb, 1993; Jack, 1997; Jernigan, 1991), post-mortem (Davis, 1999), and neurobiology studies (West, 1990; West, 1993a; West, 1993; for reviews, see Peters, 1981; Curcio, 1982; Landfield, 1984; Landfield, 1986; Adams, 1987; Coleman, 1987; Detoledo-Morrell, 1988a; Detoledo-Morrell, 1988b; Detoledo-Morrell, 1993; Flood, 1988; Haug, 1984; Kemper, 1994). Overall, age seems to have an important effect on the neural substrates of emotion processing. In particular, the observed reduction in temporal and frontal activation may underlie the difficulties in emotion processing that have been documented in the elderly. Our results suggest that the previously reported age-related differences in the processing of facial expressions also occur within the auditory modality with nonlinguistic vocalizations. The neural correlates of the age-related difference in perception of emotional nonlinguistic vocalizations found in this study have yet to be explored using neuroimaging techniques.

Emotional regulation needs also to be discussed, as what participants exactly do to infer the sender's emotional state is still unknown. One could argue that the observed age-related differences in rating emotional nonlinguistic vocalizations, that is, older participants rating

stimuli as less emotional, could be interpreted as consistent with the hypothesized increase in self-regulatory capacity associated with aging (Gross, 1997). Indeed, emotion-focused coping and understanding seem to increase with age (Diehl, 1996; Folkman, 1987). Older adults' self-report of emotional reactivity appears to decrease across life span (Lawton, 1992): they report less physiological responsiveness during emotional experiences (Gross, 1997; Lawton, 1992), and less consistent emotional expressivity (Lawton, 1992; Gross, 1997). Physiological measures support these findings: cardiovascular reactivity associated with emotional responding is reduced (Levenson, 1991; Levenson, 1994), and skin conductance to emotionally arousing stimuli is lower in older adults (Denburg, 2003).

Another possible explanation of the age-related difference observed in this study could be that older individuals use rating analog scales differently than younger people. This, however, seems unlikely given that the authenticity scale was used similarly by the two age groups (see figure 3). Finally, the wider range of age for the older group (42-72 years) compared to the younger group (19-30 years) could have an effect on age differences found in this study. This did not seem to be the case, as similar results were obtained in a separate analysis where age was introduced as a parametric covariate.

#### *Age-Gender Interaction in Authenticity Perception*

Age influenced not only ratings of valence and intensity, but of authenticity as well. However, this effect of age on authenticity ratings interacted with the sex of participants. Older women rated all stimuli as less authentic than younger women (see figure 4). Further experiments are required before any interpretation of this finding can be proposed.

#### *Sex Differences in Emotional Perception*

In this study, sex did not seem to have a significant influence on the ratings of perceived emotion. Only the above-mentioned interaction age by sex on the authenticity judgment was observed.

In general, studies on sex difference, using categorical judgment, suggest that women and men perceive facial expressions differently, especially in valence ratings. For example, Crews and Harrison (1994) reported that men identify photographs of neutral faces as being angry more often than do women. Also, Van Strien and Van Beek (2000) found, using a two-alternative forced-choice, that young women tended to rate male cartoon faces more positively than young men (however, as the authors suggested, this sex difference could have been due to the fact that only used male faces were used). Some studies have reported that women seem to be more accurate than men at identifying facial expression using hemi-field presentation method (Harrison, 1990; Adolphs, 2001). Interestingly, no such sex-related differences using prosodic stimuli have been reported in the literature. Men, however, seem to be slower than women to process emotional prosody. Schirmer et al. (2002) reported sex-related difference in the time course of the relationship between emotional prosody and word recognition (see also Besson, 2002 for a discussion). However, future work is required, isolating perception of emotional prosody the word recognition, to elucidate whether processing emotional prosody is different between men and women.

#### *Methodological considerations*

There are several important methodological differences between previous studies exploring age and sex effects of emotional perception and ours. For example, we did not compare performances but ratings on continuous scales. Thus, there were no correct or incorrect answers. Most previous studies have used forced-choice labeling, i.e. to present an emotional stimulus to a participant, who has then to indicate which emotion it signals among a limited choice of emotions. These studies usually identify as correct or incorrect the attribution of an emotional category (e.g., happiness, sadness) to the given stimulus. This method limits the measurement of ability at perceiving the emotion expressed to a

categorical choice (Frank, 1993; Pittam, 1993), which perhaps does not capture the complexity of emotional processing. This can be avoided by the use of continuous scales.

It is also important to mention that we measured the participant's response as an ability to infer emotional state of the sender. Participants had to rate what they thought the speaker was expressing, using three bipolar dimensions (valence: positive-negative; intensity: very intense-not at all intense; authenticity: very authentic-not at all authentic). However, we do not know what exactly the participants did to rate the emotional nonlinguistic vocalizations. According to Russell et coll. (Russell, 2003; p.334) "A receiver's typical response might include much more than attribution of an emotion or perhaps no such attribution at all. One well-supported possibility is that the receiver perceives the internal state of the sender in terms of broad bipolar dimensions such as valence (pleasure-displeasure) and activation (sleepy-hyperactivated)". Also, what information the participant extracts first from an emotional stimulus is still unknown: this topic has also been thoroughly addressed by Russell et coll. (Russell, 2003). The participant could obtain from an emotional stimulus information on the sender's valence, activation, etc., then he would use this information to infer the sender's emotion. In the opposite, the participant could decode a specific emotion from the stimulus, then he would infer the information of sender's valence. Thus, several participant's responses to an emotional stimulus can be measured, such as the electromyographic response while the participant makes judgments on bipolar dimensions (e.g., whether the stimulus is hospitable or hostile; see review by Cacioppo, 2000), as well as the participant's ability to identify the emotional category of a stimulus or to infer the emotional state of the sender. Future work exploring several participant's responses to emotional stimuli, such as the participant's valence and activation, and the participant's ability to infer the sender's emotional state, will provide insight into the emotional processing in humans.

### Conclusion

This study compared the perception of important emotional dimensions such as valence, intensity and authenticity of nonlinguistic vocalizations using a visual analog scale in different groups of age and sex. Overall, this study showed (1) a consistency in ratings of perceived valence in nonlinguistic vocalizations across individuals, (2) older individuals rated stimuli as less emotional and intense than the younger group, and (3) older women judged vocalizations as less authentic.

The present study can be thought of as a first step in the development of a nonlinguistic auditory set equivalent of that of facial expressions (such as the one by (Ekman, 1976), i.e. a set of stimuli which can be used to compare emotional processing across cultures and languages, as well as in infants). Such a set of nonlinguistic vocalizations could contribute to explore the hypothesis of the existence of a core of universal meaning in perceiving and expressing emotions. Also, the integration of nonlinguistic vocalizations into models of emotional perception is essential to have a more complete view of emotional processing in regard or not to the means expressing the emotion (e.g., facial expression, body posture, prosody). Future studies will explore perception of a wider range of emotions, such as grunts and contemplative hums, as well as ambiguous emotions.



### Acknowledgments

We thank Frederick Aidan Remick, Roger Fearry, and Jean-Sébastien Robitaille for providing the stimuli, Catherine Bédard and Mathieu Décary-Hétu for help with testing, and Francine Giroux for statistical advice. This study was supported by NSERC, Université de Montréal, and NATEQ.



### Figure Captions

*Figure 1.* Valence ratings on a visual analog scale (1-100 points) divided by emotional category (negative, positive, neutral) and age group.  $*p < 0.05$ .

*Figure 2.* Intensity ratings on a visual analog scale (1-100 points) divided by emotional category (negative, positive, and, neutral vocalizations) and age group.  $*p < 0.01$ .

*Figure 3.* Authenticity ratings on a visual analog scale (1-100 points) divided by emotional category (negative, positive, and, neutral vocalizations) and age group.

*Figure 4.* Age-by-sex interaction in the authenticity judgment.

Figure 1

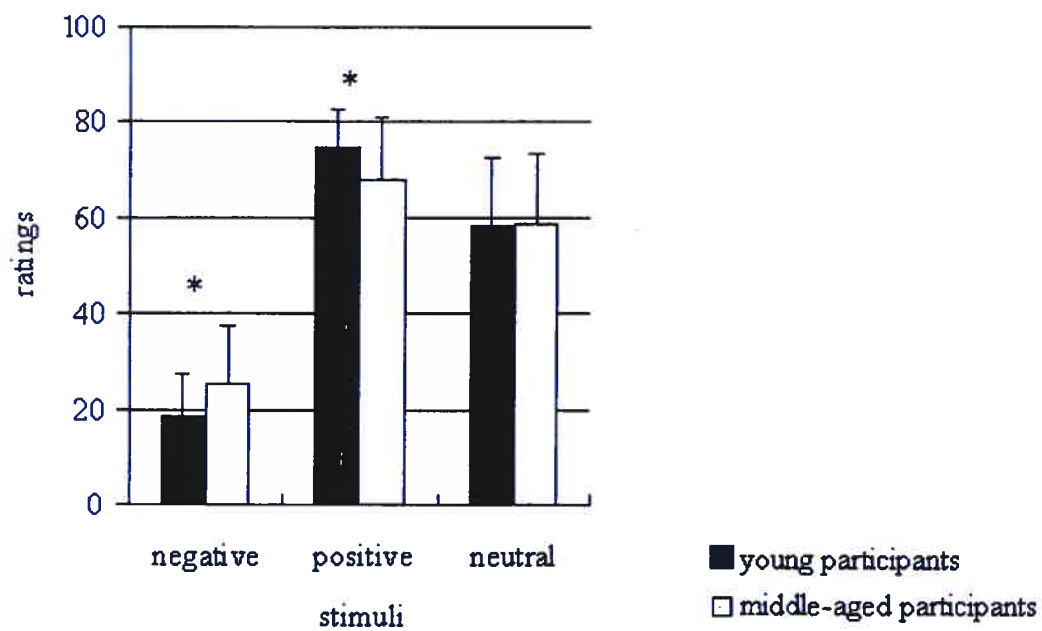


Figure 2

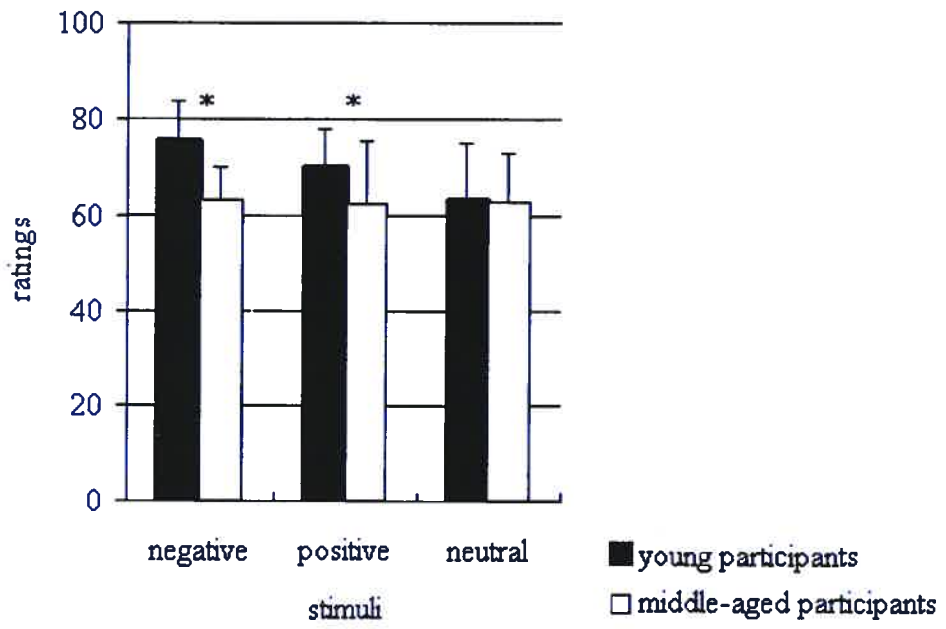


Figure 3

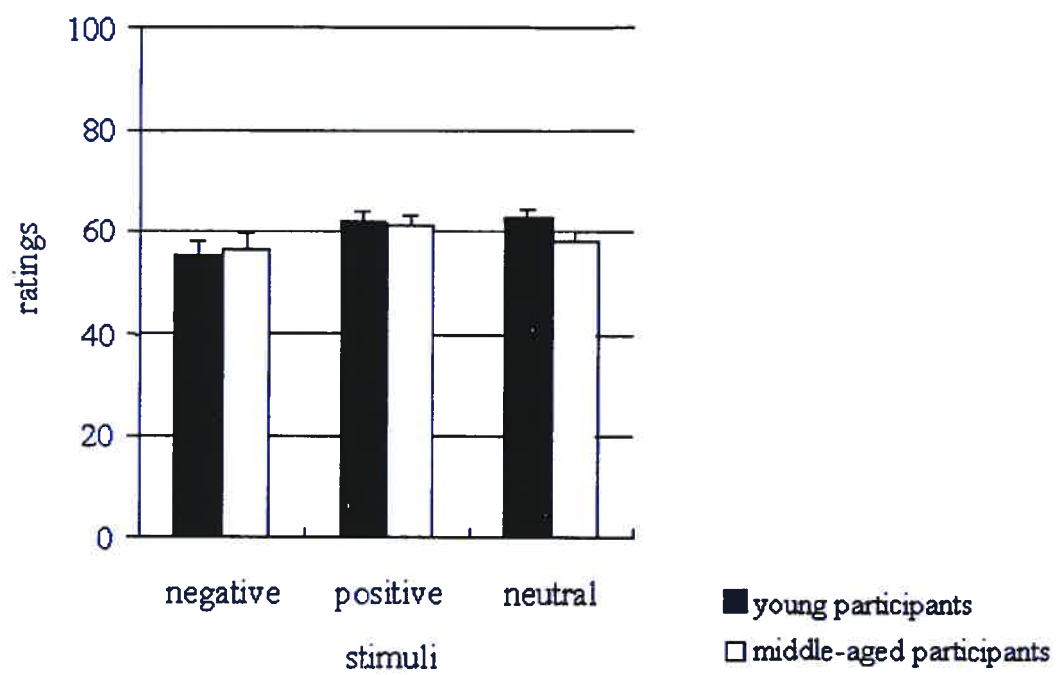
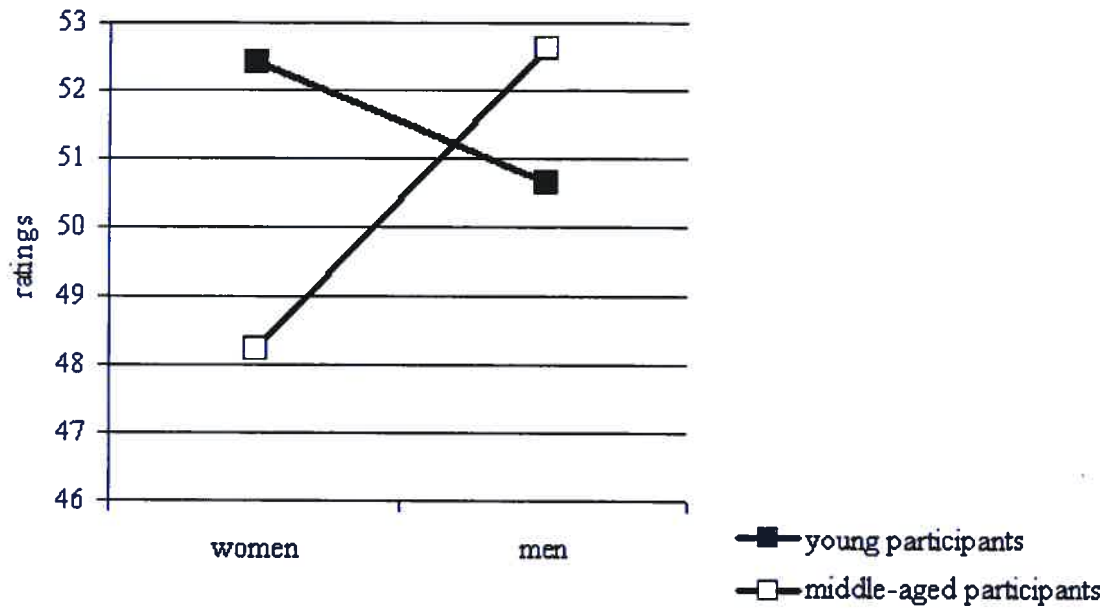


Figure 4



## References

- Adolphs R (2002). Neural systems for recognizing emotion. *Curr Opin Neurobiol* 12(2): 169-77.
- Adolphs R, Jansari A, et al. (2001). Hemispheric perception of emotional valence from facial expressions. *Neuropsychology* 15(4): 516-24.
- Albas DC, McCluskey KW, et al. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *J Cross Cult Psychol* 7: 481-89.
- Allen R, Brosigole L (1993). Facial and auditory affect recognition in senile geriatrics, the normal elderly and young adults. *Int J Neurosci* 68(1-2): 33-42.
- Anderson AK, Phelps EA (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport* 9(16): 3607-13.
- Bachorowski JA (1999). Vocal expression and perception of emotion. *Curr Dir Psychol Sci* 8: 53-7.
- Bachorowski JA, Owren MJ (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychol Sci* 6: 219-24.
- Bachorowski JA, Owren MJ (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J Acoust Soc Am* 106(2): 1054-63.
- Banse R, Scherer KR (1996). Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70(3): 614-36.
- Bard R, Gardner KH (1996). Influences on development in infant chimpanzees: Enculturation, temperament, and cognition. *Reaching into Thought: The Minds of*



- the Great Apes*. Russon A, Bard KA, Parker ST. Cambridge, Cambridge University Press: 235-56.
- Barr RG, Hopkins B, et al. (2000). *Crying as a Sign, a Symptom, & a Signal*. New York, Cambridge University Press.
- Baum SR, Pell MD (1999). The neural basis of prosody: Insight from lesion studies and neuroimaging. *Aphasiology* 13: 581-608.
- Beckman ME, Edwards J (1990). Lengthening and shortening and the nature of prosodic constituency. *Papers in Laboratory Phonology I: Between Grammar and the Physics of Speech*. Kingston J, Beckman ME. Cambridge, England, Cambridge University Press: 152-78.
- Behrens S (1985). The perception of stress and lateralization of prosody. *Brain Lang* 26: 332-48.
- Beier EG, Zautra AJ (1972). Identification of vocal communication of emotions across cultures. *J Consult Clin Psychol* 39: 166.
- Beldoch M (1964). Sensitivity to expression meaning in three modes of communication. *The Communication of Emotional Meaning*. Davitz JR. New York, McGraw-Hill: 31-42.
- Belin P, Fecteau S, et al. (2004). Thinking the voice: Neural correlates of voice perception. *Trends Cogn Sci* 8(3): 129-35.
- Belin P, Zatorre RJ, et al. (2000). Voice-selective areas in human auditory cortex. *Nature* 403(6767): 309-12.
- Bertrand M (1969). *The Behavioral Repertoire of the Stumptail Macacque*. Karger, Basel.

- Besson M, Magne C, et al. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends Cogn Sci* 6(10): 405-7.
- Blessner B (1972). Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J Speech Hear Res* 15(1): 5-41.
- Bliss TV, Collingridge GL (1993). A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* 361(6407): 31-9.
- Bliss TV, Lomo T (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J Physiol* 232(2): 331-56.
- Blonder LX, Bowers D (1991). The role of the right hemisphere in emotional communication. *Brain* 114(3): 1115-27.
- Borod JC (1993). Cerebral mechanisms underlying facial, prosodic, and lexical emotional expression: A review of neuropsychological studies and methodological issues. *Neuropsychology* 7: 445-63.
- Borod JC, Cicero BA, et al. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology* 12(3): 446-58.
- Borod JC, Koff E, et al. (1985). Channels of emotional expression in patients with unilateral brain damage. *Arch Neurol* 42: 345-8.
- Bowers D, Coslett HB (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. *Neuropsychologia* 25(2): 317-28.
- Brådvik B, Dravins C, et al. (1991). Disturbances of speech prosody following right

hemisphere infarcts. *Acta Neurol Scand* 84: 114-26.

- Braun K, Poeffel G (2001). Recognition of mother's voice evokes metabolic activation in the medial prefrontal cortex and lateral thalamus of Octodon Degus pups. *Neuroscience* 103(4): 861-4.
- Breitenstein C, Daum I, et al. (1998). Emotional processing following cortical and subcortical brain damage: contribution of the fronto-striatal circuitry. *Behav Neurol* 11(1): 29-42.
- Brown BL (1980). The detection of emotion in vocal qualities. *Language: Social Psychological Perspectives*. Giles H, Robinson PW, Smith P. Oxford, Pergamon: 237-45.
- Brown P, Fraser C (1979). Speech as a marker of situation. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 33-62.
- Brown P, Levinson S (1979). Social structure, groups and interaction. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 291-341.
- Bryan K (1989). Language prosody in the right hemisphere. *Aphasiology* 3: 285-99.
- Buchanan TW, Lutz K, et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Brain Res Cogn Brain Res* 9(3): 227-38.
- Burns KL, Beier EG (1973). Significance of vocal and visual channels in the decoding of emotional meaning. *J Comm* 23: 118-30.
- Cacioppo JT, Berntson GG, et al. (2000). The psychophysiology of emotion. *Handbook of Emotions*. Lewis M, Haviland-Jones JM. New York, Guilford Press. 2: 173-91.

- Cahill L, McGaugh JL (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends Neurosci* 21: 294-99.
- Cancelliere AE, Kertesz A (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain Cogn* 13(2): 133-47.
- Cannon WB (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *Am J Psychol* 39: 106-24.
- Carmon A, Nachshon I (1973). Ear asymmetry in perception of emotional non-verbal stimuli. *Acta Psychol (Amst)* 37(6): 351-7.
- Chevalier-Sokolnikoff S (1974). *The Ontogeny of Communication in the Stumptail Macaque (Macaca Arctoides)*. Karger, Basel.
- Chomsky N, Halle M (1968). *The Sound Pattern of English*. New York, Harper & Row.
- Coleman PD, Flood DG (1987). Neuron numbers and dendritic extent in normal aging and Alzheimer's disease. *Neurobiol Aging* 8(6): 521-45.
- Convit A, De Leon MJ, et al. (1997). Specific hippocampal volume reductions in individuals at risk for Alzheimer's disease. *Neurobiol Aging* 18(2): 131-8.
- Corwin MJ, Golub HL (1996). The infant cry: What can it tell us? *Curr Probl Pediatr* 26: 325-34.
- Cosmides L (1983). Invariances in the acoustic expression of emotion during speech. *J Exp Psychol: Hum Percept Perform* 9: 864-81.
- Crews WD, Harrison DW (1994). Sex differences and cerebral asymmetry in facial affect perception as a function of depressed mood. *Psychobiology* 22: 112-6.

- Curcio CA, Buell SJ, et al. (1982). Morphology of the aging nervous system: Not all downhill. *Advances in Neurogerontology*. Mortimer JA, Pirozzolo FJ, Maletta GJ. New York, Praeger. 3: 7-35.
- Darby DG (1993). Sensory aprosodia: A clinical clue to lesions of the inferior division of the right middle cerebral artery? *Neurology* 43(3 Pt 1): 567-72.
- Darwin C (1872/1965). *The Expression of the Emotions in Man and Animals*. Chigaco, University of Chicago Press.
- Davidson RJ (1992). Prolegomenon to the structure of emotion: Gleanings from neuropsychology. *Cogn Emotion* 6: 245-68.
- Davidson RJ (1992a). Anterior cerebral asymmetry and the nature of emotion. Special issue: The role of frontal lobe maturation in cognitive and social development. *Brain Cogn* 20: 125-51.
- Davidson RJ (1995). Cerebral asymmetry, emotion, and affective style. *Brain Asymmetry*. Davidson RJ, Hugdahl K. Cambridge (MA), MIT Press: 361-87.
- Davidson RJ, Fox NA (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science* 218(4578): 1235-7.
- Davidson RJ, Irwin W (1999). The functional neuroanatomy of emotion and affective style. *Trends Cogn Sci* 3(1): 11-21.
- Davis DG, Schmitt FA, et al. (1999). Alzheimer neuropathologic alterations in aged cognitively normal subjects. *J Neuropathol Exp Neurol* 58(4): 376-88.
- Davitz JR, Davitz LJ (1959). The communication of feelings by content-free speech. *J Comm* 9: 6-13.

- Denburg NL, Buchanan TW, et al. (2003). Evidence for preserved emotional memory in normal older persons. *Emotion* 3(3): 239-53.
- Detoleto-Morrell L, Geinisman Y, et al. (1988a). Age-dependent alterations in hippocampal synaptic plasticity: Relation to memory disorders. *Neurobiol Aging* 9(5-6): 581-90.
- Detoleto-Morrell L, Geinisman Y, et al. (1988b). Individual differences in hippocampal synaptic plasticity as a function of aging: Behavioral, electrophysiological, and morphological evidence. *Neural Plasticity: A Lifespan Approach*. Petit TL, Ivy G. New York, Liss: 283-328.
- Detoleto-Morrell L, Geinisman Y, et al. (1993). Synaptic morphological markers of age-related memory dysfunction. *Neurosci Facts* 4(3): 25-6.
- Diehl M, Coyle N, et al. (1996). Age and sex differences in strategies of coping and defense across the life span. *Psychol Aging* 11(1): 127-39.
- Dimitrovsky L (1964). The ability to identify the emotional meaning of vocal expressions at successive levels. *The Communication of Emotional Meaning*. Davitz JR. New York, McGraw-Hill: 69-86.
- Drummond JE, McBride ML, et al. (1993). The development of mothers understanding of infant crying. *Clin Nurs Res* 2: 396-413.
- Dusenbury D, Knower FH (1938). Experimental studies of the symbolism of action and voice (III): A study of the specificity of meaning in facial expression. *Q J Speech* 24: 424-35.
- Dysktra K, Gandour J, et al. (1995). Disruption of prosody after frontal lobe seizures in the non-dominant hemisphere. *Aphasiology* 9(5): 453-76.



- Ehlers L, Dalby M (1987). Appreciation of emotional expressions in the visual and auditory modality in normal and brain damaged patients. *Acta Neurol Scand* 76: 251-6.
- Ekman P (1992). An argument for basic emotions. *Cogn Emotion* 6: 169-200.
- Ekman P (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychol Bull* 115: 268-87.
- Ekman P, Friesen WV (1976). *Pictures of Facial Affect*. Paolo Alto, CA, Consulting Psychologists Press.
- Esposito G, Kirkby BS, et al. (1999). Context-dependent, neural system-specific neurophysiological concomitants of ageing: Mapping PET correlates during cognitive activation. *Brain* 122 (5): 963-79.
- Fairbanks G, Pronovost W (1939). An experimental study of the pitch characteristic of the voice during the expression of emotion. *Speech Monogr* 6: 87-104.
- Fama R, Sullivan EV, et al. (1997). Selective cortical and hippocampal volume correlates of Mattis Dementia Rating Scale in Alzheimer disease. *Arch Neurol* 54(6): 719-28.
- Fenster CA, Blake LK, et al. (1977). Accuracy of vocal emotional communications among children and adults and the power of negative emotions. *J Commun Disord* 10: 301-14.
- Ferreira F (1993). The creation of prosody during sentence production. *Psychol Rev* 100: 233-53.
- Fichtel C, Hammerschmidt K (2003). Responses of squirrel monkeys to their experimentally modified mobbing calls. *J Acoust Soc Am* 113(5): 2927-32.
- Fitch WT (1997). Vocal tract length and formant frequency dispersion correlate with body

- size in rhesus macaques. *J Acoust Soc Am* 102: 1213-22.
- Flood DG, Coleman PD (1988). Neuron numbers and sizes in aging brain: Comparisons of human, monkey, and rodent data. *Neurobiol Aging* 9(5-6): 453-63.
- Folkman S, Lazarus RS, et al. (1987). Age differences in stress and coping processes. *Psychol Aging* 2(2): 171-84.
- Frank MG, Ekman P, et al. (1993). Behavioral markers and recognizability of the smile of enjoyment. *J Pers Soc Psychol* 64(1): 83-93.
- Frick RW (1985). Communication emotion: The role of prosodic features. *Psychol Bull* 97: 412-29.
- George MS, Parekh PI, et al. (1996). Understanding emotional prosody activates right hemisphere regions. *Arch Neurol* 53(7): 665-70.
- Gerken L (1996a). Prosodic structure in young children's language production. *Language* 72(4): 683-712.
- Gerken L (1996b). Prosody's role in language acquisition and adult parsing. *J Psycholinguist Res* 25(2): 345-56.
- Ghazanfar AA, Hauser M (2001). The auditory behaviour of primates: A neuroethological perspective. *Curr Opin Neurobiol* 11: 712-20.
- Gil da Costa R, Palleroni A, et al. (2003) Rapid acquisition of an alarm response by a neotropical primate to a newly introduced avian predator. *Proc R Soc Lond B Biol Sci* 270(1515): 605-10.
- Giles H (1979). Ethnicity markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press.

- Golomb J, de Leon MJ, et al. (1993). Hippocampal atrophy in normal aging. An association with recent memory impairment. *Arch Neurol* 50(9): 967-73.
- Goodall J (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, Cambridge: Belknap Press.
- Gorelick PB, Ross ED (1987). The aprosodias: further functional-anatomical evidence for the organization of affective language in the right hemisphere. *J Neurol Neurosurg Psychiatry* 50: 553-60.
- Gouzoules H, Gouzoules S (2000). Agonistic screams differ among four species of macaques: The significance of motivation-structure rules. *Anim Behav* 59: 501-12.
- Gouzoules S, Gouzoules H, et al. (1984). Rhesus monkey (*Maccaca mulatta*) screams: Representational signaling in the recruitment of agonistic aid. *Anim Behav* 23: 182-93.
- Grady CL, McIntosh AR, et al. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science* 269(5221): 218-21.
- Grady CL, McIntosh AR, et al. (1999). The effects of age on the neural correlates of episodic encoding. *Cereb Cortex* 9(8): 805-14.
- Grady CL, Van Meter JW, et al. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8(11): 2511-6.
- Gray JA (1990). Brain systems that mediate both emotion and cognition. *Cogn Emotion* 4: 269-88.
- Green RS, Cliff N (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. *Percept Psychophys* 17: 429-38.

- Gross JJ, Carstensen LL, et al. (1997). Emotion and aging: Experience, expression, and control. *Psychol Aging* 12(4): 590-9.
- Grunwald IS, Borod JC, et al. (1999). The effects of age and gender on the perception of lexical emotion. *Appl Neuropsychol* 6(4): 226-38.
- Gunning-Dixon FM, Gur RC, et al. (2003). Age-related differences in brain activation during emotional face processing. *Neurobiol Aging* 24(2): 285-95.
- Habib M, Daquin G, et al. (1995). Mutism and auditory agnosia due to bilateral insular damage -role of the insula in human communication. *Neuropsychologia* 33(3): 327-39.
- Haggard MP, Parkinson AM (1971). Stimulus and task factors as determinants of ear advantages. *Q J Exp Psychol* 23(2): 168-77.
- Harrison DW, Gorelczenko PM, et al. (1990). Sex differences in the functional asymmetry for facial affect perception. *Int J Neurosci* 52(1-2): 11-6.
- Haug H (1984). Macroscopic and microscopic morphometry of the human brain and cortex: A survey in the light of new results. *Brain Pathol* 1: 123-49.
- Hauser M (1991). Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations. *Ethology* 89: 29-46.
- Heilman KM, Bowers D (1984). Comprehension of affective and non-affective prosody. *Neurology* 34: 917-21.
- Heilman KM, Scholes R, et al. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *J Neurol Neurosurg Psychiatry* 38(1): 69-72.
- Helfrich H (1979). Age markers in speech. *Social Markers in Speech*. Scherer KR, Giles H.

Cambridge, Cambridge University Press: 63-107.

- Herrero JV, Hillix WA (1990). Hemispheric performance in detecting prosody. A competitive dichotic listening task. *Percept Mot Skills* 71: 479-86.
- Hornak J, Rolls ET, et al. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia* 34(4): 247-61.
- Hornstein MG (1967). Accuracy of emotional communication and interpersonal compatibility. *J Pers* 35: 20-8.
- Iidaka T, Okada T, et al. (2002). Age-related differences in the medial temporal lobe responses to emotional faces as revealed by fMRI. *Hippocampus* 12(3): 352-62.
- Imaizumi S, Mori K, et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* 8(12): 2809-12.
- Izard CE (1972). *Patterns of Emotion: A New Analysis of Anxiety and Depression*. New York, Academic Press Inc.
- Jack CR Jr, Petersen RC, et al. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology* 49(3): 786-94.
- Jernigan TL, Archibald SL, et al. (1991). Cerebral structure on MRI, Part I: Localization of age-related changes. *Biol Psychiatry* 29(1): 55-67.
- Joanette Y, Goulet P, et al. (1990). *Right Hemisphere and Verbal Communication*. New York, Springer Verlag.
- Johnson WF, Emde RN, et al. (1986). Recognition of emotion from vocal cues. *Arch Gen Psychiatry* 43(3): 280-3.

- Jones B (1979). Sex and visual field effects on accuracy and decision making when subjects classify male and female faces. *Cortex* 15(4): 551-60.
- Jones B, Anuza T (1982). Effects of sex, handedness, stimulus and visual field on mental rotation. *Cortex* 18(4): 501-14.
- Kawashima R, Itoh M, et al. (1993). Changes of regional cerebral blood flow during listening to an unfamiliar spoken language. *Neurosci Lett* 161(1): 69-72.
- Kemper TL (1994). Neuroanatomical and neuropathological changes during aging and in dementia. *Clinical Neurology of Aging*. Albert ML, Knoepfel EJE. New York, Oxford University Press: 3-67.
- Kiss I, Ennis T (2001). Age-related decline in perception of prosodic affect. *Appl Neuropsychol* 8(4): 251-4.
- Knower FH (1941). Analysis of some experimental variations of simulated vocal expressions of the emotions. *J Soc Psychol* 14: 369-72.
- Kotz SA, Alter K, et al. (2000). The interface between prosodic and semantic processes: An ERP study. *Supp J Cogn Neurosci*: 121.
- Kotz SA, Meyer M, et al. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain Lang* 86: 366-76.
- Kramer E (1964). Elimination of verbal cues in judgments of emotion from voice. *J Abnorm Soc Psychol* 68: 390-96.
- Ladd DR, Silverman KEA (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker effect. *J Acoust Soc Am* 78(2): 435-44.



- Lakshminarayanan K, Ben Shalom D, et al. (2003). The effect of spectral manipulations on the identification of affective and linguistic prosody. *Brain Lang* 84(2): 250-63.
- Lalande S, Brawn CM, et al. (1992). Effects of right and left hemisphere cerebrovascular lesions on discrimination of prosodic and semantic aspects of affect in sentences. *Brain Lang* 42: 165-86.
- Lamendella JT (1977). The limbic system in human communication. *Studies in Neurolinguistics*. Whitaker H, Whitaker HA. New York, Academic Press: 157-222.
- Landfield PW, Applegate MD, et al. (1984). Nerve cell and synaptic decline in brain aging: Implications for animal models and for a hypothesis of Alzheimer's disease. *Modern Aging Research: Comparative Pathobiology of Major Age-Related Diseases: Current Status and Research Frontiers*. Scarpelli DG, Migaki G. New York, Liss. 4: 333-55.
- Landfield PW, Pitler TA, et al. (1986). The aged hippocampus. A model system for studies on mechanisms of behavioral plasticity and brain aging. *The Hippocampus*. Issakson P, Pribam KH. New York, Plenum Press. 3: 323-67.
- Lang PJ, Bradley MM, et al. (1992). A motivational analysis of emotion: Reflex-cortex connections. *Psychol Sci* 3: 44-9.
- Lawton MP, Kleban MH, et al. (1992). Dimensions of affective experience in three age groups. *Psychol Aging* 7(2): 171-84.
- Levenson RW, Carstensen LL, et al. (1991). Emotion, physiology, and expression in old age. *Psychol Aging* 6(1): 28-35.
- Levenson RW, Carstensen LL, et al. (1994). The influence of age and gender on affect, physiology, and their interrelations: A study of long-term marriages. *J Pers Soc*

*Psychol* 67(1): 56-68.

Levy PK (1964). The ability to express and perceive vocal communication of feelings. *The Communication of Emotional Meaning*. Davitz JK. New York, McGraw-Hill: 43-55.

Ley RG, Bryden MP (1982). A dissociation of right and left hemisphere effects for recognizing emotional tone and verbal content. *Brain Cogn* 1: 3-9.

Lieberman P, Michaels SB (1962). Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *J Acoust Soc Am* 34: 922-27.

MacPherson SE, Phillips LH, et al. (2002). Age, executive function, and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychol Aging* 17(4): 598-609.

Martin AJ, Friston KJ, et al. (1991). Decreases in regional cerebral blood flow with normal aging. *J Cereb Blood Flow Metab* 11(4): 684-9.

Masataka N, Kohda M (1988). Primate play vocalizations and their functional significance. *Folia Primatol* 50: 152-6.

McCluskey KW, Albas DC (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *Int J Psychol* 16: 119-32.

McCluskey KW, Albas DC, et al. (1975). Cross-cultural differences in the perception of emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Dev Psychol* 11: 15-21.

McDowell CL, Harrison DW, et al. (1994). Is right hemisphere decline in the perception of emotion a function of aging? *Int J Neurosci* 79(1-2): 1-11.

- Mesquita B, Frijda NH, et al. (1997). Culture and emotion. *Handbook of Cross-Cultural Psychology*. Berry PB, Saraswathi TS. Boston, Allyn & Bacon. 12: 255-97.
- Meunier M, Bachevalier J, et al. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J Neurosci* 13(12): 5418-32.
- Meyer E, Alter K, et al. (2002). Functional MRI evidence reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17: 73-88.
- Miller CT, Miller J, et al. (2001). Selective phonotaxis by cotton-top tamarins. *Behaviour* 138: 811-26.
- Mitchell RL, Elliott R, et al. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41(10): 1410-21.
- Monrad-Krohn GH (1947). Dysprosody or altered 'melody of language'. *Brain* 70: 405-15.
- Montepare J, Koff E, et al. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *J Nonverbal Behav* 23: 133-52.
- Moreno C, Borod JC, et al. (1993). The perception of facial emotion across the adult life span. *Dev Neuropsychol* 9: 305-14.
- Morgan JL, Meier RP, et al. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cogn Psychol* 19: 498-550.
- Morris JS, Frith CD, et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383(6603): 812-5.
- Morris RG, Anderson E, et al. (1986). Selective impairment of learning and blockade of

- long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. *Nature* 319(6056): 774-6.
- Morton EF (1977). On the occurrence and significance of motivational-structural rules in some bird and mammal sounds. *Am Nat* 111: 855-69.
- Murphy DG, DeCarli C, et al. (1996). Sex differences in human brain morphometry and metabolism: An in vivo quantitative magnetic resonance imaging and positron emission tomography study on the effect of aging. *Arch Gen Psychiatry* 53(7): 585-94.
- Murray EA (1979). Infant crying as an elicitor of parental behavior: An examination of two models. *Psychol Bull* 86: 191-215.
- Murray IR, Arnott JL (1993). Toward a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J Acoust Soc Am* 93(2): 1097-108.
- Newman J (1985). The infant cry of primates: An evolutionary perspective. *Infant Crying: Theoretical and Research Perspectives*. Lester BM, Boukydis CFZ. New York, Plenum Press: 307-23.
- Oatley K, Johnson-Laird PN (1987). Towards a cognitive theory of emotions. *Cogn Emotion* 1: 29-50.
- O'Leary DS, Andreason NC, et al. (1996). A positron emission tomography study of binaurally and dichotically presented stimuli: Effects of level of language and directed attention. *Brain Lang* 53(1): 20-39.
- Oscar-Berman M, Hancock M, et al. (1990). Emotional perception and memory in alcoholism and aging. *Alcohol Clin Exp Res* 14(3): 383-93.
- Osser HA (1964). A distinctive feature analysis of the vocal communication of emotion.

*Diss Abstr* 25: 3708.

- Ouellette G, Baum SR (1993). Acoustic analysis of prosodic cues in left and right-hemisphere-damaged patients. *Aphasiology* 8(3): 257-83.
- Pell MD (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia* 36(8): 701-15.
- Pell MD (1999). The temporal organization of affective and non-affective speech in patients with right-hemisphere infarcts. *Cortex* 35(4): 455-77.
- Pell MD, Baum SR (1997). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain Lang* 57(2): 195-214.
- Peper M, Irle E (1997). Categorical and dimensional decoding of emotional intonations in patients with focal brain lesions. *Brain Lang* 58: 233-64.
- Peters A, Vaughan DW (1981). Central nervous system. *Aging and Cell Structure*. Johnson JEJ. New York, Plenum Press. 1: 1-34.
- Petit-Taboue MC, Landeau B, et al. (1998). Effects of healthy aging on the regional cerebral metabolic rate of glucose assessed with statistical parametric mapping. *NeuroImage* 7(3): 176-84.
- Pfaff PL (1954). An experimental study of the communication of feeling without contextual material. *Speech Monogr* 21: 155-56.
- Phillips LH, MacLean RD, et al. (2002). Age and the understanding of emotions: Neuropsychological and sociocognitive perspectives. *J Gerontol B Psychol Sci Soc Sci* 57(6): P526-30.



- Phillips ML, Young AW, et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc R Soc Lond B Biol Sci* 265(1408): 1809-17.
- Phillips ML, Young AW, et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389(6650): 495-8.
- Pihan H, Ackermann H, et al. (1997). The cortical processing of perceived emotion: A DC potential study on affective prosody. *Neuroreport* 8(623-27).
- Pihan H, Altenmuller E, et al. (2000). Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system: A DC potential study. *Brain* 123: 2338-49.
- Pittam J, Scherer KR (1993). Vocal expression and communication of emotion. *The Handbook of Emotions*. Lewis M, Haviland JM. New York, Guilford Press.
- Pollack I, Rubenstein H, Horowitz A (1960). Communication of verbal modes of expression. *Lang Speech* 3: 121-130.
- Pugh KR, Offywitz BA, et al. (1996). Auditory selective attention: An fMRI investigation. *NeuroImage* 4(3 Pt 1): 159-73.
- Rama P, Martinkauppi S, et al. (2001). Working memory of identification of emotional vocal expressions: An fMRI study. *NeuroImage* 13(6 Pt 1): 1090-101.
- Raz N (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. *Handbook of Aging and Cognition*. Craik FIM, Salthouse TA. Erlbaum: NJ, Mahwah. II.
- Rendall D (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J Acoust Soc Am* 113(6): 3390-402.



- Rendall D, Owren MJ, et al. (1998). The role of vocal track filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J Acoust Soc Am* 103: 602-14.
- Rendall D, Rodman PS, et al. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim Behav* 51: 1007-15.
- Robin DA, Tranel D, et al. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang* 39: 539-55.
- Robinson BW (1976). Limbic influences on human speech. *Ann N Y Acad Sci* 280: 761-76.
- Robinson WP (1979). Speech markers and social class. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 211-49.
- Roland PE (1993). *Brain Activation*. New York, John Wiley.
- Ross ED (1981). The aprosodias: Functional anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol* 38: 561-9.
- Ross ED (1993). Nonverbal aspects of language. *Neurol Clin* 11(1): 9-23.
- Ross ED, Anderson B, et al. (1989). Crossed aprosodia in strongly dextral patients. *Arch Neurol* 46: 206-9.
- Ross ED, Edmondson JA, et al. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis. *J Phonetics* 14: 283-302.
- Ross ED, Homan RW, et al. (1994). Differential hemispheric lateralization of primary and social emotions: Implications for developing a comprehensive neurology for emotions, repression, and the subconscious. *Neuropsychiatr Neuropsychol Behav Neurol* 7(1-19).

- Ross ED, Thompson RD, et al. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain Lang* 56(1): 27-54.
- Ross M, Duffy RJ, et al. (1973). Contribution of the lower audible frequencies to the recognition of emotions. *Am Ann Deaf* 118: 37-42.
- Rubin DC, Friendly M (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Mem Cogn* 14: 79-94.
- Russell JA (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull* 115(1): 102-41.
- Russell JA (1997). Reading emotions from into faces: Resurrecting a dimensional-contextual perspectives. *The Psychology of Facial Expression*. Russell JA, Fernandez-Dols JM. New York, Cambridge University Press: 295-320.
- Russell JA, Bachorowski JA, et al. (2003). Facial and vocal expressions of emotion. *Annu Rev Psychol* 54(1): 329-49.
- Russell JA, Bullock M (1986). On the dimensions preschoolers use to interpret facial expressions of emotion. *Dev Psychol* 22: 97-102.
- Sackeim HA, Greenberg MS, et al. (1982). Hemispheric asymmetry in the expression of positive and negative emotions: Neurologic evidence. *Arch Neurol* 39: 210-8.
- Sackeim HA, Gur RC (1979). Lateral asymmetry in intensity of emotional expression. *Neuropsychologia* 163: 473-81.
- Safer MA, Leventhal H (1977). Ear difference in evaluating emotional tones of voice and verbal content. *J Exp Psychol Hum Percept Perform* 3(1): 75-82.

- Sander K, Roth P, et al. (2003). Left-lateralized fMRI in the temporal lobe of high repressive women during the identification of sad prosodies. *Cogn Brain Res* 16: 441-56.
- Sander K, Scheich H (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res Cogn Brain Res* 12(2): 181-98.
- Scherer KR (1981a). Vocal indicators of stress. *Speech Evaluation in Psychiatry*. Darby J. New York, Grune & Stralton: 171-87.
- Scherer KR (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *J Pers* 40(2): 191-210.
- Scherer KR (1979a). Non-linguistic indicators of emotion and psychopathology. *Emotions in Personality and Psychopathology*. Izard CE. New York, Plenum Press: 495-529.
- Scherer KR (1979b). Personality markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 147-209.
- Scherer KR (1981). Speech and emotional states. *Speech Evaluation in Psychiatry*. Darby J. New York, Grune & Stralton: 189-220.
- Scherer KR (1982). Methods of research on vocal communication: Paradigms and parameters. *Handbook of Methods in Nonverbal Behavior Research*. Scherer KR, Ekman P. Cambridge, Cambridge University Press: 136-98.
- Scherer KR (1985). Vocal affect signaling: A comparative approach. *Advances in the Study of Behavior*. Academic Press. 15: 189-244.
- Scherer KR (1986). Vocal affect expression: A review and a model for future research. *Psychol Bull* 99: 143-65.

- Scherer KR (1995). Expression of emotion in voice and music. *J Voice* 9(3): 235-48.
- Scherer KR (1997a). Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. *Cogn Emotion* 11: 113-50.
- Scherer KR (1997b). The role of culture in emotion-antecedent appraisal. *J Pers Soc Psychol* 73: 902-22.
- Scherer KR (2000a). Psychological models of emotion. *The Neuropsychology of Emotion*. Borod J. Oxford/New York, Oxford University Press: 137-66.
- Scherer KR, Banse R, et al. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *J Cross Cult Psychol* 32(1): 76-92.
- Scherer KR, Banse R, et al. (1991). Vocal cues in emotion encoding and decoding. *Motiv Emotion* 15: 123-48.
- Scherer KR, Ceschi G (2000). Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Pers Soc Psychol Bull* 26(3): 327-39.
- Scherer KR, Feldstein S, et al. (1985). Vocal cues to deception: A comparative channel approach. *J Psycholinguist Res* 14(4): 409-25.
- Scherer KR, Kappas A (1988). Primate vocal expression of affective states. *Primate Vocal Communication*. Todt D, Goedeeking P, Newman E. Heidelberg, Springer: 171-94.
- Scherer KR, Oshinsky JS (1977). Cue utilization in emotion attribution from auditory stimuli. *Motiv Emotion* 1(4): 331-46.
- Scherer K, Wallbott HG (1994). Evidence for universality and cultural variation of differential emotion response patterning. *J Pers Soc Psychol* 66(2): 310-28.

- Scherer KR, Zei B (1988). Vocal indicators of affective disorders. *Psychother Psychosom* 49(3-4): 179-86.
- Schirmer A, Kotz SA, et al. (2002). Sex differentiates the role of emotional prosody during word processing. *Brain Res Cogn Brain Res* 14(2): 228-33.
- Schlanger BB (1973). Identification by normal and aphasic subjects of semantically meaningful and meaningless emotional toned sentences. *Acta Symbolica* 4: 30-8.
- Schlanger BB, Schlanger P (1976). The perception of emotionally toned sentences by the right hemisphere damaged and aphasic subjects. *Brain Lang* 3: 396-403.
- Schmitt JJ, Hartje W, et al. (1997). Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositionnal speech. *Cortex* 33: 65-81.
- Scott SK, Blank CC, et al. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(12): 2400-6.
- Scott SK, Young AW, et al. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385(6613): 254-7.
- Sebeok TA (1968). *Animal Communication: Techniques of Study and Results of Research*. Bloomington, Indiana University Press.
- Sebeok TA (1977). *How Animals Communicate*. Bloomington, Indiana University Press.
- Semple S, McComb K(2000). Perception of female reproductive state form vocal cues in a mamal species. *Proc R Soc London B* 267: 707-12.
- Seron X, van der Kaa MA, et al. (1982). Decoding paralinguistic signals: Effects of semantic and prosodic cues on aphasics comprehension. *J Commun Disord* 15: 223-

31.

- Seyfarth RM, Cheney DL, et al. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science Wash DC* 210: 801-3.
- Shipley-Brown F, Dingwall W (1988). Hemispheric processing of affective linguistic intonation contours in normal subjects. *Brain Lang* 33: 16-26.
- Siegman AW (1978). The telltale voice: Nonverbal messages of verbal communication. *Nonverbal Behavior and Communication*. Siegman, AW, Feldstein S. Hillsdale, NJ, Erlbaum: 183-243.
- Smith PM (1979). Sex markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 109-46.
- Smith WJ (1977). *The Behaviour of Communicating: The Ethological Approach*. Cambridge (MA), Harvard University Press.
- Sobotka SS, Davidson RJ, et al. (1992). Anterior brain electrical asymmetries in response to reward and punishment. *Electroencephalogr Clin Neurophysiol* 83(4): 236-47.
- Sogon S (1975). A study of the personality factor which affects the judgment of vocally expressed emotions. *Japanese Journal of Psychology* 46: 247-54.
- Speedie LJ, Coslett HB, et al. (1984). Repetition of affective prosody in mixed transcortical aphasia. *Arch Neurol* 41: 268-70.
- Starkstein SE, Federoff JP, et al. (1994). Neuropsychological and neuroradiological correlates of emotional prosody comprehension. *Neurology* 44: 515-22.
- Thompson LA, Aidinejad MR, et al. (2001). Aging and the effects of facial and prosodic



- cues on emotional intensity ratings and memory reconstructions. *J Nonverbal Behav* 25(2): 101-25.
- Tompkins CA, Flowers CR (1985). Perception of emotional intonation by brain damaged adults: The influence of task processing levels. *J Speech Hear Res* 28: 527-83.
- Tosi O (1979). *Voice Identification: Theory and Legal Applications*. Baltimore, University Park Press.
- Tucker DM, Watson RT (1977). Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology* 27: 947-50.
- Tzourio N, Massioui FE, et al. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage* 5: 63-77.
- van Bezooijen R (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht: Foris.
- Van Lancker D, Sidtis JJ (1992). The identification of affective prosodic stimuli by left and right hemisphere damaged subjects: All errors are not created equal. *J Speech Hear Res* 35: 963-70.
- van Lawick-Goodall J (1968). A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. *Primates: Studies in Adaptation and Variability*. Jay PC. New York, Holt, Rinehart & Winston: 313-74.
- Van Strien JW, Van Beek S (2000). Ratings of emotion in laterally presented faces: Sex and handedness effects. *Brain Cogn* 44(3): 645-52.
- Vikingstad EM, George KP, et al. (2000). Cortical language lateralisation in right handed normal subjects using functional magnetic resonance imaging. *J Neurol Sci* 175: 17-27.

- Wallbott HG, Scherer KR (1986). Cues and channels in emotion recognition. *J Abnorm Soc Psychol* 51: 690-9.
- Walton JH, Orlikoff RF (1994). Speaker race identification from acoustic cues in the vocal signal. *J Speech Hear Res* 37(4): 738-45.
- Weintraub S, Mesulam MM, et al. (1981). Disturbances in prosody: A right-hemisphere contribution to language. *Arch Neurol* 38: 742-4.
- Weiss DJ, Garibaldi BT, et al. (2001). The production and perception of long calls in cotton-top tamarins (*Saguinus oedipus*). *J Comp Psychol* 115: 258-71.
- West MJ (1993a). New stereological methods for counting neurons. *Neurobiol Aging* 14(4): 275-85.
- West MJ (1993b). Regionally specific loss of neurons in the aging human hippocampus. *Neurobiol Aging* 14(4): 287-93.
- West MJ, Gundersen HJ (1990). Unbiased stereological estimation of the number of neurons in the human hippocampus. *J Comp Neurol* 296(1): 1-22.
- Whiteside SP (1998). Identification of a speaker's sex: A study of vowels. *Percept Mot Skills* 86(2): 579-584.
- Wildgruber D, Pihan H, et al. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage* 15(4): 856-69.
- Williams CE, Stevens KN (1972). Emotions and speech: Some acoustical correlates. *J Acoust Soc Am* 52: 1238-50.
- Williams CE, Stevens KN (1981). Vocal correlates of emotional states. *Speech Evaluation*

*in Psychiatry*. Darby JK. New York, Grune & Stratton: 221-42.

Wolf G, Gorski R, et al. (1972). Acquaintance and accuracy of vocal emotions. *J Commun* 22: 300-5.

Zatorre RJ (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* 84: 566-72.

Zatorre RJ, Belin P (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11(10): 946-53.

Zatorre RJ, Belin P, et al. (2002). Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6(1): 37-46.

Zatorre RJ, Evans AC, et al. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14: 1908-19.

Zatorre RJ, Evans AC, et al. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256(5058): 846-9.

Zatorre RJ, Samson S (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain* 114: 2403-17.

Zeskind PS, Marshall TR (1988). The relation between variation in pitch and maternal perception of infant crying. *Child Dev* 59: 193-6.

Zuckerman M, Lipets MS, et al. (1975). Encoding and decoding nonverbal cues of emotion. *J Pers Soc Psychol* 32: 1068-76.

## **CHAPITRE 3**

### **ARTICLE 2**

# **AMORÇAGE COMPORTEMENTAL DANS LA PERCEPTION DES VOCALISATIONS ÉMOTIONNELLES : L'INFLUENCE DU SEXE DU LOCUTEUR**

**ARTICLE 2****PRIMING OF NON-SPEECH VOCALIZATIONS IN MALE  
ADULTS: THE INFLUENCE OF THE SPEAKER'S GENDER****Shirley FECTEAU, Jorge L. ARMONY,****Yves JOANETTE & Pascal BELIN*****Brain and Cognition* (2004) 55(2): 300-2.**

Priming of non-speech vocalizations in male adults: the influence of the speaker's gender

Shirley Fecteau<sup>1,2,3</sup>, Jorge L. Armony<sup>4</sup>, Yves Joanette<sup>2,3,5</sup>, Pascal Belin<sup>2,3</sup>

<sup>1</sup> Département de psychiatrie, Université de Montréal, Montréal, Canada

<sup>2</sup> Institut universitaire de gériatrie de Montréal, Montréal, Canada

<sup>3</sup> Groupe de recherche en neuropsychologie expérimentale et cognitive, Université de Montréal, Montréal, Canada

<sup>4</sup> Douglas Hospital Research Centre, McGill University, Montréal, Canada

<sup>5</sup> Département d'orthophonie et d'audiologie, Université de Montréal, Montréal, Canada

Running Head: Priming of non-speech vocalizations

Correspondence concerning this article should be addressed to Shirley Fecteau,  




### Abstract

Previous research reported a priming effect for voices. However, the type of information primed is still largely unknown. In this study, we examined the influence of speaker's gender and emotional category of the stimulus on priming of non-speech vocalizations in ten male participants, who performed a gender identification task. We found a significant priming effect, which was greater for female than male vocalizations. In contrast, the emotional category of the stimuli did not appear to modulate the magnitude of priming. These data suggest that different aspects of the information conveyed in non-speech vocalizations may contribute differently to behavioral priming, possibly influenced by task relevance.

## Introduction

Priming is a type of implicit memory, a nonconscious influence of past experience on current performance or behavior. It consists of an improved identification of perceptual material, and is observed in terms of decreases of latency, or error rates in responding to repeated, as compared to novel, stimuli. Auditory studies, using speech, showed priming effects at the semantic level (Maratos et al., 2000), as well as priming of the speaker's identity (Ellis et al., 1997). Non-speech vocalizations, such as laughs or cries, play a crucial role in social interactions. Still, little is known about what factors may influence how we perceive and encode these stimuli, and therefore how we implicitly react to them. Here, we conducted a study using a repetition-priming paradigm to investigate the perception of non-speech vocalizations in normal male adults. In particular, we explored whether the gender of the speaker and the emotional categories of vocalizations influence priming.

## Method

### *Participants*

Ten right-handed male volunteers (mean age = 23.1 years, SD = 2.3) without neurological or psychiatric history participated in this study, as part of an event-related fMRI experiment. All participants had normal audition and were not under medication.

### *Stimuli*

In a previous experiment, forty participants rated 563 stimuli categorized as laughs, sexual sounds, cries, screams of fear, and non-emotional vocalizations (e.g., coughs) on valence, intensity, and authenticity of the emotion expressed, using three visual analog scales (1-100 points). Of this original set, 72 vocalizations, grouped into three categories based on valence—positive, negative, and neutral—, were selected. Each vocalization was produced by a different speaker, half of them male and half female. The positive and negative categories of sounds were equalized on intensity and authenticity of the emotion expressed

according to the rates of the 40 judges. All stimuli were equalized for maximum peak values with Mitsyn (WLH) and had a mean duration of 1.6s (SD = 0.4s).

### *Task*

We used a repetition-priming paradigm in which each sound was presented twice, using MCF (DigiVox). Stimuli were repeated after an unpredictable number of intervening trials varying between 1 and 180 items (median = 40). Participants had to perform a gender identification task, pressing one button when a man produced the stimulus, and another when a woman produced the vocalization (counterbalanced across participants). Following this task, participants had to rate each stimulus on valence, intensity, and authenticity of the emotion expressed on three visual analog scales (1-100 points).

### Results

In the gender identification task, the mean accuracy was 67.6%, with mean percentages of errors and omissions (no button press) at 30.7% and 1.7%, respectively. Numerically, errors of commissions (wrong response for one stimulus of a pair) were greater for the first occurrence of stimuli (6.0%) as compared to the second one (4.9%), but this difference was not significant ( $t(9) = 1.15, p = 0.28$ ).

Median reaction times (RTs) were calculated for correct answers only. We used the median to minimize the effects of outliers. Priming scores were calculated as the difference of median RTs between first and second stimulus presentation and entered in a repeated measures ANOVA with emotional category (positive, negative, and neutral) and gender as factors. Tests of between subjects effects revealed an overall effect of priming (98ms,  $F(1,9) = 6.72, p < 0.05$ ). This priming effect was not significantly modulated by the number of intervening items ( $p > 0.5$ ). There was a main effect of gender on priming ( $F(1,9) = 5.76, p < 0.05$ ), but no significant effect of emotional category ( $F(2,18) < 1$ ) (cf. Fig 1).

For judgment ratings, there was a main effect of gender on valence ( $F(1,9) = 31.91, p < 0.001$ ), on intensity ( $F(1,9) = 15.86, p < 0.01$ ), and on authenticity ( $F(1,9) = 7.65, p < 0.05$ ): ratings were stronger for female vocalizations than for male ones (cf. figure 1). There was also a main effect of emotional categories on valence ( $F(2,18) = 388.36, p < 0.0001$ ), on intensity ( $F(2,18) = 28.97, p < 0.0001$ ), and a trend on authenticity ( $F(2,18) = 3.25, p = 0.06$ ). Finally, we found a significant interaction between gender x emotional category on valence ( $F(2,18) = 9.14, p < 0.01$ ) and on intensity ( $F(2,18) = 11.29, p < 0.001$ ), but not on authenticity.

Insert figure 1 here

### Discussion

In this paper, we reported an effect of auditory priming using non-speech vocalizations. Priming was observed in terms of decreases of latency in gender identification to the second presentation of a stimulus, with respect to its first presentation.

We also found a gender effect, as there was greater priming for female than male voices. This could reflect a selective gender bias in the processing of voices. Consistent with this hypothesis, Vuilleumier et al. (unpublished observations) found, in a study of recognition memory of mutual gaze, that participants were better at recognizing previously seen faces from the opposite gender than from the same gender. This gender effect could also be due to physical features. For example, previous studies showed that the fundamental frequency of stimuli, the acoustic parameter that mainly distinguishes female from male voices, plays a significant role in priming (Church and Schacter, 1994). Another issue to consider is task relevance. Priming effects are particularly sensitive to the task used; that is, priming appears to depend on the cognitive resources used to perform the task. Facial expression studies showed differences in stimulus processing according to the task performed, gender or emotion identification, using the same stimuli and the same

participants (Gur et al., 2002). That is, specific acoustic features used to identify gender may contribute to the gender effect on priming, and overshadow other information conveyed by voices (e.g., emotional content).

Several studies have shown an enhanced explicit memory for emotional, particularly negative, material compared to neutral stimuli. Thus, it was somewhat surprising not to find an effect of emotional category on priming in our study. One possibility is that the valence of our stimuli was not strong enough to induce a differential priming effect. This, however, appears not to be the case based on the valence and intensity ratings of the stimuli (cf. figure 1). A more likely explanation lies on the aforementioned effects of task relevance on priming. Future studies, using both gender and emotional category identification tasks, will address this issue.

### Acknowledgments

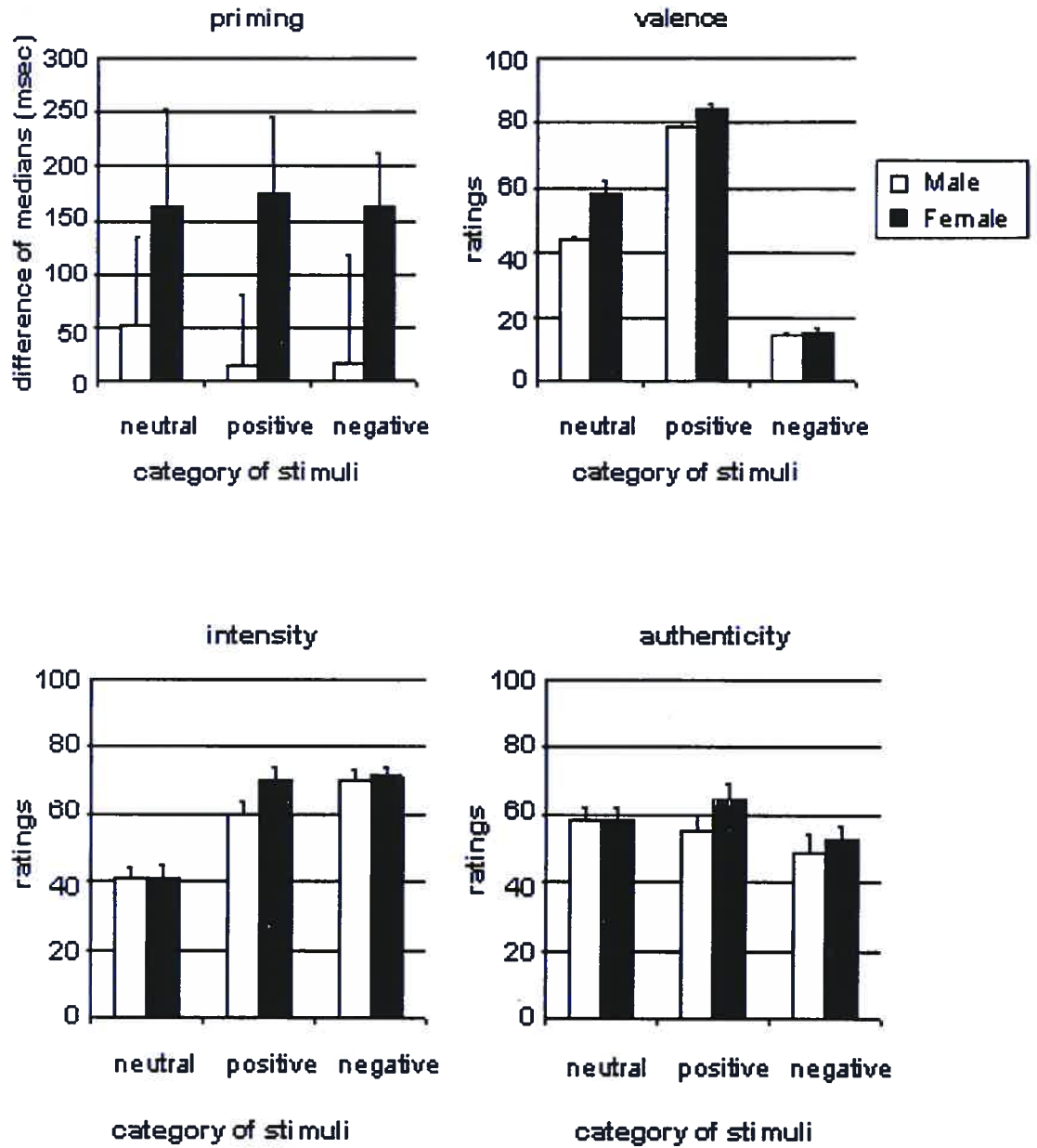
We would like to thank Frederick Aidan Remick, Roger Fearry, and Jean-Sébastien Robitaille for providing the stimuli, Pierre Ahad for his expertise in sound technology and computer programming, and Gilles Beaudoin, Boualem Mensour, and Jean-Maxime Leroux for help in developing the fMRI sequence. This study was supported by NSERC, Université de Montréal, and NATEQ.



**Figure caption**

*Figure 1.* Upper left panel shows priming effect in milliseconds, i.e., difference in median RT between 1st and 2nd occurrence of stimuli. Other panels show emotional ratings on three dimensions, split by emotional category (neutral, positive, negative) and speaker's gender.

Figure 1



## References

- Church BA, Schacter DL (1994). Perceptual specificity of auditory priming: implicit memory for voice intonation and fundamental frequency. *J Exp Psychol Learn Mem Cogn* 20: 521-33.
- Ellis H, Jones DM, Mosdell N (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol* 88(1): 143-56.
- Gur RC, Schroeder L, Turner T, McGrath C, Chan RM, Turetsky BI, Alsop D, Maldjian J, Gur RE (2002). Brain activation during facial emotion processing. *NeuroImage* 16(3 Pt 1): 651-62.
- Maratos EJ, Allan K, Rugg MD (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia* 38(11): 1452-65.

**CHAPITRE 4**

**ARTICLE 3**

**SPÉCIFICITÉ DES RÉPONSES CÉRÉBRALES À LA  
VOIX HUMAINE? UNE ÉTUDE D'IRMf  
ÉVÉNEMENTIELLE**

**ARTICLE 3**

**IS VOICE PROCESSING SPECIES-SPECIFIC IN HUMAN  
AUDITORY CORTEX ? AN fMRI STUDY**

**Shirley FECTEAU, Jorge L. ARMONY,**

**Yves JOANETTE & Pascal BELIN**

*NeuroImage* (accepté)

Is voice processing species-specific in human auditory cortex? An fMRI study

Shirley Fecteau<sup>1,2,3</sup>, Jorge L. Armony<sup>3,4</sup>, Yves Joanette<sup>2,3</sup>, Pascal Belin<sup>2,3</sup>


<sup>1</sup> Département de psychiatrie, Université de Montréal, Montreal, Canada

<sup>2</sup> Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montreal, Canada

<sup>3</sup> Centre de recherche en neuropsychologie expérimentale et cognition (CERNEC), Université de Montréal, Montreal, Canada

<sup>4</sup> Douglas Hospital Research Center, McGill University, Montreal, Canada

Correspondence should be addressed to Shirley Fecteau,  
Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville,  
Montreal, Que., H3C 3J7, Canada





### Abstract

Recent studies suggested a sensitivity of regions of the human superior temporal sulcus (STS) to the sound of the human voice. However, the question of the species-specificity of this response is still open. Healthy adult volunteers were scanned in an event-related fMRI design to compare responses in the STS to human and animal vocalizations, as well as to control nonvocal sounds (e.g., musical instruments). Bilateral activation of anterior STS was observed for human vocalizations, when contrasted with both nonvocal sounds and animal vocalizations. Animal vocalizations, compared to nonvocal sounds, elicited a more restricted left STS activation, although this region responded even more strongly to human vocalizations. This study provides the first evidence suggesting a species-specificity in STS responses to vocalizations in humans.

### Introduction

Species-specific vocalizations are social sounds upon which several species of vertebrates rely for their survival (e.g. Andrew, 1963; Winter et al., 1966; Altmann, 1967; Struhsaker, 1967; van Lawick-Goodall, 1968; Fossey, 1972; Green, 1975; Gautier and Gautier 1977; Seyfarth et al., 1980; Petersen, 1982; Smith et al., 1982; Snowdon, 1982). Evidence from neurophysiological studies on primary auditory cortex in non-human primates suggests that species-specific vocalizations are highly effective stimuli which can induce greater neural responses than other less behaviorally relevant sounds (Wang and Kadia, 2001). More recently, a neuroimaging study (Poremba et al., 2004) on primary auditory cortex in rhesus monkeys showed that species-specific vocalizations induce greater activation in the left superior temporal gyrus, relative to the right, whereas no such asymmetric responses were observed for human vocalizations and nonvocal sounds.

In humans, recent neuroimaging studies identified brain regions, located within the superior temporal sulcus (STS), that respond more strongly to human voices than to nonhuman sounds (Belin et al., 2000; Belin et al., 2002). However, the issue of species-specificity of these responses is still unresolved. Given that some of the acoustic features of animal vocalizations are similar to human vocalizations (e.g., Scherer and Kappas, 1988; Rendall, 2003), it is possible that voice-responsive areas in STS could also respond, to a comparable degree, to nonhuman animal vocalizations. Alternatively, these regions could show a species-specific response, and therefore respond more strongly to human than to animal vocalizations.

To address this issue, we conducted an event-related fMRI study to directly test the species-specificity of STS responses to vocal sounds. To that effect, we compared brain activity elicited by human and animal vocalizations. Based on evidence from previous studies on the involvement of STS on human voice processing, we focused our analysis on this region.

## Materials and methods

### *Participants*

Fifteen healthy, right-handed adults ( $n = 15$ , mean age = 22.6 years, six females) with no history of neurological or psychiatric conditions participated in this study. Informed written consent was obtained and the study was approved by the ethical committee from *Centre Hospitalier de l'Université de Montréal*.

### *Procedure*

Auditory stimuli were presented binaurally through pneumatic headphones, sealed by foam ear inserts and further shielded by plastic ear defenders (providing an attenuation of fMRI scanning noise of about 30dB) using MCF (DigiVox, Montreal), with a sound-pressure level of 85-90 dB. They consisted of 144 sounds divided into five categories (Figure 1): 24 speech sounds (e.g., vowels, words and sentences in different languages), 24 nonlinguistic vocalizations (e.g. laughs, screams, coughs), 24 cat vocalizations, 24 mixed-animal vocalizations, and 48 nonvocal sounds (e.g., mechanical sounds and musical instruments). The 24 mixed-animal vocalizations were produced by: alligator, antelope, bird, camel, chimpanzee, chinchilla, cow, dog, donkey, eagle, elephant, frog, hippopotamus, horse, kangaroo, koala, moose, pig, raccoon, rooster, sea lion, sheep, turtle, and whale (sample stimuli from all categories can be obtained at <http://www.mapageweb.umontreal.ca/belinp/>). There was no significant difference in mean duration between human vocalizations (mean = 1.21; SD = 0.22) and animal vocalizations (mean = 1.34; SD = 0.32). The mean duration for nonvocal sounds was 1.56 (SD = 0.32). The reason for presenting a substantial number of cat vocalizations was to have a category of vocalizations which was comparable, as much as possible, to the human vocalizations category in terms of semantic familiarity and homogeneity. All human and animal vocalizations were produced by different individuals and most stimuli used in this study were different from those used in Belin et al. (2000). Sounds were normalized to a common

peak of intensity with Mitsyn (WLH) and CoolEdit Pro (Syntrillium Software Corp.). Stimuli were presented in a random order (mean SOA = 5.5s, S.D = 2.4s). Null-events (i.e. scanner noise) were included and were used as baseline in the main auditory contrast (described below).

Insert Figure 1 about here

### *Imaging Details*

Scanning was performed on a 1.5 Tesla MRI system (Magnetom Vision, Siemens Electric, Erlangen, Germany) at the *Centre Hospitalier de l'Université de Montréal*. Functional scans were acquired with a single-shot echo planar gradient-echo (EPI) pulse sequence (TR = 2.6s, TE = 40 ms, flip angle = 90°, FOV = 215 mm, Matrix = 128 x 128). The 28 axial slices (resolution 3.75 x 3.75 mm in plane, 5 mm thickness) in each volume were aligned with the AC-PC line, covering the whole brain. A total of 320 volumes were acquired (the first four volume of each series were later discarded to allow for T1 saturation). After the functional scanning, T1-weighted anatomical images were obtained for each participant (1mm x 1mm x 1mm resolution). Scanner noise was continuous throughout the experiment providing a constant background (baseline).

### *fMRI Analysis*

Image processing and statistical analysis were performed using SPM99 (Wellcome Department of Cognitive Neurology; Friston et al., 1995; Worsley and Friston, 1995), as done in previous studies (Armony and Dolan, 2001, 2002). Briefly, the imaging time-series was realigned to the first volume to correct for interscan movement. Time-series were shifted using sinc interpolation to correct for differences in slice acquisition times. The functional images were then spatially normalized to a standard stereotaxic space (Talairach and Tournoux, 1988) based on a template provided by the Montreal Neurological Institute (Evans et al., 1994) to allow for group analysis. Finally, functional data were spatially

smoothed with an 8mm isotropic Gaussian kernel to compensate for residual inter-participant variability and to allow for the application of Gaussian random field theory in the statistical analysis (Friston et al., 1995). Each subject's structural scan was coregistered with the mean realigned functional image and normalized using the parameters determined for the functional images. A mean anatomical image was created from the participants' individual scans, onto which activation were overlaid for anatomical localization.

Data analysis was performed in a two-stage mixed-effects analysis (equivalent to a random effects analysis) in which BOLD responses for each subject were first modeled using a synthetic hemodynamic function in the context of the fixed-effects general linear model. Subject-specific linear contrasts on the parameter estimates were then entered into a second-level analysis to perform between-subjects analyses, resulting in a t-statistic for each voxel. These t-statistics (transformed to Z-statistics) constitute a statistical parametric map (SPM). SPM's were thresholded at  $p = 0.05$ , corrected for multiple comparisons across the brain, except for the STS, where a  $p = 0.001$  (uncorrected) was used, based on our *a priori* hypothesis on the involvement of this region in voice processing (Belin et al., 2000). All statistical comparisons involved comparisons between some of the five conditions (speech, nonlinguistic vocalizations, cat vocalizations, mixed-animal vocalizations, and nonvocal sounds), except for the main auditory contrast in which all experimental conditions were compared to null-events.

After the scanning session, participants were asked to rate the familiarity of the cat and mixed-animal vocalizations using a visual analog scale (1-100). The order of presentation of the stimuli was different from that of the scanning session and was counterbalanced across participants. The effects of stimulus familiarity on brain activation were analyzed by multiplying the HRF by the mean-corrected familiarity score for each stimulus. This was done on an individual basis, that is, using the familiarity scores from each participant.



## Results

### *Auditory activation*

When all categories of auditory stimuli were compared to baseline (i.e., scanner noise), a large part of the superior temporal plane corresponding to auditory cortex was activated bilaterally  $p < 0.001$  corrected; see Table 1a and Figure 2 for the time course of the two highest peaks). These maxima were located within the superior temporal gyrus (STG), and were activated by the 5 categories of stimuli to comparable extent.

Insert Table 1 about here

Insert Figure 2 about here

### *Human vs. nonhuman sounds*

Human vocalizations (speech and nonlinguistic vocalizations) elicited greater bilateral responses in STS, compared to nonhuman sounds (animal vocalizations and nonvocal sounds), as shown in Table 1b and Figure 3. At the individual level, eleven out of fifteen participants showed significant bilateral STS responses, one showed only left STS activation, one showed only right STS activation, and two participants did not show significant STS activation ( $p > 0.01$ ).

Insert Figure 3 about here

### *Human and animal responses*

The contrast of human vs. nonvocal sounds yielded significant bilateral activations in STS (see Table 1i and Figure 4). This STS activation was very similar to the one observed in the previous contrast (human vs. nonhuman). Although cat vocalizations, contrasted with the nonvocal sounds, activated a restricted left anterior STS region (see Table 1g and Figures 4



and 5), this region responded significantly more to human than to animal vocalizations (Table 1c and Figures 4 and 5).

Insert Figure 4 about here

Insert Figure 5 about here

We further explored the left STS activation for animal vocalizations compared to the nonvocal sounds. We investigated whether this left STS response was modulated by the familiarity ratings of cat and mixed-animal stimuli. No correlation between the familiarity ratings and the left STS response to animal vocalizations was observed ( $z$ -score  $< 1$ ).

## Discussion

### *STS responses to human voices*

In this study, we investigated neural responses to human and animal vocalizations, focusing on the region of the superior temporal sulcus (STS). Previously, Belin et al (2000,2002) showed, using an fMRI block design, that circumscribed regions along the upper bank of the STS responded selectively to human voices, whether they contained linguistic information or not. This study provides further support for the existence of voice-selective areas in the human brain, using different stimuli and different participants. In addition, by using an event-related design, we can exclude potential confounding effects due to habituation and/or stimulus expectation in these responses.

The areas of voice-selective activation observed here are fairly similar to those obtained by Belin et al. (2000) (see figure 3a). Although we observed extended activation of STS / middle temporal gyrus, our peaks of activation (table 1b) correspond fairly well to the maxima reported in Belin et al. (2000), especially in the right hemisphere, with three maxima in the middle and anterior parts of STS (see coordinates from Belin et al., 2000 in figure 3a). When overlaid onto the participants' mean MR image, the activations are clearly

located within the STS region, although Talairach coordinates suggest more inferior activation, compared to Belin et al. (2000). These differences, of the order of 1cm, which are well within the variability typically observed in neuroimaging studies, could reflect individual differences between groups (e.g., Belin et al., 2002), and/or they could be due to differences in design (event-related fMRI vs. block design) and in analysis (random vs. fixed effects).

We hypothesize that the anterior activations observed here could correspond to a processing stream specialized in source identification: the 'what' stream, i.e., source recognition. Indeed, animal studies have suggested an organization of auditory cortex in processing streams, similar to that evidenced in visual cortex. Rauschecker (1998) and Rauschecker and Tian (2000) proposed the existence of an anterolaterally directed stream for the processing of auditory objects, including conspecific vocalizations. More recently, Tian et al. (2001) showed that responses of anterolateral neurons of auditory belt cortex, relative to caudolateral neurons, seemed to be more specific for stimuli such as monkey calls. In the human brain, recent neuroimaging studies support the existence of an anterior stream for sound recognition (Alain et al., 2001; Maeder et al. 2001). Our results, as well as those from Belin et al. (2000), are consistent with these studies; that is we observed more anterior activation in the STS than in the posterior part of the STS, and these could indeed reflect processes associated with identification of the source of the voice (e.g., speaker identification, Belin and Zatorre, 2003; Belin et al., 2004).

#### *Species-specificity of STS responses*

The main objective of this study was to explicitly test whether STS responses were selective for human vocalizations. To test the species-specificity of the voice-selective STS response, we included a category of animal vocalizations from several different nonhuman species. Bilateral STS regions were significantly more responsive to human vocalizations than to either animal vocalizations (see Table 1c, e, f) or nonvocal sounds (see Table 1i).

Furthermore, compared to a common baseline (i.e., nonvocal sounds), stronger STS responses were found for human vocalizations than for animal vocalizations (see Figures 4 and 5). Thus, with the present comparison set of animal, there seems to be species-specificity in the human STS response to vocalizations. It is important to emphasize that our criterion for species-selectivity is a significantly stronger response to human vocalizations than to other types of sounds, including animal vocalizations and non-vocal sounds. Thus, we cannot rule out the possibility that these regions may also respond to other types of stimuli not included in this study. In addition, in the present study, we cannot demonstrate that STS does not respond to animal vocalizations, as this would be akin to attempting to prove the null hypothesis. However, our data also indicate that some areas in the STS show a significant response to animal vocalizations as compared to the nonvocal sounds, although of much lesser magnitude than for the human sounds.

In order to test the possibility that STS responses to human voices were simply due to stimulus familiarity or homogeneity (i.e., same species), we included in our design a stimulus category entirely comprised of cat vocalizations. A comparison between human and cat vocalizations also revealed bilateral significant STS activation, similar to that between human and mixed-animal vocalizations (see Table 1e,f), providing further support for human-specificity of STS responses to vocal stimuli. Moreover, no correlation was observed between familiarity ratings and neural activity, suggesting that this measure does not account for the STS activation observed here.

#### *Animal studies of species-specificity*

Although direct comparison between results from animal and human studies is difficult, it is nonetheless informative to examine our findings in the context of studies of species-specificity in the nonhuman primate brain. Several neurophysiological studies reported that processing species-specific vocalizations is different from artificial sounds (Tian et al., 2001; Wang et al., 1995; Newman and Wollberg, 1973; Wollberg and Newman, 1973).

Fewer studies compared species-specific to heterospecific vocalizations. Wang and Kadia (2001) showed that responses of subpopulation of neurons in A1 marmoset were selective to species-specific vocalizations over backwards vocalizations, which was not observed in the cat A1 (Wang and Kadia, 2001). More recently, Poremba et al. (2004), using PET in rhesus monkeys, showed asymmetric processing of species-specific vocalizations (left > right hemisphere) within the left superior temporal gyrus, close to regions homologous to our STS peaks, which was not observed for human vocalizations and nonvocal sounds (e.g., glass breaking, tones). Thus, although cerebral responses associated with the rhesus vocalizations were not directly compared to those observed for other categories of sounds, species-specific calls appeared to be processed differently than other sounds, including heterospecific vocalizations. Thus, there is converging evidence from studies involving non-human primates suggesting that species-specific calls evoke stronger responses in some regions of auditory cortex than other types of stimuli. Given the relatively few physiological studies addressing the question of species-specific responses to vocalizations, less is currently known about the extent of such specificity in other regions of auditory cortex, particularly STS.

*Parallel with the brain responses observed for human faces*

We observed significant STS activation for animal vocalizations when contrasted with nonvocal sounds, yet this activation was smaller than that elicited by human voices (see Figures 4 and 5). This pattern is similar to what has been observed in the visual domain for the perception of human faces. Several studies have shown that faces elicit greater activation than objects in the fusiform gyrus (e.g., Kanwisher et al., 1997; McCarthy et al., 1997). Together, these results have been interpreted as evidence for the existence of a “face area” (Fusiform Face Area; FFA; Kanwisher et al., 1997). It has been suggested (Kanwisher et al., 1999) that FFA is selective for human faces but not for animals. However, Kanwisher et al. (1999) observed a significantly stronger response in this region for animal faces and whole animals than to object stimuli. Thus, the activity in the FFA was



weakest for pictures of assorted objects, stronger for the animal stimuli, and strongest for human faces. Another study (Halgren et al., 2000), using MEG, showed an occipitotemporal peak of amplitude at 165ms for animal faces, but at a level which was below its amplitude for human faces.

This notion of 'category-specific' areas in human brain has been challenged by Haxby et al. (2001; but also see Spiridon and Kanwisher, 2002). Haxby et al. (2001) proposed that the representation of faces and different categories of objects are widely distributed and overlapping, arguing that any information, regardless of type, is processed by many different parts of the brain, and that any brain region is likely to represent many different classes of information. These different notions from Haxby and Kanwisher, trying to account for the functional architecture underlying this capacity to generate a unique representation from an unlimited variety of stimuli, need to be further investigated within both the visual and auditory domains.

Moreover, as in the case of face processing, there is the question of the role of expertise: is this preference for human stimuli innate or the result of an over-exposure to those stimuli? This question is still in debate within the visual domain (Kanwisher, 2000; Tarr and Gauthier, 2000). For example, Diamond and Carey (1986) hypothesized that the mechanisms involved in face recognition are also engaged when subjects make discriminations between structurally similar exemplars of a category for which they have gained substantial visual expertise. Indeed, Gauthier et al. (2000) observed significant FFA activity with fMRI when bird experts viewed bird stimuli (but see Kanwisher, 2000). We all are all expert in processing human voices, which is one of the earliest forms of social interaction in ontogeny and the sound category that we hear most often in our daily life. It is possible that the observed STS activation associated with human vocalizations could reflect an expertise in processing categorization at a subordinate level instead of being

specific to human voices. In that case, significant STS responses to animal vocalizations should be observed in human subjects with expertise in animals.

In summary, this is the first study to present evidence suggesting selective responses to conspecific vocalizations in STS in the human brain. Bilateral STS responses were stronger to human voices (speech and nonlinguistic vocalizations) than to animal vocalizations and nonvocal sounds. The exact nature of this selectivity needs further exploration. For example, our design does not allow us to rule out the possibility that differences in cortical responses may be due to other factors, such as differences in the acoustic structures between the human and animal vocalizations acoustics, the familiarity and / or saliency of stimulus categories, the expertise for human vocalizations, or the vocal identity. Although the human vocalizations used here were produced by different speakers, these stimulus categories, unlike the other categories, might trigger a nearly automatic search for individual vocal identity. Future work will include experts in animal vocalizations to explore whether the possible search for the identity of the animal producing the sound will result in a different pattern of activation. This pattern may be similar to what we observed here: human participants, experts in human vocalizations, showed stronger activation to human vocalizations compared to categories of vocalizations for which they are not experts. The question of whether these responses reflect a hard-wired system for social communication or they are learned is still open.



### Acknowledgments

We would like to thank Marie-Hélène Grosbras and two anonymous referees for their helpful comments on earlier versions of this article, Frederick Aidan Remick, Roger Fearry, and Jean-Sébastien Robitaille for providing the stimuli, Pierre Ahad for his expertise in sound technology and computer programming, and Gilles Beaudoin, Boualem Mensour, and Jean-Maxime Leroux for help in developing the fMRI sequence. This study was supported by NSERC, Université de Montréal, and NATEQ.

## Figure captions

*Figure 1.* Stimulus categories. Left panels: average frequency distribution. x-axis: frequency (0-5000 Hz). y-axis: amplitude (84 to 36 dB). Right panels: spectrogram of a representative sample in each category. x-axis: time (variable). y-axis: frequency (0-5000 Hz). Color indicates amplitude. a: human speech stimuli. b: human nonlinguistic vocalizations. c: cat vocalizations. d: mixed-animal vocalizations. e: non-vocal stimuli. See methods for description of stimuli. Sample stimuli can be heard at: <http://www.mapageweb.umontreal.ca/belinp/>

*Figure 2.* Group-averaged peri-stimulus time response in the main auditory > baseline contrast a) in the right STG voxel [62, -28, 2], the maximal peak in the right hemisphere, and b) in the left STG voxel [-50, -30, 4], the global maximum.

*Figure 3.* a) Human vocalizations (speech and nonlinguistic vocalizations) > Nonhuman sounds (animal vocalizations (cat and mixed-animal vocalizations) and nonvocal sounds (e.g., musical instruments, environmental sounds) contrast  $p < 0.001$ . b) Group-averaged peri-stimulus time response in the Human > Nonhuman contrast in the global maximum, the right STS voxel [60, -24, -8], and in the left STS voxel [-58, -10, -12], the maximal peak on the left hemisphere. The four white circles are the maximum peaks of the contrast Human vocalizations > Nonhuman vocalizations observed in Belin et al. (2000).

*Figure 4.* a) Animal vocalizations (cat and mixed-animal vocalizations) > Nonvocal sounds (e.g., musical instruments, environmental sounds) contrast, b) Human vocalizations (speech and nonlinguistic vocalizations) > Nonvocal sounds contrast ( $p < 0.001$ ), and c) group-averaged peri-stimulus time response in the left STS voxel [-62, -16, -12], the global maximum in the Animal vocalizations (cat and mixed-animal vocalizations) > Nonvocal sounds contrast.

*Figure 5.* a) Superposition of the three main contrasts of interest (Human vocalizations (speech and nonlinguistic vocalizations) > Nonvocal sounds, Cat vocalizations > Nonvocal sounds, and Mixed-animal vocalizations > Nonvocal sounds contrasts) in the global maximum for the Cat vocalizations > Nonvocal sounds contrast, the left STS voxel [-62, -15, -13] ( $p < 0.001$ ), b) Human vocalizations (speech and nonlinguistic vocalizations) > Cat vocalizations in the right STS [60, -20, -10] ( $p < 0.001$ ) and in the left STS [-66, -36, 0] ( $p < 0.001$ ), c) and Human vocalizations (speech and nonlinguistic vocalizations) > mixed-animal vocalizations) in the right STS [56, -34, -4] ( $p < 0.001$ ) and in the left STS [-58, -12, -12] ( $p < 0.001$ ).

**Table caption**

*Table 1.* Coordinates of significant STS responses for the different contrasts of interest: coordinates are in Talairach-space; coordinates have been chosen according to their z-scores (highest peaks in each hemisphere;  $p < 0.001$ ), their location (if two local maxima were closer than 5mm, only the greater maximum was reported), and their voxel size (minimum of 3 connected voxels).

Figure 1

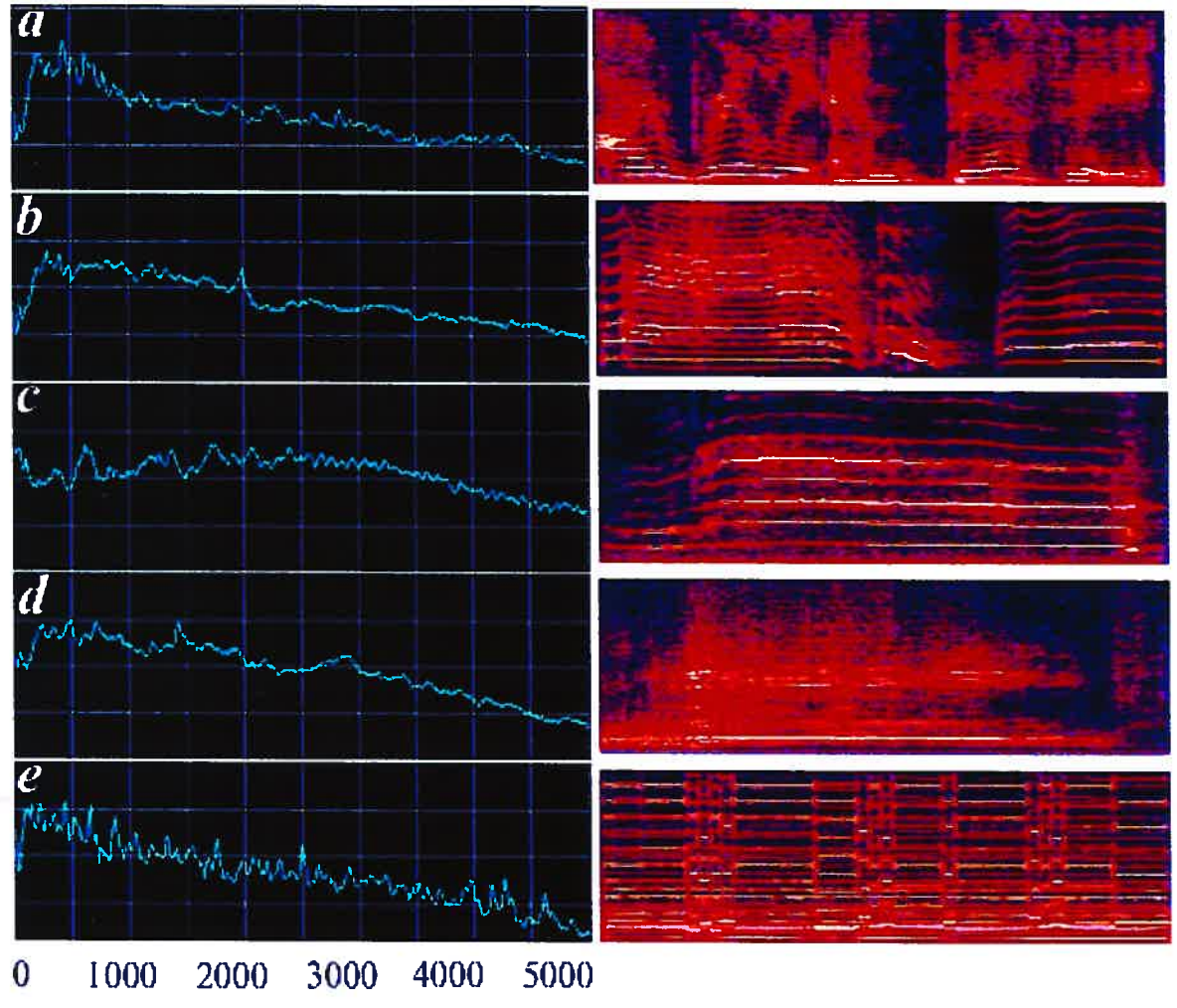


Figure 2

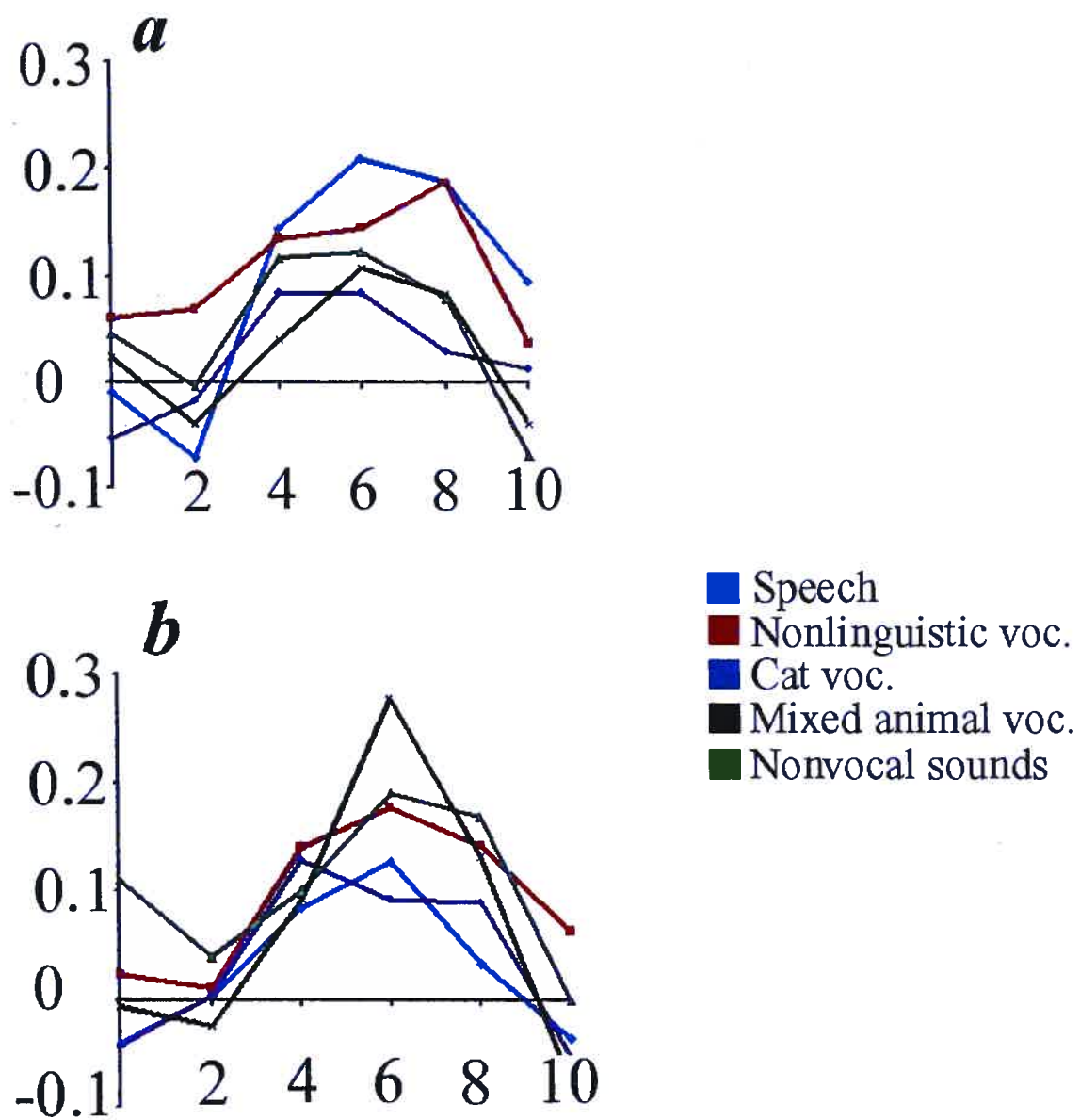




Figure 3

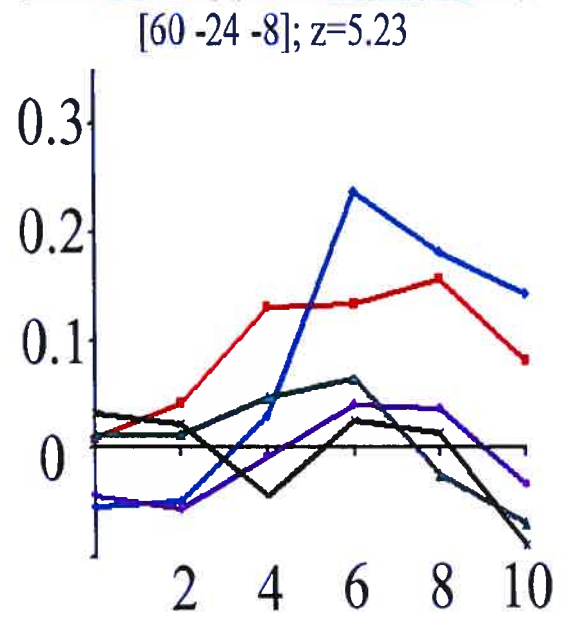
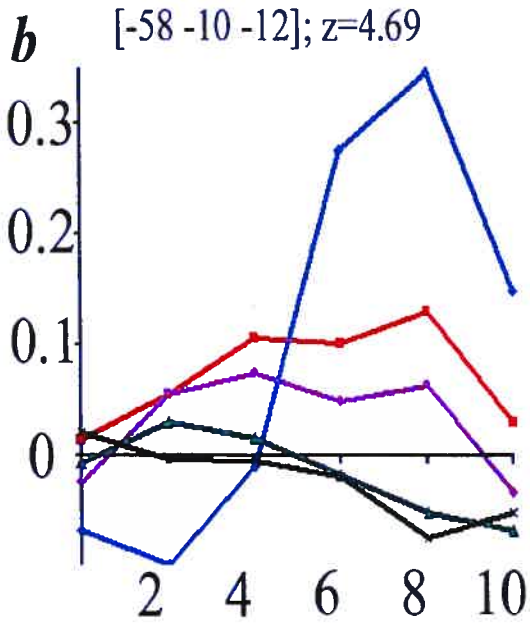
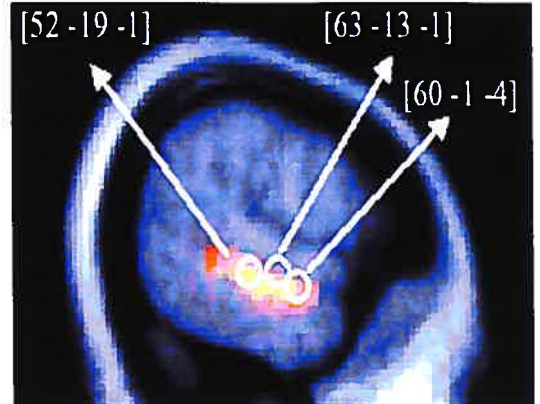
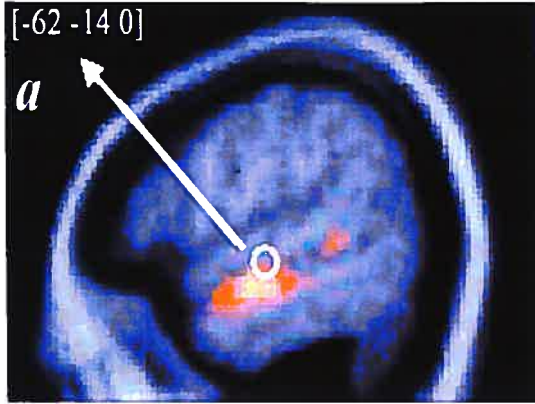


Figure 4

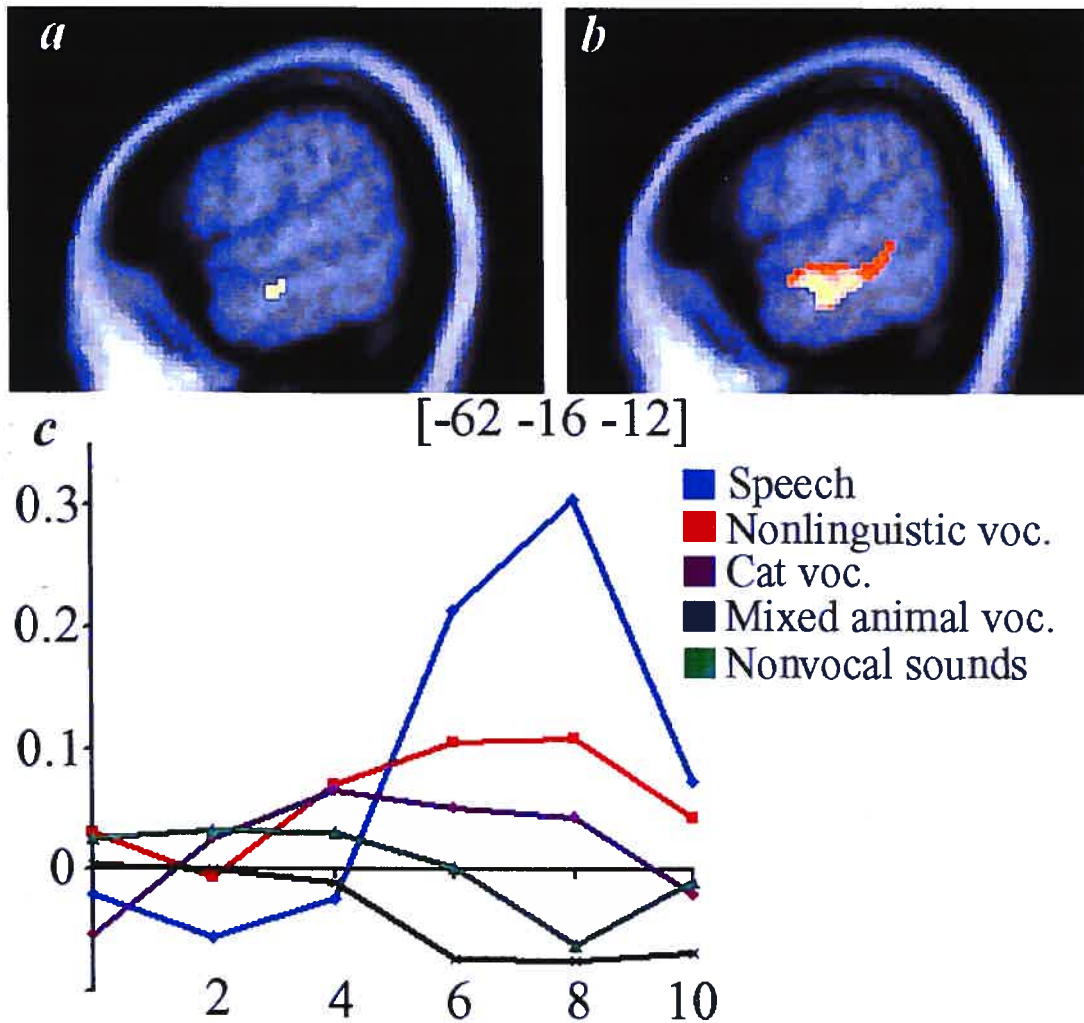


Figure 5

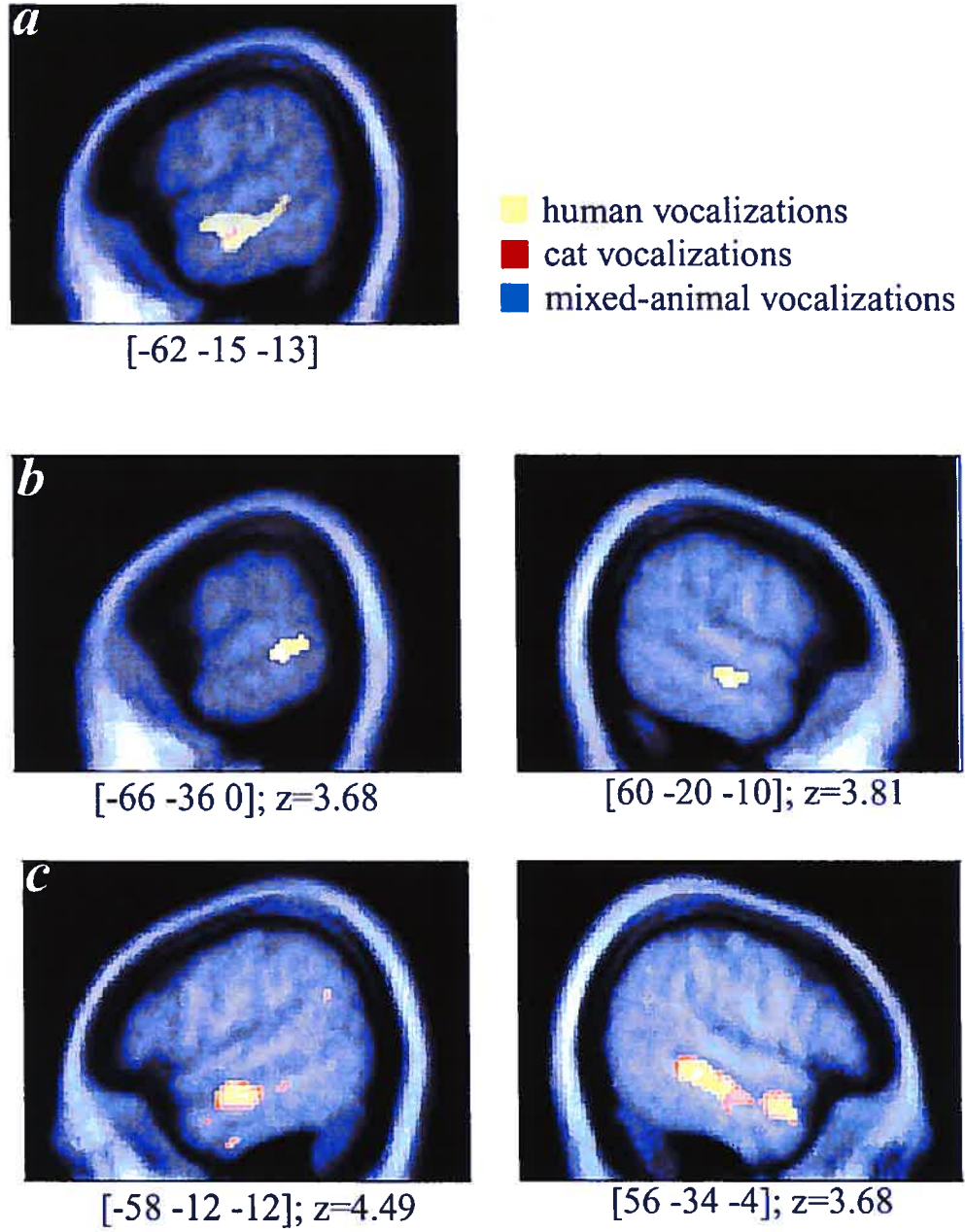


Table 1

Anatomical location	Side	Size	x	y	z	Z score
<i>(a) Main auditory activation &gt; Baseline (null-events)</i>						
superior temporal gyrus	R	3958	62	-28	2	6.19
			62	-28	-4	6.15
			64	-26	8	6.03
superior temporal gyrus	L	3509	-50	-30	4	6.34
			-54	-18	-2	5.59
			-46	-20	0	5.39
<i>(b) Human Vocalizations (speech and nonlinguistic vocalizations) &gt; Nonhuman Sounds (animal vocalizations and nonvocal sounds)</i>						
superior temporal sulcus	R	1037	60	-24	-8	5.23
			52	8	-24	5.04
			60	-14	-10	5.00
superior temporal sulcus	L	545	-58	-10	-12	4.69
			-64	-24	-10	4.41
			-66	-32	0	4.18
<i>(c) Human Vocalizations &gt; Animal Vocalizations (cat and mixed-animal vocalizations)</i>						
superior temporal sulcus	R	841	50	10	-26	4.63
			60	-24	-8	4.60
			60	-14	-10	4.34
superior temporal sulcus	L	289	-58	-10	-12	4.82
			-64	-18	-4	3.78
			-60	-20	-6	3.55
<i>(d) Animal Vocalizations (cat and mixed-animal vocalizations) &gt; Human Vocalizations</i> no significant STS response						
<i>(e) Human Vocalizations &gt; Cat vocalizations</i>						
superior temporal sulcus	R	75	60	-20	-10	3.81
			62	-24	-8	3.56
			52	-20	-16	3.41
superior temporal sulcus	L	93	-66	-36	0	3.68
			-62	-42	4	3.46
			-66	-44	6	3.35

(f) *Human Vocalizations > Mixed-animal vocalizations*

superior temporal sulcus	R	755	56	-34	-4	4.68
			60	-22	-6	4.35
			48	12	-26	4.21
superior temporal sulcus	L	286	-58	-12	-12	4.49
			-62	-14	-14	4.42
			-52	-28	-12	3.76

(g) *Cat Vocalizations > Nonvocal Sounds*

superior temporal sulcus	L	6	-62	-16	-12	3.49
--------------------------	---	---	-----	-----	-----	------

(h) *Mixed-animal Vocalizations > Nonvocal Sounds*

no significant STS response

(i) *Human Vocalizations > Nonvocal Sounds*

superior temporal sulcus	R	794	54	-24	-10	5.16
			58	-30	-8	5.12
			54	6	-20	5.10
superior temporal sulcus	L	525	-64	-26	-8	4.76
			-62	-16	-14	4.68
			-50	-38	-12	4.01

---

## References

- Alain C, Arnott SR, et al. (2000). "What" and "where" in the human auditory system. *Proc Natl Acad Sci U S A* 98(21): 12301-6.
- Altmann SA (1967). The structure of primate communication. *Social Communication Among Primates*. Altmann SA. Chicago (IL), Chicago Univ Press: 325-62.
- Andrew RJ (1963). The origins and evolution of the calls and facial expressions of the primates. *Behaviour* 20: 1-109.
- Armony JL, Dolan RJ (2001). Modulation of auditory neural responses by a visual context in human fear conditioning. *Neuroreport* 12: 3407-11.
- Armony JL, Dolan RJ (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia* 40: 817-26.
- Belin P, Fecteau S, et al. (2004). Thinking the voice: Neural correlates of voice perception. *Trends Cogn Sci* 8: 129-35.
- Belin P, Zatorre RJ (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14: 2105-9.
- Belin P, Zatorre RJ, et al. (2002). Human temporal-lobe response to vocal sounds. *Brain Res Cog Brain Res* 13: 17-26.
- Belin P, Zatorre RJ, et al. (2000). Voice-selective areas in human auditory cortex. *Nature* 403: 309-12.
- Diamond R, Carey S (1986). Why faces are and are not special: An effect of expertise. *J Exp Psychol Gen* 115: 107-17.



- Evans AC, Kamber M, et al. (1994). An MRI-based probabilistic atlas of neuroanatomy. *Magnetic Resonance Scanning and Epilepsy*. Shorvon S, Fish D, Andermann F, Bydder GM, Stefan H. New York, Plenum: 263-74.
- Fossey D (1972). Vocalizations of the mountain gorilla (*Gorilla gorilla beringei*). *Anim Behav* 20: 36-53.
- Friston KJ, Holmes AP, et al. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2: 189-210.
- Gauthier I, Skudlarski P, et al. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci* 3: 191-7.
- Gautier JP, Gautier A (1977). Communication in old world monkeys. *How Animals Communicate*. Sebeok TA. Bloomington, Indiana Univ Press: 809-964.
- Green S (1975). Variations of vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): A field study. *Developments in Field and Laboratory Research* Rosenblum LA. New York, Academic: 1-102.
- Halgren E, Raji T, et al. (2000). Cognitive response profile of the human fusiform face area as determined by MEG. *Cereb Cortex* 10: 69-81.
- Haxby JV, Gobbini MI, et al. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425-30.
- Kanwisher N (2000). Domain specificity in face perception. *Nature* 3: 759-63.
- Kanwisher N, McDermott et al. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302-11.

- Kanwisher N, Stanley D, et al. (1999). The fusiform face area is selective for faces not animals. *Neuroreport* 10: 183-7.
- Maeder PP, Meuli RA, et al. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *NeuroImage* 14: 802-16.
- McCarthy G, Puce A, et al. (1997). Face-specific processing in the human fusiform gyrus. *J Cog Neurosci* 9: 605-10.
- Newman JD, Wollberg Z (1973). Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res* 54: 287-304.
- Petersen MR (1982). The perception of species-specific vocalizations by primates: A conceptual framework. *Primate Communication*. Snowdon CT, Brown CH, Petersen MR. Cambridge (UK). Cambridge Univ Press: 171-211.
- Poremba A, Malloy M, et al. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427: 448-51.
- Rauschecker JP (1998). Parallel processing in the auditory cortex of primates. *Audiol Neurootol* 3: 86-103.
- Rauschecker JP, Tian B, et al. (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268: 111-4.
- Rauschecker JP, Tian B (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A* 97: 11800-6.
- Rendall D (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J Acoust Soc Am* 113: 3390-402.

- Scherer KR, Kappas A (1988). Primate vocal expression of affective states. *Primate Vocal Communication*. Todt D, Goedeeking P, Newman E. Heidelberg, Springer: 171-94.
- Seyfarth RM, Cheney DL, et al. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science Wash. DC* 210: 801-3.
- Smith H, Newman JD, et al. (1982). Vocal concomitants of affiliative behavior in squirrel monkeys. *Primate Communication*. Snowdon CT, Brown CH, Petersen MR. Cambridge (UK), Cambridge Univ Press: 30-49.
- Snowdon CT (1982). Linguistic and psycholinguistic approaches to primate communication. *Primate Communication*. Snowdon CT, Brown CH, Petersen MR. Cambridge (UK), Cambridge Univ Press: 212-38.
- Spiridon M, Kanwisher N (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron* 35: 1157-65.
- Struhsaker TT (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). *Social Communication Among Primates*. Altmann SA. Chicago (IL), Chicago Univ Press: 238-324.
- Talairach P, Tournoux J (1988). *Stereotaxic Coplanar Atlas of the Human Brain*. Thieme, Stuttgart (Germany).
- Tarr MJ, Gauthier I (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature* 3: 764-69.
- Tian B, Reser D, et al. (2001). Functional specialization in rhesus monkey auditory cortex. *Science* 292: 290-3.

- van Lawick-Goodall JA (1968). A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. *Primates: Studies in Adaptation and Variability*. Jay PC. (New York, Holt: Rinehart & Winston: 313-74.
- Wang X, Kadia SC (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol* 86: 2616-20.
- Wang X, Merzenich M, et al. (1995). Representation of a species-specific vocalizations in the primary auditory cortex of the common marmoset: Temporal and spectral characteristics. *J Neurophysiol* 74: 2685-2706.
- Winter P, Ploog D, et al. (1966). Vocal repertoire of the squirrel monkey (*Saimiri sciureus*), its analysis and significance. *Exp Brain Res* 1: 359-84.
- Wollberg Z, Newman JD (1973). Auditory cortex of squirrel monkey: Response patterns of single cells to species-specific vocalizations. *Science* 175: 212-4.
- Worsley KJ, Friston KJ (1995). Analysis of fMRI time-series revisited-again. *NeuroImage* 2: 173-81.

**CHAPITRE 5**

**ARTICLE 4**

**RÉPONSES NEURONALES ASSOCIÉES À LA  
PERCEPTION DE VOCALISATIONS  
ÉMOTIONNELLES: UNE ÉTUDE D'IRMf  
ÉVÉNEMENTIELLE**

**ARTICLE 4**

**NEURAL RESPONSES TO EMOTIONAL NONLINGUISTIC  
VOCALIZATIONS, AS REVEALED BY EVENT-RELATED  
FMRI**

**Shirley FECTEAU, Pascal BELIN,**

**Yves JOANETTE & Jorge L. ARMONY**

*Eur J Neurosci* (to be submitted)



Neural responses to emotional nonlinguistic vocalizations as revealed by event-related fMRI

Shirley Fecteau<sup>1,2,3</sup>, Pascal Belin<sup>2,3</sup>, Yves Joanette<sup>2,3</sup>, Jorge L. Armony<sup>3,4</sup>

<sup>1</sup> Département de psychiatrie, Université de Montréal, Montreal, Canada

<sup>2</sup> Centre de recherche de l'Institut universitaire de gériatrie de Montréal (CRIUGM), Montreal, Canada

<sup>3</sup> Centre de recherche en neuropsychologie expérimentale et cognition (CERNEC), Université de Montréal, Montreal, Canada

<sup>4</sup> Douglas Hospital Research Center, McGill University, Montreal, Canada

Correspondence should be addressed to Shirley Fecteau,  
Département de psychologie, Université de Montréal, C.P. 6128, Succ. Centre-ville,  
Montreal, Que., H3C 3J7, Canada

E-mail: shirley.fecteau@umontreal.ca

### Abstract

This study used event-related functional magnetic resonance imaging to identify the neural networks underlying emotional vocalizations processing. Fifteen healthy participants were scanned while performing a sex identification task. Stimuli consisted of emotional vocalizations: laughs, sexual vocalizations, cries, fearful screams, as well as emotionally neutral vocalizations, which were previously validated. Overall, results revealed that processing the emotional vocalizations, relative to the neutral ones, elicited activity mainly in the temporal regions. Moreover, differential patterns of activity were observed according to the type of emotional vocalization. The laughs and the sexual vocalizations induced greater responses in the amygdala, and the hypothalamus was more activated for the cries. These findings suggest that emotional information in vocalizations is processed following similar principles as for facial expressions literature.

### Introduction

Nonlinguistic vocalizations, such as laughs, gasps, and screams, represent an important source of information about the speaker, especially regarding his/her emotional state (Kreiman, 1997). Thus, the perception and interpretation of emotions conveyed through vocalizations is an essential aspect of social interactions. Vocalizations, part of our phylogeny and ontogeny, are one of the most primary means of communicating emotions (Barr et al., 2000) and the use of these stimuli allows the exploration of emotional processing of auditory stimuli that goes beyond linguistic codes.

Recently, functional neuroimaging studies have begun to explore the brain regions associated with the perception of nonlinguistic emotional vocalizations. Among other regions, emotional vocalizations were found to mainly engage temporal and frontal cortices, as well as the amygdala. Specifically, temporal areas showed increased signals in response to vocalizations of fear and disgust (Phillips et al., 1998; Morris et al., 1999), as well as to laughs and cries (Sander and Scheich, 2001), whereas frontal regions responded to vocalizations of fear and disgust (Phillips et al., 1998; Morris et al., 1999). Within the amygdala, significant activations were observed for laughs, and cries (Sander and Scheich, 2001). With respect to vocalizations of fear, one study reported increased activity (Phillips et al., 1998), whereas a second one found decreases in activity (Morris et al., 1999) vocalizations of fear.

These three studies (Phillips et al., 1998; Morris et al., 1999; Sander and Scheich, 2001) were novel and important contributions to the field of emotional processing. However, some methodological considerations have to be mentioned as they could have influenced their results. All studies used a block design and only two different speakers produced the stimuli. This may be particularly critical, as some brain regions are very sensitive to habituation when the same emotional category is presented (Fischer et al., 2003) and when the same speaker produces the vocalizations (Belin and Zatorre, 2003). Moreover, these

studies compared emotional vocalizations with mildly happy vocalizations (Phillips et al., 1998), voiced nasal sounds (Morris et al., 1999), and silence (Sander and Scheich, 2001). The use of emotionally neutral vocalizations, such as coughs, may be a better control condition to which contrasts the emotional sounds. For instance, within the visual domain, differential patterns of activity in response to facial expressions have been reported, depending on whether neutral faces or non-facial stimuli were used as the neutral condition (Iidaka et al., 2002; see Zald, 2003 for a discussion).

We therefore conducted an event-related fMRI study to further explore the brain responses to emotional nonlinguistic vocalizations. Importantly, we used four different emotional categories (laughs, sexual vocalizations, cries and fearful screams), as well as emotionally neutral vocalizations (e.g., coughs, clearing throat), with each sound produced by a different speaker.

## Methods

### *Participants*

Healthy right-handed adults ( $n = 15$ , six women; mean age of 22.6 years;  $SD = 2.7$ ) with no history of neurological or psychiatric condition performed a sex identification task of vocalizations during scanning. They were not explicitly informed of the emotional experimental variable. Informed written consent was obtained and the ethical committee from *Centre Hospitalier de l'Université De Montréal* approved the study.

### *Procedure*

Vocal stimuli were delivered binaurally through pneumatic headphones using MCF (DigiVox, Montreal) with a sound-pressure level of 85-90 dB. The headphones were sealed by foam ear inserts and further shielded by plastic ear defenders which provided an attenuation of fMRI scanning noise of about 30dB. Stimuli ( $n = 144$ ) consisted of five categories of vocalizations: laughs ( $n = 24$ ), sexual vocalizations ( $n = 24$ ), cries ( $n = 24$ ),

fearful screams ( $n = 24$ ) and neutral vocalizations ( $n = 48$ ). Stimuli were previously validated by a group of 60 individuals who rated the stimuli on emotional valence (Fecteau et al. submitted). Each vocalization was produced by a different adult speaker (half male and half female). Duration were matched across stimulus categories (mean duration of neutral vocalizations = 1.5s, SD = 0.4s; laughs = 1.6s, SD = 0.4s; sexual vocalizations = 1.6s, SD = 0.4s; cries = 1.7s, SD = 0.4s; fearful screams = 1.7s, SD = 0.4s; all pairwise comparisons  $p > 0.7$ ). All stimuli were equalized for maximum peak values with Mitsyn (WLH) and CoolEdit Pro (Syntrillium Software Corp.). Stimuli were presented twice in a pseudo-random order (mean SOA = 5.5s, S.D = 2.4s). In order to avoid habituation and/or stimulus expectation effects, we ensured that stimuli from the different categories were distributed equally through the duration of the experiment and that no more than two vocalizations from the same category (e.g., laughs) were presented together. Null-events (Schacter et al. 1997) were included to space out stimuli and to reduce the participant's expectations to hear a sound.

After the scanning session, participants were asked to rate all the vocalizations on valence of the emotion expressed using visual analog scales. As found with the independent group of participants who rated previously rated the stimuli, significant differences in valence ratings between stimuli of positive valence, those of negative valence and neutral stimuli were observed (mean rating for neutral vocalizations = 45.9, SD = 1.5; laughs = 83.1, SD = 1.7; sexual vocalizations = 78.2, SD = 1.6; cries = 14.3, SD = 1.0; fearful screams = 14.7, SD = 1.4; all pairwise comparisons  $p > 0.1$ ). The order of stimulus presentation was counterbalanced across participants.

### *Neuroimaging Procedures*

Scanning was performed on a 1.5 Tesla MRI system (Magnetom Vision, Siemens Electric, Erlangen, Germany) at the *Centre Hospitalier de l'Université de Montréal*. Functional scans were acquired with a single-shot echoplanar gradient-echo (EPI) pulse sequence (TR



= 2,6s, TE = 40 ms, flip angle = 90°, FOV = 215 mm, Matrix = 128 x 128). The 28 axial slices (resolution 3.75 x 3.75 mm in plane, 5 mm between planes) in each volume were aligned with the AC-PC line, covering the whole brain. A total of 320 volumes were acquired (the first four volume of each series were later discarded to allow for T1 saturation). After functional scanning, T1-weighted anatomical images were obtained for each participant, (1mm x 1mm x 1mm resolution). Scanner noise was continuous throughout the experiment providing a constant background.

#### *Data Analysis*

Image processing and statistical analysis were performed using SPM99 (Wellcome Department of Cognitive Neurology; Friston et al. 1995; Worsley and Friston, 1995), as done in previous studies (Armony and Dolan 2001; Armony and Dolan 2002). The imaging time-series was realigned to the first volume to correct for interscan movement. Time-series were shifted by sinc interpolation to correct for the slice acquisition times. The functional images were spatially normalized to a standard stereotaxic space (Talairach and Tournoux 1988) based on a template provided by the Montreal Neurological Institute (Evans et al. 1994) and smoothed spatially using a 8mm isotropic Gaussian kernel to compensate for residual inter-participant variability and to allow for the application of Gaussian random effect field theory in the statistical analysis (Friston et al. 1995). Each subject's structural scan was coregistered with the mean realigned functional image and normalized using the parameters determined for the functional images. A mean anatomical image was created from the participants' individual scans, onto which activation were overlaid for anatomical localization.

Data analysis was performed in a two-stage mixed-effects analysis (equivalent to a random effects analysis) in which BOLD responses for each subject for the five categories (neutral, sexual, laughter, cries, screams) were first modeled using a synthetic haemodynamic function in the context of the fixed-effects general linear model. Subject-



specific linear contrasts on the parameter estimates were then entered into a second-level analysis to perform between-subjects analyses, resulting in a t-statistic for each voxel. These t-statistics (transformed to Z-statistics) constitute a statistical parametric map (SPM). SPM's were thresholded at  $p = 0.001$  (uncorrected). Only regions with an uncorrected P value of 0.001 at the voxel-level and of 0.01 at the cluster-level were considered, in order to minimize the risk of type I errors (Friston et al., 1994). Thus all clusters were  $> 38$  voxels in size, except for the amygdala, where a  $p < 0.001$  (uncorrected) at the voxel-level was used, based on our *a priori* hypothesis on the involvement of this region in emotional vocalizations processing (Phillips et al., 1998; Morris et al., 1999; Sander and Scheich, 2001).

## Results

### *Brain responses to emotional vocalizations*

The emotional stimuli (laughs, sexual vocalizations, cries, and fearful screams), relative to the neutral vocalizations (e.g., coughs, clearing throat), induced greater activation in bilateral auditory cortex, middle part of left superior temporal gyrus (STG) (Figure 1), anterior part of both STG, left cerebellum, and right anterior cingular cortex (Table 1a). The opposite contrast did not reveal any significant activation (Table 1b).

Insert Table 1 about here

Insert Figure 1 about here

### *Brain responses to specific emotional vocalizations*

To further explore the emotional perception of vocalizations, we compared the specific categories of emotional vocalizations to a common category, the neutral vocalizations. Increases in neural activity associated with perception of the laughs were observed in anterior segment of right STG, left inferior parietal lobule, and left postcentral gyrus (Table 1c). The sexual vocalizations elicited greater activity in right STS, anterior segment of left

STG, right cuneus, left cerebellum, left inferior frontal gyrus, and left lingual gyrus. Brain regions associated with stimuli of cry were bilateral auditory cortex, right precuneus, right putamen, right claustrum, bilateral cerebellum, right hypothalamus (Figure 2c), anterior part of right STG, as well as left postcentral gyrus (Table 1d). The fearful screams induced greater activity in right superior frontal gyrus, middle part of left STG, bilateral cuneus, middle segment of right STS, dorsal frontal gyrus, left middle occipital gyrus, and left cingular gyrus (Table 1f).

Based on the literature, we expected to observe activation in amygdala for the emotional vocalizations, relative to the neutral ones. Indeed, , activation of the amygdala was observed for the laughs and for the sexual vocalizations, relative to the neutral ones, that was significant using the small volume correction compatible with our a-priori hypothesis ( $p < 0.001$ ; Figure 2a,b).

Insert Figure 2 about here

### Discussion

In this study we investigated the neural-related responses to nonlinguistic emotional vocalizations in healthy individuals. Our findings suggest that auditory cortex seems to play a role in the perception of emotional vocalizations in general, as its responses were highly significant for the emotional vocalizations (laughs, sexual vocalizations, cries, and fearful screams), compared to the emotionally neutral ones. In contrast, frontal regions, amygdala, and hypothalamus were activated by specific types of emotional vocalizations (discussed below)

#### *Temporal regions*

We observed significant responses within the auditory cortex in the comparison between emotional and the neutral vocalizations (Figure 1 and Table 1). These results suggest that

early processing of auditory information is sensitive to the emotional content of stimuli. Similar findings have been reported in the facial expression literature. The fusiform face area (FFA), which has been associated with processing basic facial information (e.g., identity; Hoffman and Haxby, 2000), seems to be modulated by their emotional expression (Dolan et al., 2001; Rotshtein et al., 2001; Vuilleumier et al., 2001). In addition, our findings are in line with data from classical conditioning studies showing enhanced auditory cortex responses to auditory stimuli (e.g., tones) with an acquired aversive value (Büchel et al., 1998; Armony and Dolan, 2001).

We also observed enhanced right middle STS responses in the contrast between emotional and neutral vocalizations (Table 1 and Figure 1). Previous studies have reported anterior and middle STS activity as being selective to human vocalizations relative to non-vocal sounds (Belin et al., 2000) and animal non-human vocalizations (Fecteau et al., submitted), and its more anterior part is sensitive to speaker adaptation (Belin and Zatorre, 2003). Our present results suggest that right middle STS is also sensitive to the emotional load expressed by human vocalizations. This middle STS activation is slightly posterior to the region that has been named the *what* stream, involved in source recognition (Alain et al., 2001; Maeder et al. 2001).

Although to a lesser extent than the STS, enhanced STG responses were observed for the emotional versus the neutral vocalizations (Figure 1 and Table 1). This region was the only one commonly activated by all four emotional categories. These findings, consistent with previous studies using vocalizations (Phillips et al., 1998) and prosody (Kawashima et al., 1993; Buchanan et al., 2000; Wildgruber et al., 2002; Mitchell et al., 2003), provide further support to the hypothesis that STG plays a role in the extraction of general emotional information, whereas other brain regions are more involved in the processing of specific emotions (Phillips et al., 1998).

### *Amygdala*

A large literature exists on the role of the amygdala in the processing of emotional stimuli. In particular, most neuroimaging studies have reported amygdala activation in response to emotional facial expressions (see review by Davidson and Irwin 1999; but see Canli et al., 1998). The involvement of the amygdala in the perception of emotional information in the auditory domain is, however, less clear. Studies using prosodic (Imaizumi et al., 1997; Royet et al., 2000) and music stimuli (Blood et al., 1999; Blood and Zatorre, 2001) failed to observe amygdala involvement in emotional processing (see the meta-analysis of Phan et al., 2002), but some studies using vocalizations (Phillips et al., 1998; Sander and Scheich, 2001) did find significant amygdala activation.

Here, we observed amygdala responses associated with emotional vocalizations, compared to neutral ones, but only for those with positive valence, right amygdala for laughs and left amygdala for sexual sounds. These findings may be of particular interest, as the general role of the amygdala in processing positive emotions is still under debate (Canli et al., 1998; Garavan et al., 2000). However, a growing literature (see Baxter and Murray 2002 and Zald 2003 for reviews) supports a role of the amygdala in the processing of positively-valenced information, using a variety of stimuli, such as dynamic facial expressions (Sato et al., 2004), static facial expressions (Breiter et al., 1996; Canli et al., 2002; Garavan et al., 2001; Gorno-Tempini et al., 2001; Killgore and Yurgelun-Todd, 2001; Gur et al., 2002; Hamann et al., 2002; Pessoa et al., 2002; Wright et al., 2002; Yang et al., 2002; Winston et al., 2003; Somerville et al., 2004), linguistic non-auditory stimuli (Hamann and Mao, 2002), olfactory stimuli (Gottfried et al., 2002a,b), taste (O'Doherty et al., 2001, 2002), laughs (Sander and Scheich, 2001), erotic film excerpts (Beauregard et al., 2001; Karama et al., 2002), rewards (Zalla et al., 2000; Knutson et al., 2001a,b), and food stimuli (LaBar et al., 2001). Clinical studies also reported deficits in processing positive-valenced stimuli in individuals with amygdala damage (Schneider et al., 1997; Whalen et al., 1998; Morris et al., 1996). The amygdala activations during exposure to the laughs and



the sexual vocalizations found in this study support its role for processing stimuli of positive valence.

In contrast, we found no significant activation within the amygdala associated with vocalizations expressing negative emotions, namely cries (sadness) and screams (fear). This is somewhat surprising, given the wealth of neuroimaging studies showing amygdala responses to negative visual stimuli, particularly fearful facial expressions (see review by Davidson and Irwin 1999). It is important to note, however, that the previous neuroimaging studies of emotional vocalizations yielded conflicting results: whereas two studies reported significant amygdala responses to fearful vocal sounds (Phillips et al., 1998) and cries (Sander and Scheich, 2001), Morris et al. (1999) found a *decrease* in amygdala activation for fearful voices, compared to happy, sad, and neutral vocalizations (voiced nasals). Thus, further research is necessary to elucidate the precise role of the amygdala in the processing of auditory information with negative emotional valence. Of particular interest will be to assess whether there is a differential involvement of the amygdala in processing emotional information based on stimulus modality.

### *Frontal regions*

The contrast of emotional vs. neutral vocalizations revealed significant activation in the anterior cingulate cortex. In addition, sexual vocalizations activated the inferior frontal gyrus, whereas fearful screams elicited responses within the superior and dorsal frontal regions. Previous studies have reported significant frontal activity associated with fearful vocalizations, specifically in the medial frontal cortex (Phillips et al., 1998) and in the superior frontal gyrus and ventral prefrontal cortex (Morris et al., 1999). Morris et al. (1999) suggested that these frontal activations reflected an inhibitory input on the amygdala (Al-Maskati et al., 1989; Zbrozyna et al., 1991), resulting in a decrease of amygdala activity for fearful screams. This hypothesis is consistent with our findings, as there was a lack of significant amygdala responses to fearful screams, accompanied by a strong frontal

activity. A similar inhibitory role of frontal regions associated with sexual arousal has also been reported. Specifically, superior frontal activity was found when individuals attempted to inhibit sexual arousal (Beauregard et al., 2001). In addition, the ventro-medial frontal cortex has been postulated to be involved in functions of control and inhibition of socially unacceptable emotional responses (see review of Gainotti, 2001).

Alternatively, the frontal activations observed in this study might represent an increased attention specific to emotional vocalizations. Previous studies have shown that emotional stimuli engage attentional resources in an automatic fashion, even when the emotional information of the stimuli is irrelevant to the task (Armony and Dolan, 2001).

#### *Hypothalamus*

We found hypothalamus activation for the cries when contrasted with the neutral vocalizations (Table 1g and Figure 3c). Although this finding was not expected, it is nevertheless consistent with previous findings. Studies using the sadness induction method (Lane et al., 1997; Damasio et al., 2000; Eugène et al., 2003) and recall of personal sad episodes (Damasio et al., 2000) reported activation in the hypothalamus. These results are also in line with the finding that hormonal changes controlled from hypothalamic nuclei are an integral part of depression, a pathological mood state in which sustained sadness is a dominant factor (Gibbons and McHugh, 1962; Sachar et al., 1970; Carpenter and Bunney, 1971; Carroll et al., 1976; Linkowski et al., 1985; also see review of Makino et al., 2002).

#### *Methodological considerations and future directions*

In contrast to the other studies exploring emotional vocalizations (Phillips et al., 1998; Morris et al., 1999; Sander and Scheich, 2001), we used an event-related design in which the stimuli were produced by different speakers and presented in a pseudo-random order to diminish habituation of cerebral signal. Nonetheless, some potential limitations of this study need to be discussed. First, the lack of physiological measures and / or mood



questionnaires prevents us from distinguishing between emotional perception and mood. It is possible that emotional stimuli have changed the emotional state of the participants, although the rapid, counterbalanced presentation of stimuli of different emotional valence makes this unlikely.

Although we used a larger number of emotional categories than previous studies, it still constitutes a fairly limited subset of all possible emotional nonlinguistic vocalizations that are used in social interactions (Kreiman, 1997). Thus, it would be interesting to include other forms of emotional vocalizations (such as those expressing anger).

Finally, we used a sex identification task, which could have had an effect on some of the observed activations. For example, STS appears to be task-dependent, that is STS has been shown to be activated during direct emotional task (e.g., judgment of the type of emotion) and not during incidental task (e.g., sex identification) (Winston et al., 2003). Thus, as we used a sex identification task, it is possible that a direct task would yield greater STS responses than those observed, and may then show sensitivity to the emotional vocalizations category. The amygdala responses have also been shown to be sensitive to the task used. Gur et al. (2002) observed greater limbic response during the emotion discrimination (indicating whether the emotion was positive or negative) than age discrimination task. Still, most studies reporting amygdala activation to emotional stimuli have employed incidental (non-emotional) tasks (Morris et al., 1998; Whalen et al., 1998; Sander and Scheich, 2001; Vuilleumier et al., 2001; Dolan, 2002; Ohman, 2002; Winston et al., 2002; Winston et al., 2003). Within the auditory domain, Sander and Scheich (2001) found no difference in the magnitude of amygdala responses between self-induction emotions and pitch-shifts detection task. Moreover, a recent meta-analysis (Baas et al., 2004) suggests that amygdala activation is not significantly related to task instructions.

### Acknowledgments

We would like to thank Frederick Aidan Remick, Roger Fearry, and Jean-Sébastien Robitaille for providing the stimuli, Pierre Ahad for his expertise in sound technology and computer programming, and Gilles Beaudoin, Boualem Mensour, and Jean-Maxime Leroux for helping developing the fMRI sequence. This study was supported by NSERC, Université de Montréal, and NATEQ.

## Figure captions

*Figure 1.* Superposition of the four main contrasts of interest (Laughs > Neutral vocalizations; Sexual > Neutral vocalizations; Cries > Neutral vocalizations; Fearful screams > Neutral vocalizations; ( $p < 0.001$ )) of the maximal peaks of the Emotional > Neutral vocalizations contrast: a) the left voxel [-36, -34, 6] and size of activation for this voxel, b) in the right voxel [54, -14, 8] and size of activation for this voxel. x -axis: stimulus categories. y-axis: effect size.

*Figure 2.* a) Laughs > Neutral vocalizations contrast of the maximal peak in the right amygdala voxel [22 -4 -26], z-score = 3.95,  $p < 0.001$ , and size of activation for this voxel. x -axis: stimulus categories. y-axis: effect size. b) Sexual > Neutral vocalizations contrast of the maximal peak in the left amygdala voxel [-26, 6, -26], z-score = 3.63,  $p < 0.001$ , and size of activation for this voxel. x -axis: stimulus categories. y-axis: effect size. c) Cries > Neutral vocalizations contrast in the maximal peak in the hypothalamus, the right voxel [4, -2, -6], z-score = 4.04,  $p < 0.001$ , group-averaged peri-stimulus time response of this right voxel. x -axis: time (second). y-axis: effect size, and size of activation for the this right voxel. x -axis: stimulus categories. y-axis: effect size.

**Table caption**

*Table 1.* Coordinates of significant brain responses of regions for the different contrasts of interest: coordinates are in Talairach-space; coordinates have been chosen according to their voxel-level ( $p > 0.001$  uncorrected) and cluster-level ( $p > 0.01$  uncorrected). L: left, R: right.

Figure 1

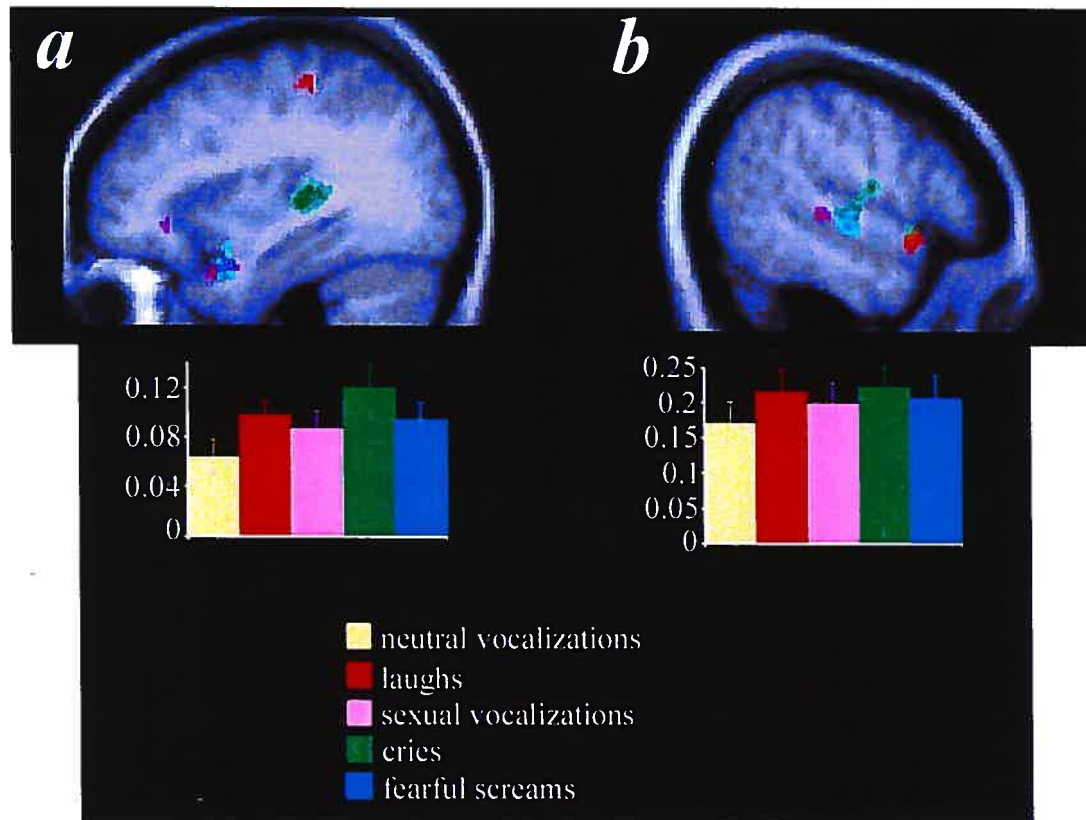


Figure 2

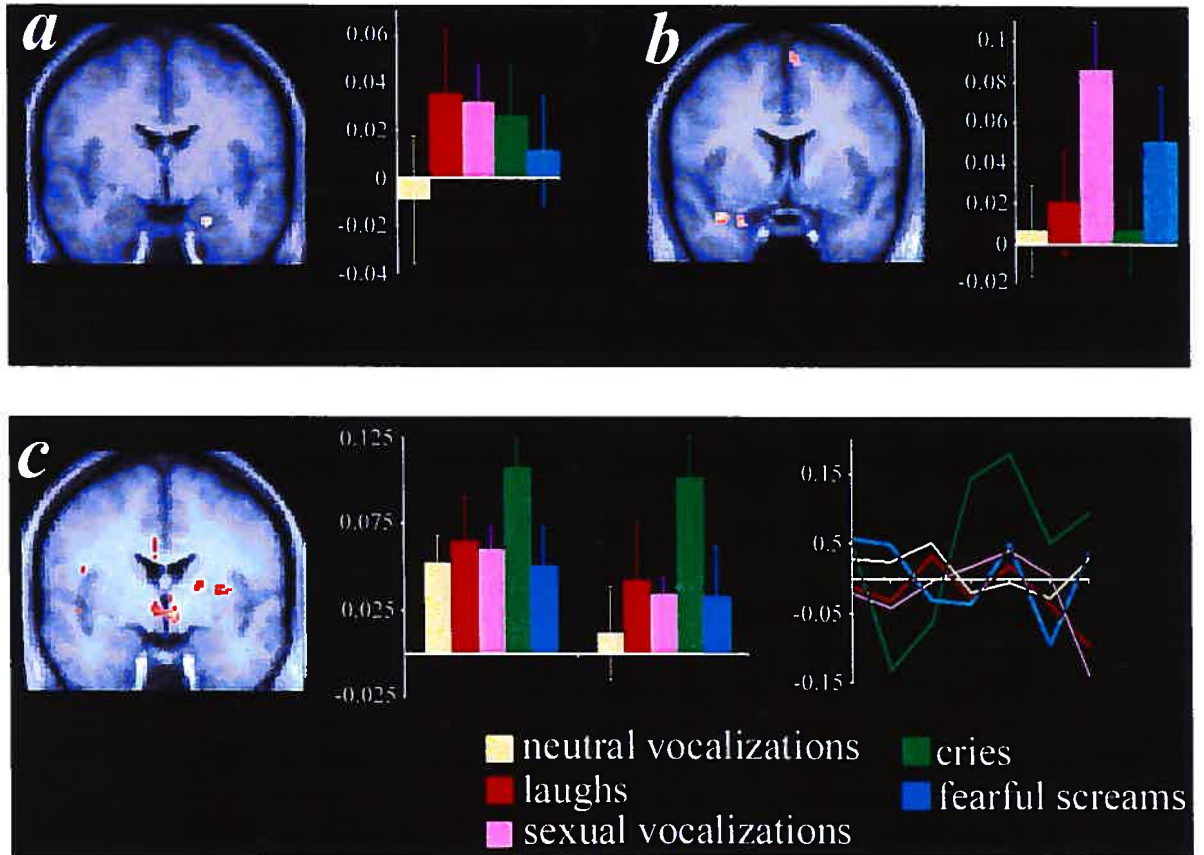




Table 1

Anatomical location	Side	Size	x	y	z	Z score
<i>(a) Emotional (laughs, sexual voc., cries, fearful screams) &gt; Neutral vocalizations</i>						
primary auditory cortex / middle superior temporal sulcus	R	536	54	-14	8	5.20
middle superior temporal gyrus	L	189	-36	-34	6	5.17
cerebellum	L	40	-10	-58	-38	4.33
anterior superior temporal gyrus	L	97	-38	6	-24	4.29
anterior superior temporal gyrus	R	101	54	10	-8	4.10
cerebellum / lingual gyrus	L	54	-10	-50	-2	4.24
primary auditory cortex	L	41	-54	-14	6	3.74
anterior cingular cortex	R	45	2	-4	38	3.60
<i>(b) Neutral &gt; Emotional vocalizations (laughs, sexual voc., cries, fearful screams)</i>						
no significant response						
<i>(c) Laughs &gt; Neutral vocalizations</i>						
anterior superior temporal gyrus	R	47	56	12	-8	4.32
inferior parietal lobule	L	76	-48	-30	54	4.29
postcentral gyrus	L	42	-40	-26	56	3.79
<i>(d) Neutral vocalizations &gt; Laughters</i>						
no significant response						
<i>(e) Sexual vocalizations &gt; Neutral vocalizations</i>						
superior temporal sulcus / primary auditory cortex	R	152	46	-28	2	4.86
anterior superior temporal gyrus / temporal pole	L	48	-38	6	-24	3.95
cuneus	R	188	12	-80	20	3.92
cerebellum	L	57	-22	-44	-22	3.78
inferior frontal gyrus	L	114	-46	32	-4	3.78
lingual gyrus	L	48	-12	-50	0	3.69

(f) *Neutral vocalizations > Sexual vocalizations*

no significant response

(g) *Cries > Neutral vocalizations*

primary auditory cortex / planum temporale	L	325	-34	-26	6	5.19
primary auditory cortex	R	509	36	-22	6	4.95
precuneus	R	124	16	-64	38	4.77
putamen	R	66	16	2	6	4.51
claustrum	R	76	34	0	8	4.27
cerebellum	L	38	-12	-58	-40	4.06
hypothalamus	R	113	4	-2	-6	4.04
anterior superior temporal gyrus	R	58	54	10	-6	3.94
gyrus postcentral	L	48	-16	-38	58	3.93
cerebellum	R	50	20	-60	-48	3.81

(h) *Neutral vocalizations > Cries*

no significant response

(i) *Fearful screams > Neutral vocalizations*

superior frontal gyrus	R	176	4	34	48	5.06
middle superior temporal gyrus	L	72	-36	6	-26	4.56
cuneus	R	119	4	-80	38	4.35
middle superior temporal sulcus	R	240	50	-20	-2	4.22
dorsal frontal gyrus	R/L	66	0	18	64	3.87
middle occipital gyrus	L	41	-28	-80	14	3.84
cuneus	L	49	-4	-86	6	3.74
cingular gyrus	L	43	8	32	24	3.60

(j) *Neutral vocalizations > Fearful screams*

no significant response

## References

- Alain C, Arnott SR, et al. (2000). "What" and "where" in the human auditory system. *Proc Natl Acad Sci U S A* 98(21): 12301-6.
- Armony JL, Dolan RJ (2001). Modulation of auditory neural responses by a visual context in human fear conditioning. *Neuroreport* 12(15): 3407-11.
- Armony JL, Dolan RJ (2002). Modulation of spatial attention by fear-conditioned stimuli: An event-related fMRI study. *Neuropsychologia* 40(7): 817-26.
- Baas D, Aleman A, et al. (2004). Lateralization of amygdala activation: A systematic review of functional neuroimaging studies. *Brain Res Brain Res Rev* 45(2): 96-103.
- Barr RG, Hopkins B, et al. (2000). *Crying as a sign, a symptom, & a signal*. New York, Cambridge University Press.
- Baxter MG, Murray EA (2002). The amygdala and reward. *Nat Rev Neurosci* 3(7): 563-73.
- Beauregard M, Levesque J, et al. (2001). Neural correlates of conscious self-regulation of emotion. *J Neurosci* 21(18): RC165.
- Belin P, Zatorre RJ (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14(16): 2105-9.
- Belin P, Zatorre RJ, et al. (2000). Voice-selective areas in human auditory cortex. *Nature* 403: 309-12.
- Blood AJ, Zatorre RJ (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci U S A* 98(20): 11818-23.

- Blood AJ, Zatorre RJ, et al. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci* 2(4): 382-7.
- Breiter HC, Etcoff NL, et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17(5): 875-87.
- Buchanan TW, Lutz K, et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Brain Res Cogn Brain Res* 9(3): 227-38.
- Buchel C, Morris J, et al. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron* 20(5): 947-57.
- Canli T, Sivers H, et al. (2002). Amygdala response to happy faces as a function of extraversion. *Science* 296: 2191.
- Canli T, Desmond JE, et al. (1998). Hemispheric asymmetry for emotional stimuli detected with fMRI. *Neuroreport* 9(14): 3233-9.
- Carpenter WT, Bunney WE (1971). Adrenal cortical activity in depressive illness. *Am J Psychiatry* 128: 31-40.
- Carroll BJ, Curtis GC, et al. (1976). Urinary free cortisol excretion in depression. *Psychol Med* 6: 43-50.
- Damasio AR, Grabowski TJ, et al. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci* 3(10): 1049-56.
- Davidson RJ, Irwin W (1999). The functional neuroanatomy of emotion and affective style. *Trends Cogn Sci* 11-21.

- Dolan R, Morris JS, et al. (2001). Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci U S A* 98(17): 10006-10.
- Dolan RJ (2002). Emotion, cognition, and behavior. *Science* 298(5596): 1191-4.
- Eugène F, Levesque J, et al. (2003). The impact of individual differences on the neural circuitry underlying sadness. *NeuroImage* 19: 354-64.
- Evans AC, Kamber M, et al. (1994). An MRI-based probabilistic atlas of neuroanatomy. In: *Magnetic resonance scanning and epilepsy*. Shorvon S, Fish D, Andermann F, Bydder GM, Stefan H (Eds.). New York, Plenum: 263-74.
- Fecteau S, Armony JL, et al. (submitted). Is voice processing species-specific in human auditory cortex? An fMRI study.
- Fecteau S, Armony JL, et al. (submitted). Judgment of emotional nonlinguistic vocalizations: Age-related differences.
- Fischer H, Furmark T, et al. (2000). Brain representation of habituation to repeated complex visual stimulation studied with PET. *Neuroreport* 11(1): 123-6.
- Friston, KJ, Worsley KJ, et al. (1994). Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapp* 1: 210-20.
- Friston, KJ, Holmes AP, et al. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2: 189-210.
- Garavan H, Pankiewicz J, et al. (2000). Cue-induced cocaine craving: neuroanatomical specificity for drug users and drug stimuli. *Am J Psychiatry* 157(11): 1789-98.
- Garavan H, Pendergrass JC, et al. (2001). Amygdala response to both positively and negatively valenced stimuli. *Neuroreport* 12(12): 2779-83.

- Gainotti G (2001). Disorders of emotional behaviour. *J Neurol* 248(9): 743-9.
- Gibbons JL, McHugh PR (1962). Plasma cortisol in depressive illness. *J Psychiatr Res* 1: 162-71.
- Gorno-Tempini ML, Pradelli S, et al. (2001). Explicit and incidental facial expression processing: An fMRI study. *NeuroImage* 14(2): 465-73.
- Gottfried JA, Deichmann R, et al. (2002a). Functional heterogeneity in human olfactory cortex: An event-related functional magnetic imaging study. *J Neurosci* 22, 10819-28.
- Gottfried JA, O'Doherty, et al. (2002b). Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic imaging study. *J Neurosci* 22, 10829-37.
- Gur RC, Schroeder L, et al. (2002). Brain activation during facial emotion processing. *NeuroImage* 16: 651-62.
- Hamann S, Mao H (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport* 13(1): 15-9.
- Hamann SB, Ely TD, et al. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychol Sci* 13(2): 135-41.
- Hoffman EA, Haxby JV (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3(1): 80-4.
- Iidaka T, Omori M, et al. (2001). Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI. *J Cogn Neurosci* 13(8): 1035-47.



- Imaizumi S, Mori K, et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* 8(12): 2809-12.
- Karama S, Lecours AR, et al. (2002). Areas of brain activation in males and females during viewing of erotic film excerpts. *Hum Brain Mapp* 16(1): 1-13.
- Kawashima R, Itoh M, et al. (1993). Changes of regional cerebral blood flow during listening to an unfamiliar spoken language. *Neurosci Lett* 161(1): 69-72.
- Killgore WD, Yurgelun-Todd DA (2001). Sex differences in amygdala activation during the perception of facial affect. *Neuroreport* 12(11): 2543-7.
- Kreiman J (1997). Listening to voices: Theory and practice in voice perception research. *Talker Variability in Speech Research*. Johnson K, Mullenix J (Eds.). New York, Academic Press: 85-108.
- Knutson B, Adams CM, Fong GW, et al. (2001a). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 21(16): RC159.
- Knutson B, Fong GW, et al. (2001b). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 12(17): 3683-7.
- LaBar KS, Gitelman DR, et al. (2001). Hunger selectively modulates corticolimbic activation to food stimuli in humans. *Behav Neurosci* 115(2): 493-500.
- Lane RD, Reiman EM, et al. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *Am J Psychiatry* 154: 926-33.
- Linkowski P, Mendlewicz J, et al. (1985). The 24 hour profile of adrenocorticotropin and cortisol in major depressive illness. *J Clin Endocrinol Methods* 61: 429-38.

- Maeder PP, Meuli RA, et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *NeuroImage* 14: 802-16.
- Makino S, Hashimoto K, et al. (2002). Multiple feedback mechanisms activating corticotropin-releasing hormone system in the brain during stress. *Pharmacol Biochem Behav* 73(1):147-58.
- Al-Maskati HA, Zbrozyna AW (1989). Stimulation in prefrontal cortex area inhibits cardiovascular and motor components of the defence reaction in rats. *J Auton Nerv Syst* 28: 117-25.
- Mitchell RLC, Elliott R, et al. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41(10): 1410-21.
- Morris JS, Scott SK, et al. (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia* 37(10): 1155-63.
- Morris JS, Friston KJ, et al. (1998). A neuromodulatory role for the human amygdala in processing emotional facial expressions. *Brain* 121: 47-57.
- Morris JS, Frith CD, et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383(6603): 812-5.
- O'Doherty J, Rolls ET, et al. (2001). Representation of pleasant and aversive taste in the human brain. *J Neurophysiol* 85:1315-21.
- O'Doherty JP, Deichmann R, et al. (2002). Neural responses during anticipation of a primary taste reward. *Neuron* 33(5): 815-26.

- Pessoa L, McKenna M, et al. (2002). Neural processing of emotional faces requires attention. *Proc Natl Acad Sci USA* 99(17): 11458-63.
- Phan KL, Wager T, et al. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage* 16: 331-48.
- Phillips ML, Young AW, et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc R Soc Lond B Biol Sci* 265(1408): 1809-17.
- Rotshtein P, Malach R, et al. (2001). Feeling or features: Different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron* 32(4): 747-57.
- Royet JP, Zald D, et al. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: A positron emission tomography study. *J Neurosci* 20(20): 7752-9.
- Sachar EJ, Hellman L, et al. (1970). Cortisol production in depressive illness. A clinical and biochemical clarification. *Arch Gen Psychiatry* 23: 289-98.
- Sander K, Scheich H (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res Cogn Brain Res* 12(2): 181-98.
- Sato W, Kochiyama T, et al. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: An fMRI study. *Brain Res Cogn Brain Res* 20(1): 81-91.
- Schacter DL, Buckner RL, et al. (1997). Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* 6(4): 259-69.
- Schneider F, Grodd W, et al. (1997). Functional MRI reveals left amygdala activation during emotion. *Psychiatry Res* 76(2-3): 75-82.

- Talairach P, Tournoux J (1988). *Stereotaxic coplanar atlas of the human brain*. Stuttgart (Germany), Thieme.
- Vuilleumier P, Armony JL, et al. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron* 30(3): 829-41.
- Whalen PJ, Rauch SL, et al. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J Neurosci* 18(1): 411-8.
- Wildgruber D, Pihan H, et al. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage* 15(4): 856-69.
- Winston JS, O'Doherty J, et al. (2003). Common and distinct neural responses during direct and incidental processing of multiple facial emotions. *NeuroImage* 20(1): 84-97.
- Winston JS, Strange BA, et al. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat Neurosci* 5(3): 277-83.
- Worsley KJ, Friston KJ (1995). Analysis of fMRI time-series revisited-again. *NeuroImage* 2(3): 173-81.
- Wright CI, Martis B, et al. (2002). Enhanced amygdala responses to emotional versus neutral schematic facial expressions. *Neuroreport* 13(6): 785-90.
- Yang TT, Menon V, et al. (2002). Amygdalar activation associated with positive and negative facial expressions. *Neuroreport* 13(14):1737-41.
- Zald DH (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Res Brain Res Rev* 41(1): 88-123.

Zalla T, Koechlin E, et al. (2000). Differential amygdala responses to winning and losing: A functional magnetic resonance imaging study in humans. *Eur J Neurosci* 12(5): 1764-70.

Zbrozyna AW, Westwood DM (1991). Stimulation in prefrontal cortex inhibits conditioned increase in blood pressure and avoidance bar pressing in rats. *Physiol Behav* 49: 705-8.

## **CHAPITRE 6**

### **DISCUSSION GÉNÉRALE**



## Discussion générale

L'objectif principal de cette thèse était d'explorer la perception des vocalisations humaines émotionnelles. Les quatre expérimentations composant la thèse ont mesuré différents processus de perception des vocalisations humaines. Plus spécifiquement, elles ont identifié l'implication de régions neuronales dans la perception de vocalisations nonlinguistiques humaines et ont exploré cette dernière à l'aide de jugement de dimensions émotionnelles et d'amorçage. Dans l'ensemble, ces quatre études montrent que la perception des vocalisations nonlinguistiques humaines semble constituer un traitement *spécifique*, comparativement aux traitements de perception d'autres stimuli auditifs, et que parmi même ces vocalisations, des traitements relativement différents semblent sous-jacents à des types de vocalisations, par exemple selon le contenu émotionnel que portent les vocalisations et selon le sexe du locuteur qui exprime les vocalisations. Plus spécifiquement, au niveau comportemental, la perception de valence émotionnelle, à savoir si le stimulus exprime une émotion positive, négative ou est non chargé émotionnellement, semble être constante entre les groupes, indépendamment du sexe et de l'âge des participants adultes. Toutefois, le jugement des dimensions de valence, d'intensité et d'authenticité des vocalisations nonlinguistiques émotionnelles semble se modifier au cours du développement adulte (article 1). Aussi, le traitement des vocalisations nonlinguistiques semble impliquer un processus d'amorçage, lequel serait influencé par le sexe du locuteur (article 2). En ce qui concerne les résultats de neuroimagerie fonctionnelle, les substrats neuronaux sous-jacents au traitement de la voix humaine ne semblent pas traiter de façon significative les vocalisations animales nonhumaines (article 3). Enfin, les bases neuronales sous-jacentes à la perception des vocalisations nonlinguistiques humaines semblent être en partie dépendantes du contenu émotionnel, au même titre que celles associées aux traitements des expressions faciales et de la prosodie émotionnelle (article 4). L'ensemble des résultats de ces études contribue à la définition de ce système du traitement de la voix humaine, lequel commence seulement à être inclus dans les modèles théoriques. Cependant,

des limites importantes restreignent l'interprétation de certains résultats de ces études. Ces limites sont soulignées dans cette discussion générale. Elles guideront possiblement les choix méthodologiques des études futures.

### **Article 1. Perception de dimensions émotionnelles des vocalisations nonlinguistiques: influence du facteur âge**

Le premier article inclus dans la thèse avait pour but d'explorer la perception de dimensions émotionnelles des vocalisations nonlinguistiques humaines. Plus spécifiquement, nous avons étudié la perception des dimensions de valence, d'intensité et d'authenticité de l'émotion exprimée par des vocalisations nonlinguistiques chez un groupe de jeunes adultes et un groupe d'adultes d'âge moyen (moyenne de 54 ans). L'ensemble des 60 participants ont jugé 563 vocalisations nonlinguistiques sur trois échelles visuelles analogues. Les stimuli regroupaient un large éventail de vocalisations représentant trois catégories émotionnelles: émotion positive (rire et vocalisation sexuelle), émotion négative (pleur et cri de peur) et vocalisations non chargées émotionnellement (p. ex., éternuement). Ces stimuli avaient préalablement été jugés selon la catégorie vocale (rire, vocalisation sexuelle, pleur, cri de peur et «autre») et l'intensité émotionnelle exprimée par le locuteur. Il est à noter que la consigne donnée aux participants était de juger l'émotion exprimée par un stimulus donné et non pas ce qu'ils ressentaient à l'écoute de ce stimulus. Ceci est important puisque le but de cette étude était d'explorer le traitement perceptuel et non le ressenti du participant suite à la présentation de vocalisations émotionnelles (une brève présentation des échelles utilisées suit plus loin).

#### *Résumé des résultats*

Les résultats suggèrent que le jugement de valence émotionnelle des vocalisations est comparable entre les participants, c'est-à-dire que les rires et les vocalisations sexuelles étaient jugés comme ayant une valence émotionnelle positive, les pleurs et les cris de peur ayant une valence négative et les autres vocalisations de type raclement de la gorge, toux et

éternuement étaient jugées comme étant neutre émotionnellement. Nous avons aussi observé une différence entre les deux groupes d'âge sur les jugements de valence et d'intensité de l'émotion. Le plus jeune groupe a jugé les vocalisations émotionnelles (rires, vocalisations sexuelles, pleurs et cris de peur) comme étant significativement plus positives / négatives et plus intenses que le groupe plus âgé. Aussi, les femmes plus âgées ont jugé l'ensemble des vocalisations comme étant moins authentiques que les trois autres groupes de participants. Mis à part ce dernier résultat, aucune autre différence entre les femmes et les hommes était observée.

*Échelles utilisées pour explorer les dimensions de valence, d'intensité et d'authenticité*

Avant d'aborder les interprétations possibles qu'on peut tirer de ces résultats, il est nécessaire de discuter du type d'échelles utilisées dans cette étude. Comme mentionné antérieurement, nous étions intéressés à explorer la perception des vocalisations émotionnelles et non pas le ressenti des participants vis-à-vis ces vocalisations. Nous avons spécifiquement demandé aux participants si l'émotion exprimée était positive ou négative, émotionnellement intense ou pas du tout intense et authentique ou pas du tout authentique (voir figure 1).

Figure 1. Échelles utilisées dans cette étude pour juger les vocalisations humaines sur les dimensions de valence, d'intensité et d'authenticité

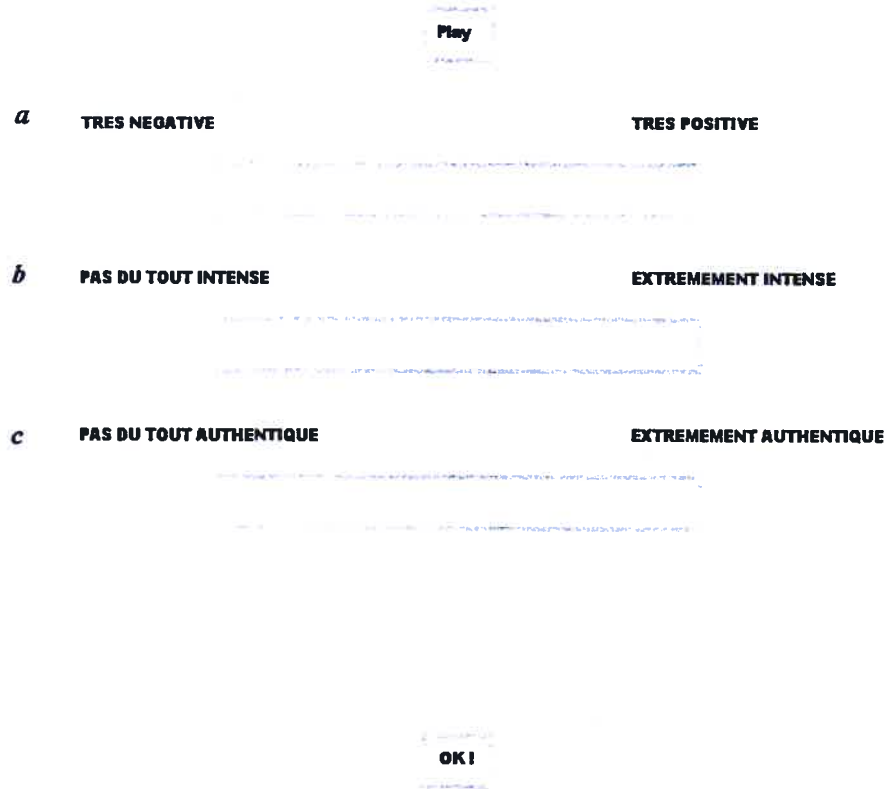


Figure 1. Différents jugements de l'émotion exprimée pour une même vocalisation. Jugement a) de valence, b) d'intensité et c) d'authenticité.

Nous avons considéré l'exploration du ressenti des participants et le pouvoir d'induction d'émotion en utilisant les vocalisations comme étant une étape ultérieure à celle d'examiner l'accord des participants sur la valence de l'émotion exprimée via une vocalisation, ainsi que des jugements quant à l'intensité et l'authenticité de l'émotion. Le jugement d'authenticité est moins courant dans la littérature (Scherer, Banse et al., 2001). Toutefois, même si la grande majorité des stimuli étaient produits par des acteurs professionnels, nous avons jugé important d'inclure cette mesure d'authenticité afin d'identifier si les émotions exprimées semblaient vraies, mais surtout parce que nous

croyons que la perception de l'authenticité de l'émotion exprimée est essentielle dans la perception générale des émotions et qu'il est possible qu'elle module cette perception générale ainsi que le ressenti des participants.

Nos stimuli, catégorisés et décrits en termes de valence, d'intensité et d'authenticité, sont maintenant adéquats à être utilisés pour étudier le ressenti et examiner s'ils induisent des émotions. La méthode classique est de demander aux participants si le stimulus était plaisant ou déplaisant ainsi que son niveau d'activation («arousal»). La figure 2 est tirée de l'article de Russell (2003). Il serait également intéressant d'étudier, par exemple l'impact du degré d'intensité et d'authenticité sur l'induction d'émotion.

Figure 2. Modèle du noyau affectif proposé par Russell (2003)

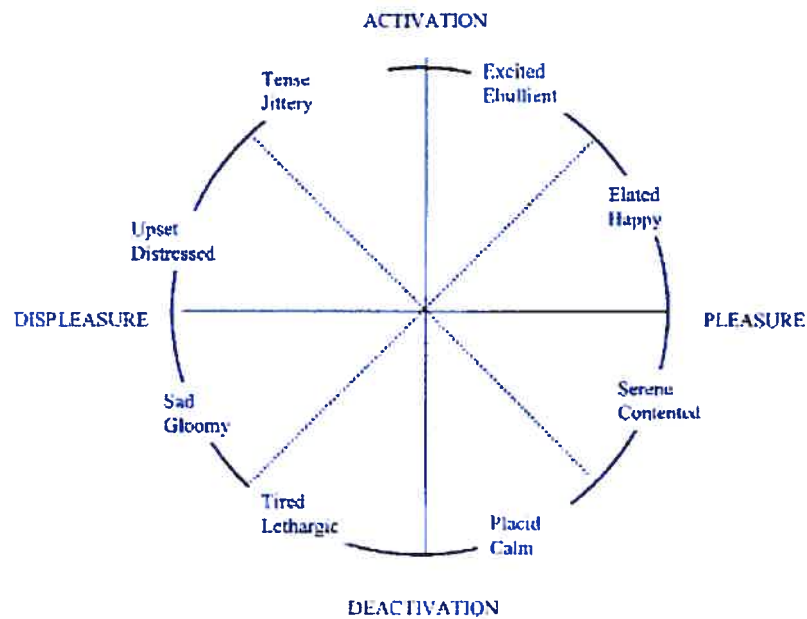


Figure 2. La dimension horizontale, plaisant – déplaisant, s'étend d'un extrême (p. ex. agonie) à l'autre (p. ex., extase), en passant par un point neutre (niveau d'adaptation). La dimension verticale, l'activation, s'étend du *sommeil* à l'excitation frénétique, en passant par différents stades d'alerte et par l'appesantissement.



*Interprétations des résultats et limitations de l'étude*

Les différences d'âge dans les jugements de valence, d'intensité et d'authenticité peuvent être interprétées de différentes façons. Premièrement, il est possible que cette différence d'âge soit bien réelle, c'est-à-dire que le jugement de différentes dimensions émotionnelles se modifie au cours du temps. La littérature sur la perception des émotions suggère entre autres que les expressions faciales sont jugées comme étant moins émotionnelles chez l'adulte plus âgé comparativement à un adulte plus jeune (p. ex., Sullivan et Ruffman, 2004). Certaines études proposent que ces modifications de perception des émotions seraient en partie expliquées par des changements anatomiques et fonctionnels du cerveau (Gunning-Dixon, Gur et al., 2003). D'autres études montrent que la régulation des émotions et la capacité d'adaptation («coping») augmente avec l'âge. Aussi, il sera important de considérer différents facteurs, tel l'éducation et les valeurs, propres aux générations. Afin de mieux comprendre cette possible différence liée à l'âge dans la perception des vocalisations émotionnelles, les futures études devront d'une part, explorer si des corrélations existent entre les bases neuronales impliquées dans la perception des vocalisations émotionnelles et les jugements émotionnels de ces mêmes stimuli et voir si des différences entre les groupes d'âge sont observées et d'autre part, mesurer chez ces mêmes participants leur capacité d'adaptation.

Une limite considérable de cette étude est la nature de la tâche que le participant devait effectuer. Il ne nous est pas possible de savoir exactement quelle stratégie les participants ont utilisée pour juger les dimensions émotionnelles des vocalisations nonlinguistiques. Sans mesures électrophysiologiques (p. ex., conductance électrodermale, dilatation de la pupille), il n'est pas possible de connaître le degré d'induction d'émotion des participants lorsqu'ils jugeaient les stimuli. Le débat à savoir ce que nous faisons exactement pour juger un stimulus émotionnel n'est toujours pas résolu. Certains proposent que nous devons préalablement ressentir l'émotion exprimée par autrui afin de la juger, d'autres suggèrent que nous jugeons l'émotion sans que notre niveau d'induction d'émotion en soit affecté de



façon significative (Russell, Bachorowski et al, 2003). Dans le cas de notre étude, il n'est donc pas possible d'identifier à quel niveau les différences de jugement des dimensions émotionnelles entre les groupes d'âge des participants se situent. Il se peut que les deux groupes d'âge aient utilisé une seule et même stratégie. S'ils ont utilisé la stratégie qui requiert un ressenti de l'émotion exprimée par autrui afin de juger l'émotion, les différences de jugement reliées à l'âge observées dans cette étude s'expliqueraient par des habiletés de régulation et d'adaptation des émotions plus fortes chez les adultes d'âge moyen comparativement aux jeunes adultes, qui eux auraient un niveau d'induction d'émotion plus sensible. Il est également possible que l'ensemble des participants ait, au contraire, jugé les stimuli sans que leur niveau d'induction d'émotion soit affecté. Dans ce cas, les différences de jugement émotionnel selon l'âge suggèrent une simple modification de la perception des émotions au cours du développement adulte. Une autre possibilité est que les deux groupes d'âge n'aient pas utilisé la même stratégie pour juger les stimuli émotionnels. Dans ce cas, nos résultats suggéreraient un changement de stratégie au cours du développement dans le jugement de dimensions de valence, d'intensité et d'authenticité de l'émotion exprimée via des stimuli vocaux émotionnels.

## **Article 2. Effet d'amorçage des vocalisations nonlinguistiques humaines: influence du sexe du locuteur**

Le but du deuxième article était d'explorer s'il existait un effet d'amorçage dans la perception de vocalisations nonlinguistiques. Plus spécifiquement, nous avons testé si la valence de l'émotion exprimée par la vocalisation et si le sexe du locuteur modulaient l'effet d'amorçage. Les vocalisations exprimaient des émotions à valence positive (rire et vocalisation sexuelle), négative (pleur et cri de peur) et neutre (p. ex., raclement de la gorge, éternuement). La moitié des stimuli était produite par des femmes, l'autre moitié par des hommes. Chaque stimulus était présenté deux fois, dans un ordre pseudo-aléatoire. Les participants, dix hommes, ont effectué une tâche d'identification du sexe du locuteur. L'effet d'amorçage était calculé en termes de temps de réponse et de nombre d'erreurs : la

moyenne des temps de réponse serait plus courte et le nombre d'erreurs moins important à la seconde occurrence des stimuli comparativement à la première.

### *Résumé des résultats*

Les résultats ont montré un effet d'amorçage en termes de temps de réponse pour l'ensemble des stimuli vocaux, toutefois cet effet n'était pas significatif en termes de nombre d'erreurs. Le sexe du locuteur semble moduler l'effet d'amorçage: les participants étaient plus rapides à identifier les vocalisations produites par des femmes lors de la seconde occurrence comparativement à la première occurrence, alors que cet effet n'était pas significatif pour les stimuli prononcés par des hommes. La valence émotionnelle, quant à elle, n'a pas influencé l'effet d'amorçage général observé, alors que nous nous attendions à trouver un effet. En d'autres termes, l'effet d'amorçage pour les vocalisations était comparable indépendamment si la vocalisation portait une charge émotionnelle ou non ou si l'émotion était positive ou négative. En résumé, les résultats de cette étude suggèrent que les hommes mémorisent implicitement mieux les vocalisations nonlinguistiques produites par les femmes que par les hommes.

### *Interprétations des résultats et limitations de l'étude*

Certaines limites des résultats de la deuxième étude doivent être mentionnées. L'effet d'amorçage pour les vocalisations nonlinguistiques semble être important en termes de temps de réponse. Toutefois, la raison pour laquelle l'effet d'amorçage est significatif pour les voix de femmes et ne l'est pas pour les voix d'hommes doit être davantage exploré. Il est possible que les caractéristiques acoustiques propres aux voix de femmes soient responsables de l'effet d'amorçage observé. Les paramètres acoustiques des voix de femmes et d'hommes diffèrent surtout en termes de fréquence: les hautes fréquences sont plus caractéristiques de voix de femmes comparativement aux voix d'hommes. Il se peut que l'effet d'amorçage observé soit dépendant de façon significative des caractéristiques fréquentielles. Church et Schacter (1994) suggèrent que la F0 est exclusivement

responsable de l'effet d'amorçage qu'ils ont observé, lequel était plus important lorsque les stimuli vocaux étaient produits par le même locuteur, comparativement à deux locuteurs différents. Il est aussi possible que d'autres différences acoustiques entre les voix de femmes et d'hommes soient en partie responsables de cet effet d'amorçage plus important pour les voix de femmes (voir Klatt et Klatt, 1990 pour une comparaison des paramètres acoustiques entre les voix de femmes et d'hommes). Donc, comme il a été mentionné dans les articles précédents, l'attribution d'une valeur acoustique aux vocalisations produites par des femmes et des hommes contribuerait à une meilleure définition de l'effet d'amorçage, à savoir si la structure acoustique est responsable ou influence en partie l'effet d'amorçage.

Une autre question qui demeure en suspend est celle de l'existence d'un effet d'amorçage chez des participantes. Il serait intéressant de voir si les femmes mémorisent implicitement mieux les voix d'hommes comparativement à celles des femmes ou si elles montrent, comme dans le cas de notre étude, un effet d'amorçage pour les voix de femmes et non pas pour les voix d'hommes. Dans tel cas, il serait suggéré que les caractéristiques acoustiques soient davantage responsables de l'effet d'amorçage que le sexe du locuteur. Church et Schacter (1994) ne rapportent pas d'effet de sexe du locuteur, toutefois leurs groupes participants étaient composés de femmes et d'hommes. Il aurait été intéressant de voir si l'effet d'amorçage selon le sexe du locuteur était plus important chez les participants de sexe opposé.

Alors qu'un effet d'amorçage selon le sexe du locuteur semble être un résultat relativement solide même si on ne connaît pas encore à quoi exactement il est attribuable, un effet d'amorçage selon la valence émotionnelle de stimuli auditifs est beaucoup moins clair. Comme il a déjà été mentionné dans l'introduction de la thèse, il ne semble pas avoir été rapporté dans la littérature d'effet d'amorçage selon l'émotion de stimuli auditifs. Par exemple, Church et Schacter (1994) n'ont pas trouvé d'effet d'amorçage plus important pour les patrons prosodiques émotionnels comparativement à ceux linguistiques. Ils

proposent que les caractéristiques acoustiques des stimuli soient responsables des effets d'amorçage, alors que le contenu sémantique et émotionnel des stimuli soit responsable des meilleures performances dans les tâches impliquant la mémoire explicite. Il est donc possible que l'effet d'amorçage soit différent selon la modalité des stimuli: alors qu'un effet d'amorçage significatif pour les stimuli émotionnels est rapporté dans les études lexicales (p. ex., Bradley, Mogg et al., 1994; Luo, Peng, et al., 2004), cet effet semble être de moindre importance pour les stimuli auditifs. Ceci reste à être exploré par des études utilisant des stimuli de différentes modalités.

### **Article 3. Réponses neuronales spécifiques associées à la perception des vocalisations conspécifiques chez l'humain**

Le troisième article de la thèse avait pour but d'étudier la spécificité des bases neuronales qui reçoivent un changement du signal hémodynamique associé à la perception des vocalisations humaines. Nous avons scanné le cerveau d'adultes sains en utilisant l'IRMf événementielle alors qu'ils écoutaient passivement des stimuli auditifs. Trois catégories de stimuli étaient présentées: (1) des vocalisations humaines (p. ex., rire, raclement de la gorge, éternuement, phonème américain, phrase finnoise), (2) des vocalisations animales nonhumaines (ex., un aboiement, un coassement, un meuglement) et (3) des sons nonvocaux (p. ex., un klaxon, un instrument de musique, une chasse d'eau de toilette). Les résultats de cette première étude ont montré des réponses neuronales plus importantes chez les participants lorsqu'ils entendaient des vocalisations humaines, comparativement aux stimuli auditifs nonhumains (vocalisations animales nonhumaines et stimuli nonvocaux). Les contrastes qui nous intéressent davantage afin d'explorer la spécificité des réponses neuronales aux vocalisations humaines sont bien sûr ceux des vocalisations humaines vs. vocalisations animales et son opposé, soit vocalisations animales vs. vocalisations humaines.



### *Résumé des résultats*

Des réponses significativement plus fortes dans les régions du STS des deux hémisphères étaient observées lorsqu'on soustrayait les activations cérébrales associées aux vocalisations humaines de celles des vocalisations animales. Ces réponses plus importantes pour les vocalisations humaines nous ont amenés à proposer que ces augmentations du signal hémodynamique dans ces régions soient sélectives aux vocalisations conspécifiques chez l'humain, comparativement aux vocalisations hétérospécifiques. Toutefois, les raisons de ces réponses sélectives aux stimuli conspécifiques demeurent à être explorées.

### *Interprétations des résultats et limitations de l'étude*

Maintenant que nos résultats suggèrent des réponses du STS sélectives aux vocalisations conspécifiques chez l'humain, la prochaine étape est d'identifier les facteurs qui rendent compte de cette sélectivité. Parce que les vocalisations humaines utilisées dans cette étude portent certainement des caractéristiques particulières qui ne sont pas partagées avec les vocalisations animales nonhumaines -puisque nous avons observé des activations sélectives du STS pour les vocalisations humaines- ces caractéristiques constituent donc des éléments qui contribuent à ces réponses cérébrales spécifiques. Ces éléments doivent être explorés et manipulés dans les études futures afin de mieux déterminer leur importance dans la caractérisation des vocalisations humaines attribuée par le cerveau humain.

Différents éléments ne nous permettent donc pas de considérer l'interprétation de sélectivité aux vocalisations conspécifiques comme étant la seule possible et qui rende compte à elle seule de cette sélectivité. Nous avons noté quelques facteurs qui sont susceptibles de contribuer aux résultats observés dans cette étude. Ils constituent en fait des caractéristiques intrinsèques dépendantes de stimuli naturels et qui sont, par conséquent, difficilement comparables entre différentes catégories de stimuli. Nous soulignons plus particulièrement l'importance de la structure acoustique, de la signification et de la charge émotionnelle, ainsi que de l'identité du locuteur.

Premièrement, même si le but de cet article était d'explorer si les régions du STS répondaient plus fortement aux vocalisations humaines comparativement aux vocalisations animales nonhumaines, et ce en utilisant des stimuli naturels (non modifiés acoustiquement), la structure acoustique des stimuli a possiblement contribué à la sélectivité des réponses du STS pour les voix humaines comparativement aux voix animales. Malgré que certaines études suggèrent que les vocalisations humaines et animales partagent plusieurs paramètres vocaux (p. ex., le singe-écureuil (*Saimiri sciureus*) présente un répertoire vocal étonnant: des vocalisations harmoniquement structurées, une fréquence constante ou non, des sons modulés, des énoncés monosyllabique et multisyllabiques; Jürgens, 1998; voir aussi Rendall, 2003; Riede et Zuberbuhler, 2003), des différences acoustiques entre elles demeurent et ne sont certainement pas négligeables. Toutefois, définir un stimulus auditif (ou une catégorie de stimuli) par une ou plusieurs valeurs acoustiques et déterminer qu'une réponse cérébrale soit spécifiquement associée à cette définition acoustique donnée ou partagée avec d'autres facteurs qui rende ce stimulus unique n'est pas trivial, mais toutefois important à faire. Des travaux en cours comme ceux de Theunissen (p. ex., Singh et Theunissen, 2003) vont éventuellement contribuer à mieux définir acoustiquement les stimuli auditifs, en termes de mesure de modulation de l'amplitude.

Un autre type d'aspects qui contribue certainement à caractériser les vocalisations humaines comme étant spéciales comparativement aux vocalisations animales est celui de la signification et de la charge émotionnelle. De par notre expertise en voix humaine, comparativement à d'autres stimuli auditifs, la voix humaine porte certainement une signification particulière et une charge émotionnelle chez l'humain qui n'est pas comparable à celle des vocalisations animales nonhumaines. À l'âge de 20 ans, nous avons baigné dans un environnement sonore composé de 10 000 à 50 000 heures de voix humaine. Il serait donc important de tester si des réponses cérébrales spécifiques sont associées à des vocalisations animales nonhumaines chez des ornithologues,



primatologistes, etc., c'est-à-dire des humains qui, par exemple, identifient une signification et une charge émotionnelle particulière aux chants d'oiseau comparativement à d'autres vocalisations animales nonhumaines. Basé sur les travaux de Gauthier et coll. (p. ex., Tarr et Gauthier, 2000; Gauthier et Nelson, 2001; Grill-Spector, Knouf et al., 2004), il est fort possible que des augmentations significatives du signal hémodynamique dans certaines régions du cortex auditif soient associées aux vocalisations animales d'expertises. Une autre avenue qui demeure à être explorée pour déterminer davantage l'importance de la signification et de la charge émotionnelle des vocalisations conspécifiques dans les réponses neuronales est celle de la possibilité d'une sélectivité des réponses pour les vocalisations conspécifiques chez les primates nonhumains. Il est fort possible que des réponses sélectives aux vocalisations conspécifiques soient aussi observées chez les primates nonhumains, sauvages ou captifs. Bien que peu d'études chez l'animal ont directement comparé les vocalisations conspécifiques et hétérospécifiques, plusieurs données sur la perception des vocalisations suggèrent une spécificité dans les réponses neuronales associées aux vocalisations conspécifiques (p. ex., Poremba, Malloy et al., 2004), et ce même comparativement à des sons synthétiques appariés en propriétés acoustiques (p. ex., Grace, Amin et al., 2003). Maintenant il est nécessaire de comparer directement les images de neuroimagerie fonctionnelle d'un groupe d'humains avec celles d'un groupe de singes, ce type d'étude contribuera certainement à une meilleure définition des réponses sélectives du cortex auditif pour les vocalisations conspécifiques à tout le moins chez les primates, humain et nonhumain. Il est aussi possible que la réponse sélective du STS pour les vocalisations conspécifiques observée dans notre étude soit unique à l'espèce humaine. Ceci pourrait s'expliquer par le fait que les vocalisations hétérospécifiques qui occupent notre quotidien, tel les miaulements et les aboiements, sont moins nombreuses, moins variées et de moindre importance dans notre survie comparativement, par exemple, au cas des singes rhésus sauvages. Une autre voie à explorer dans les prochaines études qui dépasse toutefois un peu la question de la spécificité des réponses neuronales de la perception des vocalisations conspécifiques est

celle de la perception de sa propre voix. La littérature sur les propres chants des oiseaux («bird's own songs») suggère même que des neurones du cortex auditif du «système de chant» chez les oiseaux chanteurs mâles répondent sélectivement et préférentiellement à ses propres chants comparativement aux chants conspécifiques (Margoliash, 1986; Doupe, 1997).

Un autre élément qui doit être éventuellement exploré afin de connaître son importance dans l'association de réponses cérébrales sélectives à des stimuli vocaux humains est celui de l'identité du locuteur. Il est bien connu que la voix humaine dévoile plusieurs caractéristiques du locuteur, tel l'âge (Hartman et Danahuer, 1976; Linville, 1996) et le sexe (Lass, Hughes et al., 1976; Mullenix, Johnson et al., 1995) et différentes régions cérébrales seraient impliquées selon si le participant entend le même stimulus prononcé par différentes voix versus s'il entend différents stimuli prononcés par la même voix. En effet, la partie plus antérieure du STS de l'hémisphère droit semble être davantage activée lorsque le participant entend différents locuteurs produire le même stimulus (p. ex., une syllabe) que lorsqu'il entend le même locuteur prononcer différentes syllabes (Belin et Zatorre, 2003). Enfin, il est aussi important d'étudier si des réponses neuronales sont associées à des catégories spécifiques de vocalisations humaines, par exemple un rire, et ceci est l'objet de la quatrième étude de la thèse.

#### **Article 4. Bases neuronales de la perception des vocalisations humaines nonlinguistiques émotionnelles**

Le quatrième article qui compose cette thèse avait comme objectif d'explorer les bases neuronales de différentes catégories de vocalisations humaines: rires, vocalisations sexuelles, pleurs, cris de peur et vocalisations émotionnellement neutres. La méthodologie de cette expérimentation était similaire à celle du troisième article. En utilisant un paradigme d'IRMf événementielle, nous avons scanné des participants sains alors qu'ils

entendaient des vocalisations nonlinguistiques. Les participants devaient faire une tâche d'identification du sexe des locuteurs.

#### *Résumé des résultats*

Nous avons observé que certaines régions cérébrales sont impliquées dans le traitement de vocalisations émotionnelles, indépendamment de la catégorie, alors que d'autres structures sont associées à des catégories spécifiques (p. ex., rires). Plus spécifiquement, les résultats suggèrent des réponses cérébrales plus importantes pour les stimuli émotionnels (pires, vocalisations sexuelles, pleurs et cris de peur) comparativement aux vocalisations émotionnellement neutres. Ces activations étaient principalement situées dans les régions temporales (cortex auditif primaire, STS et gyri temporaux supérieurs (GTS)) et frontales. Nous avons également observé que des catégories spécifiques de vocalisations émotionnelles ont induit une augmentation du signal hémodynamique dans certaines régions cérébrales, comparativement aux vocalisations émotionnellement neutres. Des réponses plus importantes dans la région de l'amygdale étaient observées lorsque les participants entendaient les rires et les vocalisations sexuelles. La catégorie des stimuli de pleur a, quant à elle, induit des réponses plus importantes dans la région de l'hypothalamus.

#### *Interprétations des résultats et limitations de l'étude*

Nos résultats suggèrent que les régions temporales sont impliquées de façon importante dans le traitement des émotions véhiculées par les vocalisations nonlinguistiques. Il semble que le cortex auditif primaire, connu pour traiter l'information auditive de base (p. ex. le rythme, voir les revues de Thaut, 2003 et de Samson, Belin, et al., 2001), soit sensible à la charge émotionnelle que portent les vocalisations nonlinguistiques. La FFA («fusiform face area»), connue pour traiter l'information de base dans le traitement des visages, est également sensible aux émotions qu'expriment les expressions faciales (Dolan, Morris et al., 2001; Rotshein, Malch et al., 2001; Vuilleumier, Dolan et al., 2001). Aussi, nos résultats vont dans le même sens que certaines données sur le conditionnement classique.

Des augmentations de réponse du cortex auditif primaire ont été associées à des stimuli auditifs auxquels on leur avait attribué une valeur aversive (Büchel, Morris, et al., 1998; Armony et Dolan, 2001). Il est possible que les régions temporales jouent un rôle général dans le traitement des émotions via les vocalisations émotionnelles, alors que d'autres structures seraient impliquées à un niveau ultérieur du traitement des émotions qui répondent à des émotions spécifiques (p. ex., hypothalamus).

Nous avons également observé une augmentation de l'activité de l'amygdale associée à la présentation des rires et des vocalisations sexuelles. Ces résultats supportent les données suggérant que l'amygdale soit impliquée dans le traitement des stimuli émotionnels de valence positive (voir revues de Baxter et Murray, 2002 et de Zald, 2003). Toutefois, cette augmentation d'activité de l'amygdale était de moins grande importance pour les vocalisations exprimant des émotions négatives (pleurs et cris de peur). Ces résultats semblent aller à l'encontre des études rapportant que l'amygdale soit principalement impliquée dans le traitement de stimuli exprimant la peur. En effet, une augmentation du signal de l'amygdale associée aux expressions faciales exprimant la peur est bien documentée (p. ex., Morris, Frith et al., 1996). Cependant, il semble que cette association ne soit pas aussi claire lorsque les stimuli sont auditifs. Certaines études rapportent une augmentation de la réponse de l'amygdale (p. ex., Sander et Scheich, 2001), d'autres une diminution de sa réponse (Morris, Scott, et al., 1999), enfin plusieurs études n'observent pas de changement dans la réponse de l'amygdale. Par exemple, aucune étude ne rapporte de réponse de l'amygdale pour les émotions négatives exprimées par des stimuli auditifs dans la méta-analyse de Phan, Wager et al. (2002). Il est par conséquent possible que les réponses de l'amygdale ne soient pas comparables entre les différentes modalités (p. ex., expressions faciales versus stimuli auditifs vocaux).

Concernant les réponses des régions frontales associées aux vocalisations émotionnelles comparativement à celles non-émotionnelles, il est possible qu'elles aient joué un rôle



inhibiteur. Des études (Beauregard et al., 2001; et voir revue de Gainotti, 2001) suggèrent que les régions frontales jouent un rôle inhibiteur important dans le traitement de stimuli émotionnels. De plus, d'autres études proposent que ce rôle des régions frontales affecte particulièrement l'amygdale (Al-Maskati et al., 1989; Zbrozyna et al., 1991). Même que Morris, Scott et al. (1999) proposent que l'augmentation du signal des régions frontales à la présentation de vocalisations de peur a joué un rôle inhibiteur sur l'activation de l'amygdale, pour laquelle ils ont observé une diminution du signal. Dans le cas de notre étude, les cris de peur ont particulièrement induit une augmentation de l'activité des régions frontales (en fait, elle constitue la réponse la plus importante). Il est donc possible que cette importante activité des régions frontales pour les cris de peur ait joué un rôle inhibiteur expliquant en partie la relative augmentation (non-significative) de l'amygdale pour ces stimuli.

Une des limites considérables de cette étude est l'absence de mesures physiologiques ou de questionnaire évaluant le niveau de l'affect des participants afin de savoir si une modification de l'affect des participants était associée aux stimuli émotionnels (ou à une catégorie spécifique de stimuli). Par conséquent, il n'est pas possible de déterminer avec certitude si seule la perception des vocalisations était mesurée ou bien s'il y avait induction d'émotion chez les participants. Ce même problème est observé dans les études antérieures sur le traitement des vocalisations émotionnelles. Ces études ont utilisé une approche par bloc (voir ci bas pour une discussion des approches par bloc versus événementielle). Par exemple, dans l'étude de Sander et Scheich (2001), des blocs de 60 secondes de rire continu produit par deux locuteurs étaient présentés. Cette approche module fort possiblement l'état affectif des participants mais aucune mesure montrant le niveau d'induction chez les participants n'a été prise. Les futures études devront utiliser par exemple des mesures électrodermales et de la dilatation de la pupille afin de mieux identifier le processus étudié (p. ex., perception versus induction).

Une autre limite de cette étude est le nombre limité de catégories vocales utilisées. Par exemple, il est nécessaire que les futures études incluent des vocalisations exprimant, par exemple des émotions sociales, lesquelles ont récemment grandement contribué à mieux comprendre le traitement des stimuli émotionnels (p. ex., Moll, Oliveira-Souza et al., 2003).

Enfin, une force considérable des deux études de neuroimagerie dont nous venons de discuter est la méthodologie utilisée. Nous avons utilisé une approche événementielle, inclus des événements nuls («null-events» ou silence de la durée d'un stimulus) et l'ordre de présentation des stimuli a été construite afin d'éliminer les séquences de plusieurs stimuli d'une même catégorie. Ces choix ont été faits dans le but de diminuer l'effet d'habituation de certaines régions cérébrales suite à la présentation de stimuli d'une même catégorie (pour une discussion complète sur ce sujet, voir Josephs et Henson, 1999). Dans le cas de l'étude de la spécificité de la réponse neuronale des STS à la voix humaine comparativement aux stimuli auditifs nonhumains (article 3), le design événementiel est nécessaire afin d'éliminer l'habituation de la réponse des STS. En effet, Belin et Zatorre (2003) ont montré une diminution du signal du segment antérieur du STS de l'hémisphère droit associée à la présentation de stimuli vocaux produits par le même locuteur. En ce qui concerne l'étude sur des substrats neuronaux associés à la perception de vocalisations émotionnelles, le design événementiel est aussi important puisque des régions sont très sensibles à l'habituation induite par la présentation de stimuli émotionnels de même catégorie (Fischer et al., 2003). Cette approche permet également de diminuer (ou d'éliminer) l'induction d'émotion chez les participants.

L'ensemble des résultats des quatre études suggèrent que la perception des vocalisations nonlinguistiques requiert des bases neuronales lesquelles se distinguent en partie de celles associées à la perception d'autres stimuli auditifs et qui impliquent différentes régions selon la catégorie des vocalisations. La perception des vocalisations



nonlinguistiques semble aussi être influencée selon l'âge des participants et le sexe des locuteurs. Malgré le peu d'études sur la perception des vocalisations nonlinguistiques et des limites méthodologiques des études présentées dans cette thèse, la perception des vocalisations nonlinguistiques mérite son module dans les modèles théoriques de traitement de l'information.

### **Modèle de la perception de la voix**

Belin, Fecteau et al. (2004) ont proposé un modèle de la perception de la voix (voir figure 3). Ce modèle, basé sur celui de la perception des visages de Bruce et Young (1986), suggère que pour traiter la voix, il est nécessaire de faire en premier lieu une analyse de bas niveau de la structure acoustique, ensuite l'information vocale est analysée de façon partiellement indépendante selon le type de traitement exécuté, soit le traitement du langage, du contenu émotionnel ou de l'identité du locuteur. Les régions cérébrales sous-jacentes à ces traitements interagiraient avec les homologues cérébraux impliqués dans le traitement des visages.

Figure 3. Modèle de la perception de la voix proposé par Belin, Fecteau et al. (2004)

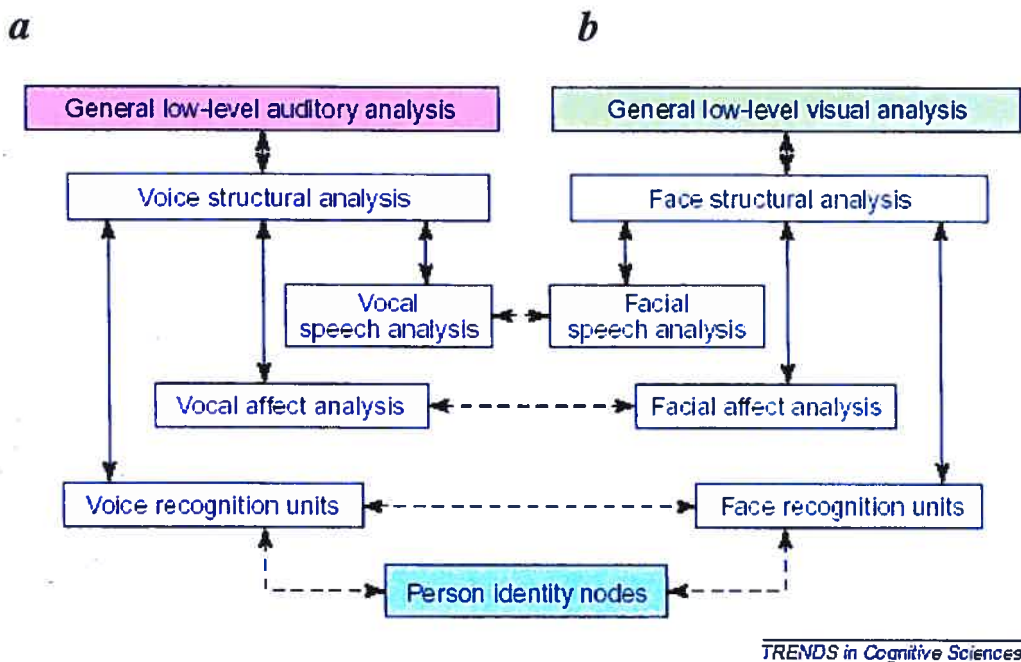


Figure 3. a) Organisation fonctionnelle du traitement de la voix, similaire à celle proposée pour le traitement des visages. b) Adaptation du modèle de perception des visages de Bruce et Young (1986). Les flèches en pointillées indiquent les interactions multimodales.

Éventuellement, il sera important de détailler chacune des trois composantes, l'analyse du langage, de l'identité du locuteur et du contenu émotionnel. Dans le cadre de ce travail, nous nous inspirons de modèles déjà proposés dans la littérature sur chacune de ces composantes afin de donner un aperçu de comment est traitée l'information que porte la voix humaine, ainsi que rapporter les régions qui semblent être significativement impliquées. Il est à noter que cet exercice ne se veut pas exhaustif, mais souligne seulement les points importants qui ont été rapportés dans la littérature concernant le traitement de la voix humaine au niveau du langage, plus particulièrement au niveau de l'identité du locuteur et de l'émotion.

### *Analyse du langage oral*

Étant donné que le but de la thèse concerne les vocalisations nonlinguistiques, nous nous limiterons à mentionner que plusieurs disciplines, tel la linguistique (p. ex., Chomsky, 1959), les neurosciences (p. ex., Doupe et Kuhl, 1999; Westermann et Reck Miranda, 2004), l'anthropologie (p. ex., Ferguson, 1964) et les sciences computationnelles (p. ex., Guenther, 1994), proposent autant de modèles théoriques de perception du langage que de façons qu'elles ont de concevoir le langage (voir revue de Kuhl, Tsao et al., 2001 et de Diehl, Lotto et al., 2004). En ce qui concerne les régions cérébrales associées au traitement du langage, les études en neuroimagerie (p. ex., Poeppel et Hickok, 2004; Hickok et Poeppel 2000; Zatorre et Binder, 2000; Scott et Johnsrude, 2003) rapportent principalement la participation des régions du cortex temporal supérieur des deux hémisphères, incluant le gyrus d'Heschl, le planum temporal ainsi que les parties médiane et antérieure du STS. Il semble que certaines régions soient davantage impliquées dans des analyses particulières. Par exemple, au niveau du STG, les régions antérieure et médiane répondraient aux phonèmes (Petersen, Fox et al., 1988; Démonet, Chollet et al., 1992; Zatorre, Evans et al., 1994), alors que les parties supérieures seraient impliquées dans l'organisation des phonèmes (p. ex., segmentation, séquence; Creutzfeldt, Ojemann et al., 1989). Au niveau du STS, les régions médianes seraient associées à la perception du langage sans qu'il y ait traitement sémantique (p. ex., phrase présentée temporellement inversée), alors que les régions plus antérieures de l'hémisphère gauche seraient impliquées dans le traitement du langage même lorsque son signal acoustique est dégradé.

### *Analyse de l'identité du locuteur par la voix*

En ce qui concerne l'analyse de l'identité du locuteur, il ne semble pas exister jusqu'à maintenant de modèle. Quelques études rapportent que nous ayons une bonne habileté à reconnaître les voix familières (Van Lancker, Kreiman et al., 1985a,b) et les voix connues (Schweinberger, Herholz et al., 1997). Aussi, des effets d'amorçage lorsqu'on identifie des voix familières et non familières ont été observés (Ellis 1997; Schweinberger, Herholz et

al., 1997) et cet effet était plus important pour les voix familières comparativement aux voix non familières (Schweinberger, 2001). Bien que le traitement de l'identité du locuteur ait été jusqu'à maintenant peu étudié et qu'aucun modèle ne semble encore exister, différents types de résultats laissent croire à l'existence d'un module de reconnaissance du locuteur par sa voix relativement indépendant de celui de la reconnaissance du langage.

Premièrement, il semble que l'identification du locuteur se fait indépendamment de ce qu'il prononce. En effet, nous sommes capables de reconnaître le locuteur même s'il produit un mot que nous n'avons pas encore entendu prononcé par le locuteur en question. Schweinberger (2001) a observé un effet d'amorçage pour les voix familières et non familières lorsque l'extrait langagier amorcé était différent de celui testé et cet effet d'amorçage était aussi plus important pour les voix familières comparativement aux voix non familières. Il semble donc que ce sont en fait les propriétés idiosyncratiques de la voix du locuteur qu'on identifie afin de le reconnaître. Ces propriétés idiosyncratiques ne sont pas encore clairement identifiées et on ne connaît pas non plus comment ces attributs sont liés à la connaissance sémantique de l'identité du locuteur. Nous savons que plusieurs paramètres acoustiques sont à la fois impliqués dans le traitement de la reconnaissance du locuteur et du langage (Remez, Fellowes et al., 1997). Toutefois, il semble que certaines propriétés acoustiques fourniraient des indices particulièrement importants dans la reconnaissance du locuteur, par exemple les caractéristiques fréquentielles seraient davantage nécessaires contrairement à l'information phonétique (Kuwabara et Ohgushi, 1987; Schweinberger, 2001). Aussi, Van Lancker et Canter (1982) rapportent une double dissociation entre processus de reconnaissance du locuteur et celui du langage. Ils ont observé chez des individus avec aphasie une reconnaissance intacte de la voix, ainsi que des individus avec phonagnosie dont le traitement du langage était intact. Donc, l'ensemble de ces résultats suggère que le processus de reconnaissance du locuteur par sa voix soit relativement indépendant du traitement du langage.

Une autre avenue de recherche qui propose un module d'analyse de l'identité du locuteur relativement indépendant de celui du langage est l'étude des bases neuronales associées à la perception du locuteur. Même si les systèmes impliqués dans la perception des mots et celui de la reconnaissance du locuteur sont très fortement interreliés (Knosche, Lattner et al, 2002), des différences d'activation sont rapportées.

Malgré la difficulté de mettre ensemble les quelques études de neuroimagerie sur la perception du locuteur, de par d'importantes différences entre les tâches utilisées, il semble tout de même que les parties antérieures des régions temporales supérieures de l'hémisphère droit soient particulièrement associées à la perception du locuteur. Plus spécifiquement, on a rapporté une activité hémodynamique plus importante dans les régions antérieures des lobes temporaux (Imaizumi, Mori et al., 1997), dans la partie antérieure du lobe temporal droit (Nakamura, Kawashima et al., 2001) et de la partie antérieure du STS droit (Von Kriegstein, Eger et al., 2003; Von Kriegstein et Giraud, 2004). De plus, en utilisant un paradigme d'adaptation, Belin et Zatorre (2003) ont montré une diminution significative de l'activité hémodynamique dans la région antérieure du STS droit. Enfin, on a aussi observé l'implication de la partie postérieure du STS droit (Von Kriegstein et Giraud, 2004) et du cortex enthorinal (Nakamura, Kawashima et al., 2001). Ces études suggèrent donc l'existence de populations neuronales du cortex auditif, au-delà de l'architecture fonctionnelle, sous-jacentes au traitement des propriétés idiosyncratiques de la voix du locuteur qui lui donne son identité unique et qui nous permet de la reconnaître. Elles ont identifié des régions associées à la reconnaissance du locuteur différentes de celles associées à la perception du langage (Belin et Zatorre, 2003; Von Kriegstein, Eger et al., 2003), du contenu sémantique (Von Kriegstein et Giraud, 2004), de l'émotion (Imaizumi, Mori et al., 1997), ainsi que des régions associées à la perception des voix familières différentes de celles impliquées dans la perception des voix non familières (Nakamura, Kawashima et al., 2001). L'ensemble des régions cérébrales associées à l'identification du locuteur serait différent de celui impliqué dans la perception



phonologique et de l'identité des mots, tel que proposé par les études sur la phonagnosie (Van Lancker, Cummings et al., 1988). Des individus avec lésion à l'hémisphère droit montraient des déficits importants dans des tâches de reconnaissance de voix familières, alors que des tâches de discrimination de voix non familières et de traitement des sons de l'environnement étaient intactes (Van Lancker, Kreiman et al. 1989). Le traitement de discrimination de voix non familières était aussi déficitaire chez des individus avec lésion à l'hémisphère droit ou gauche comparativement à un groupe d'individus neurologiquement sains, observant une corrélation entre un déficit dans la reconnaissance de la voix et la présence de lésion au niveau du lobe pariétal droit (Van Lancker, Kreiman et al. 1989). Donc, les résultats des quelques études sur la phonagnosie laissent aussi croire à un module du traitement de la reconnaissance du locuteur relativement indépendant de celui du langage. Basé sur les travaux explorant les voies visuelles chez l'humain et les études en audition chez le singe, Mesulam (1998) propose une aire spécialisée dans l'identification des patrons vocaux du locuteur, l'aire v (voir à la page 1019, figure 2c). Selon cet auteur, ce groupe de neurones spécialisés fait partie des aires «downstreams», lesquelles traiteraient les éléments plus complexes des stimuli auditifs, contrairement aux aires «upstreams», lesquelles procéderaient au traitement de traits plus élémentaires, tel la fréquence et le ton. S'il est vrai que les caractéristiques fréquentielles sont nécessaires à l'identification du locuteur, cette aire v est possiblement spécialisée dans l'identification de patrons fréquentiels uniques au locuteur.

Aussi, les études chez le nouveau-né suggèrent que le traitement de la reconnaissance du locuteur constitue entre autres un apprentissage implicite qui débute in utero. Les résultats rapportent que le fœtus humain a cette capacité de reconnaître la voix de sa mère (p. ex., DeCasper et Fifer, 1980; Kisilevsky, Hains et al., 2003) en montrant une préférence à écouter la voix de sa propre mère à celles d'autres femmes non familières. Les chercheurs proposent par conséquent qu'un processus d'apprentissage implicite est impliqué dans la reconnaissance du locuteur (DeCasper et Fifer, 1980; Spence et DeCasper, 1987; Spence et



Freeman, 1996). Ceci constitue un autre type de données suggérant un traitement de la reconnaissance du locuteur relativement indépendant de celui du langage, ce dernier étant effectif seulement beaucoup plus tard que celui de l'identité du locuteur dans le parcours développemental de l'humain.

*Analyse de l'émotion exprimée par la voix*

Enfin, le troisième type de traitement perceptuel proposé par le modèle de Belin, Fecteau et al. (2004) est celui de l'émotion exprimée par la voix. Plusieurs modèles de perception des émotions ont été proposés depuis le début des recherches sur les émotions, laquelle est grandement centrée sur la perception de stimuli visuels (p. ex., «International Affective Picture Series» de Lang, Bradley et al., 1999; visages d'Ekman et Friesen, 1978). Peu d'études ont discuté de la perception que nous avons, par exemple, d'un pleur, d'un rire ou d'un cri de peur. L'intégration de l'analyse de l'émotion exprimée par la voix dans les modèles théoriques est nécessaire afin de compléter la caractérisation de la perception des émotions.

Adolphs (2000a,b) propose un modèle de traitement des expressions faciales en fonction du temps en y présentant les régions impliquées et le type de traitement concerné (voir figure 4). Ce modèle caractérise l'expression faciale comme manifestation d'une émotion en réaction à la présentation d'un stimulus visuel saillant. Nous proposons une contrepartie auditive et vocale à son modèle. Il est à noter qu'Adolphs lui-même mentionne la possibilité que certains mécanismes impliqués dans son modèle peuvent être les mêmes que ceux associés au traitement de la prosodie émotionnelle. Nous suggérons la situation suivante: un participant entend une vocalisation nonlinguistique de saillance significative (p. ex., un cri de peur) et qu'en réaction, à son tour, il produit une vocalisations (p. ex. une vocalisation de peur).

La première étape du modèle d'Adolphs concerne les structures impliquées dans le traitement perceptuel d'un stimulus visuel de saillance importante. Elle correspond à l'étape d'analyse acoustique de bas niveau d'un stimulus vocal du modèle de Belin, Fecteau et al. (2004), laquelle impliquerait une activité hémodynamique plus importante du cortex auditif primaire.

Figure 4. Modèle de perception d'une expression faciale proposé par Adolphs (2002a,b)

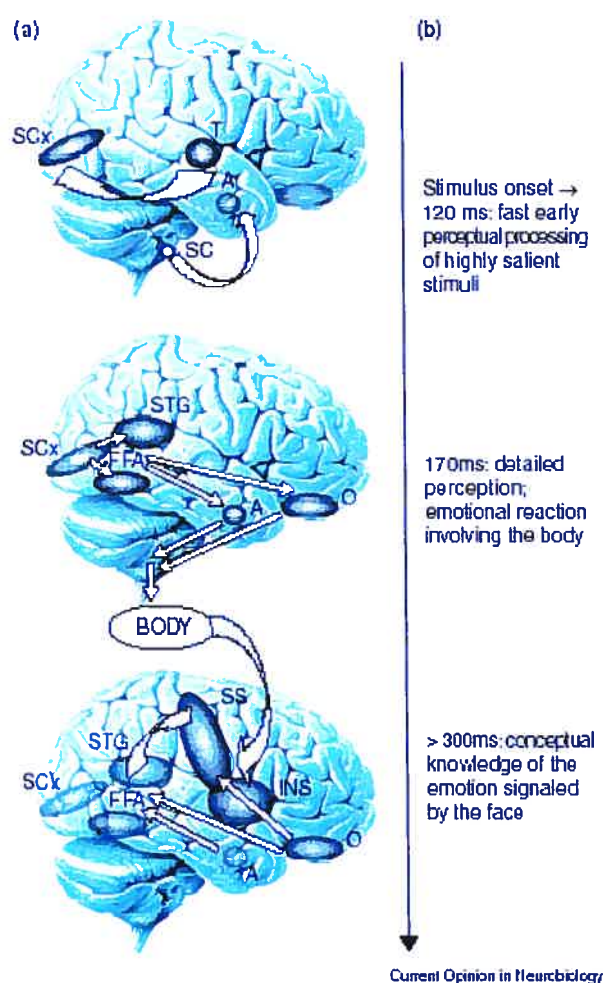


Figure 4. Traitement des expressions faciales en fonction du temps. a) Structures impliquées dans la reconnaissance des émotions à différentes périodes du temps. A, amygdale; FFA, «fusiform face area»; INS, insula; O, cortex orbitofrontal; SC, colliculus supérieur; SCx, cortex strié; SS, cortex somatosensoriel; STG, gyrus temporal supérieur; T, thalamus. b) Le cours temporel de la reconnaissance de l'émotion, du début du

stimulus jusqu'à la reconnaissance de l'émotion. Il est possible que plusieurs de mécanismes soulignés ici sont les mêmes impliqués dans la reconnaissance de l'émotion via d'autres classes de stimuli, tel la prosodie.

La seconde étape du modèle d'Adolphs propose une perception détaillée du stimulus en question et implique une réaction émotionnelle impliquant le corps. Dans le cas de l'analyse détaillée du stimulus vocal, les régions du cortex temporal supérieur (STS et STG), l'amygdale, l'insula, le noyau caudé, le gyrus cingulaire ainsi que les régions du cortex frontal et préfrontal sont particulièrement sollicitées (Phillips, Young et al., 1998; Morris, Scott et al., 1999; Sander et Scheich, 2001; voir tableau 2 de l'introduction de la thèse pour plus de détails sur ces études). Pour ce qui est de la réaction émotionnelle impliquant le corps, nous suggérons que ce mécanisme est comparable indépendamment de la modalité du stimulus saillant, visuelle ou auditive. Une fois l'évaluation du stimulus complétée, les systèmes autonome et somatique s'enclenchent et cette réponse peut être décrite par des mesures électrophysiologiques (p. ex., cardiaques, électrodermales). Il a été suggéré que les patrons d'activité du système nerveux autonome soient distinctifs selon l'émotion, plus particulièrement pour la colère, la peur et le dégoût (Ekman, Levenson et al., 1983) et pour la tristesse (Levenson, Carstensen et al., 1991). Pour plus de détails sur les réponses physiologiques spécifiques selon l'émotion, voir Schwartz, Weinberger et al. (1981), Roberts et Weerts (1982), Stemmler (1989), Levenson, Ekman et al., (1992) et Boiten (1996). Il serait toutefois intéressant d'explorer si des stimuli de différentes valences mais de même intensité enclenchent des réponses physiologiques similaires.

La troisième étape du modèle d'Adolphs concerne la reconnaissance conceptuelle de l'émotion signalée par une expression faciale. Dans notre cas, cette reconnaissance serait signalée par une vocalisation de peur. Deux importants processus sont impliqués: la production d'un stimulus vocal et l'induction d'émotion. En ce qui concerne la production vocale, certainement dû aux limites techniques des instruments de neuroimagerie (p. ex., mouvements de la tête), nous connaissons peu les substrats neuronaux qui y sont associés (voir les récentes études sur la fluidité verbale de Amunts, Weiss et al., 2004 et de Gaillard,

Sachs et al., 2003). Dans le cas de l'induction d'émotion, nous suggérons que les mêmes régions sont impliquées indépendamment de la modalité du stimulus saillant responsable de cette induction. Comme dans le cas des réponses physiologiques, l'induction d'émotion semble solliciter des régions cérébrales en partie spécifiques selon l'émotion (p. ex., George, Ketter et al., 1995 et voir les méta-analyses de Phan, Wager et al., 2002 et de Phan, Wager et al., 2004). Par exemple, des réponses de l'insula (Lane, Reiman et al., 1997), du cortex cingulaire antérieur (Barrett, Pike et al., 2004), du cortex orbitofrontal et du cortex frontal inférieur (Pardo, Pardo et al., 1993) sont particulièrement associées à l'induction de tristesse. Un autre exemple, une réponse hémodynamique importante de l'amygdale est souvent associée à l'induction de peur (voir la revue de Aggleton, 2000). Il est aussi suggéré que l'intensité de l'émotion, par conséquent du niveau d'activation («arousal») du participant, est important. Certaines régions cérébrales sont sollicitées par une émotion indépendamment de la valence (Breiter, Etcoff et al., 1996) ou par une émotion de forte intensité indépendamment de la valence (Anderson, Christoff et al., 2003; Small, Gregory et al., 2003; voir aussi les revues d'Anderson et Sobel, 2003b et de Sander, Grafman et al., 2003).

Les résultats rapportant des activations associées à la présentation de stimuli saillants divergent pour plusieurs raisons. Premièrement, la plupart des études de neuroimagerie n'utilisent pas de mesures électrophysiologiques ou de questionnaires évaluant le niveau d'induction émotionnelle du participant, par conséquent il est difficile d'identifier le processus mesuré (induction ou perception d'un stimulus saillant). Nous ne détaillerons pas davantage cette question -à savoir s'il y a induction d'émotion pour identifier l'émotion exprimée par si le stimulus ou l'identification de l'émotion peut se faire sans induction- puisqu'elle a déjà été abordée dans le premier article (page 53). Un autre facteur qui rend compte en partie de la divergence des résultats est la méthode d'induction utilisée. Différentes réponses hémodynamiques selon si la méthode est externe ou interne ont été observées (Reiman, Lane et al., 1997). Une autre raison pour laquelle les résultats divergent



est que l'induction d'une émotion à un participant implique (ou la présentation d'un stimulus saillant), non seulement le processus de perception ou d'induction, mais aussi d'autres processus, particulièrement l'attention. Les travaux de Vuilleumier et ses collègues visent entre autres à mesurer l'effet de l'attention sur le traitement de stimuli émotionnels (Bentley, Vuilleumier et al., 2003; Holmes, Vuilleumier et al., 2003; Vuilleumier, Armony et al., 2001; voir aussi les revues de Dolan et Vuilleumier, 2003 et de Pessoa et Ungerleider, 2004). Enfin, la variabilité des réponses hémodynamiques entre les participants semble être suffisamment importante pour en discuter ici (Eugène, Lévesque et al., 2003). Cette variabilité est due à différents facteurs autant au niveau méthodologique (p. ex., type d'analyse, le nombre de participants; Nadeau, Williamson et al., 1998; Friston, Holmes et al., 1999) qu'au niveau du sexe des participants (Pardo, Pardo et al., 1993; George, Ketter et al., 1996; Schneider, Habel et al., 2000; Cahill, Haier et al., 2001; Sander Brechmann, et al., 2003; Canli, Desmond, et al. (2002). Karama, Lecours et al., 2002; Wager, Phan et al., 2003) ou de leur personnalité (Abercombie, Schaefer et al., 1998; Johnson, Wiebe et al., 1999; Canli, Zhao et al., 2001; Fisher, Tillfors et al., 2001; Canil, Sivers et al., 2002; Schaefer, Jackson et al., 2002; Phan, Taylor et al., 2004; Somerville, Kim et al., 2004).

La contrepartie auditive et vocale que nous proposons au modèle d'Adolphs (2002a,b) se veut seulement une esquisse. En résumé, nous faisons l'hypothèse que certains processus sont les mêmes dans les deux mises en scène, celle illustrée par le modèle Adolphs et celle que nous avons proposé, alors que d'autres processus sont différents. La mise en scène du modèle d'Adolphs est la suivante: un individu perçoit un stimulus, en fait une analyse détaillée, considère que le stimulus en question est saillant et enfin réagit au stimulus par des manifestations physiologiques et des expressions faciales. Nous avons proposé une mise en scène similaire, mais représentant un pendant auditif et vocal: un individu entend un cri de peur et exprime son émotion par une vocalisation de peur. Plus spécifiquement, nous faisons l'hypothèse que l'analyse perceptuelle du stimulus, la première étape du

modèle, sollicite des régions cérébrales différentes (cortex visuel (FFA) vs. cortex auditif (STS)) selon la mise en scène. Alors que la seconde étape, l'analyse détaillée, impliquent aussi des régions différentes selon la modalité du stimulus, mais certains substrat sous-jacents au traitement qui rend compte de la saillance du stimulus sont les mêmes pour les deux mises en scène (p. ex., l'amygdale, le thalamus, l'hypothalamus, l'hippocampe, l'insula, le cortex cingulaire, le cortex préfrontal médian et latéral, les ganglions de la base), ainsi que les réactions physiologiques associées à la perception du stimulus saillant. Enfin, la troisième étape du modèle, la prise de conscience de la saillance du stimulus sollicite les mêmes régions cérébrales, alors que l'expression de l'individu (expression vocale vs. vocalisation) implique différents substrats neuronaux selon la mise en scène. Ceci n'est qu'une hypothèse formulée à partir des différentes études. Il est encore trop tôt pour lui donner une valeur significative de par la divergence des résultats rapportés dans la littérature. Différentes expérimentations doivent directement tester les nombreux processus impliqués dans cette simple mise en scène.

### *Évolution du langage interne*

Pour conclure la thèse, il s'avère intéressant de discuter de l'évolution du langage. L'évolution du langage humain (langage interne ou «*language-I*»), qui aurait débuté au moment de notre divergence avec notre ancêtre chimpanzé commun il y a approximativement 6 million d'années, est toujours non décrite. Afin de caractériser son évolution et d'identifier ce qui a permis à l'humain de développer le langage interne, Hauser, Chomsky et al. (2002) mentionnent la nécessité de conduire des études comparatives incluant tous les vertébrés (et même plus!) en y incluant l'étude d'autres facultés que celle du langage (p. ex., la navigation, la quantification des nombres et les relations sociales). Nous suggérons que les études comparatives doivent également élargir leur corpus de langage interne aux vocalisations nonlinguistiques humaines puisque d'une



part, elles constituent le plus proche analogue des vocalisations animales<sup>1</sup> et d'autre part, le répertoire vocal chez les animaux est uniquement composé de vocalisations (qui ne constitue pas un langage interne).

La comparaison des vocalisations entre les espèces caractériserait des éléments importants, lesquels contribueraient certainement à identifier ce qui a permis à l'humain de développer le langage interne. Plus spécifiquement, elle permettrait d'identifier et de caractériser les vocalisations qui sont partagées entre les espèces et celles qui sont partagées entre certaines espèces seulement (p. ex., une vocalisation utilisée par l'humain et le chimpanzé, mais inexistante du répertoire du chien des prairies).

En faisant l'hypothèse que les vocalisations nonlinguistiques utilisées par l'humain sont plus nombreuses, complexes, raffinées et subtiles que celles utilisées par les animaux, nous suggérons que les composantes uniques à l'humain ne sont pas spécifiques au langage mais s'élargissent au moins à la faculté du langage au sens large du terme.

La comparaison des vocalisations entre les espèces permettrait donc de savoir si, en effet, des vocalisations sont uniques à l'humain (ou à une autre espèce). Si tel est le cas, il serait nécessaire de les caractériser afin d'extraire ce que ces vocalisations ont en commun et de connaître pourquoi ces vocalisations ne font pas parties du répertoire vocal des autres espèces. Par exemple, nécessitent-elles un seuil minimum de théorie de l'esprit (p. ex., vocalisation exprimant de l'empathie), impliquent-elles des émotions morales? De plus, la liaison entre les vocalisations (partagées et non partagées entre les espèces) avec les autres facultés (p. ex., la navigation) permettrait d'identifier les caractéristiques communes sous-jacentes à ces différentes facultés (p. ex., la théorie de l'esprit). Enfin, ces comparaisons

---

<sup>1</sup> Il est à noter que, malgré que les vocalisations animales incluent les vocalisations humaines, l'utilisation du terme «vocalisations animales» exclue les vocalisations humaines pour plus de clarté dans le texte (idem pour le terme «animaux», qui exclue dans le cadre de ce travail les humains).

entre vocalisations et les liaisons entre facultés fourniraient des indices quant aux différentes hypothèses suggérant ce qui a permis à l'humain de développer le langage interne (p. ex., existence de la récursivité, d'une théorie de l'esprit, de la capacité d'imitation). Par exemple, si seuls les humains produisent des vocalisations exprimant de l'empathie, ceci irait dans le sens de l'hypothèse que l'absence d'une théorie de l'esprit chez les animaux ne leur permette pas de développer un langage interne.

Enfin, il est important de décrire les vocalisations nonlinguistiques chez l'humain: leur parcours développemental, la continuité (ou la discontinuité) entre les vocalisations nonlinguistiques et linguistiques, ainsi que les différences culturelles des vocalisations nonlinguistiques chez l'humain. Les études comparatives entre les espèces doivent également inclure ces variables.

En bref, la mise en commun de l'identification des vocalisations uniques à l'humain et celles partagées entre les espèces, ainsi que la description de la continuité entre vocalisation nonlinguistiques et linguistiques permettra certainement d'identifier des mécanismes sous-jacents en partie responsables de la composante unique à l'humain qui lui a permis de développer la faculté de langage interne. Les vocalisations nonlinguistiques représentent une source importante d'information sur l'évolution de la faculté du langage au sens large du terme, tant au niveau phylogénique qu'ontogénique.

*Pourquoi essayons-nous d'apprendre aux singes notre langage? Essayons plutôt de communiquer avec eux par des vocalisations nonlinguistiques. Nous pourrions ainsi partager un même mode de communication.*

## Bibliographie

- Abercrombie HC, Schaefer SM, et al. (1998). Metabolic rate in the right amygdala predicts negative affect in depressed patients. *Neuroreport* 9(14): 3301-7.
- Adams II, Jones DG (1987). Effects on normal and pathological aging on brain morphology: Neurons and synapses. *Current Topics in Research on Synapses*. Jones DG. New York, Liss. 4: 1-84.
- Adolphs R (2002a). Neural systems for recognizing emotion. *Curr Opin Neurobiol* 12(2): 169-77.
- Adolphs R (2002b). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behav Cogn Neurosci Rev* 1: 21-61.
- Aggleton JP (2000). *The amygdala: A functional analysis*. New York, Oxford University Press.
- Albas DC, McCluskey KW, et al. (1976). Perception of the emotional content of speech: A comparison of two Canadian groups. *J Cross Cult Psychol* 7: 481-89.
- Albert MS (1996). Cognitive and neurobiologic markers of early Alzheimer disease. *Proc Natl Acad Sci U S A* 93(24): 13547-51.
- Amunts K, Weiss PH et al. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitecturally defined stereotaxic space-The roles of Brodmann areas 44 and 45. *NeuroImage* 22(1): 42-56.
- Anderson AK, Phelps EA (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport* 9(16): 3607-13.
- Anderson AK, Christoff K et al. (2003a). Dissociated neural representations of intensity and valence in human olfaction. *Nature Neurosci* 6(2): 196-202.

- Anderson AK, Sobel N. (2003b). Dissociating intensity from valence as sensory inputs to emotion. *Neuron* 39(4): 581-3.
- Bachorowski JA (1999). Vocal expression and perception of emotion. *Curr Dir Psychol Sci* 8: 53-7.
- Bachorowski JA, Owren MJ (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychol Sci* 6: 219-24.
- Bachorowski JA, Owren MJ (2003). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Ann N Y Acad Sci*, 1000: 244-65.
- Bachorowski JA, Owren MJ (1999). Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *J Acoust Soc Am* 106(2): 1054-63.
- Banse R, Scherer KR (1996). Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70(3): 614-36.
- Bard R, Gardner KH (1996). Influences on development in infant chimpanzees: Enculturation, temperament, and cognition. *Reaching into Thought: The Minds of the Great Apes*. Russon A, Bard KA, Parker ST. Cambridge, Cambridge University Press: 235-56.
- Barr RG, Hopkins B, et al. (2000). *Crying as a Sign, a Symptom, & a Signal*. New York, Cambridge University Press.
- Barrett J, Pike GB, et al. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *Eur J Neurosci* 19(2): 458-64.
- Baum SR, Pell MD (1999). The neural basis of prosody: Insight from lesion studies and

- neuroimaging. *Aphasiology* 13: 581-608.
- Baxter MG, Murray EA (2002). The amygdala and reward. *Nat Rev Neurosci* 3(7): 563-73.
- Beauregard M, Levesque J, Bourgoin P (2001). Neural correlates of conscious self-regulation of emotion. *J Neurosci* 21(18): RC165.
- Beckman ME, Edwards J (1990). Lengthening and shortening and the nature of prosodic constituency. *Papers in Laboratory Phonology I: Between Grammar and the Physics of Speech*. Kingston J, Beckman ME. Cambridge, England, Cambridge University Press: 152-78.
- Behrens S (1985). The perception of stress and lateralization of prosody. *Brain Lang* 26: 332-48.
- Beier EG, Zautra AJ (1972). Identification of vocal communication of emotions across cultures. *J Consult Clin Psychol* 39: 166.
- Beldoch M (1964). Sensitivity to expression meaning in three modes of communication. *The Communication of Emotional Meaning*. Davitz JR. New York, McGraw-Hill: 31-42.
- Belin P, Fecteau S, Bédard C (2004). Thinking the voice: Neural correlates of voice perception. *Trends Cogn Sci* 8(3): 129-35.
- Belin P, Zatorre RJ (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14(16): 2105-9.
- Belin P, Zatorre RJ, et al. (2000). Voice-selective areas in human auditory cortex. *Nature* 403(6767): 309-12.
- Bentley P, Vuilleumier P, et al. (2003). Effects of attention and emotion on repetition



priming and their modulation by cholinergic enhancement. *J Neurophysiol* 90(2): 1171-81.

- Bertrand M (1969). *The Behavioral Repertoire of the Stumptail Macaque*. Karger, Basel.
- Blessner B (1972). Speech perception under conditions of spectral transformation. I. Phonetic characteristics. *J Speech Hear Res* 15(1): 5-41.
- Blonder LX, Bowers D (1991). The role of the right hemisphere in emotional communication. *Brain* 114(3): 1115-27.
- Boiten F (1996). Autonomic response patterns during voluntary facial action. *Psychophysiology* 33(2): 123-31.
- Borod JC (1993). Cerebral mechanisms underlying facial, prosodic, and lexical emotional expression: A review of neuropsychological studies and methodological issues. *Neuropsychology* 7: 445-63.
- Borod JC, Cicero BA, et al. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology* 12(3): 446-58.
- Borod JC, Koff E, et al. (1985). Channels of emotional expression in patients with unilateral brain damage. *Arch Neurol* 42: 345-8.
- Bowers D, Coslett HB (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. *Neuropsychologia* 25(2): 317-28.
- Bradley P, Mogg K, et al. (1994). Implicit and explicit memory for emotional information in non-clinical subjects. *Behav Res Ther* 32(1): 65-78.
- Brådvik B, Dravins C, et al. (1991). Disturbances of speech prosody following right



hemisphere infarcts. *Acta Neurol Scand* 84: 114-26.

- Braun K, Poeffel G (2001). Recognition of mother's voice evokes metabolic activation in the medial prefrontal cortex and lateral thalamus of Octodon Degus pups. *Neuroscience* 103(4): 861-4.
- Breitenstein C, Daum I, et al. (1998). Emotional processing following cortical and subcortical brain damage: Contribution of the fronto-striatal circuitry. *Behav Neurol* 11(1): 29-42.
- Breiter HC, Etcoff NL, et al. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron* 17(5): 875-87.
- Brown BL (1980). The detection of emotion in vocal qualities. *Language: Social Psychological Perspectives*. Giles H, Robinson WP, Smith PM. Oxford, Pergamon: 237-45.
- Brown P, Fraser C (1979). Speech as a marker of situation. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 33-62.
- Brown P, Levinson S (1979). Social structure, groups and interaction. *Social Markers in Speech*. KR Scherer, Giles H. Cambridge, Cambridge University Press: 291-341.
- Bruce V, Young A (1986). Understanding face recognition. *Br J Psychol* 77: 305-27.
- Bryan K (1989). Language prosody in the right hemisphere. *Aphasiology* 3: 285-99.
- Buchanan TW, Lutz K, et al. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Brain Res Cogn Brain Res* 9(3): 227-38.
- Burns KL, Beier EG (1973). Significance of vocal and visual channels in the decoding of

- emotional meaning. *J Comm* 23: 118-30.
- Cahill L, McGaugh JL (1998). Mechanisms of emotional arousal and lasting declarative memory. *TINS* 21: 294 -99.
- Cahill L, Haier RJ, et al. (2001). Sex-related difference in amygdala activity during emotionally influenced memory storage. *Neurobiol Learn Mem* 75(1): 1-9.
- Cancelliere AE, Kertesz A (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain Cogn* 13(2): 133-47.
- Canli T, Zhao Z, et al. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behav Neurosci* 115(1): 33-42.
- Canli T, Desmond JE, et al. (2002). Sex differences in the neural basis of emotional memories. *PNAS* 99(16): 10789-94.
- Canli T, Sivers H, et al. (2002). Amygdala response to happy faces as a function of extraversion. *Science* 296(5576): 2191.
- Carmon A, Nachshon I (1973). Ear asymmetry in perception of emotional non-verbal stimuli. *Acta Psychol (Amst)* 37(6): 351-7.
- Chevalier-Sokolnikoff S (1974). *The Ontogeny of Communication in the Stumptail Macaque (Macaca Arctoides)*. Karger, Basel.
- Chomsky N (1959). A review of BF Skinner's verbal behavior. *Language* 35: 26-58.
- Chomsky N, Halle M (1968). *The Sound Pattern of English*. New York, Harper & Row.
- Church BA, Schacter DL (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *J Exp Psychol Learn Mem Cogn* 20(3): 521-33.

- Corwin MJ, Golub HL (1996). The infant cry: What can it tell us? *Curr Probl Pediatr* 26: 325-34.
- Cosmides L (1983). Invariances in the acoustic expression of emotion during speech. *J Exp Psychol Hum Percept Perform* 9: 864-81.
- Creutzfeldt O, Ojemann G, et al. (1989). Neuronal activity in the human lateral temporal lobe. I. Responses to speech. *Exp Brain Res* 77: 451-75.
- Darby DG (1993). Sensory aprosodia: A clinical clue to lesions of the inferior division of the right middle cerebral artery? *Neurology* 43(3 Pt 1): 567-72.
- Darwin C (1872/1965). *The Expression of the Emotions in Man and Animals*. Chicago, University of Chicago Press.
- Davidson RJ (1992a). Anterior cerebral asymmetry and the nature of emotion. Special issue: The role of frontal lobe maturation in cognitive and social development. *Brain Cogn* 20: 125-51.
- Davidson RJ (1992b). Prolegomenon to the structure of emotion: Gleanings from neuropsychology. *Cogn Emotion* 6: 245-68.
- Davidson RJ (1995). Cerebral asymmetry, emotion, and affective style. *Brain Asymmetry*. Davidson RJ, Hugdahl K. Cambridge (MA), MIT Press: 361-87.
- Davidson RJ, Fox NA (1982). Asymmetrical brain activity discriminates between positive and negative affective stimuli in human infants. *Science* 218(4578): 1235-7.
- Davidson RJ, Irwin W (1999). The functional neuroanatomy of emotion and affective style. *Trends Cogn Sci* 3(1): 11-21.
- Davitz JR, Davitz LJ (1959). The communication of feelings by content-free speech. *J*

*Comm* 9: 6-13.

DeCasper AJ, Fifer WP (1980). Of human bonding: Newborns prefer their mothers' voices. *Science* 208(4448): 1174-6.

Démonet JF, Chollet F, et al. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain* 115: 1753-68.

Diehl RL, Lotto AJ, Holt LL (2004). Speech perception. *Annu Rev Psychol* 55: 149-79.

Dimitrovsky L (1964). The ability to identify the emotional meaning of vocal expressions at successive levels. *The Communication of Emotional Meaning*. Davitz JR. New York, McGraw-Hill: 69-86.

Dolan RJ, Vuilleumier P (2003). Amygdala automaticity in emotional processing. *Ann N Y Acad Sci* 985: 348-55.

Doupe AJ (1997). Song- and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *J Neurosci* 17(3): 1147-67.

Doupe AJ, Kuhl P (1999). Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567-631.

Drummond JE, McBride ML, et al. (1993). The development of mothers understanding of infant crying. *Clin Nurs Res* 2: 396-413.

Dusenbury D, Knower FH (1938). Experimental studies of the symbolism of action and voice (III): A study of the specificity of meaning in facial expression. *Q J Speech* 24: 424-35.

Dysktra K, Gandour J, et al. (1995). Disruption of prosody after frontal lobe seizures in the non-dominant hemisphere. *Aphasiology* 9(5): 453-76.

- Ehlers L, Dalby M (1987). Appreciation of emotional expressions in the visual and auditory modality in normal and brain damaged patients. *Acta Neurol Scand* 76: 251-6.
- Eisenberg N (2000). Emotion, regulation, and moral development. *Annu Rev Psychol* 51: 665-97.
- Ekman P (1992a). An argument for basic emotions. *Cogn Emotion* 6: 169-200.
- Ekman P (1992b). Are there basic emotions? *Psychol Rev* 99(3): 550-3.
- Ekman P (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. *Psychol Bull* 115: 268-87.
- Ekman P, Levenson RW, Friesen W (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221(4616): 1208-10.
- Ellis HD, Jones DM, et al. (1997). Intra- and inter-modal repetition priming of familiar faces and voices. *Br J Psychol* 88: 143-56.
- Eugène F, Lévesque J et al. (2003). The impact of individual differences on the neural circuitry underlying sadness. *NeuroImage* 19, 354-64.
- Fairbanks G, Pronovost W (1939). An experimental study of the pitch characteristic of the voice during the expression of emotion. *Speech Monographs* 6: 87-104.
- Fenster CA, Blake LK, et al. (1977). Accuracy of vocal emotional communications among children and adults and the power of negative emotions. *J Comm Disord* 10: 301-14.
- Ferguson CA (1964). Baby talk in six languages. *Am Anthropol* 66: 103-14.
- Ferreira F (1993). The creation of prosody during sentence production. *Psychol Rev* 100:

233-53.

- Fichtel C, Hammerschmidt K (2003). Responses of squirrel monkeys to their experimentally modified mobbing calls. *J Acoust Soc Am* 113(5): 2927-32.
- Fischer H, Tillfors M, et al. (2001). Dispositional pessimism and amygdala activity: A PET study in healthy volunteers. *Neuroreport* 12(8): 1635-8.
- Fischer H, Wright CI, et al. (2003). Brain habituation during repeated exposure to fearful and neutral faces: a functional MRI study. *Brain Res Bull* 59(5): 387-92.
- Fitch WT (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J Acoust Soc Am* 102: 1213-22.
- Frank MG, Stennett J (2001). The forced-choice paradigm and the perception of facial expressions of emotion. *J Pers Soc Psychol* 80(1): 75-85.
- Frick RW (1985). Communication emotion: The role of prosodic features. *Psychol Bull* 97: 412-29.
- Friston KJ, Holmes AP, et al. (1999). How many subjects constitute a study? *NeuroImage* 10(1): 1-5.
- Gaillard WD, Sachs BC, et al. (2003). Developmental aspects of language processing: fMRI of verbal fluency in children and adults. *Hum Brain Mapp* 18(3): 176-85.
- Gainotti G (2001). Disorders of emotional behaviour. *J Neurol* 248(9): 743-9.
- Gauthier I, Nelson CA (2001). The development of face expertise. *Curr Opin Neurobiol* 11(2): 219-24
- George MS, Ketter TA, et al. (1995). Brain activity during transient sadness and happiness in healthy women. *Am J Psychiatry* 152(3): 341-51.



- George MS, Ketter TA, et al. (1996a). Gender differences in regional cerebral blood flow during transient self-induced sadness or happiness. *Biol Psychiatry* 40(9): 859-71.
- George MS, Parekh PI, et al. (1996b). Understanding emotional prosody activates right hemisphere regions. *Arch Neurol* 53(7): 665-70.
- Gerken L (1996a). Prosodic structure in young children's language production. *Language* 72(4): 683-712.
- Gerken L (1996b). Prosody's role in language acquisition and adult parsing. *J Psycholinguist Res* 25(2): 345-56.
- Ghazanfar AA, Hauser M (2001). The auditory behaviour of primates: A neuroethological perspective. *Curr Opin Neurobiol* 11: 712-20.
- Giles H (1979). Ethnicity markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press.
- Giles H, Scherer K, et al. (1979). Speech markers in social interaction. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 343-81.
- Goodall J (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge, Belknap Press.
- Gorelick PB, Ross ED (1987). The aprosodias: Further functional-anatomical evidence for the organization of affective language in the right hemisphere. *J Neurol Neurosurg Psychiatry* 50: 553-60.
- Gouzoules H, Gouzoules S (2000). Agonistic screams differ among four species of macaques: The significance of motivation-structure rules. *Anim Behav* 59: 501-12.
- Gouzoules S, Gouzoules H, et al. (1984). Rhesus monkey (*Maccaca mulatta*) screams:

- Representational signaling in the recruitment of agonistic aid. *Anim Behav* 23: 182-93.
- Grace JA, Amin N, et al. (2003). Selectivity for conspecific song in the zebra finch auditory forebrain. *J Neurophysiol* 89(1): 472-87.
- Grady CL, Van Meter JW, et al. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport* 8(11): 2511-6.
- Gray JA (1990). Brain systems that mediate both emotion and cognition. *Cogn Emotion* 4: 269-88.
- Green RS, Cliff N (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. *Percept Psychophys* 17: 429-38.
- Greene JD, Sommerville RB, et al. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537): 2105-8.
- Grill-Spector K, Knouf N, et al. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neurosci* e-pub.
- Guenther FH (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biol Cybernet* 72: 43-53.
- Gunning-Dixon FM, Gur RC, et al. (2003). Age-related differences in brain activation during emotional face processing. *Neurobiol Aging* 24(2): 285-95.
- Habib M, Daquin G, et al. (1995). Mutism and auditory agnosia due to bilateral insular damage- role of the insula in human communication. *Neuropsychologia* 33(3): 327-39.
- Haggard MP, Parkinson AM (1971). Stimulus and task factors as determinants of ear

advantages. *Q J Exp Psychol* 23(2): 168-77.

Hartman DE, Danahuer JL (1976). Perceptual features of speech for males in four perceived age decades. *J Acoust Soc Am* 59: 713-5.

Hauser M (1991). Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations. *Ethology* 89: 29-46.

Hauser MD, Chomsky N, Fitch WT (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598): 1569-79.

Haxby JV, Hoffman EA, et al. (2002). Human neural systems for face recognition and social communication. *Biol Psychiatry* 51(1): 59-67.

Heilman KM, Bowers D (1984). Comprehension of affective and non-affective prosody. *Neurology* 34: 917-21.

Heilman KM, Scholes R, et al. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *J Neurol Neurosurg Psychiatry* 38(1): 69-72.

Helfrich H (1979). Age markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 63-108.

Herrero JV, Hillix WA (1990). Hemispheric performance in detecting prosody. A competitive dichotic listening task. *Percept Mot Skills* 71: 479-86.

Hickok G, Poeppel D (2000). Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4: 131-8.

Holmes A, Vuilleumier P, et al. (2003). The processing of emotional facial expression is gated by spatial attention: Evidence from event-related brain potentials. *Brain Res Cogn Brain Res* 16(2): 174-84.

- Hornak J, Rolls ET, et al. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia* 34(4): 247-61.
- Hornstein MG (1967). Accuracy of emotional communication and interpersonal compatibility. *J Pers* 35: 20-8.
- Imaizumi S, Mori K, et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport* 8(12): 2809-12.
- Ivry RB, Leiby PC (1993). Hemispheric differences in auditory perception are similar to those found in visual perception. *Psychol Sci* 4: 41-5.
- Izard CE (1972). *Patterns of Emotion: A New Analysis of Anxiety and Depression*. New York, Academic Press Inc.
- Izard CE (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol Rev* 99(3): 561-5.
- Joanette Y, Goulet P, et al. (1990). *Right Hemisphere and Verbal Communication*. New York, Springer Verlag.
- Johnson DL, Wiebe JS, et al. (1999). Cerebral blood flow and personality: A positron emission tomography study. *Am J Psychiatry* 156(2): 252-7.
- Johnson WF, Emde RN, et al. (1986). Recognition of emotion from vocal cues. *Arch Gen Psychiatry* 43(3): 280-3.
- Jones B (1979). Sex and visual field effects on accuracy and decision making when subjects classify male and female faces. *Cortex* 15(4): 551-60.
- Jones B, Anuza T (1982). Effects of sex, handedness, stimulus and visual field on mental

rotation. *Cortex* 18(4): 501-14.

Josephs O, Henson RNA (1999). Event-related functional magnetic resonance imaging : Modelling, inference and optimization. *Phil Trans R Soc Lond B* 354: 1215-28.

Jürgens U (1986). The squirrel monkey as an experimental model in the study of cerebral organization of emotional vocal utterances. *Eur Arch Psychiatry Neurol Sci* 236: 40-3.

Jürgens U (1998). Neuronal control of mammalian vocalization, with special reference to the squirrel monkey. *Naturwissenschaften* 85: 376-88.

Karama S, Lecours AR, et al. (2002). Areas of brain activation in males and females during viewing of erotic film excerpts. *Hum Brain Mapp* 16(1): 1-13.

Kawashima R, Itoh M, et al. (1993). Changes of regional cerebral blood flow during listening to an unfamiliar spoken language. *Neurosci Lett* 161(1): 69-72.

Kempey T, Morton J (1982). The effects of priming with regularity and irregularity related words in auditory word recognition. *Br J Psychol* 73: 441-54.

Kisilevsky BS, Hains SMJ et al. (2003). Effects of experience on fetal voice recognition. *Psychol Sci* 14(3): 220-4.

Klatt DH (1975). Voice onset time, frication, and aspiration in word-initial consonant cluster. *J Speech Hear Res* 18(4): 686-706.

Klatt DH, Klatt LC (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am* 87(2): 820-57.

Knosche TR, Lattner S, et al. (2002). Early parallel processing of auditory word and voice information. *NeuroImage* 17: 1493-1503.

- Knower FH (1941). Analysis of some experimental variations of simulated vocal expressions of the emotions. *J Soc Psychol* 14: 369-72.
- Kotz SA, Alter K, et al. (2000). The interface between prosodic and semantic processes: An ERP study. *J Cogn Neurosci (suppl.)*: 121.
- Kotz SA, Meyer M, et al. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain Lang* 86: 366-76.
- Kramer E (1964). Elimination of verbal cues in judgments of emotion from voice. *J Abnorm Psychol* 68: 390-6.
- Kreiman J (1997). Listening to voices: Theory and practice in voice perception research. *Talker Variability in Speech Research*. Johnson K, Mullenix J. New York, Academic Press: 85-108.
- Kuhl PK, Tsao FM, et al. (2001). Language/culture/mind/brain. Progress at the margins between disciplines. *Ann N Y Acad Sci* 935: 136-74.
- Kuwabara H, Ohgushi K (1987). Contributions of vocal tract resonant frequencies and bandwidths to the personal perception of speech. *Acoustica* 63: 120-8.
- Ladd DR (1996). *Intonational Phonology*. Cambridge University Press.
- Ladd DR, Silverman KEA (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker effect. *J Acous Soc Am* 78(2): 435-44.
- Lakshminarayanan K, Ben Shalom D, et al. (2003). The effect of spectral manipulations on the identification of affective and linguistic prosody. *Brain Lang* 84(2): 250-63.
- Lalande S, Brawn CM, et al. (1992). Effects of right and left hemisphere cerebrovascular



lesions on discrimination of prosodic and semantic aspects of affect in sentences. *Brain Lang* 42: 165-86.

Lamendella JT (1977). The limbic system in human communication. *Studies in Neurolinguistics*. Whitaker H, Whitaker HA. New York, Academic Press: 157-222.

Lane RD, Reiman EM, et al. (1997). Neuroanatomical correlates of happiness, sadness, and disgust. *Am J Psychiatry* 154(7): 926-33.

Lane RD, Reiman EM, et al. (1998). Neural correlates of levels of emotional awareness. Evidence of an interaction between emotion and attention in the anterior cingulate cortex. *J Cogn Neurosci* 10(4): 525-35.

Lang PJ, Bradley MM, et al. (1997). A motivational analysis of emotion: Reflex-cortex connections. *Psychol Sci* 3: 44-9.

Lang PJ, Bradley MM et al. (1999). International Affective Picture System (IAPS): Instruction Manual and Affective Ratings. *Technical Report A-4*, The Center for Research in Psychophysiology, University of Florida.

Lass NJ, Hughes KR, et al. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *J Acoust Soc Am* 59(3): 675-8.

Levenson RW, Carstensen LL, et al. (1991). Emotion, physiology, and expression in old age. *Psychol Aging* 6(1): 28-35.

Levenson RW, Ekman P, et al. (1992). Emotion and autonomic nervous system activity in the Minangkabau of west Sumatra. *J Pers Soc Psychol* 62(6): 972-88.

Levy PK (1964). The ability to express and perceive vocal communication of feelings. *The Communication of Emotional Meaning*. Davitz JK. New York, McGraw-Hill: 43-55.

- Ley RG, Bryden MP (1982). A dissociation of right and left hemisphere effects for recognizing emotional tone and verbal content. *Brain Cogn* 1: 3-9.
- Linville SE (1996). The sound of senescence. *J Voice* 10: 190-200.
- Luo Q, Peng D, et al. (2004). Emotional valence of words modulates the subliminal repetition priming effect in the left fusiform gyrus: An event-related fMRI study. *NeuroImage* 21: 414-21.
- Maratos EJ, Allan K, Rugg MD (2000). Recognition memory for emotionally negative and neutral words: An ERP study. *Neuropsychologia* 38(11): 1452-65.
- Margoliash D (1986). Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *J Neurosci* 6(6): 1643-61.
- Masataka N, Kohda M (1988). Primate play vocalizations and their functional significance. *Folia Primatol* 50: 152-6.
- Al-Maskati HA, Zbrozyna AW (1989). Stimulation in prefrontal cortex area inhibits cardiovascular and motor components of the defence reaction in rats. *J Auton Nerv Syst* 28: 117-25.
- McClelland AGR et Pring L (1991). An investigation of cross-modality effects in implicit and explicit memory. *Q J Exp Psychol: Learn Mem Cogn* 19: 309-28.
- McCluskey KW, Albas DC (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *Int J Psychol* 16: 119-32.
- McCluskey KW, Albas DC, et al. (1975). Cross-cultural differences in the perception of emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Dev Psychol* 11: 15-21.

- Mesquita B, Frijda NH (1992). Cultural variations in emotions: A review. *Psychol Bull* 112(2): 179-204.
- Mesquita B, Frijda NH, et al. (1997). Culture and emotion. *Handbook of Cross-Cultural Psychology*. Berry PB, Saraswathi TS. Boston, Allyn & Bacon. 12: 255-97.
- Mesulam MM (1998). From sensation to cognition. *Brain* 121: 1013-52.
- Meyer E, Alter K, et al. (2002). Functional MRI evidence reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum Brain Mapp* 17: 73-88.
- Miller CT, Miller J, et al. (2001). Selective phonotaxis by cotton-top tamarins. *Behaviour* 138: 811-26.
- Mitchell RL, Elliott R, et al. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* 41(10): 1410-21.
- Moll J, de Oliveira-Souza R, et al. (2003). The Morals and the human brain: A working model. *Neuroreport* 14(3): 299-305.
- Moll J, de Oliveira-Souza R, et al. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *J Neurosci* 22(7): 2730-6.
- Monrad-Krohn GH (1947). Dysprosody or altered 'melody of language. *Brain* 70: 405-15.
- Morgan JL, Meier RP, et al. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cogn Psychol* 19: 498-550.
- Morris JS, Frith CD, et al. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* 383(6603): 812-5.

- Morris JS, Scott SK, Dolan RJ (1999). Saying it with feeling: Neural responses to emotional vocalizations. *Neuropsychologia* 37(10): 1155-63.
- Morton EF (1977). On the occurrence and significance of motivational-structural rules in some bird and mammal sounds. *Am Nat* 111: 855-69.
- Mullennix JW, Johnson KA, et al. (1995). The perceptual representation of voice gender. *J Acoust Soc Am* 98(6): 3080-95.
- Murray EA (1979). Infant crying as an elicitor of parental behavior: An examination of two models. *Psychol Bull* 86: 191-215.
- Murray IR, Arnott JL (1993). Toward a simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *J Acoust Soc Am* 93(2): 1097-108.
- Nadeau SE, Williamson DJ, et al. (1998). Functional imaging: Heterogeneity in task strategy and functional anatomy and the case for individual analysis. *Neuropsychiatry Neuropsychol Behav Neurol* 11(2): 83-96.
- Nakamura K, Kawashima R, et al. (2001) Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia* 39(10): 1047-54.
- Newman J (1985). The infant cry of primates: An evolutionary perspective. *Infant Crying: Theoretical and Research Perspectives*. Lester BM, Boukydis CFZ. New York, Plenum Press: 307-23.
- Oatley K, Johnson-Laird PN (1987). Towards a cognitive theory of emotions. *Cogn Emotion* 1: 29-50.
- O'Leary DS, Andreason NC, et al. (1996). A positron emission tomography study of binaurally and dichotically presented stimuli: Effects of level of language and

directed attention. *Brain Lang* 53(1): 20-39.

Ortony A, Turner TJ (1990). What's basic about basic emotions? *Psychol Rev* 97(3): 315-31.

Osser HA (1964). A distinctive feature analysis of the vocal communication of emotion. *Diss Abstr* 25: 3708.

Ouellette G, Baum SR (1993). Acoustic analysis of prosodic cues in left and right-hemisphere-damaged patients. *Aphasiology* 8(3): 257-83.

Pakosz M (1983). Attitudinal judgments in intonation: Some evidence for a theory. *J Psycholinguist Res* 12: 311-26.

Panksepp J (1992). A critical role for "affective neuroscience" in resolving what is basic about basic emotions. *Psychol Rev* 99(3): 554-60.

Pardo JV, Pardo PJ, et al. (1993). Neural correlates of self-induced dysphoria. *Am J Psychiatry* 150(5): 713-9.

Pell MD (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia* 36(8): 701-15.

Pell MD (1999). The temporal organization of affective and non-affective speech in patients with right-hemisphere infarcts. *Cortex* 35(4): 455-77.

Pell MD, Baum SR (1997). Unilateral brain damage, prosodic comprehension deficits, and the acoustic cues to prosody. *Brain Lang* 57(2): 195-214.

Peper M, Irle E (1997). Categorical and dimensional decidong of emotional intonations in patients with focal brain lesions. *Brain Lang* 58: 233-64.

- Pessoa L, Ungerleider LG (2004). Neuroimaging studies of attention and the processing of emotion-laden stimuli. *Prog Brain Res* 144: 171-82.
- Petersen SE, Fox PT, et al. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331: 585-9.
- Poepfel D, Hickok G (2004). Towards a new functional anatomy of language. *Cognition* 92(1-2): 1-12.
- Pfaff PL (1954). An experimental study of the communication of feeling without contextual material. *Speech Monographs* 21: 155-6.
- Phan KL, Wager T, et al. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage* 16(2): 331-48.
- Phan KL, Taylor SF, et al. (2004a). Neural correlates of individual ratings of emotional salience: A trial-related fMRI study. *NeuroImage* 21(2): 768-80.
- Phan KL, Wager TD, et al. (2004b). Functional neuroimaging studies of human emotions. *CNS Spectr* 9(4): 258-66.
- Phillips ML, Young AW, et al. (1998). Neural responses to facial and vocal expressions of fear and disgust. *Proc R Soc Lond B Biol Sci* 265(1408): 1809-17.
- Phillips ML, Young AW, et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* 389(6650): 495-8.
- Pihan H, Ackermann H, et al. (1997). The cortical processing of perceived emotion: A DC-potential study on affective prosody. *Neuroreport* 8: 623-27.
- Pihan H, Altenmuller E, et al. (2000). Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system: A DC-



- potential study. *Brain* 123: 2338-49.
- Pittam J, Scherer KR (1993). Vocal expression and communication of emotion. *Handbook of Emotions*. Lewis M, Haviland JM. New York, London, The Guilford Press: 185-97
- Pollack I, Rubenstein H, et al. (1960). Communication of verbal modes of expression. *Lang Speech* 3: 121-30.
- Poremba A, Malloy M, et al. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427: 448-51.
- Pugh KR, Offywitz BA, et al. (1996). Auditory selective attention: An fMRI investigation. *NeuroImage* 4(3 Pt 1): 159-73.
- Rama P, Martinkauppi S, et al. (2001). Working memory of identification of emotional vocal expressions: An fMRI study. *NeuroImage* 13(6 Pt 1): 1090-101.
- Reiman EM, Lane RD, et al. (1997). Neuroanatomical correlates of externally and internally generated human emotion. *Am J Psychiatry* 154(7): 918-25.
- Remez RE, Fellowes JM, et al. (1997). Talker identification based on phonetic information. *J Exp Psychol Hum Percept Perform* 23: 651-66.
- Rendall D (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J Acoust Soc Am* 113(6): 3390-402.
- Rendall D, Owren MJ, et al. (1998). The role of vocal track filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *J Acoust Soc Am* 103: 602-14.
- Rendall D, Rodman PS, et al. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim Behav* 51: 1007-15.

- Riede T, Zuberbuhler K (2003). The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. *J Acoust Soc Am* 114:1132-42.
- Roberts RJ, Weerts TC (1982). Cardiovascular responding during anger and fear imagery. *Psychol Rep* 50(1): 219-30.
- Robin DA, Tranel D, et al. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain Lang* 39: 539-55.
- Robinson BW (1976). Limbic influences on human speech. *Ann N Y Acad Sci* 280: 761-76.
- Robinson WP (1979). Speech markers and social class. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 211-49.
- Roland PE (1993). *Brain Activation*. New York, John Wiley.
- Ross ED (1981). The aprosodias: Functional anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol* 38: 561-9.
- Ross ED (1993). Nonverbal aspects of language. *Neurol Clin* 11(1): 9-23.
- Ross ED, Anderson B, et al. (1989). Crossed aprosodia in strongly dextral patients. *Arch Neurol* 46: 206-9.
- Ross ED, Edmondson JA, et al. (1986). The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis. *J Phonetics* 14: 283-302.
- Ross ED, Homan RW, et al. (1994). Differential hemispheric lateralization of primary and social emotions: Implications for developing a comprehensive neurology for emotions, repression, and the subconscious. *Neuropsychiatr Neuropsychol Behav Neurol* 7: 1-19.

- Ross ED, Thompson RD, et al. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain Lang* 56(1): 27-54.
- Ross M, Duffy RJ, et al. (1973). Contribution of the lower audible frequencies to the recognition of emotions. *Am Ann Deaf* 118: 37-42.
- Rubin DC, Friendly M (1986). Predicting which words get recalled: Measures of free recall, availability, goodness, emotionality, and pronunciability for 925 nouns. *Mem Cogn* 14: 79-94.
- Ruch W, Ekman P (2001). The Expressive Pattern of Laughter. *Emotion, Qualia and Consciousness*. Kaszniak A. Tokyo: World Scientific: 426-43.
- Russell JA (1980). A circumplex model of affect. *J Pers Sociol Psychol* 39: 1161-78.
- Russell JA (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychol Bull* 115(1): 102-41.
- Russell JA (2003). Core affect and the psychological construction of emotion. *Psychol Rev* 110(1): 145-72.
- Russell JA, Bachorowski JA, et al. (2003). Facial and vocal expressions of emotion. *Annu Rev Psychol* 54(1): 329-49.
- Russell JA, Barrett LF (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J Pers Soc Psychol* 76(5): 805-19.
- Russell JA, Fehr B (1987). Relativity in the perception of emotion in facial expressions. *J Exp Psychol: Gen* 116: 223-37.
- Sackeim HA, Greenberg MS, et al. (1982). Hemispheric asymmetry in the expression of positive and negative emotions: Neurologic evidence. *Arch Neurol* 39: 210-8.

- Sackeim HA, Gur RC (1979). Lateral asymmetry in intensity of emotional expression. *Neuropsychologia* 163: 473-81.
- Safer MA, Leventhal H (1977). Ear difference in evaluating emotional tones of voice and verbal content. *J Exp Psychol Hum Percept Perform* 3(1): 75-82.
- Samson Y, Belin P et al. (2001). Auditory perception and language: Functional imaging of speech sensitive auditory cortex. *Rev Neurol (Paris)* 157(8-9): 837-46.
- Sander D, Grafman J et al. (2003). The human amygdala: An evolved system for relevance detection. *Rev Neurosci* 14(4): 303-16.
- Sander K, Brechmann A, et al. (2003a). Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Res Brain Res Protoc* 11(2): 81-91.
- Sander K, Roth P, et al. (2003b). Left-lateralized fMRI in the temporal lobe of high repressive women during the identification of sad prosodies. *Cogn Brain Res* 16: 441-56.
- Sander K, Scheich H (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Brain Res Cogn Brain Res* 12(2): 181-98.
- Schacter DL, Tulving E (1994). What are the memory systems of 1994? *Memory System*. Schacter DL, Tulving E. Cambridge (MA), MIT Press: 1-38.
- Schaefer SM, Jackson DC, et al. (2002). Modulation of amygdalar activity by the conscious regulation of negative emotion. *J Cogn Neurosci* 14(6): 913-21.
- Scherer KR (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *J Pers* 40(2): 191-210.

- Scherer KR (1979a). Personality markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 147-209.
- Scherer KR (1981a). Speech and emotional states. *Speech Evaluation in Psychiatry*. Darby J. New York, Grune & Stralton: 189-220.
- Scherer KR (1981b). Vocal indicators of stress. *Speech Evaluation in Psychiatry*. Darby J. New York, Grune & Stralton: 171-87.
- Scherer KR (1982). Methods of research on vocal communication: Paradigms and parameters. *Handbook of Methods in Nonverbal Behavior Research*. Scherer KR, Ekman P. Cambridge, Cambridge University Press: 136-98.
- Scherer KR (1985). Vocal affect signaling: A comparative approach. *Advances in the Study of Behavior*. Rosenblatt J, Beer C, Busnel MC Slater PJB. New York, Academic Press. 15: 189-244.
- Scherer KR (1986). Vocal affect expression: A review and a model for future research. *Psychol Bull* 99: 143- 65.
- Scherer KR (1995). Expression of emotion in voice and music. *J Voice* 9(3): 235-48.
- Scherer KR (1997a). Profiles of emotion-antecedent appraisal: Testing theoretical predictions across cultures. *Cogn Emotion* 11: 113-50.
- Scherer KR (1997b). The role of culture in emotion-antecedent appraisal. *J Pers Soc Psychol* 73: 902-22.
- Scherer KR (2000). Psychological models of emotion. *The Neuropsychology of Emotion*. Borod J. Oxford/New York, Oxford University Press: 137-66.
- Scherer KR, Banse R, et al. (2001). Emotion inferences from vocal expression correlate

across languages and cultures. *J Cross Cult Psychol* 32(1): 76-92.

- Scherer KR, Kappas A (1988). Primate vocal expression of affective states. *Primate Vocal Communication*. Todt D, Goedecking P, Newman E. Heidelberg, Springer: 171-94.
- Scherer KR, Oshinsky JS (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation Emotion* 1(4): 331-46.
- Scherer KR, Wallbott HG (1994). Evidence for universality and cultural variation of differential emotion response patterning. *J Pers Soc Psychol* 66(2): 310-28.
- Scherer KR, Zei B (1988). Vocal indicators of affective disorders. *Psychother Psychosom* 49(3-4): 179-86.
- Schlanger BB (1973). Identification by normal and aphasic subjects of semantically meaningful and meaningless emotional toned sentences. *Acta Symbolica* 4: 30-8.
- Schlanger BB, Schlanger P (1976). The perception of emotionally toned sentences by the right hemisphere damaged and aphasic subjects. *Brain Lang* 3: 396-403.
- Schmitt JJ, Hartje W, et al. (1997). Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositionnal speech. *Cortex* 33: 65-81.
- Schneider F, Habel U, et al. (2000). Gender differences in regional cerebral activity during sadness. *Hum Brain Mapp* 9(4): 226-38.
- Schwartz GE, Weinberger DA, Singer JA (1981). Cardiovascular differentiation of happiness, sadness, anger, and fear following imagery and exercise. *Psychosom Med* 43(4): 343-64.
- Schweinberger SR, Herholz A, et al. (1997). Recognizing famous voices: Influence of



stimulus duration and different types of retrieval cues. *J Speech Lang Hear Res* 40: 453-63.

Schweinberger SR (2001). Human brain potential correlates of voice priming and voice recognition. *Neuropsychologia* 39: 921-36.

Scott SK, Blank CC, et al. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123(12): 2400-6.

Scott SK, Johnsrude IS (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci* 26: 100-7.

Scott SK, Young AW, et al. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature* 385(6613): 254-7.

Sebeok TA (1968). *Animal Communication: Techniques of Study and Results of Research*. Bloomington, Indiana University Press.

Sebeok TA (1977). *How Animals Communicate*. Bloomington, Indiana University Press.

Semple S, McComb K (2000). Perception of female reproductive state from vocal cues in a mammal species. *Proc R Soc London B* 267: 707-12.

Seron X, van der Kaa MA, et al. (1982). Decoding paralinguistic signals: Effects of semantic and prosodic cues on aphasic comprehension. *J Comm Disord* 15: 223-31.

Seyfarth RM, Cheney DL, et al. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science Wash DC* 210: 801-3.

Shiple-Brown F, Dingwall W (1988). Hemispheric processing of affective linguistic

intonation contours in normal subjects. *Brain Lang* 33: 16-26.

Siegmán AW (1978). The telltale voice: Nonverbal messages of verbal communication. *Nonverbal Behavior and Communication*. Siegmán AW, Feldstein S. Hillsdale, NJ, Erlbaum: 183-243.

Singh NC, Theunissen F (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114(6 Pt 1): 3394-411.

Small DM, Gregory MD, et al. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron* 39(4): 701-11.

Smith PM (1979). Sex markers in speech. *Social Markers in Speech*. Scherer KR, Giles H. Cambridge, Cambridge University Press: 109-46.

Smith WJ (1977). *The Behaviour of Communicating: The Ethological Approach*. Cambridge (MA), Harvard University Press.

Sobin C, Alpert M (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *J Psycholinguist Res* 28: 347-65.

Sobotka SS, Davidson RJ, et al. (1992). Anterior brain electrical asymmetries in response to reward and punishment. *Electroencephalogr Clin Neurophysiol* 83(4): 236-47.

Somerville LH, Kim H et al. (2004). Human amygdala responses during presentation of happy and neutral faces: correlations with state anxiety. *Biol Psychiatry* 55(9): 897-903.

Speedie LJ, Coslett HB, et al. (1984). Repetition of affective prosody in mixed transcortical aphasia. *Arch Neurol* 41: 268-70.

SpenceMJ, DeCasper AJ (1987) Prenatal experience with low-frequency maternal voice

- sounds influences neonatal perception of maternal voice samples. *Infant Behav Dev* 10: 133-42.
- Spence MJ, Freeman MS (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behav Dev* 19: 199-212.
- Starkstein SE, Federoff JP, et al. (1994). Neuropsychological and neuroradiological correlates of emotional prosody comprehension. *Neurology* 44: 515-22.
- Stemmler G (1989). The autonomic differentiation of emotions revisited: convergent and discriminant validation. *Psychophysiology* 26(6):617-32.
- Sullivan S, Ruffman T (2004). Emotion recognition deficits in the elderly. *Int J Neurosci* 114(3): 403-32.
- Tarr MJ, Gauthier I (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci* 3(8): 764-9.
- Thaut MH (2003). Neural basis of rhythmic timing networks in the human brain. *Ann N Y Acad Sci* 999: 364-73.
- Tompkins CA, Flowers CR (1985). Perception of emotional intonation by brain damaged adults: The influence of task processing levels. *J Speech Hear Res* 28: 527-83.
- Tosi O (1979). *Voice Identification: Theory and Legal Applications*. Baltimore, University Park Press.
- Turner TJ, Ortony A (1992). Basic emotions: Can conflicting criteria converge? *Psychol Rev* 99(3): 566-71.
- Tucker DM, Watson RT, et al. (1977). Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology* 27: 947-50.

- Tzourio N, Massioui FE, et al. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage* 5: 63-77.
- Van Bezooijen R, Otto, SA, et al. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *J Cross Cult Psychol* 14: 387-406.
- Van Bezooijen R (1984). *Characteristics and Recognizability of Vocal Expressions of Emotion*. Dordrecht, Foris.
- Van Hoof JA (1972). A comparative approach to the phylogeny of laughter and smiling. *Non-Verbal Communication*. Hinde RA. London, Cambridge University Press: 209-41.
- Van Lancker D, Canter GJ (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn* 1(2): 185-95.
- Van Lancker D, Cummings JL, et al. (1988). Phonagnosia: A dissociation between familiar and unfamiliar voices. *Cortex* 24: 195-209.
- Van Lancker D, Kreiman J, et al. (1985a). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *J Phonetics* 13: 19-38.
- Van Lancker D, Kreiman J, et al. (1985b). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *J Phonetics* 13: 39-52.
- Van Lancker D, Kreiman J, et al. (1989). Voice perception deficits: Neuroanatomical correlates of phonagnosia. *J Clinical Exp Neuropsychol* 11: 665-74.
- Van Lancker D, Sidtis JJ (1992). The identification of affective prosodic stimuli by left and right hemisphere damaged subjects: All errors are not created equal. *J Speech Hear*

*Res* 35: 963-70.

- Van Lawick-Goodall J (1968). A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. *Primates: Studies in Adaptation and Variability*. Jay PC. New York, Holt, Rinehart & Winston: 313-74.
- Vikingstad EM, George KP, et al. (2000). Cortical language lateralisation in right handed normal subjects using functional magnetic resonance imaging. *J Neurol Sci* 175: 17-27.
- Von Kriegstein K, Giraud AL (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage* e-pub.
- Von Kriegstein K, Eger E, et al. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res* 17: 48-55.
- Vuilleumier P, Armony JL, et al. (2001). Effects of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron*, 30(3): 829-41.
- Wager TD, Phan KL, et al. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: A meta-analysis of findings from neuroimaging. *NeuroImage* 19(3): 513-31.
- Wallbott HG, Scherer KR (1986). Cues and channels in emotion recognition. *J Abnorm Soc Psychol* 51: 690-99.
- Walton JH, Orlikoff RF (1994). Speaker race identification from acoustic cues in the vocal signal. *J Speech Hear Res* 37(4): 738-45.
- Weintraub S, Mesulam MM, et al. (1981). Disturbances in prosody: A right-hemisphere contribution to language. *Arch Neurol* 38: 742-4.

- Weiss DJ, Garibaldi BT, et al. (2001). The production and perception of long calls in cotton-top tamarins (*Saguinus oedipus*). *J Comp Psychol* 115: 258-71.
- Westermann G, Reck Miranda E (2004). A new model of sensorimotor coupling in the development of speech. *Brain Lang* 89(2): 393-400.
- Whiteside SP (1998). Identification of a speaker's sex: A study of vowels. *Percept Mot Skills* 86(2): 579-84.
- Wildgruber D, Pihan H, et al. (2002). Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. *NeuroImage* 15(4): 856-69.
- Williams CE, Stevens KN (1972). Emotions and speech: Some acoustical correlates. *J Acoust Soc Am* 52: 1238-50.
- Williams CE, Stevens KN (1981). Vocal correlates of emotional states. *Speech Evaluation in Psychiatry*. Darby J. New York, Grune & Stratton: 221-42.
- Wolf CG (1977). The processing of fundamental frequency in a dichotic matching task. *Brain Lang* 4: 70-7
- Wolf G, Gorski R, et al. (1972). Acquaintance and accuracy of vocal emotions. *J Comm* 22: 300-5.
- Zald DH (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Res Brain Res Rev* 41(1): 88-123.
- Zatorre RJ (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* 84: 566-72.
- Zatorre RJ, Belin P (2001). Spectral and temporal processing in human auditory cortex.



*Cereb Cortex* 11(10): 946-53.

Zatorre RJ, Belin P, et al. (2002). Structure and function of auditory cortex: Music and speech. *Trends Cogn Sci* 6(1): 37-46.

Zatorre RJ, Binder JR (2000). Functional and structural imaging of the human auditory system. *Brain Mapping: The Systems*. Academic Press: 365-402.

Zatorre RJ, Evans AC, et al. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* 14: 1908-19.

Zatorre RJ, Evans AC, et al. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256(5058): 846-9.

Zatorre RJ, Samson S (1991). Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain* 114: 2403-17.

Zbrozyna AW, Westwood DM (1991). Stimulation in prefrontal cortex inhibits conditioned increase in blood pressure and avoidance bar pressing in rats. *Physiol Behav* 49: 705-8.

Zeskind PS, Marshall TR (1988). The relation between variation in pitch and maternal perception of infant crying. *Child Dev* 59: 193-6.

Zuckerman M, Lipets MS, et al. (1975). Encoding and decoding nonverbal cues of emotion. *J Pers Soc Psychol* 32: 1068-76

