Université de Montréal

Ontogénèse et spécificité de la voix humaine

Par
Maude Beauchemin

Département de Psychologie
Faculté des Arts et des Sciences

Thèse présentée à la Faculté des Études Supérieures et Postdoctorales
en vue de l'obtention du grade de *Philosophiae* doctor (Ph.D.)
en Psychologie – Recherche et Intervention
option Neuropsychologie clinique

Juillet 2011

Université de Montréal

Faculté des Études Supérieures et Postdoctorales

Cette thèse intitulée :

Ontogénèse et spécificité de la voix humaine

Présentée par :

Maude Beauchemin

A été évaluée par un jury composé des personnes suivantes :

Michelle McKerral
Président-rapporteur

Maryse Lassonde
Directrice de recherche

Isabelle Peretz
Membre du jury

Dave Saint-Amour
Examinateur externe

Thérèse Cabana
Représentante du doyen de la FES

RÉSUMÉ

La voix est un stimulus auditif omniprésent dans notre environnement sonore. Elle permet non seulement la parole, mais serait aussi l'équivalent d'un visage auditif transmettant notamment des informations identitaires et affectives importantes. Notre capacité à discriminer et reconnaître des voix est socialement et biologiquement importante et elle figure parmi les fonctions les plus importantes du système auditif humain. La présente thèse s'intéressait à l'ontogénèse et à la spécificité de la réponse corticale à la voix humaine et avait pour but trois objectifs : (1) mettre sur pied un protocole électrophysiologique permettant de mesurer objectivement le traitement de la familiarité de la voix chez le sujet adulte; (2) déterminer si ce même protocole pouvait aussi objectiver chez le nouveau-né de 24 heures un traitement préférentiel d'une voix familière, notamment la voix de la mère; et (3) mettre à l'épreuve la robustesse d'une mesure électrophysiologique, notamment la *Fronto-Temporal Positivity to Voices*, s'intéressant à la discrimination pré-attentionnelle entre des stimuli vocaux et non-vocaux. Les résultats découlant des trois études expérimentales qui composent cette thèse ont permis (1) d'identifier des composantes électrophysiologiques (*Mismatch Negativity* et P3a) sensibles au traitement de la familiarité d'une voix; (2) de mettre en lumière un patron d'activation corticale singulier à la voix de la mère chez le nouveau-né, fournissant le premier indice neurophysiologique de l'acquisition du langage, processus particulièrement lié à l'interaction mère-enfant; et (3) de confirmer l'aspect pré-attentionnel de la distinction entre une voix et un stimulus non-vocal tout en accentuant la sélectivité et la sensibilité de la réponse corticale réservée au traitement de la voix.

Mots-clés : voix, familiarité, discrimination, timbre, électrophysiologie, N1, MMN, FTPV, P3a, analyse de sources, nouveau-nés.

ABSTRACT

Voice is a very prominent auditory stimulus in our acoustic environment. It is not only the carrier of speech, but would also be an auditory face that conveys important affective and identity information. Our ability to discriminate and recognize voices is socially and biologically important as it is amongst the most important functions of the human auditory system. This thesis was interested in the ontogenesis and specificity of the cortical response to human voice and had three objectives: (1) to develop an electrophysiological protocol to objectively measure the processing of voice familiarity in adult subjects; (2) to assess whether the same electrophysiological protocol could also objectify preferential processing of a familiar voice in 24-hour-old newborns, in particular the mother's voice; and (3) to test the robustness of an electrophysiological measure, more specifically the Fronto-Temporal Positivity to Voices, interested in pre-attentional discrimination between vocal and non-vocal stimuli. Results from these three experimental designs have enabled (1) to identify electrophysiological components (Mismatch Negativity and P3a) sensitive to the processing of voice familiarity; (2) to highlight a singular pattern of cortical activation to the mother's voice in newborns, providing the first neurophysiological evidence of language acquisition, a process especially related to the mother-child interaction; and (3) to confirm that vocal/non-vocal discrimination is a pre-attentional process, while enhancing the selectivity and the specificity of voice processing cortical response.

Keywords : voice, familiarity, discrimination, timbre, electrophysiology, N1, MMN, FTPV, P3a, source analysis, newborns.

TABLE DES MATIÈRES

LISTE DES TABLEAUX

Article 2 : Mother and Stranger: An Electrophysiological Study of Voice Processing in Newborns

LISTE DES FIGURES

Article 1 : Electrophysiological Markers of Voice Familiarity

Article 2 : Mother and Stranger: An Electrophysiological Study of Voice Processing in Newborns

Article 3 : The Adult Brain Rapidly Distinguishes Between Vocal and Non-Vocal Stimuli

LISTE DES SIGLES ET DES ABRÉVIATIONS

ERP	*Event-related potentials*
FTPV	*Fronto-Temporal Positivity to Voices*
ms	milliseconde
STS	Sillon temporal supérieur
VSR	*Voice Specific Response*

Un défi de relevé parmi tant d'autres!

REMERCIEMENTS

Je tiens d'abords à remercier ma directrice, Maryse Lassonde, sans qui cette thèse aurait été tout autre. Je dois en quelque sorte mon doctorat à un concours de circonstances : j'ai eu la chance que Maryse me choisisse comme étudiante au baccalauréat, qu'elle accepte que je poursuive mes études graduées sous sa supervision et qu'elle me fasse ultimement suffisamment confiance pour mettre entre mes mains ce qui est devenu ma thèse doctorale, ces trois beaux projets. Elle a aussi su m'entourer d'une équipe dont chacun des membres a contribué de près ou de loin à la concrétisation de mes projets. C'est avec leur aide inestimable qu'ont vu le jour ces travaux de recherche.

Un gros MERCI aussi à tous ceux et celles avec qui j'ai partagé mon quotidien des dernières six (ouch!) années. Vous avez su rendre ce long, et parfois pénible, parcours doctoral des plus agréables. Je n'échangerais ces années passées en votre compagnie pour rien au monde et souhaite de tout cœur continuer à vous côtoyer. Un doctorat ne sert pas uniquement à en apprendre sur la science, ça forge également de belles amitiés. Un merci plus particulier à Isabelle dont le courage m'épate à tout égard, à Mélissa et Jean-François dont la détermination est exemplaire, à Madeleine pour son humour, à Julie pour son écoute, à Francis qui m'a donné l'exemple d'aller au bout de ses rêves, à Sylvain que je connais depuis si longtemps et à Magalie et Lyssa pour le sprint final. À Stéphane et Maria, vous êtes irremplaçables et c'est pour cette raison que vous nous survivez tous. Merci également à tous les professeurs et chercheurs qui m'ont entourée et ont su m'inspirer à différents moments de ce périple (j'ai même appris ce qu'était une bande passante!).

Merci à ma famille, à mes parents et à ma petite sœur aussi qui ont, sans contredit, énormément contribué à la personne que je suis devenue. Ils ont toujours été dernière moi, admirant les petits et grands pas que j'ai faits et que je continue de faire. Ils sont source d'inspiration, me poussant toujours à aller plus loin. Un dernier merci, mais non le moindre, à Louis et Léonie. Louis qui a su démystifier (et parfois mystifier!) l'ampleur d'une thèse doctorale, m'épaulant à chacune des étapes. Merci de m'encourager dans tout ce j'entreprends et désire entreprendre. Léonie qui, quant à elle, me ramène à ce qui est vraiment important. Merci pour ton amour pur et débordant, tu es un vrai rayon de soleil qui a illuminé mes journées les plus ennuagées. Tu es certainement (et de loin) mon projet le plus ambitieux, ma plus grande réussite doctorale. Je suis fière d'être ta maman et je t'aime.

INTRODUCTION

Au cours des dernières décennies, il y a eu un engouement des études en neurosciences cognitives afin d'identifier les aires et systèmes cérébraux qui sont dédiés au traitement de certaines catégories de stimuli perceptifs. À titre d'exemple, maintes études ont proposé que la perception des visages est distincte et ségréguée de la perception visuelle d'autres objets et de leur identification (Allison, Puce, Spencer, & McCarthy, 1999; Kanwisher, McDermott, & Chun, 1997). En accord avec l'hypothèse selon laquelle des principes communs d'organisation fonctionnelle devraient exister entre les modalités sensorielles, une telle spécialisation devrait aussi exister en perception auditive de stimuli d'origine humaine (Belin & Zatorre, 2000). La voix humaine serait de fait un bon candidat pour ce type de traitement spécialisé, entre autres en raison de sa prédominance dans notre environnement sonore, de son rôle général au sein des interactions humaines, et du fait qu'elle permet le langage.

L'être humain est un être social remarquable, mais les mécanismes neuronaux sous-jacents qui permettent cette socialisation demeurent encore largement méconnus. Notre capacité d'analyser et de catégoriser l'information contenue dans la voix joue un rôle clé dans les interactions sociales humaines (Belin, Fecteau, & Bedard, 2004). En effet, la voix permet non seulement la parole, mais elle serait aussi selon certains auteurs l'équivalent d'un visage auditif (Belin, et al., 2004), transmettant notamment des informations identitaires et affectives importantes. Ainsi, mise à part la capacité à analyser et décortiquer la parole, le système auditif effectue aussi plusieurs opérations complexes sur les sons vocaux, nous dotant de capacités allant de la simple détection d'une voix humaine dans un environnement sonore bruyant à l'extraction de certaines informations sur des locuteurs inconnus ou encore nous permettant de reconnaître des voix familières. Or, le son émis par

la voix humaine devient significatif en soi, indépendamment de sa valence phonétique. En fait, la capacité à traiter avec aisance les caractéristiques para-phonétiques de la voix est cruciale pour une identification rapide du locuteur (Van Dommelen, 1990). De plus, le timbre de la voix fournit d'importants indices sur le genre, l'âge, le statut et l'état affectif de l'orateur (Fellowes, Remez, & Rubin, 1997; Linville, 1996; Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995; Remez, Fellowes, & Rubin, 1997; Scherer, 1995). L'humain, comme plusieurs autres espèces d'ailleurs, est particulièrement doué pour extraire les informations paralinguistiques de la voix (Belin, et al., 2004). Par contre, pour ce faire, nous devons d'abord et d'emblée établir si une structure acoustique harmonique entendue a été produite ou non par des cordes vocales, à savoir s'il s'agit réellement d'un stimulus vocal, d'une voix. Une question se pose donc : cette discrimination est-elle faite par un système particulièrement sensible et spécifique à la voix humaine, ou ne résulterait-elle pas d'un traitement fait par le système acoustique plus général?

Plusieurs études se sont penchées sur la spécificité neuronale dédiée au traitement de la voix. Ainsi, les données émanant d'études cliniques de même que divers résultats expérimentaux ont permis de statuer que la voix est réellement « spéciale » pour le cerveau humain. La première section de cette thèse vise donc à parcourir l'amalgame de ces évidences pour ensuite aboutir à trois études expérimentales qui en constituent le cœur et qui supportent le fait que la voix est spéciale.

LA PHYLOGÉNÈSE DU TRAITEMENT DE LA VOIX

Comprendre comment le cerveau traite les sons émis pour communiquer est important puisque l'environnement auditif des primates (humains ou non) est bombardé de vocalisations

provenant des pairs. Ces vocalisations, émanant d'un répertoire vocal riche, sont utilisées dans des contextes variés, tel des interactions agonistes ou affiliatives entre les membres d'un même groupe social, des cris d'alarme ou des cris territoriaux (Belin, 2006; Winter, Ploog, & Latta, 1966). Chaque individu est par conséquent exposé quotidiennement à un nombre important de vocalisations. Notamment, chez l'humain et dans sa société moderne, les voix sont partout : provenant d'individus physiquement présents ou encore de sources plutôt virtuelles telles la radio, la télévision, etc. La perception des vocalisations constitue donc un aspect important du comportement auditif de nombreuses espèces animales et demeure cruciale pour leurs interactions sociales, leur succès de reproduction et de survie (Wang, 2000).

La parole, portée par la voix, n'est pas fondamentalement différente, au sens acoustique, des vocalisations animales, quoique moins stéréotypée (Suga, 1992). Selon Wang (2000), il existe trois facteurs qui peuvent biaiser les réponses corticales en réponse à des vocalisations chez une espèce particulière. Ces facteurs sont : la prédisposition de l'évolution, la plasticité reliée au développement et la plasticité dépendant de l'expérience.

Dans le courant des années 70, l'hypothèse qui prévalait était que les vocalisations spécifiques aux espèces étaient représentées au niveau cortical par l'activité de neurones spécifiques appelés des « détecteurs d'appel ». Certains auteurs ont en effet constaté qu'un petit nombre de cellules répondaient uniquement aux vocalisations (Winter & Funkenstein, 1973). Ces mêmes auteurs ont également démontré, en utilisant un ensemble plus important de vocalisations, que plus de la moitié des cellules qui s'étaient montrées sensibles aux vocalisations étaient, en plus,

en quelque sorte sélectives dans leur réponse, ne répondant pas à plus de deux vocalisations acoustiquement similaires. Cependant, il a, par la suite, été démontré que les neurones individuels du cortex auditif répondaient souvent à plus d'un type de vocalisations ou à des vocalisations ayant différentes fonctions (Newman & Wollberg, 1973a, 1973b). L'hypothèse des neurones « détecteurs d'appel » a, par conséquent, ensuite été remplacée par l'idée que les espèces codent les sons complexes selon les patrons de décharge des populations neuronales (Creutzfeldt, Hellweg, & Schreiner, 1980).

La perception et le traitement de la parole sont peut-être uniques à l'Homme, mais d'autres capacités de perception de la voix existent chez d'autres espèces. Les primates semblent entre autres avoir une sensibilité accrue à des fréquences correspondant à la gamme inclue dans les vocalisations spécifiques de leur espèce (Aitkin, Merzenich, Irvine, Clarey, & Nelson, 1986; Wang, 2000), à savoir que les singes ont une meilleure sensibilité (en termes de seuils auditifs absolus plus petits) avec les hautes fréquences comparativement aux basses, contrairement aux humains (Owren, Hopp, Sinnott, & Petersen, 1988). Par ailleurs, les macaques arrivent à identifier les individus de leur espèce en fonction des vocalisations (Rendall, Owren, & Rodman, 1998; Rendall, Rodman, & Emond, 1996). Des études neurophysiologiques dans le cortex auditif primaire des primates non-humains démontrent aussi que des vocalisations spécifiques à l'espèce sont des stimuli très efficaces pour induire une forte réponse neuronale (Poremba et al., 2004; Wang & Kadia, 2001).

Il semble raisonnable de supposer que les humains ont probablement été soumis à une pression similaire à celle de l'évolution des primates non-humains pour développer des mécanismes spécialisés dans l'extraction exacte d'informations vocales (Belin, 2006). D'un point de vue évolutif, l'importance primordiale de la parole dans les interactions sociales de l'Homme rend le raffinement des mécanismes corticaux dédiés au traitement de la voix très adaptatif.

LA PHONAGNOSIE : UN TROUBLE DU TRAITEMENT DE LA VOIX

La phonagnosie fait référence à une altération de la reconnaissance d'une voix et/ou de la discrimination entre deux voix (Van Lancker, Cummings, Kreiman, & Dobkin, 1988). Bien que ces deux troubles soient dénommés phonagnosie, nous savons qu'il s'agit de deux fonctions différentes et indépendantes et qu'elles peuvent être sélectivement atteintes (Van Lancker & Kreiman, 1987), en plus d'être dissociables des troubles de la parole et du langage (Assal, Aubert, & Buttet, 1981). En effet, en utilisant des voix familières et inconnues à titre de stimuli, ces auteurs ont pu identifier une différence marquée entre la capacité à reconnaître une voix familière et la capacité à distinguer deux voix inconnues. Les résultats cliniques et radiologiques suggèrent que la reconnaissance d'une voix familière est altérée suite à une lésion des régions inférieures et latérales du lobe pariétal droit (Van Lancker, et al., 1988) tandis que l'altération de la capacité à discriminer entre deux voix survient suite à une lésion temporale de l'un ou l'autre hémisphère (Van Lancker, et al., 1988; Van Lancker, Kreiman, & Cummings, 1989). Cette double dissociation entre la reconnaissance et la discrimination de la voix humaine suggère que ces deux fonctions sont médiées par différentes structures corticales qui contribuent de façon différentielle au syndrome clinique.

D'ailleurs, une étude clinique subséquente utilisant la tomodensitométrie (CT-scan) a montré qu'un lobe pariétal droit intact était nécessairement présent dans tous les cas de reconnaissance vocale normale, alors qu'un lobe pariétal droit endommagé était significativement corrélé avec un déficit de la reconnaissance de la voix (Van Lancker, et al., 1989). Puisque la reconnaissance vocale est aussi observée chez les patients aphasiques (y compris les graves aphasiques globaux ne présentant pratiquement aucun langage fonctionnel expressif ou réceptif), cette dernière est une capacité dissociable des fonctions langagières de l'hémisphère gauche (Van Lancker & Canter, 1982). De plus, il a été suggéré que la reconnaissance de la voix est également dissociable de la reconnaissance des sons de l'environnement (Van Lancker & Kreiman, 1987), suggérant que la voix serait traitée par un système indépendant et analogue à celui impliqué dans le traitement des visages (Bruce & Young, 1986). D'ailleurs, Belin et collaborateurs (2004) ont utilisé le modèle de perception des visages de Bruce et Young (1986) comme cadre théorique et l'ont adapté pour comprendre les processus perceptifs et cognitifs impliqués dans la perception des voix.

LES ÉTUDES DE LOCALISATION DES AIRES ASSOCIÉES AU TRAITEMENT DE LA VOIX

Pendant longtemps, l'étude de patients cérébrolésés était la seule façon d'investiguer le traitement cortical de la voix. Tel que mentionné ci-dessus, les lésions cérébrales chez l'humain nous ont suggéré l'existence de mécanismes spécifiques au traitement des voix dans le lobe temporal. Les études de neuroimagerie fonctionnelle ont depuis supplémenté les données cliniques antérieures sur le traitement neuronal de la voix. On reconnaît à ce jour que des régions corticales sont dédiées au traitement de stimuli vocaux. Belin et collaborateurs (2000) ont été parmi les premiers à démontrer une activation bilatérale accrue des sillons temporaux supérieurs (STS), mais particulièrement au

sein de l'hémisphère droit, en réponse aux stimuli vocaux (i.e. vocalisations humaines incluant de la parole ou non) comparativement à des stimuli non-vocaux (tels des sons environnementaux, des sons humains non-vocaux, du bruit blanc modulé) ou des stimuli vocaux embrouillés. Il est aussi reconnu que cette activation préférentielle des STS n'est pas entièrement due à la présence de parole dans les stimuli vocaux (Belin, Zatorre, & Ahad, 2002; Belin, et al., 2000). De fait, le STS droit, notamment dans sa région antérieure, ne nécessite aucun contenu linguistique pour s'activer en réponse à des sons vocaux (Belin, et al., 2002), devenant alors une région potentiellement candidate pour le traitement d'autres aspects paralinguistiques de la voix. D'ailleurs, cette même région, le STS droit antérieur, s'est également montrée sensible à l'identité vocale (Belin & Zatorre, 2003; Kriegstein & Giraud, 2004; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003), permettant entre autres la discrimination entre divers locuteurs et l'identification d'une voix familière (Imaizumi et al., 1997; Nakamura et al., 2001). Tout comme l'avait démontré une étude animale (Poremba, et al., 2004), il se trouve que la réponse préférentielle du STS droit antérieur aux sons vocaux serait spécifique aux vocalisations de l'espèce, dans ce cas-ci la voix humaine comparativement à des vocalisations animales (Fecteau, Armony, Joanette, & Belin, 2004). Ces données (Belin & Zatorre, 2003; Belin, et al., 2000; Fecteau, et al., 2004) sont aussi compatibles avec les études de neuroimagerie soutenant l'existence d'un traitement antérieur permettant la reconnaissance des sons (Alain, Arnott, Hevenor, Graham, & Grady, 2001; Belin & Zatorre, 2000; Kaas & Hackett, 1999; Maeder et al., 2001; Rauschecker & Tian, 2000; Romanski et al., 1999; Romanski et al., 2000), reflétant dans le cas présent des processus associés avec l'identification de la source vocale ou en d'autres mots avec l'identification du locuteur (Belin, et al., 2004; Belin & Zatorre, 2003). La série d'études menées par

ces chercheurs nous a fourni des preuves solides quant à l'existence de régions corticales non seulement sensibles, mais aussi sélectives à la voix humaine.

La voix comme source sonore

Les voix sont des sources sonores et le traitement cortical de la voix implique donc nécessairement l'analyse des différentes propriétés que possèdent une source sonore, telle la hauteur, l'intensité, la localisation spatiale (Warren, Scott, Price, & Griffiths, 2006). Plusieurs paramètres acoustiques sont couramment utilisés pour décrire les informations contenues dans la voix. Selon plusieurs auteurs, deux de ces paramètres sont particulièrement pertinents au plan perceptif : (1) la fréquence fondamentale (ou f0, communément appelée la tonalité) et (2) la fréquence des divers formants du spectre (Lavner, Gath, & Rosenhouse, 2000; Van Dommelen, 1990). La fréquence fondamentale (f0), résultant de la vibration des cordes vocales dans le larynx, détermine la hauteur perçue d'une voix entendue; le plus souvent, elle est plus élevée chez les femmes que chez les hommes (Titze, 1989). De plus, le conduit vocal, qui peut être considéré comme un filtre mobile complexe, permet d'amplifier certaines fréquences émises par le larynx et d'en atténuer d'autres. Les fréquences amplifiées, aussi appelées formants, sont caractéristiques de l'individu puisqu'elles dépendent à la fois de la taille du conduit vocal – fortement corrélée avec la taille du corps de l'individu, contrairement à f0 (Fitch, 1997)– et de sa forme, déterminée par la configuration des articulateurs (Klatt & Klatt, 1990). Les formants établissent le timbre distinctif de la voix de chaque individu, de sorte qu'un auditeur peut identifier des voix familières et discriminer ou catégoriser des voix inconnues. Les troisième et quatrième formants seraient principalement responsables de la perception du genre du locuteur puisqu'ils dépendent de la forme de la cavité du

pharynx, qui est généralement plus importante chez les hommes (Lavner, et al., 2000). Par conséquent, la fréquence fondamentale tout comme les formants nous transmettent de l'information importante quant à l'identité du locuteur. D'ailleurs, nous pouvons reconnaître des locuteurs qui nous sont familiers à partir de stimuli à ondes sinusoïdales analogues à leurs vocalisations (Remez, et al., 1997).

Une distinction fonctionnelle existe au sein du cortex temporal supérieur droit pour le traitement des différents paramètres de la voix (Lattner, Meyer, & Friederici, 2005). La tonalité de la voix serait traitée dans les régions antérieures et entourant le gyrus de Heschl. L'analyse d'une voix nécessite aussi l'encodage de l'identité vocale comme source sonore particulière (Warren, et al., 2006). Ainsi, le timbre, défini opérationnellement par la propriété permettant de distinguer deux sons de hauteur, durée et intensité identiques (ANSI, 1973), est un aspect critique permettant l'identification de voix et d'autres sources sonores (McAdams & Cunible, 1992). L'analyse du timbre est donc nécessaire et les informations spectrales de la voix seraient traitées dans la partie postérieure du gyrus temporal supérieur de même que bilatéralement dans les régions avoisinant le planum temporal et le planum pariétal (Lattner, et al., 2005; Menon et al., 2002; Warren, Jennings, & Griffiths, 2005). La prototypicalité serait quant à elle principalement traitée dans la partie antérieure du gyrus temporal supérieur droit (Lattner, et al., 2005).

Le traitement neuronal de la voix se ferait selon un schéma hiérarchique (Warren, et al., 2006) qui impliquerait trois étapes d'analyse : (1) la détection du changement de la source vocale; (2) l'extraction des caractéristiques vocales; et (3) l'analyse détaillée de l'identité vocale. Des données

récentes en neuroimagerie appuient le fait que des mécanismes cérébraux distincts seraient dédiés aux différentes étapes du traitement de la voix (Warren, et al., 2006), à savoir que la détection d'un changement dans la source sonore vocale se fait par des mécanismes généraux d'analyse de la source sonore, situés dans la partie postérieure du lobe temporal supérieur (planum temporal postérieur droit s'étendant inférieurement jusqu'à la partie postérieure du STS), et que l'analyse de l'information spectrale présente dans la voix active davantage les régions plus antérieures (STS médian et antérieur bilatéraux). Bien que ces régions soient impliquées et se soient révélées sensibles à l'analyse de différents paramètres acoustiques pertinents dans le traitement de la voix, ces auteurs ne prétendent pas qu'elles sont limitées au traitement de la voix ou en d'autres termes, sélectives à la voix.

L'expertise joue-t-elle un rôle?

Nous sommes tous experts à traiter les voix humaines, lesquelles sont une des premières formes d'interaction sociale dans l'ontogénèse en plus d'être la catégorie sonore à laquelle nous sommes le plus exposés au quotidien. Il a été suggéré que l'activation préférentielle du STS en réponse à la voix reflèterait plutôt une expertise dans le traitement de la catégorisation sonore à un niveau subordonné et ne serait pas nécessairement spécifique à la voix humaine. Advenant cette possibilité, une plus grande activation du STS devrait être aussi observée chez des personnes ayant notamment une autre expertise auditive, ce que plusieurs auteurs ont effectivement démontré (Chartrand & Belin, 2006; Chartrand, Filion-Bilodeau, & Belin, 2007; Chartrand, Peretz, & Belin, 2008; Fecteau, et al., 2004). Il existe donc un corrélat neuronal de l'expertise auditive redéfinissant le rôle des régions antérieures du STS qui sont considérées comme étant très sélectives à la voix

(Belin, et al., 2002; Belin, et al., 2000). De fait, il semble que ces régions ne traitent pas exclusivement les stimuli vocaux, mais contribuent également au traitement expert d'autres catégories sonores.

LES ÉTUDES DU DÉCOURS TEMPOREL DU TRAITEMENT DE LA VOIX

Bien que la résolution spatiale des études de neuroimagerie est séduisante et nous informe des régions corticales impliquées dans le traitement de divers stimuli, notamment la voix, cette technique ne fournit pas d'informations précises quant au déroulement temporel du traitement cortical des stimuli. Il est donc important de complémenter les connaissances acquises en localisation spatiale avec des mesures permettant une meilleure résolution temporelle, tels que les potentiels évoqués (*event-related potentials*, ERP) qui se sont montrés utiles pour mettre en évidence des changements dans l'activité corticale notamment lorsque de l'information auditive est présentée. Très peu d'études électrophysiologiques se sont encore penchées sur le traitement de la voix.

Certains auteurs ont tenté d'élucider le statut particulier de la voix dans l'activité corticale auditive. Levy et collaborateurs (2001) ont cherché à caractériser les ERP suscités par des stimuli vocaux non-phonétiques. Afin de contrôler pour la variété de facteurs pouvant intervenir et être responsables des différentes réponses électrophysiologiques, les auteurs ont comparé les stimuli vocaux avec des sons d'instruments de musique (provenant de quatre catégories distinctes) de fréquence fondamentale correspondante. Ils ont identifié une composante positive qui semblait dévoiler un traitement différentiel pré-phonologique de la voix humaine. Cette composante ERP qu'ils

ont dénommé *Voice Specific Response* (VSR) culmine à 320 ms après la présentation du stimulus et est éminente en réponse aux voix humaines, alors que l'amplitude et la latence de cette même composante en réponse à différents instruments de musique ne se distinguent pas entre elles. Toutefois, la polarité de la composante VSR, de même que sa latence et sa distribution topographique plutôt frontale sont similaires à ceux de la composante P3a induite par des distracteurs dans un paradigme *oddball*. En effet, la P3a est généralement considérée comme une réponse d'orientation à des stimuli qui exige une allocation de l'attention même si les stimuli ne sont pas pertinents à la tâche (Friedman, Cycowicz, & Gaeta, 2001). Elle reflète donc une réponse corticale non intentionnelle à des stimuli nouveaux. La P3a peut être évoquée lorsqu'un stimulus est (1) exceptionnel (Grillon, Courchesne, Ameli, Elmasian, & Braff, 1990); (2) rare par opposition aux autres distracteurs (Katayama & Polich, 1996); ou (3) facile à distinguer des autres distracteurs (Comerchero & Polich, 1999). Pour en revenir à la composante VSR identifiée par Levy et collaborateurs (2001), elle ne peut s'expliquer par sa « réelle » rareté puisque tous les distracteurs étaient équiprobaux. Toutefois, les auteurs sont conscients et soulèvent l'argument qu'il se peut que tous les instruments de musique aient été regroupés sémantiquement et subjectivement dans une catégorie conceptuelle par rapport aux voix humaines et que la composante VSR reflète la rareté de la catégorie « voix ». Le cas échéant, on pourrait dire que l'effet mesuré par la VSR est parallèle à celui d'une P3a. Dans la même étude, Levy et collaborateurs (2001) ont donc mené une seconde expérience dans laquelle il n'y avait que deux groupes de distracteurs, soit des voix et un seul type d'instruments, les cuivres. La réplication de la différence entre les ERP suscités par des voix humaines et les cuivres, présentés selon la même probabilité, exclut, selon les auteurs, la possibilité que les voix humaines aient suscité une plus ample VSR en raison de la rareté de leur catégorie.

Toutefois, il ne peut être complètement exclu que la composante VSR mesurée soit effectivement une réponse d'orientation, semblable à la P3a. En effet, étant donné la saillance écologique de la voix, il se peut que celle-ci soit toujours considérée comme étant catégoriquement différente, suscitant une réponse d'orientation de l'attention en raison du caractère exceptionnel que l'on accorde à la voix humaine, et ce indépendamment de sa rareté, ou d'une différence dans son intensité, sa fréquence fondamentale, ou son enveloppe spectrale.

On peut aussi supposer qu'avec l'apparition tardive de la VSR, il est peu probable qu'elle reflète des processus perceptifs de base. Elle pourrait vraisemblablement être un membre d'une famille de composantes, incluant la P3a, étant toutes des manifestations différentes d'un mécanisme cognitif général d'attention (Escera, Alho, Schroger, & Winkler, 2000; Escera, Alho, Winkler, & Naatanen, 1998; Friedman, et al., 2001). D'ailleurs, Levy et collaborateurs (2003) ont par la suite démontré les effets modulateurs de l'attention et du niveau de traitement sur la composante VSR. Ainsi, le fait de ne pas porter attention ou de porter attention sur une caractéristique autre que le timbre n'engendre pas de VSR. De plus, l'évocation d'une composante positive en réponse à des stimuli instrumentaux montre bien que sous certaines conditions, la VSR peut tout aussi être évoquée par d'autres stimuli ayant une structure harmonique complexe, outre les voix (Levy, et al., 2003). D'ailleurs, une étude magnétoencéphalographique ayant utilisé le même protocole et les mêmes stimuli (Gunji et al., 2003) n'a point été en mesure d'identifier la contrepartie magnétique de la VSR électrique.

Quoi qu'il en soit, ces résultats sont importants puisqu'ils tentent d'identifier une contrepartie électrophysiologique à l'activation du STS en imagerie. Il serait tentant de suggérer que les générateurs de cette positivité tardive puissent être situés le long des STS antérieurs bilatéraux, pourtant, 320 ms est une durée considérable pour voir apparaître une réponse différentielle à une catégorie de sons aussi biologiquement importante que sont les voix. D'ailleurs, on reconnaît une réponse électrophysiologique beaucoup plus précoce en réponse aux visages (N170) (Bentin, Allison, Puce, Perez, & McCarthy, 1996) et comme exprimé plus haut, en accord avec l'hypothèse selon laquelle des principes communs d'organisation fonctionnelle devraient exister entre les différentes modalités sensorielles, une telle réponse putative précoce devrait aussi existée en perception auditive de stimuli d'origine humaine.

Holeckova et collaborateurs (2006), dans un but différent et clinique chez une population comateuse, se sont intéressés à : 1) comparer les réponses corticales de sujets à leur propre nom, auquel ils ne portent pas attention, à celles induites par des stimuli non-vocaux ayant des caractéristiques physiques comparables; et 2) comparer les réponses corticales de sujets à leur propre nom en fonction de la familiarité de l'orateur. Les stimuli vocaux et non vocaux se sont avérés être traités différemment dès les premières étapes de l'analyse sensorielle (composante N1). Les topographies démontraient de plus grandes amplitudes au sein de l'hémisphère gauche qu'au sein de son homologue droit pour les deux stimuli vocaux (voix familière et non-familière) tandis que l'activité semblait davantage bilatéralisée en réponse aux stimuli non-vocaux, ce qui est peu surprenant compte tenu de la nature verbale des stimuli vocaux (Binder et al., 2000). De plus, les voix familière et étrangère ont aussi suscité une plus grande réponse d'orientation que les stimuli

non-vocaux. Les réponses corticales aux deux voix ne diffèrent pas entre elles jusqu'à 300 ms après la présentation du stimulus. C'est plutôt au cours de la phase tardive de la P3a que les auteurs ont remarqué une différence : à savoir que les topographies demeurent semblables pour les deux stimuli vocaux, mais de plus grandes amplitudes sont notées en réponse à la voix familière. Par ailleurs, la plus grande différence entre la réponse aux voix familières et celle aux voix étrangères est apparue comme une activité soutenue pariétale au-delà de 450 ms suivant la présentation d'une voix familière, suggérant, selon les auteurs, un traitement cortical plus profond des voix familières.

Une étude électrophysiologique, plus récemment publiée, a aussi démontré de plus grandes amplitudes en réponse aux voix par rapport à d'autres catégories sonores aux régions fronto-temporales, émergeant aussi tôt que 164 ms, composante qu'ils ont dénommé « *Fronto-Temporal Positivity to Voices* » (FTPV) (Charest et al., 2009). Par conséquent, de l'activité neuronale dans une fenêtre temporelle entourant les 200 ms semble sensible aux différences vocale/non-vocale, en plus d'être sensible à des processus cognitifs de plus haut niveau, tel le traitement du genre (Zaske, Schweinberger, Kaufmann, & Kawahara, 2009) et l'identité vocale (Titova & Naatanen, 2001). Il est intéressant de noter la coïncidence temporelle entre la FTPV en réponse aux voix et la fameuse N170 en réponse aux visages. En effet, puisqu'en interaction sociale un même éventail d'informations, notamment l'identité et l'affect, est transmis simultanément par la voix et le visage, ces informations devraient s'intégrer de façon parcimonieuse (Campanella & Belin, 2007). De cette façon, Charest et collaborateurs (2009) suggèrent que la FTPV serait l'analogue auditif de la bien connue N170. Par contre, est-ce que cette FTPV serait également dépendante de l'attention portée aux stimuli présentés, tout comme la VSR s'est avérée l'être? En effet, l'étude de Charest et

collaborateurs (2009) ne peut répondre à cette question de par le paradigme qu'ils ont utilisé qui nécessitait une réponse comportementale de la part des sujets. Il demeurait donc irrésolu si l'activité mesurée est dépendante de processus attentionnels. La question de savoir si la discrimination vocale/non-vocale est un processus cortical pré-attentionnel pose un défi intéressant : si les sujets ne portent pas attention à une séquence sonore, comment pouvons-nous savoir si la discrimination en jeu a réellement eu lieu? Une façon d'y arriver est de mesurer l'activité corticale en réponse à des stimuli sonores présentés dans le cadre d'un paradigme auditif *oddball* passif au sein duquel des stimuli rares sont intercalés dans une séquence de stimuli standards identiques et pendant lequel les sujets portent leur attention à une modalité sensorielle différente. Rogier et collaborateurs (2010) ont donc mené une étude auprès d'enfants qui étaient exposés à des séquences de stimuli auditifs : (1) des stimuli vocaux à travers lesquels étaient insérés des stimuli non-vocaux, et (2) des stimuli non-vocaux à travers lesquels étaient insérés des stimuli vocaux. Ces auteurs ont mis en évidence, en analysant les stimuli standards fréquents (vocaux et non vocaux), un effet de voix suggérant une spécialisation précoce et pré-attentionnelle des mécanismes cérébraux impliqués dans le traitement de la voix (Rogier, et al., 2010).

L'ONTOGÉNÈSE DU TRAITEMENT DE LA VOIX

Avec le développement des techniques non-invasives d'imagerie ou d'enregistrement cortical qui peuvent être utilisées en toute sécurité auprès des enfants et des nourrissons, les bases cérébrales sous-jacentes aux premières étapes de développement peuvent maintenant être étudiées de façon empirique, questions qui auparavant auraient été examinées par pure spéculation théorique.

La voix permet la parole et le langage. Les études in utero ou chez le nouveau-né se sont justement jusqu'à présent surtout intéressées au développement du langage. Avant même que l'enfant ne prononce son premier mot, il est bombardé de stimuli linguistiques. Ceux-ci sont traités de manière préférentielle (Ramus, Hauser, Miller, Morris, & Mehler, 2000; Vouloumanos & Werker, 2007). Dès la naissance, les nourrissons sont capables de reconnaître les propriétés de leur langue maternelle (Dehaene-Lambertz, Hertz-Pannier, & Dubois, 2006; Dehaene-Lambertz et al., 2006; P. K. Kuhl, 2004) et de discriminer de façon pré-attentionnelle, tout comme les adultes, des voyelles différentes (Alho et al., 1998; Cheour-Luhtanen et al., 1995). Une étude en imagerie fonctionnelle a démontré que la parole présentée dans la langue maternelle activait des régions très circonscrites, au niveau de l'hémisphère gauche périsylvien, similaires à celles trouvées chez l'adulte (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002). En effet, d'un point de vue fonctionnel, les quelques études d'imagerie cérébrale suggèrent une asymétrie fonctionnelle précoce, dès les premiers mois de vie. L'amplitude des potentiels évoqués à des stimuli auditifs est plus grande dans l'hémisphère gauche qu'au sein de son homologue droit chez le nourrisson de deux mois (Dehaene-Lambertz, 2000), et la réponse hémodynamique à la langue maternelle est nettement asymétrique dans la partie postérieure du lobe temporal supérieur du nouveau-né (Pena et al., 2003).

Les nouveau-nés ne peuvent parler ou comprendre la parole et pourtant ils sont en mesure de reconnaître la voix. DeCasper et Fifer (1980) ont été les pionniers à démontrer qu'en utilisant le conditionnement opérant, les nouveau-nés, dans les trois jours suivant leur naissance, pouvaient non seulement discriminer entre la voix de leur mère et celles d'étrangères, mais aussi apprendre à

modifier leurs patrons de tétées de façon à leur permettre d'entendre la voix de leur mère plus fréquemment. Toujours en effectuant des études comportementales, d'autres auteurs ont mis en évidence qu'à moins de deux heures de vie, ils étaient en mesure d'observer davantage de mouvements d'orientation lorsque la mère prononçait le nom de son poupon comparativement à lorsqu'une voix étrangère le faisait (Querleu et al., 1984). Des études mesurant les changements dans la fréquence cardiaque des nouveau-nés lors de la présentation de différentes voix ont aussi démontré qu'ils possèdent la capacité de discriminer entre plusieurs voix et la capacité à reconnaître la voix de leur mère et de leur père (Ockleford, Vince, Layton, & Reader, 1988). Cette capacité serait même présente in utero, chez le fœtus à terme (DeCasper, Lecanuet, Busnel, Granier-Deferre, & Maugeais, 1994; Fifer & Moon, 1994; Kisilevsky et al., 2003). D'ailleurs, il a été démontré dans une étude comportementale que, dans un environnement sonore gêné par de la parole distractrice, les très jeunes enfants écoutaient davantage et portaient plus attention lorsqu'ils entendaient la voix de leur mère (Barker & Newman, 2004). De plus, une autre étude, électrophysiologique cette fois-ci, comparant la réponse corticale à un mot prononcé par la voix de la mère versus une voix étrangère, a démontré que les nourrissons allouaient plus d'attention au traitement de la voix de leur mère (Purhonen, Kilpelainen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2004). D'ailleurs, le même groupe de chercheurs a également démontré qu'à quatre mois, le poupon aurait un modèle mnésique de la voix de sa mère (Purhonen, Kilpelainen-Lees, Valkonen-Korhonen, Karhu, & Lehtonen, 2005).

Certains auteurs ont de plus cherché à explorer les activations induites par la voix de la mère au-delà de celles induites par le traitement de la parole seule, permettant de faire un parallèle avec

les données chez l'adulte démontrant une latéralisation hémisphérique gauche du traitement linguistique classiquement opposée à un avantage de l'hémisphère droit pour l'identification et la discrimination des voix et de leur contenu émotionnel (Belin, et al., 2000; Ethofer et al., 2006; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Ils ont entre autres démontré que l'asymétrie fonctionnelle déjà répertoriée ne s'étend pas à tous les stimuli auditifs, et que le lobe temporal gauche est particulièrement sensible à la parole (exercée par la voix de la mère ou une voix étrangère), par rapport à la musique (Dehaene-Lambertz et al., 2010). Les auteurs argumentent que ces stimuli diffèrent de par leur familiarité de même que sur de nombreuses propriétés acoustiques, entre autres la vitesse des transitions temporelles, pouvant être un facteur essentiel. En effet, le discours à rebours est souvent utilisé comme stimuli contrôle pour la parole puisqu'il contient les mêmes transitions rapides que l'on retrouve dans la parole et certaines données tendent à démontrer que, chez les poupons de trois mois, aucune différence significative de latéralisation n'est observée dans les régions temporales lorsque la parole est comparée au discours renversé, les deux activant davantage l'hémisphère gauche (Dehaene-Lambertz, et al., 2002). De plus, des réponses corticales plus fortes ont été mesurées en réponse à la voix de la mère, entraînant, de façon impromptue, des différences importantes dans la partie postérieure du lobe temporal gauche, de même qu'au niveau du cortex préfrontal antérieur bilatéralement, pointant potentiellement vers un réseau linguistique et émotionnel pouvant jouer un rôle important dans l'apprentissage chez le nouveau-né (Dehaene-Lambertz, et al., 2010). En effet, l'activation postérieure gauche soulève la question de l'impact de la voix de la mère sur le traitement linguistique puisqu'il s'agit d'une région impliquée dans les représentations phonologiques chez l'adulte (Caplan, Gow, & Makris, 1995). Ces données suggèrent néanmoins que le traitement phonétique chez les nourrissons est sensible aux caractéristiques de la

voix du locuteur, pouvant être améliorée par l'écoute d'une voix très familière comme celle de la mère. L'ensemble de ces données tendent à expliquer pourquoi la clarté du discours maternel a une forte incidence sur les capacités de discrimination phonémique des jeunes enfants (Liu, Kuhl, & Tsao, 2003). Mais pourquoi la voix de la mère est-elle si spéciale?

Le système auditif du fœtus est morphologiquement mature environ 13 semaines avant la naissance et on serait à même d'enregistrer une certaine réactivité fœtale aux stimulations sonores au cours du dernier trimestre de gestation (Querleu, Renard, & Crepin, 1981a, 1981b). Par contre, les sons externes qui atteignent le fœtus sont grandement atténués, surtout en ce qui concernent les hautes fréquences (Ockleford, et al., 1988). Bien que plusieurs sons de l'environnement (voix du père ou de la fratrie, balayeuse, etc.) puissent être entendus par le fœtus quotidiennement, la voix de la mère est susceptible d'être la plus saillante en partie parce qu'elle est plus fréquemment entendue et aussi parce qu'elle est plus forte. De fait, la voix de la mère est transportée jusqu'au fœtus par conduction à travers le corps de même que par la voie aérienne. Il y a donc amplement d'occasions pour que les caractéristiques de la voix de la mère deviennent particulièrement familières pour le fœtus avant la naissance, même si celles-ci sont légèrement altérées. Par ailleurs, puisque la relation entre une mère et son enfant est la prémisse de base pour assurer des soins adéquats et la sécurité, cette relation a une importance particulière pour la survie biologique, non seulement d'un enfant en particulier, mais aussi au sens plus large, pour la survie de l'espèce humaine.

La voix, partie intégrante des interactions sociales, a un rôle adaptatif important très tôt dans le développement humain. Puisque la voix est perçue aussi tôt que pendant la vie fœtale et que les

études comportementales ont montré que les nouveau-nés arrivent très bien à discriminer la voix de leur mère de celles d'étrangères, ceci pointe vers une spécialisation précoce des régions du cerveau impliquées dans le traitement des sons vocaux, bien que ces zones cérébrales demeurent non identifiées chez les enfants. Par contre, des chercheurs se sont intéressés à étudier les corrélats électrophysiologiques du traitement de la voix chez les enfants d'âge pré-scolaire en comparant les potentiels évoqués auditifs en réponse aux sons vocaux à ceux en réponse aux sons de l'environnement (Rogier, et al., 2010). Ils ont mesuré une réponse temporelle aux stimuli vocaux morphologiquement très différente de la réponse aux stimuli non vocaux et les auteurs établissent un lien avec la composante FTPV identifiée chez l'adulte, suggérant que l'activation de processus cérébraux distincts serait responsable du traitement des deux types de stimuli. L'effet vocal mis en évidence par ces auteurs dès les 60 ms aux sites fronto-temporaux droits indiqueraient une spécialisation précoce (pré-attentionnelles) des mécanismes cérébraux impliqués dans le traitement de la voix.

OBJECTIFS ET HYPOTHÈSES DE RECHERCHE

La présente thèse s'intéresse à l'ontogénèse et à la spécificité de la réponse corticale à la voix humaine. Le premier objectif était donc de mettre sur pied un protocole électrophysiologique permettant de mesurer objectivement le traitement de la familiarité de la voix chez le sujet adulte. Nous souhaitions ensuite utiliser ce même protocole afin de déterminer si une telle mesure pouvait aussi objectiver chez le nouveau-né de 24 heures un traitement préférentiel d'une voix familière, notamment la voix de la mère. Finalement, nous avons voulu mettre à l'épreuve la robustesse d'une

mesure électrophysiologique, notamment la FTPV, s'intéressant à la discrimination pré-attentionnelle entre des stimuli vocaux et non-vocaux.

Premier article : Marqueurs électrophysiologiques de la familiarité d'une voix chez le sujet adulte

Admettant que notre capacité à discriminer et reconnaître des voix est socialement et biologiquement importante et qu'elle figure parmi les fonctions les plus importantes du système auditif humain, cette première étude visait à déterminer si des marqueurs électrophysiologiques pouvaient être utilisés comme mesures objectives du traitement de la familiarité d'une voix chez le sujet adulte. Nous avons par conséquent mesuré deux composantes électrophysiologiques (la MMN et la P3a) en réponse à une voix familière et à une voix inconnue, toutes deux présentées en tant que stimuli rares dans le cadre d'un paradigme *oddball*.

Hypothèses : Nous supposions que les composantes électrophysiologiques en réponse à une voix familière seraient de plus grandes amplitudes lorsque comparées aux composantes électrophysiologiques en réponse à une voix inconnue.

Deuxième article : Étude électrophysiologique du traitement de la familiarité de la voix chez le nouveau-né

Sachant que des régions corticales semblent dédiées au traitement de la voix chez l'adulte, que des mesures électrophysiologiques se sont montrées sensibles au traitement préférentiel de la familiarité de la voix et que nous connaissons très peu sur la façon dont le cerveau du nouveau-né

traite les informations vocales, nous avons voulu, par le biais de cette deuxième étude, étudier, au moyen de l'électrophysiologie et des analyses de sources, le traitement cortical dédié à la voix de la mère par rapport à celui réservé au traitement d'une voix étrangère chez le nouveau-né de 24 heures.

Hypothèses : Cette étude se voulait davantage exploratoire. Par contre, étant donné les données comportementales soutenant que le nouveau-né et même le fœtus savent reconnaître et réagissent préférentiellement à la voix de leur mère, nous avions tout de même émis l'hypothèse selon laquelle les composantes électrophysiologiques mesurées (MMN et P3a) seraient, tout comme chez l'adulte, de plus grandes amplitudes en réponse à une voix familière, dans le cas présent, à la voix de la mère lorsque comparées à celles en réponse à une voix inconnue. En ce qui a trait aux analyses de sources, nous postulions que les générateurs de l'activité mesurée en électrophysiologie se situeraient probablement au sein des régions dédiées au traitement de la voix chez l'adulte, notamment le cortex temporal droit.

Troisième étude : Le cerveau adulte discrimine-t-il rapidement les stimuli vocaux des stimuli non-vocaux?

Les études de neuroimagerie ont fourni des preuves pour le traitement spécifique et localisé de la voix, mais le décours temporel de ce traitement demeure mal compris. L'utilisation de l'électrophysiologie a su démontrer de l'activité neuronale, dans une fenêtre temporelle entourant les 200 ms, sensible aux différences vocale/non-vocale. Par ailleurs, il semble aussi que le traitement cortical de la voix serait un processus pré-attentionnel. Toutefois, dans ce dernier contexte, la

robustesse de la réponse électrophysiologique mesurée demeure précaire, car jusqu'à présent, seule une recherche s'est intéressée à étudier le traitement cortical de la voix à l'aide d'un paradigme soulignant son aspect pré-attentionnel. Or, cette étude, d'un point de vue méthodologique, analysait les sons standards fréquents. En utilisant le paradigme électrophysiologique élaboré et utilisé au sein des deux études antérieures, cette troisième étude expérimentale cherchait donc, d'une part, à consolider le fait que la discrimination entre des stimuli vocaux et non-vocaux se fait de façon pré-attentionnelle et, d'autre part, à vérifier si une réponse électrophysiologique ne pourrait pas être enregistrée en analysant seulement les stimuli rares vocaux et non vocaux présentés équiprobablement dans un paradigme *oddball* passif ayant comme standards une séquence de tons purs identiques. De cette façon, nous mettions l'accent sur la sélectivité et la sensibilité de la réponse corticale réservée au traitement de la voix en obtenant un corrélat de traitement cortical différentiel à un nombre plus limité de stimuli sans nécessiter que les sujets ne répondent aux stimuli.

Hypothèses : Les réponses neuronales suscitées par les sons déviants rares vocaux demeureraient de plus grandes amplitudes lorsque comparées à celles en réponse à des sons déviants rares non vocaux étant donné la sélectivité et la spécificité de la voix au niveau cortical.

ARTICLES DE LA THÈSE

ARTICLE 1 :

ELECTROPHYSIOLOGICAL MARKERS OF VOICE FAMILIARITY

Publié dans European Journal of Neuroscience (2006), 23 (11), 3081-3086.

Receiving Editor: Matthew Rushworth

Running Head: ERPs and voice familiarity

Electrophysiological markers of voice familiarity

Keywords: event-related potentials, voice familiarity discrimination, mismatch negativity, P3a, long-term memory

Maude Beauchemin[1,2], Louis De Beaumont[2], Phetsamone Vannasing[1], Aline Turcotte[1,2], Claudine Arcand[1,2], Pascal Belin[1,3] and Maryse Lassonde[1,2]

[1]Centre de Recherche, Hôpital Ste-Justine

[2]Centre de Recherche en Neuropsychologie et Cognition (CERNEC), Université de Montréal

[3]Department of Psychology, University of Glasgow

Correspondence:
Dr Maryse Lassonde
Canada Research Chair in Developmental neuropsychology
Universite de Montreal
CP 6128, Succ. Centre-Ville
Montreal Qc
Canada H3C 3J7
E-mail :
Fax :

Total number of pages: 23
Total number of figures: 3
Number of words in manuscript: 3236
Number of words in Abstract: 143
Number of words in Introduction: 813

Abstract

Our ability to discriminate and recognize human voices is amongst the most important functions of the human auditory system. The current study sought to determine whether electrophysiological markers could be used as objective measures of voice familiarity, by looking at the electrophysiological responses (Mismatch Negativity and P3a) when the infrequent stimulus presented is a familiar voice as opposed to an unfamiliar voice. Results indicate that the MMN elicited by a familiar voice is greater than that elicited by an unfamiliar voice at FCz. The familiar voice also produced a greater P3a wave than that triggered by the unfamiliar voice at Fz. Since both the MMN and the P3a were elicited as participants were instructed not to pay attention to incoming stimulation, these findings suggest that voice recognition is a particularly potent pre-attentive process whose neural representations can be objectively described through electrophysiological assessments.

Introduction

The human voice is a very prominent stimulus in our auditory environment as it plays a critical role in most human interactions, particularly as the carrier of speech. Our ability to discriminate and recognize human voices is amongst the most important functions of the human auditory system, especially in the context of speaker identification (Belin et al., 2004; van Dommelen, 1990). Several neuropsychological and neuroimaging studies have attempted to investigate the underlying neural mechanisms involved in voice processing. Neuropsychological studies have demonstrated the existence of a specific disability in recognizing human voices, a disorder called phonagnosia (Assal et al., 1981; Malone et al., 1982; van Lancker et al., 1989). Patients with phonagnosia can either show deficits in the ability to discriminate between voices, reflecting perceptual impairments in human voice processing, or deficits in speaker identification, which might reflect an echoic sensory memory dysfunction (van Lancker & Kreiman, 1987). As neuropsychological dissociations are commonly used to assess for neurofunctional modularity, phonagnosia may imply the existence of a perceptual brain mechanism dedicated to human voice processing, and more precisely, to voice identity (Belin et al., 2004). Moreover, several neuroimaging studies have provided robust evidence for neuronal systems that were specifically activated in human voice processing. Specific bilateral regions along the superior temporal sulcus (STS) were found to be voice-selective (Belin et al., 2000; Binder et al., 2000). These and several other studies provide a robust justification for the specificity of particular brain areas dedicated to human voice processing (Assal et al., 1981; Imaizumi, 1997; van Lancker & Canter, 1982; van Lancker et al., 1988). However, few electrophysiological studies have yet provided a reliable demonstration for the existence of distinctive neural mechanisms associated with human voice discrimination (Titova & Näätänen, 2001; Levy et al., 2003).

ERPs have been useful to highlight changes in cortical activity when auditory information is presented. Auditory oddball paradigms, which involve the presentation of infrequent stimuli embedded among frequent stimuli, have commonly been used to generate the ERP component known as the MMN (Näätänen et al., 1978; Näätänen & Alho, 1995). Since auditory perception involves the discrimination of a vast array of sound features such as pitch, rhythm, timbre, loudness and formant transitions, the mismatch negativity (MMN), a change-specific component of the auditory event-related potentials (ERPs), seems to be an appropriate tool to investigate such auditory perceptual discrimination abilities. Furthermore, the MMN can be recorded without having the participant paying attention to the incoming stimulation. Indeed, the MMN is a versatile measure that can pre-attentively discriminate the smallest alterations when any one parameter differs between two consecutive stimuli.

According to Näätänen and Alho's model (1995), the discrimination of two successive stimuli differing in only one parameter reflects the involvement of two different neural representations. In other words, a frequently presented stimulus forms a neural trace in the echoic sensory memory, which can last up to 8 to10 seconds (Böttcher-Gandor & Ullsperger, 1992). The sensory input from the infrequent stimulus does not fit with the existing neural trace, therefore resulting in a negative deflection, the MMN component. Thus, the MMN is elicited by any discriminable change in some repetitive aspects of auditory stimulation stored in echoic memory (Cowan et al., 1993). With regard to its neuroanatomical substrates, the MMN is topographically distributed over fronto-central recording sites as MMN dipoles, which at least in part give rise to this distribution, lie in the superior temporal auditory cortex (Girard et al., 1990; Sams et al., 1991; Scherg et al., 1989). Moreover, this MMN occurs roughly 200 ms after stimulus onset (McGee et al., 1997). In previous studies, the MMN component has been elicited by using a change in a wide range of sounds including simple

sinusoidal tones (Tervaniemi et al., 1997), musical tones (Alho & Sinervo, 1997), phonemes and syllables (Näätänen et al., 1997; Näätänen, 1995). More recently, Titova and Näätänen (2001) explored whether the MMN could be used as an objective measure of voice discrimination using a female voice as a standard stimulus while infrequent stimuli included one male voice and three female voices. Significant positive correlations were established between the MMN amplitude and the dissimilarity ratings. Interestingly, according to these authors, the MMN amplitude proved to be a more reliable indicator of voice identity than the behavioural dissimilarity ratings.

Hence, based on those findings, it appears possible that the MMN could be used as a reliable and objective measure of voice familiarity, rather than using attention-dependent behavioural measures. The MMN can in fact, not only be used as a measure of speaker discrimination, but based on previous studies suggesting that long-term memory may have a significant impact on short-term memory functioning (Näätänen et al., 1997; Winkler et al., 1999), it could also be used as a sensitive measure of voice familiarity. Thus, we hypothesized that the MMN elicited by a familiar voice should be distinct from the MMN evoked by an unfamiliar voice.

Methods

Participants

Fifteen French-speaking right-handed adults (18-25 years) participated in the present study. Exclusion criteria included hearing, attentional or neurological disorders. Moreover, participants were asked to withhold from any drug or alcohol consumption at least 24 hours before testing. Participants needed to display an exogenous N1 component in their ERP waveform to be part of the experimental group, as it reflects normal auditory system functioning. One participant was excluded from this study as excessive artefacts and technical problems during testing contaminated the ERP recordings.

Stimuli

Stimuli were samples of the French vowel /a/ pronounced by the natural voice of different speakers. People were instructed to pronounce /a/ as in the word “allô”, the French word for hello, to control for undesired variation in pitch when pronouncing the vowel. The amplitude envelope of the beginning of the recorded vowel was kept intact in order to keep the natural rising envelope of the sound, thus maintaining the voice as natural and identifiable as possible. Stimuli were 212 ms in duration (including 10 ms fall time) and were presented at 75 dB sound pressure level (SPL) at the participant’s head. The stimulus onset asynchrony (time interval between the onsets of two consecutive stimuli) was 800 ms, thus keeping the inter-stimulus interval (ISI) at 588 ms. Stimuli were presented in a pseudorandom oddball sequence composed of three different stimuli: (a) a standard unfamiliar voice (85% of the total number of presented stimuli), (b) an infrequent unfamiliar voice (7.5% of the total number of presented stimuli), and (c) a second infrequent stimulus (7.5% of the total number of presented stimuli) being a familiar voice. Any infrequent stimulus was always preceded by at least three standard stimuli to ensure that a neural trace for the standard stimulus had formed. Familiar voices were obtained as each participant was required to ask a close relative or a long-time friend to visit our laboratory in order to record their voice. Unfamiliar voices (for both standard and infrequent stimuli) were selected according to pitch similarity from a pool of available voices recorded for the purpose of the present study.

Four blocks of 390 stimuli were presented. The sequence of stimuli was generated by the E-Prime Psychology Software on a DELL computer located in an adjacent room. Stimuli were binaurally delivered through speakers (XTS-24 model from Bookshelves), positioned 20 cm left and right from the participant’s head, at auricular height.

Procedure

The experiments took place in the sensory and cognitive electrophysiology laboratory of the Centre Hospitalier Universitaire Mère-Enfant (Ste-Justine Hospital). Participants first had to carefully read and sign a consent form providing detailed information about the experimental procedure, which was approved by the Ste-Justine Hospital's ethics committee. Participants were rewarded with a financial compensation of $30 for taking part in our study.

The cortical responses were acquired with a high-density recording system, the Geodesic 128-Sensor Net (Electrical Geodesics Inc., Eugene, OH) at a sampling rate of 250 Hz, a band-pass filter of 0.1-100 Hz, and Cz referenced. Participants were seated in a comfortable chair located in a semi-obscure Faraday room. Impedance was kept below 50 kΩ, which is within recommended range when using a powerful amplifier such as the one used in this study (Net Amps 200). Evoked potentials were recorded with the Net Station program on a Macintosh G4 computer. Because cortical responses vary with awareness state, subjects were required to stay awake throughout the whole testing session. Participants were instructed beforehand that they would hear a familiar voice (which was identified) and a stranger's voice among the standard stimuli. To ensure that they did not pay attention to the sequence of stimuli during testing, all participants were instructed to focus their attention on a silent subtitled movie of their choice presented on a monitor at a distance of about 2 meters. Participants were monitored through an infrared camera equipped with an integrated speaker, allowing them to communicate at all times with the experimenters located in an adjacent room. The total duration of the session was kept within one hour.

Once ERP recordings were completed, five participants were randomly selected to perform a behavioural discrimination between the three voices (one familiar voice and two unfamiliar voices) they were presented with during the ERP study. They were instructed to press on the space bar of a computer when they heard a familiar voice and to ignore the presentation of an unfamiliar voice.

Data Analysis

EEG recordings were analyzed using the BrainVision Analyzer program on an IBM computer. Various pre-processing filters were applied on the data, such as a raw data inspector, ocular correction, artefacts rejection and baseline correction. Electrodes AFz, Fz and FCz were referenced to both left and right linked mastoids with a frequency band-pass filter of 1-30 Hz at 24 dB/oct. EEG epochs of 800 ms (including 200 ms pre-stimulus period) were averaged after artefact rejection (threshold of ±100 μV), independently for each stimulus type. No ERP averages were based on less than 90 trials. MMN responses were computed according to the following procedure: ERPs evoked by a standard stimulus (unfamiliar voice I) were subtracted from ERPs evoked by the presentation of an infrequent stimulus (either unfamiliar voice II or familiar voice) for each participant. Responses to standard stimuli that immediately followed the presentation of infrequent stimuli were excluded from the standard stimulus average as different brain activation could have resulted from the dynamic formation of a new memory trace. Latencies of the most negative peaks in the individual difference waves were measured between 80 and 280 ms after stimulus onset. The MMN amplitude and the duration of the temporal window in which the MMN waveform takes place varied fairly across participants. Therefore, the MMN component was obtained using the area under the curve contained within a 50 ms time window in which the midpoint had previously been identified in a peak amplitude detection manipulation performed for each participant (Kujala et al., 2005; Nenonen et al., 2005; Restuccia et al., 2005; Takegata et al., 2005). The MMN component values obtained when

presented with an unfamiliar voice were then compared to those elicited by a familiar voice. Thus, the dependent variable is the area under the MMN curve, with one within-subject factor: familiar/unfamiliar voice.

Control experiment

In order to ensure that no other sound features could be used to distinguish the familiar stimulus from the two unfamiliar stimuli (frequent and infrequent), we included a control group consisting of eight right-handed participants (18-25 years old) for whom none of the stimuli were familiar. These participants listened to the same set of three stimuli that had been presented to eight participants from the experimental group. The procedure and data analyses were the same as those reported above.

Results

The grand-average ERPs illustrated in Figure 1, obtained when averaging the responses elicited by each stimulus for all participants, show that the infrequent familiar voice elicited a negative displacement (MMN) of greater amplitude than the infrequent unfamiliar voice when taking the standard stimulus for reference.

Include Figure 1 about here

Midline electrodes anterior to Cz were further analysed as the MMN has been found to be topographically distributed in fronto-central regions (Ilvonen et al., 2004; Näätänen et al., 2004; van Zuijen et al., 2005; Ylinen et al., in press). As illustrated in Figure 2A, the MMN response elicited by the familiar voice was significantly greater than that of the unfamiliar voice at the fronto-central site (FCz) [F (1,13) = 5.368, $p < 0.05$]. Although not reaching significance, electrodes anterior to FCz,

namely Fz and Afz, tended to record a greater MMN in response to a familiar voice than to an unfamiliar voice (Figure 2A). In control subjects for whom none of those stimuli were familiar, the MMN responses elicited by the two infrequent voices were statistically equivalent at all three electrode sites (Afz, Fz, FCz, all p values > 0.05), as shown in figure 2B. It is noteworthy that the unfamiliar voices elicited similar response amplitudes in both the experimental and control groups (see Fig. 2A and B), whereas the familiar voice clearly produced higher response amplitude in the experimental group (Fig 2A).

Include Figure 2 about here

Brain activity topographies elicited when hearing a familiar voice as opposed to an unfamiliar voice seem to confirm the observed trend. Indeed, as illustrated in Figures 3A and B, the familiar voice elicited more negative activity at 200 ms post-onset in the fronto-central region of the cerebral cortex when compared to that elicited by an unfamiliar voice. A t-test conducted by BrainVision, which assumed that, in order for activity distribution to be significantly different across voice familiarity conditions, its value would have to be superior to 2.65, given the Student's t-distribution table, revealed no significant difference [$t_{0.01}(13) < 2.65$]. Thus, this result suggests that both topographies differed only in amplitude and that their distributions were not spatially different.

Insert Figure 3 about here

Further inspection of the averaged ERPs curves also revealed the presence of significantly greater P3a amplitude when hearing a familiar voice when compared to that elicited by an unfamiliar voice (Figure 1A). Thus, the positive peak of the P3a wave was a posteriori identified with a peak amplitude detection manipulation performed for each participant between 240 and 320 ms after stimulus onset. The Fz electrode site [$F(1,13) = 6.821$, $p < 0.05$] showed a significantly greater P3a

amplitude to a familiar voice than to an unfamiliar voice (Figure 2C). When similar statistical analyses were computed in control subjects for whom all three voices were unfamiliar, we found that the P3a component elicited by the two infrequent stimuli were statistically equivalent at all three electrode sites (Afz, Fz, FCz, all p values > 0.05), as shown in figure 2D).

Brain activity topographies elicited when hearing a familiar voice as opposed to an unfamiliar voice also confirm this pattern. As illustrated in Figure 3, the grand-average P3a wave elicited by a familiar voice (C) was more positive in the frontal region of the cerebral cortex when compared to the grand-average P3a wave elicited by an unfamiliar voice (D). A t-test, also conducted by BrainVision, contrasting both P3a topographies revealed no significant difference across voice familiarity conditions [$t_{0.01}(13) < 2.65$], again suggesting that both topographies differed only in amplitude and that their distributions were not spatially different.

The behavioural task revealed that, on average, participants could accurately recognized 97 % (± 0.07%) of familiar voices while correctly ignoring 97% (± 0.04%) of unfamiliar voices.

Discussion

The main result of the present study is the significant difference between the electrophysiological responses (MMN and P3a) elicited by a familiar voice when compared to those of an unfamiliar voice. These findings provide preliminary support suggesting that specialized areas for voice processing are especially tuned to familiar voices as opposed to unfamiliar voices. Indeed, although scalp distributions for both electrophysiological components (MMN and P3a) were not found to be significantly different across voice familiarity conditions, thus suggesting that underlying

activation dipoles have similar position and orientation, they did differ in terms of their activation strength. Dipole source analyses would, however, further be required to validate this interpretation. Moreover, both the MMN and the P3a were elicited despite the fact that participants were instructed not to pay attention to incoming stimulation, which suggests that there is some degree of pre-attentive voice familiarity evaluation, and those results were corroborated behaviourally. Furthermore, these results could not merely be due to subtle acoustic parameters differences between the familiar voice and the unfamiliar voice as control subjects for whom none of the three voices were familiar did not show significantly different MMN and P3a responses across voice familiarity conditions.

In his original MMN theory, Näätänen (1990) suggested that long-term memory was connected to short-term memory only via conscious and attention-dependent processes. However, a more recent study conducted by Näätänen and colleagues (1997) challenged previous beliefs as it showed that long-term memory traces have a considerable impact on MMN elicitation. In that study, the MMN was enhanced when a frequently repeated phoneme was replaced by a native-language phoneme as opposed to another phoneme belonging to a foreign language for which no long-term memory trace had been established. Thus, long-term memory has a significant impact on auditory short-term memory. Our MMN results confirm this interpretation: the detection of a familiar voice, which requires its retrieval in long-term memory, argues in favour of a direct connection from long-term memory to the feature-analysis system of the short-term memory used to detect the acoustic features of stimuli (Huotilainen et al., 2001). Indeed, since greater brain activation was evoked by the infrequent familiar voice as opposed to the equiprobable infrequent unfamiliar voice, and since this brain activation difference can only be attributed to the previous formation of a long-term memory

trace for the familiar voice, our MMN results provide compelling evidence for the implication of long-term memory in the feature-analysis system for voice processing.

The present study also identified a positive component peaking at about 300 ms from stimulus onset. The polarity of this component, its latency as well as its mainly frontal activation reflect the presence of a Novelty P3 component elicited by outstanding distracters in an oddball paradigm. The Novelty P3 (usually referred to as P3a) is considered to reflect unintentional brain response to novel stimuli. A P3a component is evoked when a stimulus is (a) outstanding (hence novel) (Grillon et al., 1990); (b) rare as opposed to the other distracters (Katayama & Polich, 1996); or (c) easy to distinguish from the frequent distracters (Comerchero & Polich, 1999). Within the context of our experimental paradigm, a familiar voice would appear to be both outstanding and easier to distinguish from unfamiliar voices, be they frequent or rare, thereby eliciting a greater P3a amplitude. Thus, our findings indicate that a familiar voice is a sufficiently compelling stimulus to distract listeners from their ongoing activity.

Interestingly, similar results arise from studies of face recognition, which is considered to be an automatic process that allows the identification of familiar individuals (Bobes et al., 2000). The processing of familiar and unfamiliar faces has frequently been used to understand face recognition. Similarly to our findings, Bobes and colleagues (2000) identified a larger late positive component (LPC) related to the P300 component for the familiar faces when contrasted with that of unfamiliar faces.

Conclusion

Our electrophysiological results confirm that specific neural mechanisms are particularly tuned to voice discrimination. Our results further underline the prominent role played by the human voice in our auditory environment, not only as the carrier of speech but as importantly, as an index of familiarity.

Acknowledgements

This study was supported by the Canada Research Chair program awarded to Maryse Lassonde as well as from the National Science and Engineering Research grants awarded to Maryse Lassonde and Pascal Belin.

References

Alho, K., & Sinervo, N. (1997) Preattentive processing of complex sounds in the human brain. *Neurosci Lett*, **233**, 33-36.

Assal, G., Aubert, C., & Buttet, J. (1981) Asymétrie cérébrale et reconnaissance de la voix. *Rev Neurol*, **137**, 255-268.

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

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

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex*, **10**, 512-528.

Bobes, M. A., Martín, M., Olivares, E., & Valdés-Sosa, M. (2000) Different scalp topography of brain potentials related to expression and identity matching of faces. *Brain Res Cogn Brain Res*, **9**, 249-260.

Böttcher-Gandor, C., & Ullsperger, P. (1992) Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiol*, **29**, 546-550.

Comerchero, M. D., & Polich J. (1999) P3a and P3b from typical auditory and visual stimuli. *Clin Neurophysiol*, **110**, 24-30.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993) Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn Mem Cogn*, **19**, 909-921.

Girard, M. H., Perrin, F., Pernier, J., & Bouchet, P. (1990) Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. *Psychophysiol*, **27**, 627-640.

Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., & Braff, D. (1990) Effects of rare non-target stimuli on brain electrophysiological activity and performance. *Int J Psychophysiol*, **9**, 257-267.

Huotilainen, M., Kujala, A., & Alku, P. (2001) Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neurosci Lett*, **310**, 133-136.

Ilvonen, T., Kujala, T., Kozou, H., Kiesilainen, A., Salonen, O., Alku, P., & Näätänen, R. (2004) The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neurosci Lett*, **366**, 235-240.

Imaizumi S. (1997) Vocal identification of speaker and emotion activates different brain regions. *Neuroreport*, **8**, 2809-2812.

Katayama, J., & Polich, J. (1996) P300 from one-, two-, three-stimulus auditory paradigms. *Int J Psychophysiol*, **23**, 33-40.

Kujala, T., Lepisto, T., Nieminen-vonWendt, T., Näätänen, P., & Näätänen, R. (2005) Neurophysiological evidence for cortical discrimination impairment of prosody in Asperger syndrome. *Neurosci Lett*, **383**, 260-265.

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

Malone, D. R., Morris, H. H., Kay, M. C., & Levin, H. S. (l982) Prosopagnosia: a dissociation between recognition of familiar and unfamiliar faces. *J Neurol Neurosurg Psychiatry*, **45**, 820-822.

McGee, T., Kraus, N., & Nicol, T. (1997) Is it really a mismatch negativity? An assessment of methods for determining response validity in individual subjects. *Electroencephalogr Clin Neurophysiol*, **104**, 359-368.

Näätänen, R. (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci*, **13**, 201-288.

Näätänen, R. (1995) The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear*, **16**, 6-18.

Näätänen, R., & Alho, K. (1995) Mismatch negativity – A unique measure of sensory processing in audition. *Int J Neurosci*, **80**, 317-337.

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

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, **385**, 432-434.

Näätänen, R., Syssoeva, O., & Takegata, R. (2004) Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. *Psychophysiol*. **41**, 660-663.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005) Speech-sound duration processing in a second language is specific to phonetic categories. *Brain Lang*, **92**, 26-32.

Restuccia, D., Della Marca, G., Marra, C., Rubino, M., & Valeriani, M. (2005) Attentional load of the primary task influences the frontal but not the temporal generators of mismatch negativity. *Brain Res Cogn Brain Res*, **25**, 891-899.

Sams, M., Kaukoranta, E., Hamalainen, M., & Näätänen, R. (1991) Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiol*, **28**, 21-28.

Scherg, M., Vajsar, J., & Picton, T. W. (1989) A source analysis of the human auditory evoked potential. *J Cogn Neurosci*, **1**, 336-355.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005) Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Brain Res Cogn Brain Res*, **25**, 169-179.

Tervaniemi, M., Schröger, E., & Näätänen, R. (1997) Pre-attentive processing of spectrally complex sounds with asynchronous onsets: an event-related potential study with human subjects. *Neurosci Lett*, **227**, 197-200.

Titova, N., & Näätänen, R. (2001) Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett*, **308**, 63-65.

van Dommelen, W. A. (1990) Acoustic parameters in human speaker recognition. *Lang Speech*, **33**, 259-272.

van Lancker, D., & Canter, G. J. (1982) Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn*, **1**, 185-195.

van Lancker, D., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988) Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex*, **24**, 195-209.

van Lancker, D., & Kreiman, J. (1987) Voice discrimination and recognition are separate abilities. *Neuropsychologia*, **25**, 829-834.

van Lancker, D., Kreiman, J., & Cummings, J. (1989) Voice perception deficits: Neuroanatomical correlates of phonagnosia. *J Clin Exp Neuropsychol*, **11**, 665-674.

van Zuijen, T. L., Sussman, E., Winkler, I., Näätänen, R., & Tervaniemi, M. (2005) Auditory organization of sound sequences by a temporal or numerical regularity--a mismatch negativity study comparing musicians and non-musicians. *Brain Res Cogn Brain Res*, **23**, 270-276.

Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csépe, V., et al. (1999) Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cogn Brain Res*, **7**, 357-369.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Näätänen, R. (in press). Mismatch negativity (MMN) elicited by changes in phoneme length: A cross-linguistic study. *Brain Res*.

Figure legends

Figure 1. Grand-average ERPs elicited by a familiar voice (dashed) and by an unfamiliar voice (dotted), referenced to the standard stimulus (black) in the experimental group.

Figure 2. Areas under the curve during a 50 ms time window centered at the peak latency for both MMN elicited by infrequent stimuli, for the experimental group (A) and the control group (B). Amplitudes at the peak latency for both P3a elicited by infrequent stimuli, for the experimental group (C) and the control group (D).

Figure 3. Grand-average MMN (ERPs to the infrequent – ERPs to the standard) for three midline anterior electrodes and their topographies at around 200 ms post-stimulus onset in response to (A) a familiar voice and (B) an unfamiliar voice. Grand-average ERPs for the same three midline electrodes and their topographies at around 300 ms post-stimulus onset in response to (C) a familiar voice and (D) an unfamiliar voice.

Figure 1

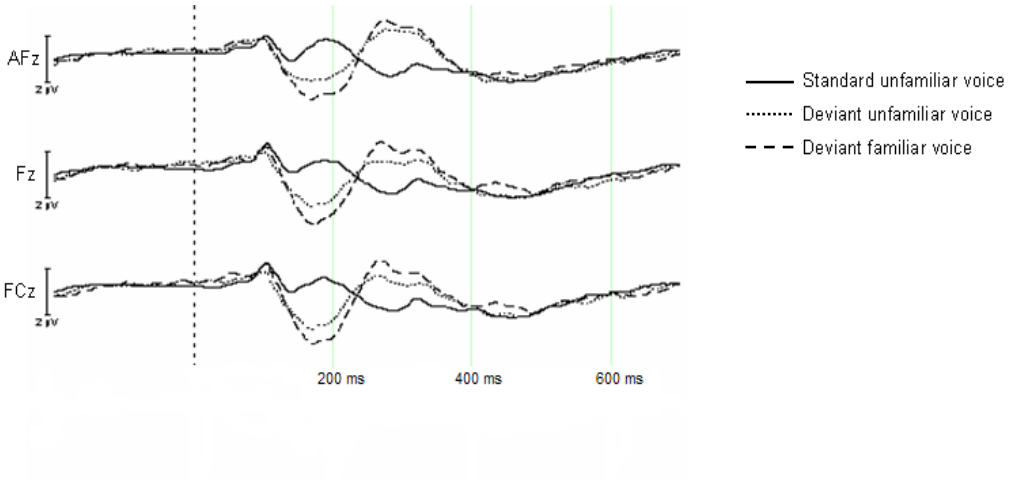


Figure 2

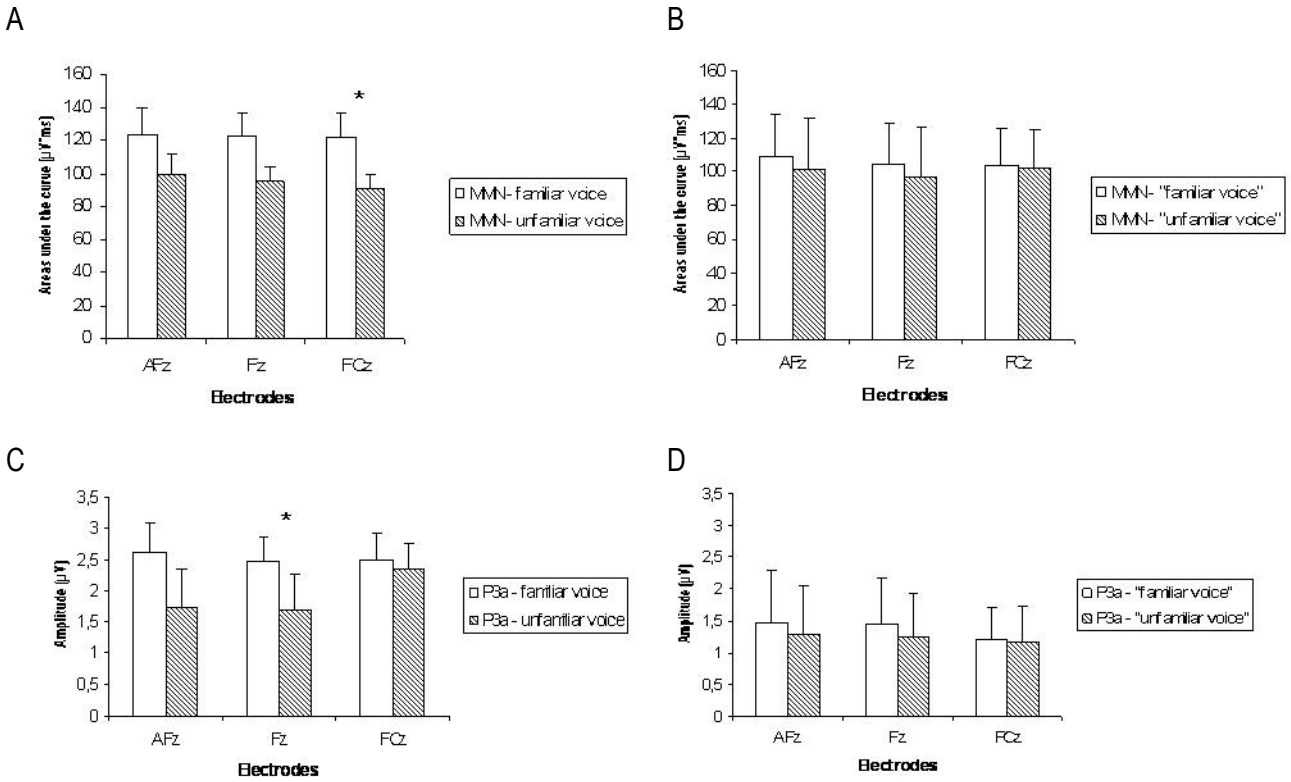
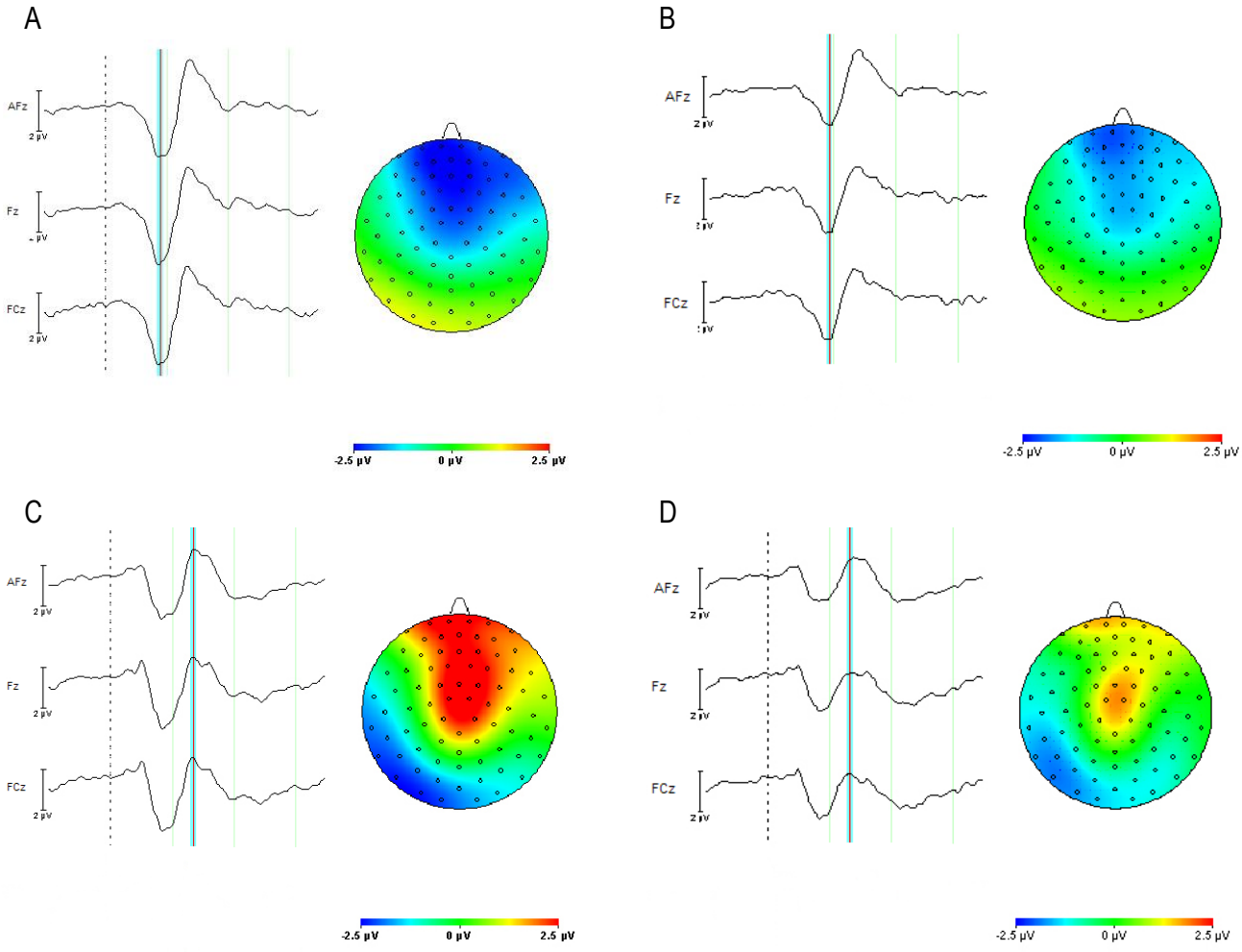


Figure 3


A
AFz
Fz
FCz
2 µV
-2.5 µV
0 µV
2.5 µV
B
AFz
Fz
FCz
-2.5 µV
0 µV
2.5 µV
C
AFz
Fz
FCz
2 µV
-2.5 µV
0 µV
2.5 µV
D
AFz
Fz
FCz
2 µV
-2.5 µV
0 µV
2.5 µV

ms

ARTICLE 2 :

MOTHER AND STRANGER: AN ELECTROPHYSIOLOGICAL STUDY OF VOICE PROCESSING IN NEWBORNS

Publié dans Cerebral Cortex (2011), 21, 1705-1711.

Mother and Stranger: an electrophysiological study of voice processing in newborns

Running Title: Mother's voice processing in newborns

Maude Beauchemin[1,2] B.Sc., Berta González-Frankenberger[1,2] Ph.D., Julie Tremblay[1] Jr.Eng., Phetsamone Vannasing[1] EPM, Eduardo Martínez-Montes[3] Ph.D., Pascal Belin[2,4] Ph.D., Renée Béland[1,2] Ph.D., Diane Francoeur[1] M.D., Ana-Maria Carceller[1] M.D., Fabrice Wallois[5] M.D./Ph.D. and Maryse Lassonde[1,2] Ph.D.

1–Centre de recherche, CHU Sainte-Justine, Montréal, Canada

2–Centre de recherche en neuropsychologie et cognition (CERNEC), Université de Montréal, Montréal, Canada

3–Cuban Neuroscience Center, Havana, Cuba

4–Department of Psychology, University of Glasgow, Scotland, United Kingdom

5–Groupe de Recherches sur l'Analyse Multimodale de la Fonction Cérébrale (GRAMFC, EA 4293), Université de Picardie Jules Verne, CHU Nord, Amiens, France

Corresponding author:

Maryse Lassonde, Ph.D.
Department of Psychology
Pavillon Marie-Victorin
90, Ave Vincent d'Indy
Montreal, Quebec
H2V 2S9
Email :
Phone :
Fax :

ABSTRACT

In the mature adult brain, there are voice-selective regions that are especially tuned to familiar voices. Yet, little is known about how the infant's brain treats such information. Here, we investigated, using electrophysiology and source analyses, how newborns process their mother's voice compared to that of a stranger. Results suggest that, shortly after birth, newborns distinctly process their mother's voice at an early, pre-attentional, level, and at a later, presumably cognitive level. Activation sources revealed that exposure to the maternal voice elicited early language-relevant processing whereas the stranger's voice elicited more voice-specific responses. A central, probably motor response was also observed at a later time, which may reflect an innate auditory-articulatory loop. The singularity of left-dominant brain activation pattern together with its ensuing sustained, greater central activation in response to the mother's voice, may provide the first neurophysiologic index of the preferential mother's role in language acquisition.

Keywords: Mismatch negativity (MMN), Newborns, Source analyses, Voice processing

Voices have proven to be special within the human auditory cortex (Belin et al., 2000). Using functional magnetic resonance imaging in human volunteers, voice-selective regions can be found bilaterally along the upper bank of the superior temporal sulcus (STS), with predominant right hemisphere activation. This particularly salient stimulus in our auditory environment is also known to play a prominent role in most human interactions, even more so in the context of speaker identification (Belin et al., 2004; van Dommelen, 1990). These specialized areas for voice processing are especially tuned to familiar voices as opposed to unfamiliar ones (Beauchemin et al., 2006). Behavioural evidence shows that this ability is present very early in life: foetuses and newborns react preferentially to their mother's voice over that of a female stranger (DeCasper and Fifer, 1980; Kisilevsky et al., 2003; Ockleford et al., 1988; Querleu et al., 1984). However, the means by which infants acquire this ability remain ambiguous: the neurophysiology of infant-mother interaction is still poorly understood as no study has ever looked at newborns' brain responses to their mother's voice. We therefore thoroughly investigated cortical activity in response to voices (mother and female stranger) as well as the brain regions that generate such activity in 16 newborn babies (mean age: 21 hours). We found that not only do newborns process their mother's voice more actively than that of a stranger, but they also process it differently. The maternal voice (the sound /a/, 212 msec) preferentially activated language-relevant cortical areas whereas the stranger's voice predominantly activated voice-specific areas. Additionally, the mother's and to a lesser extent the stranger's voice, later activated central, motor areas: a finding we interpret as reflecting an innate auditory-articulatory loop, more specifically tuned to the mother's voice.

METHODS

Participants

Sixteen newborns (8–27 hours, mean: 21 hours) participated in the present study. They were full-termed (>37 weeks gestational age) and did not encounter any problem during pregnancy, labour or delivery. Their birth record and neonatal exam indicated that they were healthy and their physiological parameters at birth were considered normal (APGAR, weight, length, head circumference). Mothers gave informed written consent, which was approved by the Ste-Justine Hospital's ethics committee, for their infant and were present during testing. All newborns explicitly showed an exogenous stimulus-related response (infantile N1 as labelled by (Novak et al., 1989)) that reflects normal auditory system functioning. Additionally, families of the infants did not have a history of hearing, language or reading pathology.

Stimuli and Procedure

The stimuli and procedure were inspired from previous work in our laboratory (Beauchemin et al., 2006). Briefly, voice familiarity can objectively be measured with a modified oddball paradigm, using a frequent unfamiliar voice and two rare voices, only one of which is highly familiar (the mother's voice in the present case). A cerebral mismatch process (MMN) is therefore triggered by the two rare stimuli in an automatic comparison process with the neuronal memory trace left by the repetitive unfamiliar frequent stimulus. The MMN has also been said to develop rather early in comparison to other event-related potential (ERP) waves. It has even been suggested to be the ontogenetically earliest discriminative response of the human brain (Cheour-Luhtanen et al., 1996).

Stimuli were samples of the French vowel /a/ pronounced by the natural voice of different speakers, all females. People were instructed to pronounce /a/ as in the word "allô", the French word

for hello, to control for undesired variation in pitch when pronouncing the vowel. The amplitude envelope of the beginning of the recorded vowel was kept intact in order to keep the natural rising envelope of the sound, thus maintaining the voice as natural and identifiable as possible. All voices were segmented using Cool Edit 2000. Stimuli were 212 ms in duration (including 10 ms rise and fall time) and were presented at a constant 75 dB sound pressure level (SPL) at the newborn's head so that loudness would not be responsible for any difference in brain activation. The stimulus onset asynchrony (time interval between the onsets of two consecutive stimuli) was 800 ms, thus keeping the inter-stimulus interval (ISI) at 588 ms. Stimuli were presented in a pseudorandom oddball sequence composed of three different stimuli: (a) a frequent unfamiliar voice (85% of the total number of presented stimuli), (b) an infrequent unfamiliar voice (7.5% of the total number of presented stimuli), and (c) a second infrequent stimulus (7.5% of the total number of presented stimuli) being the mother's voice. Any infrequent stimulus was always preceded by at least three frequent stimuli to increase the likelihood that a neural trace for the frequent stimulus had formed. The familiar voice was recorded when mothers visited our laboratory during their pregnancy. The unfamiliar infrequent voice was that of the attending nurse who accompanied both the mother and the newborn to our laboratory, as it has been demonstrated that newborns' brain's response is larger to a novel rare stimulus than to a known rare stimulus (Sambeth et al., 2006). Therefore, as we did not want to get novelty detection confused with familiar voice "recognition", we recorded the voice of the babies' attending nurse with whom the mother had met at her monthly appointments and then every two weeks near the end of her pregnancy. The unfamiliar infrequent voice (the attending nurse) was the same for all newborns. It is worth mentioning that the attending nurse is herself a mother, which allowed us to control for any "motherese" effect. Caretakers in most cultures use a special speech register when talking to infants that has a unique acoustic signature, called motherese. This infant-directed speech is slower, has a higher average pitch and contains exaggerated pitch contours (Kuhl

et al., 1997). Finally, the frequent unfamiliar voice was selected among a pool of available voices according to pitch similarity to both rare voices.

For the nurse's voice, fundamental frequency (F0) and center frequencies formants for F1 and F2 were 200 Hz, 865 Hz, 1413 Hz respectively. For mothers' voices, mean F0 and mean center frequencies formants were F0 = 202 Hz (SD = 17 Hz); F1 = 944 Hz (SD = 85 Hz); F2 = 1608 Hz (SD = 133 Hz). For unfamiliar frequent voices, mean F0 and mean center frequencies formants were F0 = 210 Hz (SD = 24 Hz); F1 = 975 Hz (SD = 53 Hz); F2 = 1580 Hz (SD = 134 Hz). To ensure that voices were acoustically comparable, we calculated the mean F0 difference between the nurse's voice and both mothers' voices and unfamiliar frequent voices prior to statistically comparing those differences. A paired t-test on F0 differences did not reveal any significant difference [$t(15) = 0.87$, $p = 0.39$]. Similarly, we computed and statistically contrasted mean differences between F2/F1 ratios of the nurse's voice relative to both mothers' voices and unfamiliar frequent voices. Again, a paired t-test did not show any significant difference [$t(15) = -1.57$, $p = 0.13$]. Both analyses confirmed that the mean inter-vowel space's distance between utterances of mothers and the nurse was not larger than that between utterances of unfamiliar frequent voices and the nurse's.

Four blocks of 390 stimuli were presented for a total of 1560 stimuli (117 infrequent mother's voice stimuli, 117 infrequent stranger's voice stimuli and 1326 frequent unfamiliar voice stimuli) and for a total duration of 24 minutes. The sequence of stimuli was generated by the E-Prime Psychology Software on a DELL computer located in an adjacent room. Stimuli were binaurally delivered through speakers, positioned 20 cm left and right from the newborn's head, at auricular height.

Event-related potentials (ERPs) are small voltage fluctuations reflecting, with high temporal resolution, patterns of neuronal activity evoked by a stimulus. Brain responses were acquired with a high-density recording system, the Geodesic 128-Sensor Net (Electrical Geodesics Inc., Eugene, OH) at a sampling rate of 250 Hz, a band-pass filter of 0.1-100 Hz, and Cz referenced. Mothers were seated in a comfortable chair located in a semi-obscure Faraday room, with their newborn lying on their laps. All babies were in active sleep state during recording as, even when asleep, neonates are able to process external information actively (Cheour et al., 2002a; Fifer et al., 2010) and as the MMN can be observed in infants throughout all sleep stages as well as when they are awake (Cheour et al., 1998a; Cheour et al., 1998b; Cheour et al., 2000). Sleep stage classification was scored off-line according to the EEG signal, EMG and EOG (Anders, 1979) and during all recordings, newborns were in active state. It is noteworthy that all mismatch responses (mother/frequent and stranger/frequent) were simultaneously recorded, thereby excluding differential sleep patterns as being responsible for any differences observed in mismatch responses. Electrode impedance was kept below 50 kΩ, which is within recommended range when using a powerful amplifier such as the one used in this study (Net Amps 200). Time-locked evoked potentials (EPs) were recorded with the Net Station program on a Macintosh G4 computer. Mothers and newborns were monitored through an infrared camera equipped with an integrated speaker, allowing them to communicate at all times with the experimenters located in an adjacent room. The attending nurse was also always present in the Faraday room, together with mothers and newborns. All infants were fed immediately before testing to increase the likelihood they would sleep through the entire procedure. Pauses were provided between blocks to monitor the baby's temperature and respiratory state, thereby preventing them from reaching the quiet sleep stage.

Data Analysis

Electroencephalogram (EEG) time series of 600 ms with a 100 ms pre-stimulus interval were edited off-line by visual inspection using the BrainVision AnalyzerTM program (Brain products, Munich, Germany). Baseline corrections were performed, together with several pre-processing filters and ocular corrections (EOG). Artefacts were also manually removed. EEG time series with amplitudes over 100 μV were withdrawn from the analysis. Electrodes were referenced to both linked mastoids with a frequency bandpass filter of 1-30 Hz at 24 dB/octave. For each subject, EEG time series were edited in response to the three types of stimuli: (1) Mother's voice (mean = 97.2 stimuli remaining from the 117 presented ones), (2) Stranger's voice (mean = 99.5 stimuli remaining from the 117 presented ones), (3) Frequent voice (mean = 839.8 stimuli remaining from the 1326 presented ones). The edited EEG time series for each subject and for each condition were exported to MatlabTM v7.0.4 (The MathWorks, Inc., Massachusetts, U.S.A) in order to apply the statistical analyses.

Statistical Analyses

Mismatch negativity responses (MMN) of the newborns to the mother's voice and to the stranger's voice were computed by subtracting the EPs obtained in the frequent condition from the EPs in response to the mother's and stranger's voices respectively. A non-parametric permutation test was applied to find the time points with significant differences in each of the following comparisons: (1) mother's voice against frequent, (2) stranger's voice against frequent, and (3) mother's MMN against stranger's MMN. A total of 500 permutations under the null hypothesis of no-difference between conditions were carried out separately on the EEG time series of 3 midline electrodes (Fz, FCz, Cz), where differences in the electrophysiological response (i.e. MMN) were expected (Kazemi et al., 2007a; Kazemi et al., 2007b; Riera and Fuentes, 1998). The results of the

permutation test were shown as plots of probabilities for accepting the null hypothesis for each sampled time point and for each electrode, defining as significant those time points with probability below the significance level of 0.05 (Figure 1C).

Source analyses

In order to identify the generators of the MMN responses, source analyses on the 117 electrodes (i.e. excluding the 11 EOG and electromyogram (EMG) electrodes from the 128-electrode set) were computed with low resolution electromagnetic tomography analysis (LORETA, (Pascual-Marqui et al., 1994)) in the 600 ms EEG time series, rendering a solution every 4 ms, and using the NEURONICTM Source Localizer® and Tomographic Viewer® programs (Neuronic Inc., Havana, Cuba). The newborn MRI image used as template for the source analysis was created by Kazemi and collaborators (Kazemi et al., 2007a; Kazemi et al., 2007b) and pre-processed with the NEURONICTM iMagic Pro® software (Neuronic Inc., Havana, Cuba) for extracting the surface of the head and fitting the electrodes to this surface. An isotropic and piecewise homogeneous three-sphere head model with 6,368 (generators) voxels inside the newborn's brain with a resolution of 4 mm, was used for obtaining the electric lead field (Riera and Fuentes, 1998). LORETA solutions were first calculated for each condition in every individual subject, and these were averaged across subjects to find the mean LORETA solution for each condition.

RESULTS

Figure 1A illustrates the grand-average event-related potentials (ERPs) in response to the frequent unfamiliar voice, the rare mother's voice and the resulting subtraction (mismatch negativity or MMN). Permutations (Lage-Castellanos et al.) (Figure 1C) showed two statistically significant time windows (an early pre-attentional component and a later, presumably cognitive component) in which

the response to the mother's voice differed from the frequent unfamiliar voice. The mother's voice elicited a significantly greater response than the frequent unfamiliar voice from 204 ms [$t(1,15)=2.050$; $p<.05$] to 284 ms [$t(1,15)=1.994$; $p<.05$] and then again from 364 ms [$t(1,15)=2.039$; $p<.05$] until 548 ms [$t(1,15)=2.880$; $p<.05$], peaking at 524 ms [$t(1,15)=5.261$; $p<.05$]. These results were significant on all three midline electrodes (Fz, FCz and Cz) where MMN is found to be of maximum amplitude (Beauchemin et al., 2006; Ilvonen et al., 2004; Naatanen et al., 2004).

Figure 1B illustrates the grand-average ERPs in response to the frequent unfamiliar voice, the rare unfamiliar voice (the research nurse) and the resulting MMN. As shown on Figure 1C, the rare unfamiliar voice also tended to be different from the frequent unfamiliar voice at around 176 ms [$t(1,15)=1.587$; $p>.05$], but this difference never reached significance. Permutations analyses did reveal a significant late component highly similar to that found for the mother's voice: the rare unfamiliar voice was significantly different from the frequent unfamiliar voice between 428 ms [$t(1,15)=1.991$; $p<.05$] and 528 ms [$t(1,15)=2.236$; $p<.05$], peaking at 464 ms [$t(1,15)=2.657$; $p<.05$]. At the maximum peaking amplitude, a much larger t-value obtained in the mother's voice condition ($t=5.261$ versus $t=2.657$ for the stranger's voice condition) implies that this late component is more robust in response to the mother's (0.8 μV) than to the stranger's voice (0.3 μV).

Additional permutations analyses were performed to compare the mother's voice MMN with the stranger's (Figure 1C). Statistically significant time windows were found to closely match those of previous permutations computed on both rare voice conditions relative to the frequent one. The mismatch evoked by the mother's voice differed from that of the stranger's starting at 176 ms [$t(1,15)=1.935$, $p<.05$] until 224 ms [$t(1,15)=2.075$, $p<.05$], with a maximal differentiation at 204 ms [$t(1,15)=2.369$, $p<.05$]. Permutations yielded a second significant time window from 504 ms

[t(1,15)=2.310, p<.05] to 544 ms [t(1,15)=1.921, p<.05], with a maximal peak at 528 ms [t(1,15)=2.573, p<.05]. All infants showed this specific pattern as the amplitude of their response to their mother's voice and to the nurse's voice remained within two standard deviations from the mean according to z-score calculations.

Insert Figure 1 about here

We then compared the total amplitude of brain activity elicited by both MMNs. BrainVision-assisted t-tests were computed on amplitude topographies during the two time windows showing maximal differences between the responses evoked by the mother's and the stranger's voice (Figure 2). At the early latency (i.e. at 204 ms), the mother's voice elicited greater activation over several fronto-temporal electrode sites [t.05(1,15)>2.1] that correspond to the MMN topographical distribution reported in adults (Beauchemin et al., 2006). At the late latency (i.e. at 528 ms), the mother's voice also produced greater activity over several brain regions, particularly in the right hemisphere [t.05(1,15)>2.1].

Insert Figure 2 about here

For the very first time, source analyses were computed on the newborn brain using a neonatal atlas template provided by Kazemi and collaborators (Kazemi et al., 2007a; Kazemi et al., 2007b). The mean LORETA solutions corresponding to the newborns' response to the mother's voice and to the stranger's voice are shown in Figure 3. We illustrated source distribution at 4 specific time points: 1) 100 ms post-stimulus presentation (just prior to significant differentiation of mother's vs. stranger's voice); 2) 200 ms (during the time window that revealed a maximal significant difference between the mother's vs. stranger's voice); 3) 300 ms (time point that just follows the end of the significant time window); and 4) 525 ms (time point where the mother's voice is most different from the stranger's in this late component).

Insert Figure 3 about here

Looking at average source distributions (Fig. 3), we found that mother's voice processing initially activated the left posterior temporal lobe (100 ms), an area particularly involved in language comprehension in the adult brain. This source generation was maintained in the left temporal areas during the time window that revealed a significant difference (200 ms) and even subsequently (300 ms). Source generators then shifted to right central regions at the second time point where a significant difference was found (525 ms). As for the processing of the stranger's voice, right temporal lobe processing was rather found in its early stage (100 ms), followed by a brief left temporal processing (200 ms). Source generation then switched back to the right temporal lobe (300 ms) and finally, in a similar way as the response to the mother's voice, the newborn's response to the stranger's voice was later found over right central sites (525 ms).

Include Table 1 here

Table 1 indicates the percentage of newborns having their maximal source distribution in different regions, namely temporal right, temporal left, central right and others, for each time point illustrated in Figure 3. In response to the mother's voice, the most frequent maximal activations were Temporal Left (100 ms, as in Fig. 3), Temporal Left (200 ms, as in Fig. 3), Central Right (300 ms, vs Temporal Left in Fig. 3) and Central Right (525 ms, as in Fig. 3). In response to the stranger's voice, the most frequent maximal activations were Temporal Right (100 ms, as in Fig. 3), Temporal Right (200 ms, vs Temporal Left in Fig. 3), Temporal Right (300 ms, as in Fig. 3) and Temporal Left (525 ms, vs Central Right in Fig. 3). Hence, when the rare voices were compared to the standard voice, the mother's voice elicited, in general, a maximal left temporal activation until 525 ms where a right central activation was observed. In contrast, the stranger's voice elicited right temporal activations until 525 ms where a left temporal activation was observed. Although differing slightly from grand averages provided in Figure 3, these percentages further confirm the distinct processing of both, mother's and stranger's, voices in terms of left / right hemisphere activation. Differences between

these percentages (Table 1) and grand averages (Figure 3) are observed since most babies show a source distribution as illustrated in Figure 3, albeit not a maximal one.

To better support the singularity of left dominant brain activation pattern in response to the mother's voice, we have looked at the maximal LORETA solutions at every sample point (every 4 ms) for a total of 149 samples. When looking at the lateralization of those solutions, we found that the mother's voice activated the left hemisphere in 116 samples (77.85%) and the right hemisphere in 33 samples (22.14%). The stranger's voice activated the left hemisphere in 61 samples (40.93%) and the right hemisphere in 87 samples (59.06%).

DISCUSSION

Our findings suggest that, shortly after birth, newborns process their mother's voice differently and more actively than that of strangers at both an early and late stage of processing. The time course of activation sources revealed mother's-voice specific brain activation patterns. Indeed, the mother's voice was found to be preferentially processed in the left temporal lobe at early latencies before activating central right regions. Since the voice stimulus was the pronounced vowel /a/, activation of the left temporal lobe could suggest that exposure to the maternal voice elicited language-relevant stimulus processing. In contrast, the stranger's voice elicited predominant right temporal lobe voice-specific response (Belin et al., 2002) prior to stimulating central right brain areas. In addition to providing further support to previously demonstrated tuning of specialized voice processing brain areas for familiar voices (Beauchemin et al., 2006), these results also suggest that this tuning is functional from birth, or at least within the first 24 hours following birth.

General perceptual and cognitive abilities may account for infants' distinctive mother's voice processing. The first distinction we found in the ERPs waveform is thought to reflect the mismatch process, commonly known as a pre-attentive sensory memory detection of changes in a sound stream, arguing in favour that sleeping neonates are forming representations of specific stimuli and distinguishing between them during sleep (Alho et al., 1998; Cheour et al., 1998a; Cheour et al., 1998b; Cheour et al., 2002b; Sambeth et al., 2008).

The MMN amplitude usually increases with an increasing acoustic difference between deviant and standard stimuli (Tiitinen et al., 1994). However, Cheour and collaborators (1998b) have demonstrated that brain memory traces for speech sounds override a greater acoustical difference between two stimuli: MMN amplitudes are larger in response to a native vowel than they are for a non-native one, even if that non-native vowel is acoustically more different from the standard stimulus. This assumption is also supported by previous work done in our laboratory showing greater cortical activity in response to a familiar voice when compared to that of an unfamiliar voice (Beauchemin et al., 2006). According to the contention that long-term memory traces exert a marked impact on auditory short-term memory, and therefore on MMN elicitation (Beauchemin et al., 2006; Huotilainen et al., 2001; Naatanen et al., 1997), our results also provide preliminary evidence that long-term memory is efficient from birth.

In terms of polarity, the MMN obtained in response to the mother's voice tends to be positive when compared to that obtained in response to the nurse's voice, which in the contrary tends to be negative. Kushnerenko and collaborators (2000) have proposed that, in infants, a response of positive polarity might be an analogue of the adult P3a component, indexing an involuntary attention switch to the deviant stimuli. These authors suggest that, because it emerges at the same latency as

the MMN in response to a deviant stimulus, the large amplitude P3a can mask the MMN. Earlier findings have also demonstrated that the P3a amplitude increases as a function of the magnitude of stimulus change (Yago et al., 2001) and we have shown that it is particularly modulated by voice familiarity in adults (Beauchemin et al., 2006). We therefore suggest that the MMN in response to the deviant mother's voice might have been partly overlapped by the subsequent positivity, making it positive in polarity. This also suggests that the familiar mother's voice elicit an involuntary attention deployment but that the less familiar nurse's voice fails to do so, arguing further in favour of the specificity of the response to the mother's voice.

Even if monolingual French speakers were used in the current study, it is well known that infants are able to discriminate almost all phonetic contrasts (Eimas et al., 1971; Streeter, 1976; Werker et al., 1981). Many acoustical features can differentiate between two voices, even pronouncing the same vowel: the fundamental frequency of phonation, for example. But even when f0 are similar between two voices, other acoustical cues such as formant frequencies (related to the vowel being spoken, but also to the speaker's vocal tract), or aspects of voice quality such as harmonicity (that can be measured by the harmonics-to-noise ratio) and time/frequency irregularities (captured by measures such as jitter or shimmer) contribute to the perceived differentiation of the two voices.

This being said, because of the many potential candidate acoustical features for the distinction between the mother's voice and other voices, it is unclear at present which one(s) contribute to the effect we report. This important question can be addressed in the future by using acoustical manipulations of the different voices used such that they are equated along one or more acoustical features (e.g., equating voices in both f0 and first formant frequencies, using morphing).

Then it will be possible to examine which of the different potential acoustical cues to voice identity are being used by the newborn's brain. But this question is beyond the scope of the present study that focuses on the first report of cerebral response to voice identity differences in the newborn.

A second later difference was also found which, we postulate, reveals a higher-level cognitive process. Sleeping neonates are able to process external information actively (Fifer et al., 2010) and authors have demonstrated their ability to learn, even when asleep (Cheour et al., 2002b; Fifer et al., 2010). In fact, Sambeth and collaborators (2008) have shown that newborns process structural aspects of language while sleeping. The distinct, left-dominant brain activation pattern that was found in the present study for the vowel /a/ when pronounced by the newborn's mother as opposed to right-dominant brain response for that of strangers may shed some light on the uniqueness of the mother's voice for language acquisition in infants. Regardless of culture, language and speech are acquired quickly and seemingly without effort. Social interaction assists language learning as revealed by Kuhl and collaborators (2003) in their study comparing live social interactions with televised foreign-language learning. Social interaction might play a more significant role in early language development than previously thought (Kuhl, 2007). In both speech perception and speech production, the presence of a human being interacting with a child has a strong influence on learning (Kuhl, 2004). Cases of children raised in social isolation have proven the severe and negative impact of social deprivation on language development, to the extent that normal language skills are never acquired (Kuhl, 2004; Kuhl et al., 2005).

If language and speech acquisition is learning- and environment-dependent, then there must be shared neural systems that oversee perception and action, "mirror systems" (Kuhl and Rivera-Gaxiola, 2008; Kuhl and Meltzoff, 1996). Although it was known that prenatal exposure to native-

language prosody influences newborns' perception (Mehler et al., 1988; Moon et al., 1993), Mampe and collaborators (2009) have recently shown a tendency for infants to utter melody contours similar to those perceived prenatally. These data are suggesting that not only did infants memorize the main intonation patterns of their respective surrounding language, but also that they were able to reproduce these patterns in their own production. Similarly, Chen and collaborators (2004) have demonstrated that newborns can perform corresponding mouth movements to both vowel and consonant vocal models. These authors believe that it would be more plausible and parsimonious to account for their findings in terms of a unified underlying intermodal mapping, especially considering that there was no difference in the performance of matching mouth movements between infants who closed their eyes and those who had their eyes open. Overall, their data are suggesting that newborns can map perceived sounds onto corresponding mouth movements, even if they have not seen these mouth movements in others, arguing in favour that infants possess some kind of auditory-articulatory map from birth (Kuhl and Meltzoff, 1996).

Our finding of subsequent significant central right brain activation to the mother's voice, and to a lesser extent to the stranger's voice, is compatible with the implication of premotor/supplementary motor areas as underlying neurobiological substrates of the late motor production component of the previously proposed innate auditory-articulatory map (Kuhl and Meltzoff, 1996). In sum, the singularity of left-dominant brain activation pattern to the mother's voice together with its ensuing sustained, greater right central activation could very well provide the first neurophysiologic index of language acquisition occurring through imitation, a process which would be particularly linked to the special mother-infant interaction.

REFERENCES

Alho K, Connolly JF, Cheour M, Lehtokoski A, Huotilainen M, Virtanen J, Aulanko R, Ilmoniemi RJ. 1998. Hemispheric lateralization in preattentive processing of speech sounds. Neurosci Lett. 258:9-12.

Anders TF. 1979. Night-waking in infants during the first year of life. Pediatrics. 63:860-864.

Beauchemin M, De Beaumont L, Vannasing P, Turcotte A, Arcand C, Belin P, Lassonde M. 2006. Electrophysiological markers of voice familiarity. Eur J Neurosci. 23:3081-3086.

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

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

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

Chen X, Striano T, Rakoczy H. 2004. Auditory-oral matching behavior in newborns. Dev Sci. 7:42-47.

Cheour-Luhtanen M, Alho K, Sainio K, Rinne T, Reinikainen K, Pohjavuori M, Renlund M, Aaltonen O, Eerola O, Naatanen R. 1996. The ontogenetically earliest discriminative response of the human brain. Psychophysiology. 33:478-481.

Cheour M, Alho K, Ceponiene R, Reinikainen K, Sainio K, Pohjavuori M, Aaltonen O, Naatanen R. 1998a. Maturation of mismatch negativity in infants. Int J Psychophysiol. 29:217-226.

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

Cheour M, Ceponiene R, Leppanen P, Alho K, Kujala T, Renlund M, Fellman V, Naatanen R. 2002a. The auditory sensory memory trace decays rapidly in newborns. Scand J Psychol. 43:33-39.

Cheour M, Leppanen PH, Kraus N. 2000. Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. Clin Neurophysiol. 111:4-16.

Cheour M, Martynova O, Naatanen R, Erkkola R, Sillanpaa M, Kero P, Raz A, Kaipio ML, Hiltunen J, Aaltonen O, Savela J, Hamalainen H. 2002b. Speech sounds learned by sleeping newborns. Nature. 415:599-600.

DeCasper AJ, Fifer WP. 1980. Of human bonding: newborns prefer their mothers' voices. Science. 208:1174-1176.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. 1971. Speech perception in infants. Science. 171:303-306.

Fifer WP, Byrd DL, Kaku M, Eigsti IM, Isler JR, Grose-Fifer J, Tarullo AR, Balsam PD. 2010. Newborn infants learn during sleep. Proc Natl Acad Sci U S A. 107:10320-10323.

Huotilainen M, Kujala A, Alku P. 2001. Long-term memory traces facilitate short-term memory trace formation in audition in humans. Neurosci Lett. 310:133-136.

Ilvonen T, Kujala T, Kozou H, Kiesilainen A, Salonen O, Alku P, Naatanen R. 2004. The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). Neurosci Lett. 366:235-240.

Kazemi K, Grebe R, Moghaddam AH, Lagadec P, Gondry-Jouet C, Wallois F. 2007a. Design of a digital phantom of the neonatal brain. Conf Proc IEEE Eng Med Biol Soc. 2007:5509-5512.

Kazemi K, Moghaddam HA, Grebe R, Gondry-Jouet C, Wallois F. 2007b. A neonatal atlas template for spatial normalization of whole-brain magnetic resonance images of newborns: preliminary results. Neuroimage. 37:463-473.

Kisilevsky BS, Hains SM, Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang Z. 2003. Effects of experience on fetal voice recognition. Psychol Sci. 14:220-224.

Kuhl P, Rivera-Gaxiola M. 2008. Neural substrates of language acquisition. Annu Rev Neurosci. 31:511-534.

Kuhl PK. 2004. Early language acquisition: cracking the speech code. Nat Rev Neurosci. 5:831-843.

Kuhl PK. 2007. Is speech learning 'gated' by the social brain? Dev Sci. 10:110-120.

Kuhl PK, Andruski JE, Chistovich IA, Chistovich LA, Kozhevnikova EV, Ryskina VL, Stolyarova EI, Sundberg U, Lacerda F. 1997. Cross-language analysis of phonetic units in language addressed to infants. Science. 277:684-686.

Kuhl PK, Coffey-Corina S, Padden D, Dawson G. 2005. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. Dev Sci. 8:F1-F12.

Kuhl PK, Meltzoff AN. 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. J Acoust Soc Am. 100:2425-2438.

Kuhl PK, Tsao FM, Liu HM. 2003. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc Natl Acad Sci U S A. 100:9096-9101.

Lage-Castellanos A, Martinez-Montes E, Hernandez-Cabrera JA, Galan L. False discovery rate and permutation test: an evaluation in ERP data analysis. Stat Med. 29:63-74.

Mampe B, Friederici AD, Christophe A, Wermke K. 2009. Newborns' cry melody is shaped by their native language. Curr Biol. 19:1994-1997.

Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. 1988. A precursor of language acquisition in young infants. Cognition. 29:143-178.

Moon C, Cooper R, Fifer W. 1993. Two-day-olds prefer their native language. Infant Behav Dev. 16:495-500.

Naatanen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, Allik J, Sinkkonen J, Alho K. 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 385:432-434.

Naatanen R, Syssoeva O, Takegata R. 2004. Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. Psychophysiology. 41:660-663.

Novak GP, Kurtzberg D, Kreuzer JA, Vaughan HG, Jr. 1989. Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. Electroencephalogr Clin Neurophysiol. 73:295-305.

Ockleford EM, Vince MA, Layton C, Reader MR. 1988. Responses of neonates to parents' and others' voices. Early Hum Dev. 18:27-36.

Pascual-Marqui RD, Michel CM, Lehmann D. 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int J Psychophysiol. 18:49-65.

Querleu D, Lefebvre C, Titran M, Renard X, Morillion M, Crepin G. 1984. [Reaction of the newborn infant less than 2 hours after birth to the maternal voice]. J Gynecol Obstet Biol Reprod (Paris). 13:125-134.

Riera JJ, Fuentes ME. 1998. Electric lead field for a piecewise homogeneous volume conductor model of the head. IEEE Trans Biomed Eng. 45:746-753.

Sambeth A, Huotilainen M, Kushnerenko E, Fellman V, Pihko E. 2006. Newborns discriminate novel from harmonic sounds: a study using magnetoencephalography. Clin Neurophysiol. 117:496-503.

Sambeth A, Ruohio K, Alku P, Fellman V, Huotilainen M. 2008. Sleeping newborns extract prosody from continuous speech. Clin Neurophysiol. 119:332-341.

Streeter LA. 1976. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. Nature. 259:39-41.

Tiitinen H, May P, Reinikainen K, Naatanen R. 1994. Attentive novelty detection in humans is governed by pre-attentive sensory memory. Nature. 372:90-92.

van Dommelen WA. 1990. Acoustic parameters in human speaker recognition. Lang Speech. 33 (Pt 3):259-272.

Werker JF, Gilbert JH, Humphrey K, Tees RC. 1981. Developmental aspects of cross-language speech perception. Child Dev. 52:349-355.

TABLE

	TIME	REGION			
		TEMP LEFT	TEMP RIGHT	CENT RIGHT	OTHER
MOTHER	100	**43**	18	25	12
	200	**50**	18	25	6
	300	25	25	**43**	6
	525	18	25	**50**	6
STRANGER	100	37	**43**	0	18
	200	25	**37**	18	18
	300	**31**	**31**	18	18
	525	**56**	25	28	0

Table 1: Percentage of newborns with maximal LORETA solution in different regions

The highest percentage for the 4 different latencies illustrated in Figure 3 is indicated in bold.

FIGURES

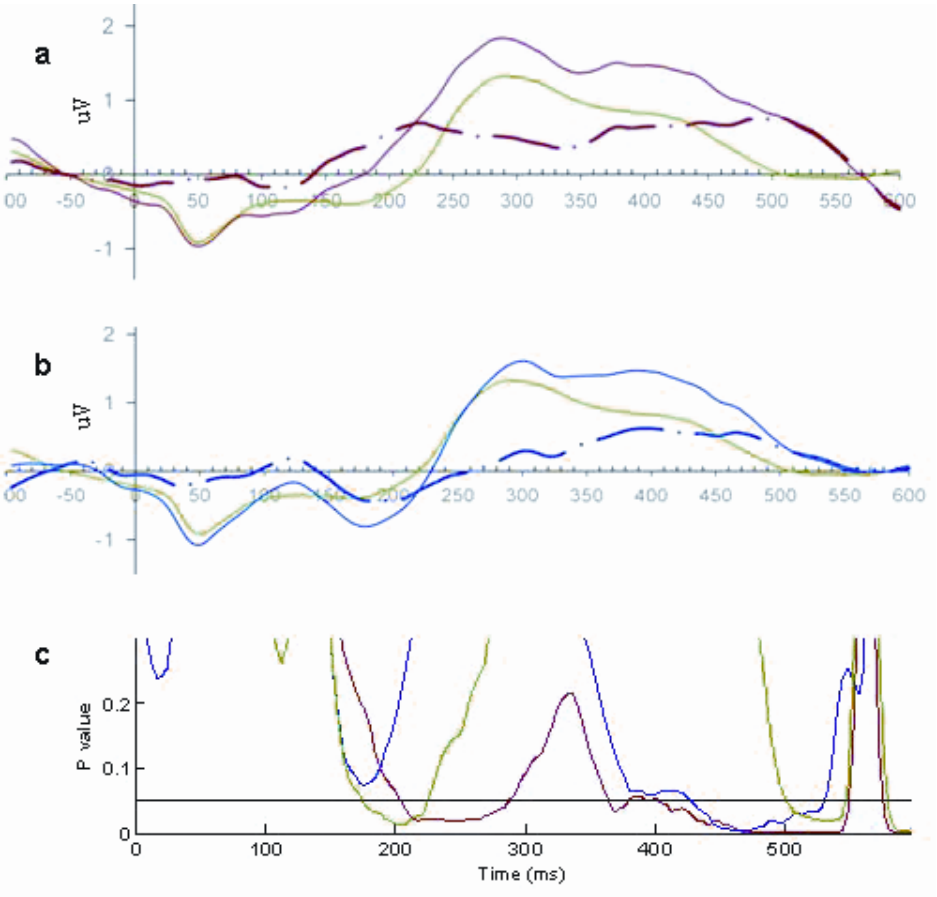


Figure 1: Event-related brain potentials and their correspondent permutations analyses.

(A) Grand-average event-related potentials in response to the frequent unfamiliar voice (green), the rare mother's voice (red) and the resulting subtraction (MMN, dashed). For purposes of clarity, only the response recorded over FCz is represented here but similar results were obtained on all electrodes that are usually activated during MMN processing (Fz, FCz, Cz).

(B) Grand-average event-related potentials in response to the frequent unfamiliar voice (green), the rare stranger's voice (blue) and the resulting subtraction (MMN, dashed). Again, we report only the result obtained over FCz.

(C) Permutations analyses showing significant time windows (below significance level of .05, horizontal line on the figure) for the mother's voice response (red), the stranger's voice response (blue) and for the comparison between both MMNs (green) on electrode FCz.

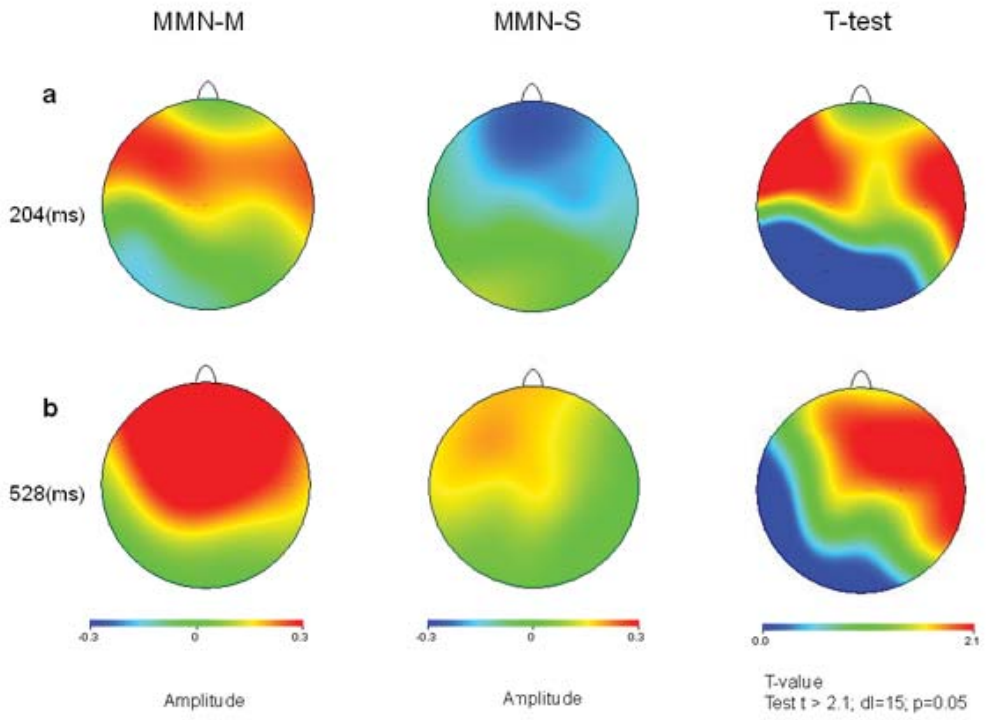


Figure 2: Brain topographies.

Time course of brain activity topographies elicited when comparing both MMNs (MMN-M: mother's and MMN-S: stranger's) in the first pre-attentional time window (A) and in the later time window (B). Note that the mother's voice elicited a much greater amplitude of activation over both time points in several electrode sites.

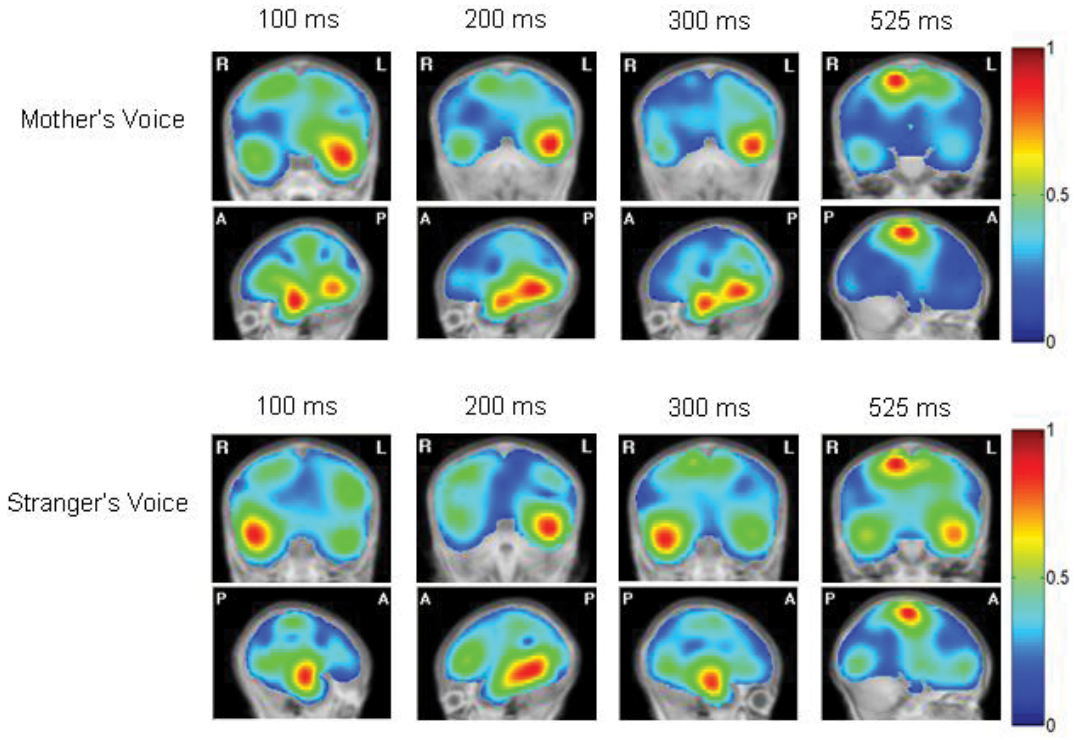


Figure 3: Source distribution.

Mean LORETA solutions (represented in coronal and sagittal planes) for the MMNs induced by the mother's and the stranger's voices at 4 different latencies. These images are normalized in order to compare just the source distribution and not the amplitudes of activation (which are presented in Figure 2). Note that the orientation of the sagittal planes - (A)nterior / (P)osterior - might switch in order to illustrate the lateralization of the maximal source distribution (in the corresponding hemisphere).

ARTICLE 3 :

THE ADULT BRAIN RAPIDLY DISTINGUISHES BETWEEN VOCAL AND NON-VOCAL STIMULI

Soumis à Neuropsychologia (2011)

The adult brain rapidly distinguishes between vocal and non-vocal stimuli

Maude Beauchemin[a,b], Berta González-Frankenberger[a,b], Pascal Belin[c,d] and Maryse Lassonde[a,b]

[a] Centre de Recherche, Centre Hospitalier Universitaire Sainte-Justine, 3175 Chemin de la Côte-Sainte-Catherine, Montréal, Québec, Canada, H3T 1C5
[b] Centre de Recherche en Neuropsychologie et Cognition (CERNEC), Département de psychologie, Université de Montréal, 90 Avenue Vincent d'Indy, Montréal, Québec, Canada, H2V 2S9
[c] Voice Neurocognition Laboratory, Institute of Neuroscience and Psychology, College of Medical, Veterinary and Life Sciences, University of Glasgow, 58 Hillhead Street, Glasgow, Scotland, G12 8QB
[d] International Laboratories for Brain, Music and Sound (BRAMS), Université de Montréal & McGill University, 1420 Boulevard Mont-Royal, Montréal, Québec, Canada, H3C 3J7

Correspondence:
Dr. Maryse Lassonde, Ph.D.
Département de psychologie
Université de Montréal
C.P.6128, succursale Centre-Ville
Montréal, Québec
Canada, H3C 3J7
Phone number :
Fax number :
E-mail :

Abstract

Neuroimaging studies have provided evidence for specific and localized processing of human voice, but the time course of this processing remains poorly understood. Using electrophysiology, previous work has reported that human voices evoked a positive component peaking around 200 ms, distinct from that elicited by other sound categories, which was termed the Fronto-Temporal Positivity to Voices (FTPV). However, it remains unclear whether this FTPV component is sufficiently robust to be elicited at a pre-attentive level, i.e. when subjects attend to another stimulation. The present study investigated neural responses to sound sequences while subjects were presented with rare deviant vocal and non-vocal sounds in a sequence of otherwise identical pure tone standards while watching a silent movie. We observed a clear FTPV at a latency of around 200 ms, thus replicating previous work, and we also found that human voices elicited significantly different ERPs morphologies, as early as 80 ms post-onset at midline electrodes. Our results indicate a robust voice-selective brain response and provide the first evidence of pre-attentive cortical specialization to voices.

Keywords: Event-related potential; Voice; Timbre; Fronto-Temporal Positivity to Voices (FTPV); N1; and P3a.

1.1 Introduction

Our ability to analyze and categorize information contained in voices plays a key role in human social interactions as the human voice is the carrier of speech (Belin, Fecteau, & Bedard, 2004). However, as suggested by the above-mentioned authors, there is more to voice than simply speech. Voice would also be an "auditory face" that conveys important affective and identity information (Belin, Zatorre, & Ahad, 2002). Thus, the sound of the human voice becomes significant irrespective of its phonetic valence. Indeed, the ability to process the paralinguistic characteristics of the human voice is important for speaker identification (Van Dommelen, 1990). Voice timbre also carries important cues about the gender, status and affective state of the speaker and we are endowed with the ability to extract such "paralinguistic" information in voices (Belin, et al., 2004). Thus, in order to do so, the brain first has to differentiate whether or not an acoustic harmonic structure was produced by a human vocal apparatus. Several studies have looked at neural specificity of voice processing.

Clinical evidence and various experimental findings have provided strong support for the contention that voice is special to the human brain. For decades, human lesion evidence (Van Lancker, Cummings, Kreiman, & Dobkin, 1988) supported the existence of specific bilateral temporal lobe mechanisms for voice analysis. More recently, functional imaging was used to investigate the neural processing of voice information. Belin and colleagues (2000) have provided robust evidence that the brain contains several regions that are sensitive and selective to voices. They have demonstrated, using fMRI, that the superior temporal sulcus (STS) was, bilaterally, more activated by human vocalisations (speech, such as isolated words, and non speech, such as laughs and coughs) than by other sound categories (such as animal cries or mechanical sounds), and that this voice-sensitive response was not entirely due to the presence of speech. Although Belin and colleagues

(2000) were among the first to demonstrate that neurons in a region of the temporal cortex, the upper bank of the STS, are selectively activated by voices, there seems to be a consensus in the literature on the topic (Belin, 2006; Belin & Zatorre, 2003; Belin, et al., 2002; Fecteau, Armony, Joanette, & Belin, 2004; Kriegstein & Giraud, 2004; Samson et al., 2001; von Kriegstein, Eger, Kleinschmidt, & Giraud, 2003). These regions were also found to be species-specific as they elicited stronger responses to human vocalizations compared to non-human vocalizations (Fecteau, et al., 2004).

Although the spatial resolution of the fMRI is appealing, this technique does not provide precise information regarding the time course of voice processing. Therefore, it is important to complement the neuroimaging evidence with measures providing better temporal resolution, such as event-related potentials (ERPs), which have shown to be useful to highlight changes in cortical activity when auditory information is presented. However, very few electrophysiological studies on voice processing have been conducted to date. Although not openly studying vocal/non-vocal discrimination, previous researches have suggested early correlates of voice processing, such as effects of voice familiarity (Beauchemin et al., 2006), speaker identity (Titova & Naatanen, 2001), voice priming (Schweinberger, 2001), voice gender adaptation (Zaske, Schweinberger, Kaufmann, & Kawahara, 2009), and speech vs. tones (Tiitinen, Sivonen, Alku, Virtanen, & Naatanen, 1999). Levy, Granot and Bentin (2001) precisely sought to characterize the ERPs elicited by non-linguistic vocal stimuli when compared to fundamental-frequency matched musical instrument sounds. They have identified a conspicuous positive component, which they termed Voice Specific Response (VSR), displaying differential processing of human voices. Peaking at about 320 ms post stimulus onset, this ERP component was prominent in response to human voices. However, the late appearance of the VSR makes it unlikely that it reflects basic perceptual processes. Undeniably, the 320 ms latency needed to show a differential response to voice as opposed to non-vocal stimuli, as measured by the

VSR, is at odd with the well-established 170 ms taken by the occipitotemporal cortex to distinguish between faces and non-face objects (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Indeed, 320 ms seems to be a considerable time to show a differential response to such a biologically important sound category, as voices. Thus, one might reasonably make the hypothesis that components differentially sensitive to vocal and non-vocal information might be observed at earlier latencies. Further, the polarity of the VSR component, its latency and frontal scalp distribution are similar to those of the Novelty P3 – also referred to as P3a – component elicited by outstanding distracters in an oddball paradigm (Comerchero & Polich, 1999; Friedman, Cycowicz, & Gaeta, 2001; Grillon, Courchesne, Ameli, Elmasian, & Braff, 1990; Katayama & Polich, 1996). Conducting a second experiment, Levy and collaborators (2003) have demonstrated the modulatory effects of attention on the VSR component.

More recently, one study has investigated the speed of voice processing by measuring ERPs in response to voices, bird songs and environmental sounds (Charest et al., 2009) and has identified an earlier electrophysiological component, termed the Fronto-Temporal Positivity to Voices (FTPV), emerging as early as 164 ms after stimulus onset and peaking around 200 ms. This interesting finding provides evidence for the existence of an electrophysiological counterpart of the STS activation observed in fMRI. Such a specific and early electrophysiological response, reflecting perceptual processing of vocal stimuli, concurs with the latency of the face-preferential N170 ERP component as well as with the time course of various above-mentioned voice processing correlates (between 150 and 200 ms) (Beauchemin, et al., 2006; Schweinberger, 2001; Tiitinen, et al., 1999; Titova & Naatanen, 2001; Zaske, et al., 2009). Previous work in our laboratory has demonstrated that healthy adults and even newborn infants can discriminate and recognize a familiar voice from an unfamiliar one (Beauchemin, et al., 2006; Beauchemin et al., 2010). Those results, obtained while

using passive oddball paradigms suggest some degree of pre-attentive voice familiarity evaluation. If voice familiarity processing, which requires the detection of a familiar voice and its retrieval from long-term memory, is pre-attentionnal, then, following basic principles of logic and hierarchy, voice processing alone, which requires the discrimination of voices among other sound stimuli, should also be pre-attentive.

However, is this FTPV attention-dependent as the VSR proved to be? The issue of whether vocal/non-vocal discrimination is a pre-attentive cortical process raises an interesting challenge: if subjects are not attending to a sound sequence, how can we find out whether the latter differentiation has occurred? One approach is to measure neural responses to sound sequences during a passive oddball listening paradigm in which rare deviant sounds are presented in a sequence of otherwise identical standards, while subjects attend to a different sensory modality. Rogier and collaborators (2010) have conducted an experiment where children were exposed, while watching a silent video, to auditory sequences of (1) vocal sounds in which non-vocal stimuli were interspersed; and (2) non-vocal sounds in which vocal stimuli were interspersed, all stimuli being extracted from Belin et al.'s (2002; 2000). These authors have evidenced, analyzing standard vocal and non-vocal sounds, a voice effect indicating early specialization of brain mechanisms involved in voice processing (Rogier, et al., 2010). In the present study, we further are wondering if such a specific response to voices could similarly be obtained analyzing rare infrequent vocal and non-vocal stimuli both presented equiprobably in an oddball paradigm. Indeed, measuring neural responses to sound sequences while listening to a passive oddball paradigm in which rare deviant vocal and non-vocal sounds are presented in a sequence of otherwise identical pure tone standards, one can then obtain a correlate of cortical differential processing without requiring participants to respond to the stimuli. It would be undeniably interesting to investigate if robust neural responses elicited by rare vocal stimuli would still

differ from those elicited in response to rare non-vocal stimuli, as it would emphasize the selectivity of the FTPV component (even being elicited by a limited number of stimuli) over and above enlightening if voice processing is really pre-attentive.

1.2 Material and Methods

1.2.1 Participants

Eighteen healthy adult subjects (age 19-23, nine males and nine females), all right-handed, gave written informed consent to participate in this study. All participants explicitly showed an exogenous stimulus-related response (N1 component) in their ERP waveform that reflects normal auditory system functioning. Furthermore, they all reported normal audition and no neurological problem.

1.2.2 Stimuli and Procedure

Stimuli were composed of 36 different stimuli (12 pure tones, 12 piano tones and 12 adult male voices pronouncing the vowel /a/). Each pure tone was matched to a piano tone and a voice in fundamental frequency therefore creating 12 different triads of same-pitch stimuli (f0 ranging between 98 Hz and 245 Hz). Stimuli were edited to a 212 ms duration and normalized in energy (RMS). The procedure was inspired from previous work in our laboratory (Beauchemin, et al., 2006; Beauchemin, et al., 2010). Stimuli were thus presented in a pseudorandom oddball sequence composed of three different stimuli from a pitch-matched triad: (a) a standard pure tone (85% of the total number of presented stimuli), (b) an infrequent natural timbre non-vocal piano tone (7.5% of the total number of presented stimuli), and (c) a second infrequent stimulus (7.5% of the total number of presented stimuli) being a natural timbre human voice. Thus deviant stimuli only differed from the standard in perceived timbre. Each participant was randomly presented with 6 out of 12 triads of stimuli, each of

which had its own exemplars of voice, piano tone and pure tone. A cerebral mismatch process could therefore be triggered by the two infrequent stimuli in an automatic comparison process with the neuronal memory trace left by the repetitive pure tone. Stimuli were presented at 75 dB sound pressure level (SPL) at the participant's head so that loudness would not be responsible for any difference in brain activation. The stimulus onset asynchrony (time interval between the onsets of two consecutive stimuli) was 800 ms, thus keeping the inter-stimulus interval (ISI) at 588 ms. Any infrequent stimulus was always preceded by at least three standard stimuli to increase the likelihood that a neural trace for the standard stimulus had formed (Matuoka et al., 2006).

The experiment took place in the Sensory and Cognitive Electrophysiology Laboratory of the Sainte-Justine's University Hospital Center. Participants first had to carefully read and sign a consent form providing detailed information about the experimental procedure, which was approved by the Ste-Justine Hospital's ethics committee. Participants were rewarded with a financial compensation of $30 for taking part in our study.

Participants were seated in a comfortable chair located in a semi-obscure Faraday room. Six blocks of 400 stimuli were presented for a total of 2400 stimuli (2040 standard pure tone stimuli, 180 infrequent natural timbre non-vocal piano tone stimuli and 180 infrequent natural timbre human voice stimuli) and for a total duration of 30 minutes. Each block had its own exemplars or triad of stimuli. The sequence of stimuli was generated by the E-Prime Psychology Software on a DELL computer located in an adjacent room. Stimuli were binaurally delivered through speakers, positioned 20 cm left and right from the participant's head, at auricular height. Cortical responses were acquired with a high-density recording system, the Geodesic 128-Sensor Net (Electrical Geodesics Inc., Eugene, OH) at a sampling rate of 250 Hz, an analog band-pass filter of 0.1-100 Hz, and Cz referenced.

Impedance was kept below 50 kΩ, which is within recommended range when using a powerful amplifier such as the one used in this study (Net Amps 200). Time-locked evoked potentials were recorded with the Net Station program on a Macintosh G4 computer. Because cortical responses vary with awareness state, subjects were required to stay awake throughout the whole testing session. To increase the likelihood that they would not pay attention to the presented sequence of stimuli during testing, all participants were instructed to focus their attention on a silent subtitled movie of their choice presented on a monitor at a distance of about 2 meters. Participants were monitored through an infrared camera equipped with an integrated speaker, allowing them to communicate at all times with the experimenters located in the adjacent room.

1.2.3 Data Analyses

EEG recordings were analyzed off-line using the BrainVision Analyzer program (Brain products, Munich, Germany) on an IBM computer. Various pre-processing filters were applied on the data, such as a raw data inspector, ocular correction, artefacts rejection and baseline correction. Trials contaminated by EOG and/or EEG artefacts were excluded from the average by an automatic rejection algorithm with threshold amplitude of ±100 μV. Midline electrodes (AFz, Fz and FCz) were specifically analysed as the cerebral mismatch process has been found to be topographically distributed in frontocentral regions (Beauchemin, et al., 2006; Ilvonen et al., 2004; Naatanen, Syssoeva, & Takegata, 2004; van Zuijen, Sussman, Winkler, Naatanen, & Tervaniemi, 2005; Ylinen, Shestakova, Huotilainen, Alku, & Naatanen, 2006). Electrodes FC5 and FC6 were also analysed as a specific cerebral response to voices has previously been recorded among those electrodes (Charest, et al., 2009). Electrodes were further referenced to both left and right linked mastoids and digitally filtered with a frequency band-pass of 1-30 Hz at 24 dB/octave. ERPs resulted from averaging EEG epochs of 800 ms starting 200 ms prior to stimulus onset. Average waveforms were then computed

for each subject independently for each stimulus type (standard pure tone, infrequent piano tone and infrequent human voice). No ERP averages were based on less than 90 trials. A semiautomatic peak detection was performed on selected electrodes, namely Afz, Fz, FCz, FC5 and FC6. The amplitude of the most positive peak between 100 and 200 ms was measured on two selected fronto-temporal electrodes (FC5 and FC6). As for the midline electrodes, the time windows in which electrodes were peaked were determined by looking at the averaged ERPs curves of all subjects. The amplitude of the most positive peak between 50 and 150 ms after stimulus onset, the most negative peak between 100 and 200 ms and the most positive peak between 175 and 300 ms were measured on electrodes Afz, Fz and FCz. Difference between cortical responses to vocal and non-vocal stimuli were assessed by Student's t-tests comparing the amplitude of several pre-determined peaks on electrodes of interest.

1.3 Results

The purpose of the present study was to assess if specific cortical activity in response to voices could also be elicited using a passive oddball paradigm in which vocal and non-vocal sounds are used as equiprobable rare deviant stimuli. Figure A.1 illustrates grand average waveforms across all 18 subjects at two selected fronto-temporal electrodes (FC5 and FC6) showing preferential cortical response to voices as opposed to non-vocal stimuli as previously reported in Charest and collaborators (2009). Indeed, cortical activity in response to voices is of greater amplitude when compared to non-vocal stimuli. This component, that we also refer to as the FTPV, peaks at 182 ms in the right hemisphere [$t(1,17) = 5.260$, $p<0.01$] and at 188 ms in the left hemisphere [$t(1,17) = 5.153$, $p<0.01$], with no significant difference in latency.

Figure A.2 illustrates grand average waveforms across all 18 subjects at three midline electrode sites (AFz, Fz and FCz), in response to standard pure tones, infrequent natural timbre non-vocal piano tones and infrequent natural timbre human voices. By only looking at the waveforms, one can see that human voices elicited greater cortical activity in the 100 ms latency range and similarly, around the 200 ms latency range. On the contrary, smaller cortical activity was recorded in response to voices in the 130 ms latency range. Statistically, infrequent human voices first elicited a positive component that peaks earlier [AFz: $t(1,17) = 2.933$, $p<0.01$; Fz: $t(1,17) = 2.893$, $p<0.05$; and FCz: $t(1,17) = 2.689$, $p<0.05$] and is of greater amplitude [AFz: $t(1,17) = 3.915$, $p<0.01$; and Fz: $t(1,17) = 2.541$, $p<0.05$] when compared to that in response to infrequent non-vocal piano tones. This positive component peaks on midline electrodes between 80 and 86 ms in response to voices whereas it peaks between 90 and 97 ms in response to piano tones. Moreover, human voices elicited significant smaller N1-related activity [AFz: $t(1,17) = 3.557$, $p<0.01$; and Fz: $t(1,17) = 3.721$, $p<0.01$] that rose substantially earlier (129-132 ms) than that of piano tones (132-141 ms) [AFz: $t(1,17) = 2.650$, $p<0.05$; and Fz: $t(1,17) = 2.283$, $p<0.05$]. Finally, human voices elicited a second greater positive component [AFz: $t(1,17) = 4.143$, $p<0.01$; Fz: $t(1,17) = 4.339$, $p<0.01$; and FCz: $t(1,17) = 9.021$, $p<0.01$], namely the P3a, which was elicited similarly in terms of latency by both deviant rare stimuli (vocal: 194-199 ms and non-vocal: 192-200 ms). Brain activity topographies elicited when hearing human voices as opposed to piano tones illustrate the above-mentioned observed trends (Figure A.3).

1.4 Discussion

Event-related potentials were recorded in 18 healthy adults listening to a passive oddball paradigm in order to investigate whether voice processing is a pre-attentive cortical process in addition to assess the robustness of this processing.

On the one hand, our data replicate previous findings of a greater fronto-temporal positivity in responses to voices (FTPV) (Charest, et al., 2009; Rogier, et al., 2010). Indeed, analysing only rare vocal and non-vocal stimuli, we observed significantly larger ERP amplitudes to voices compared to piano tones at temporal electrodes between 182 and 188 ms, suggesting a robust voice selective brain response that does not need to be elicited by the repetitive presentation of frequent vocal stimuli. Since the FTPV component is still strongly generated by the scarce presentation of vocal stimuli interspersed in a sequence of standard non-vocal stimuli and that several examplars of voices, piano tones and pure tones have been presented, one can suppose that this component is very sensitive and selectively elicited by voices.

On the other hand, unlike Charest and collaborators (2009), we have further recorded differential ERP responses to human voices at midline electrodes as early as 80 ms. The distinction between our results and data from previous work can probably be accounted for by differences in experimental designs. In fact, in order to explore the putative pre-attentive facet of voice processing, we needed to trigger a cerebral mismatch process usually obtained from a passive oddball paradigm, which has been found to be topographically distributed in fronto-central regions (Beauchemin, et al., 2006; Ilvonen, et al., 2004; Naatanen, et al., 2004; van Zuijen, et al., 2005; Ylinen, et al., 2006). This triggered cerebral mismatch process, extracted from analysing rare deviant vocal and non-vocal sounds presented in a sequence of otherwise identical pure tone standards and resulting in a correlate of cortical differential processing without requiring participants to respond to the stimuli, provides the first evidence of pre-attentive cortical specialization to voices. Indeed, not only do we know that “temporal voice areas” (TVA) (Belin, et al., 2000) are sensitive but also highly selective to voices, but we have now identified that voice processing occurs at a pre-attentive level. Our results are thus additionally stating that the auditory cortex needs far less than the already-proposed 320 ms

taken by the VSR component (Levy, et al., 2001, 2003) to show a distinctive response to voices as opposed to other sound categories. Indeed, not only did we find quite different P1 and N1 components morphologies in response to voices and piano tones, but voices actually elicited P1 and N1 components that peaked earlier when compared to those of piano tones, suggesting rapid sensory discrimination between both sound categories (Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006). We thus agree with Rogier and collaborators (2010) in saying that the processing of voices and other sound categories (in the present case, piano tones) might be subtended by neural networks differentially activated from the initial stages of sound processing.

Although previous studies have shown an attention effect on N1 amplitude (Hillyard, Woldorff, Mangun, & Hansen, 1987; Naatanen & Picton, 1987) with attended stimuli generating larger N1 components when compared to unattended stimuli, both of our deviant stimuli (voices and piano tones) are unattended and presented in a passive oddball paradigm. Therefore, the difference in N1 amplitude between both of our deviant conditions cannot be explained by such an attentional effect. However, we believe that both of our deviant stimuli (voices and piano tones) triggered involuntary attention switching to the stimulus train. We thus suppose that the conspicuous positive component following the N1, which reflects the process of involuntary attention switching itself and which proved to be larger in response to voices, might have influenced the amplitude of the preceding N1, leaving a smaller N1 in response to voices. Another plausible explanation accounting for the N1 amplitude difference is that, generally speaking, the difference in N1 amplitude elicited by various deviant stimuli is known to vary as a function of acoustic distance between each deviant and standard stimuli (Deguchi et al., 2010). Indeed, even though all presented triads of stimuli have been composed from pitch-matched stimuli, all normalized in energy and edited to the same duration, stimuli contained within a triad still differed in terms of timbre, which is a critical factor in our study. We thus believe that

the smaller N1 amplitude in response to voices could be accounted for by acoustical difference in timbre, reflecting perhaps acoustical distance or cruder voice/non voice distinction. One could also suppose that the timbre of piano tones might be more complex than that of voices, although this would need further support.

We have further identified a positive component peaking just prior to 200 ms post stimulus onset that stands significant in terms of amplitude when elicited by voices compared to non-vocal stimuli. The polarity of this component, its latency as well as its mainly frontal activation reflect the presence of an early Novelty P3 component elicited by outstanding distracters in an oddball paradigm. The Novelty P3 (usually referred to as P3a) is considered to reflect unintentional brain response to salient stimuli. A P3a component is evoked when a stimulus is (a) outstanding, hence novel (Grillon, et al., 1990); (b) rare as opposed to the other distracters (Katayama & Polich, 1996); or (c) easy to distinguish from the other distracters (Comerchero & Polich, 1999). Within the context of our experimental paradigm, human voices would appear to be both outstanding and easier to distinguish from other auditory stimuli, be they pure tones or piano tones, thereby eliciting greater P3a amplitude. Thus, our findings indicate that voices are sufficiently compelling stimuli to distract listeners from their ongoing activity. Indeed, the P3a component identified in our study could very be similar and analogous to the VSR component reported by Levy et al. (2001, 2003), indicative of the capture of attention and reflecting the shift of attention to the respective eliciting stimulus category, namely voices. In concrete terms, this orientation response would thereafter benefit not only subsequent phonological processing, but also parallel processes necessary for speaker identification.

Differential processing of voices thus seems to occur as early as 80 ms post-stimuli presentation and is also extending to later latencies. Apart from a greater activation elicited by voices,

there is no evidence of differences in topographies of electric fields at the scalp and by extension, of the configuration of active brain areas, although localization source analyses would be further needed to fully support this hypothesis.

Our results highlight why normal listeners can effortlessly extract, evaluate and categorize vast amounts of information contained in voices (Belin, et al., 2004; Belin & Zatorre, 2003; Belin, et al., 2002; Fecteau, et al., 2004; Van Dommelen, 1990; Warren, Jennings, & Griffiths, 2005; Warren, Scott, Price, & Griffiths, 2006). Knowing that voices are processed at a pre-attentional level would seem rational given that voices are one of the most biologically important stimuli, playing a key role in social interaction, which is at the heart of human life. They are our acoustical signature, carrying our unique sound characteristic patterns. From a phylogenetic perspective, cerebral voice processing has a long evolutionary history and is not specific to humans (Petkov et al., 2008; Petkov, Logothetis, & Obleser, 2009). Humans have presumably been subjected to similar evolutionary pressure as non-human primates to develop mechanisms specialized in accurately extracting information in voice (Belin, 2006). Indeed, with evidence of acquired deficits restricted to human voice perception but not to speech came suspicions of human brain mechanisms selectively dedicated to the extraction of paralinguistic information in voice (Belin, et al., 2004; Van Lancker, et al., 1988). Furthermore, a growing body of evidence argue in favour that voice cognition and perception abilities appear early in in human development (Beauchemin, et al., 2010; Friederici, 2005; Kisilevsky et al., 2003; Ockleford, Vince, Layton, & Reader, 1988). Ontogenetically speaking, voice is perceived as early as during foetal life (Kisilevsky, et al., 2003) and it is known that newborn infants clearly discriminate between their own mother's voice from other voices (Beauchemin, et al., 2010). Furthermore, Grossmann and collaborators (2010) have recently demonstrated that temporal regions specialize in processing voices very early in development, by the age of 7 months. We therefore become "expert" at voice

processing, having devoted parts of our brain processing voices. The expertise that humans have developed for rapid processing of stimuli with strong social significance such as voices is supported by our data.

Future work should attempt to determine if the current approach could be used to assess vocal/non-vocal discrimination in infants. A similar paradigm indeed proved to be helpful in evaluating the processing of voice familiarity in newborns (Beauchemin, et al., 2010). Studying voice perception abilities in infants and young children may contribute to the early diagnosis and treatment of voice communication impairments, such as autism.

1.5 Appendices

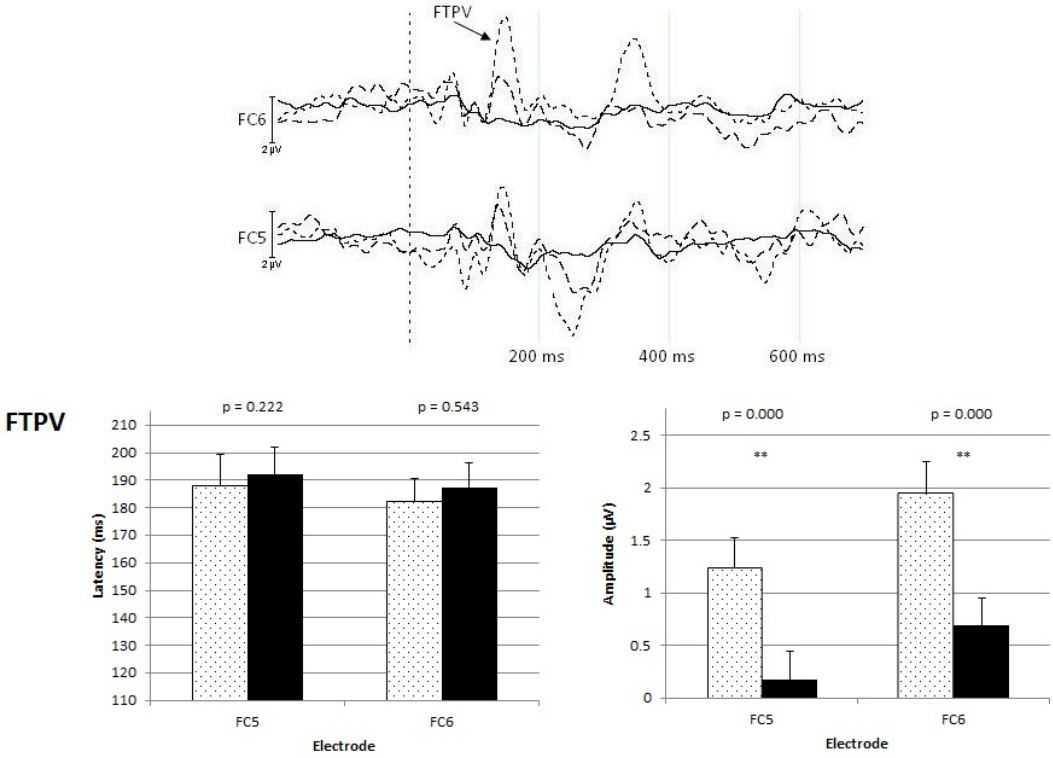


Figure A.1 Grand average waveforms at FC5 and FC6: The upper panel illustrates grand-averaged ERPs elicited by human voices (dotted) and by piano tones (dashed), referenced to the standard pure tones stimuli (black) at two temporal electrodes, namely FC5 and FC6. The lower panel shows quantitative differences between ERPs elicited by voices (dotted) and piano tones (black), in terms of latency (left) and amplitude (right).

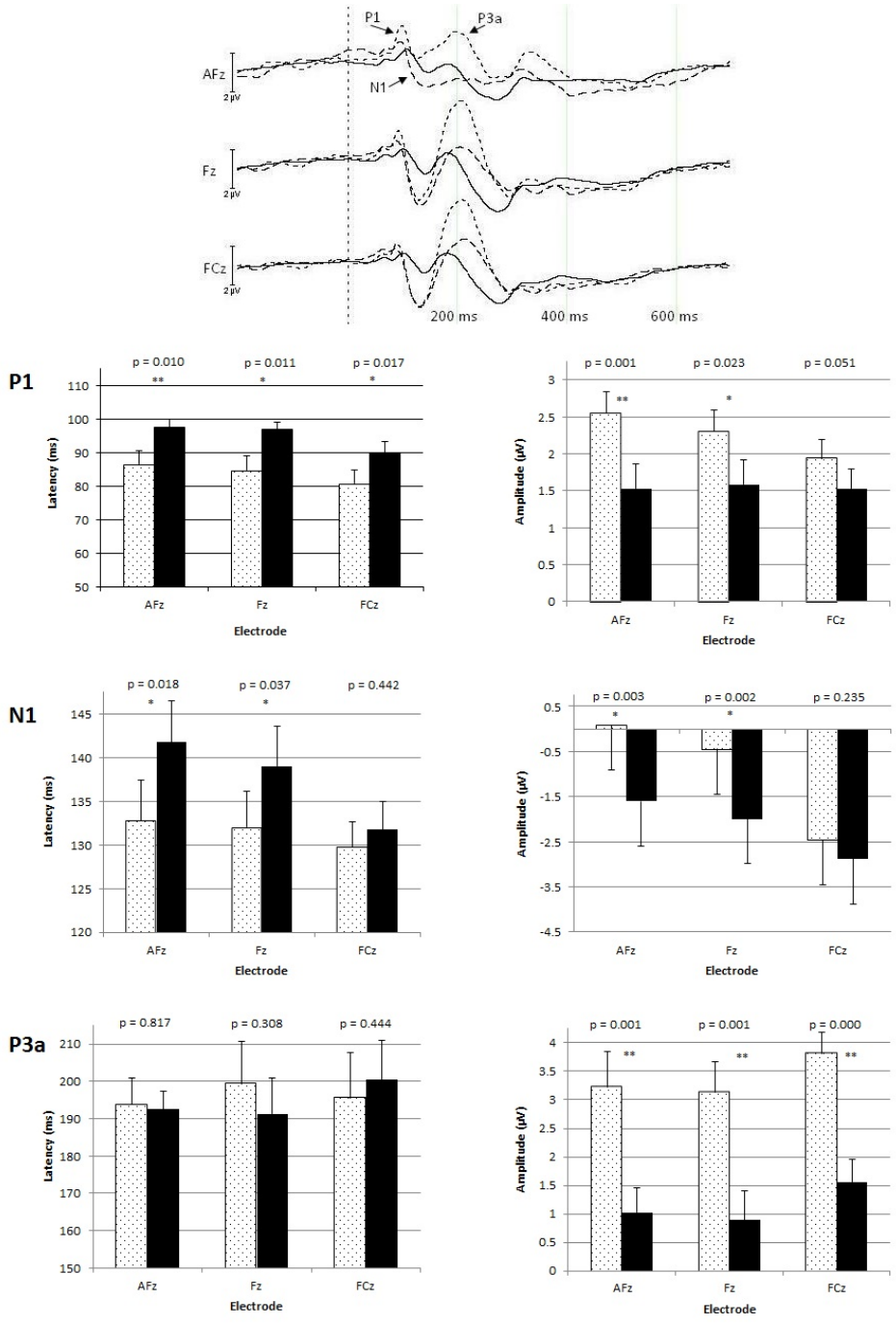


Figure A.2 Grand average waveforms at AFz, Fz and FCz: The upper panel illustrates grand-averaged ERPs elicited by human voices (dotted) and by piano tones (dashed), referenced to the standard pure tones stimuli (black) at three midline electrodes, namely AFz, Fz and FCz. The lower panels show quantitative differences between three ERPs components (P1, N1 and P3a) elicited by voices (dotted) and piano tones (black), in terms of latency (left) and amplitude (right).

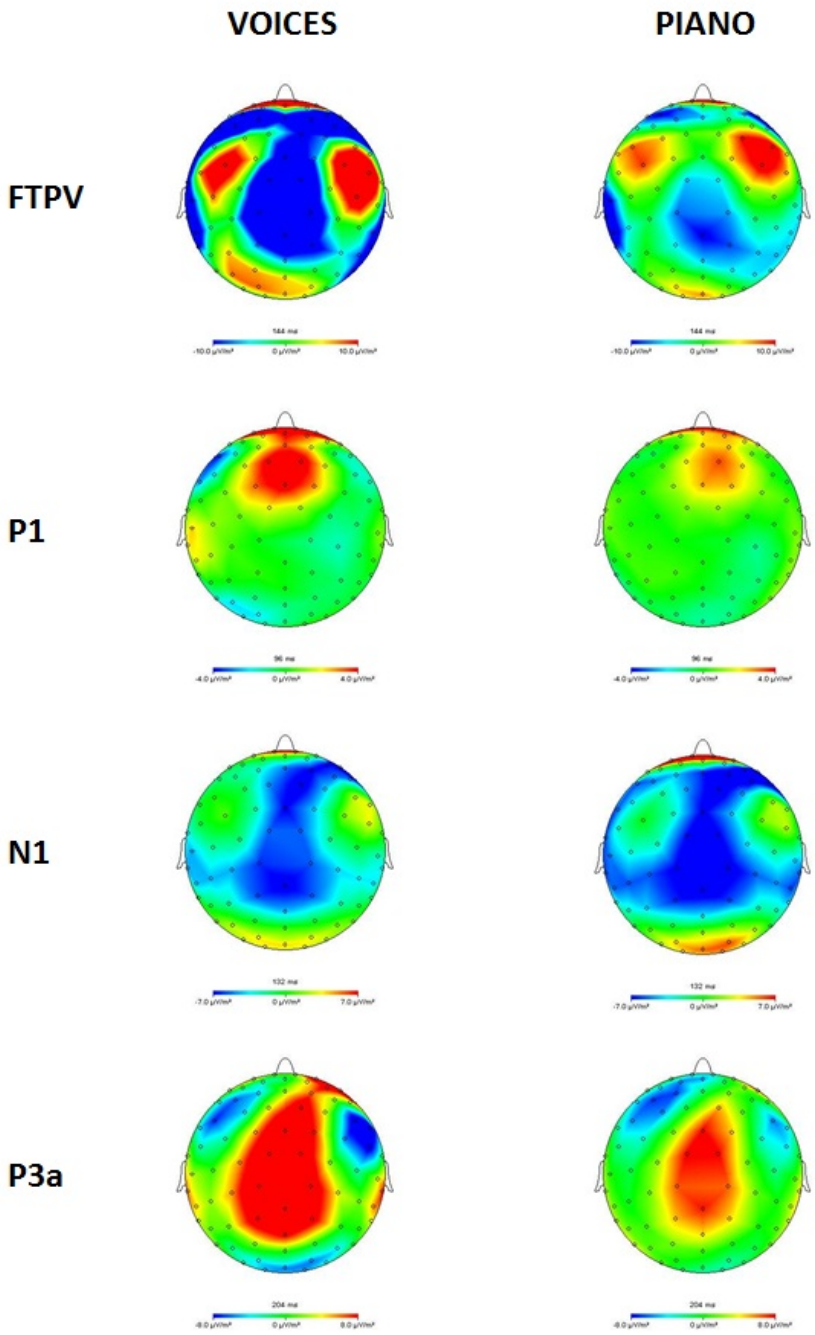


Figure A.3 Brain activity topographies: Brain topographies elicited either by human voices (left) or piano tones (right) underlying our four .ERP components, namely FTPV, P1, N1 and P3a.

1.6 References

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., et al. (2006). Electrophysiological markers of voice familiarity. Eur J Neurosci, 23(11), 3081-3086.

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., et al. (2010). Mother and Stranger: An Electrophysiological Study of Voice Processing in Newborns. Cereb Cortex.

Belin, P. (2006). Voice processing in human and non-human primates. Philos Trans R Soc Lond B Biol Sci, 361(1476), 2091-2107.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. Trends Cogn Sci, 8(3), 129-135.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport, 14(16), 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res, 13(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature, 403(6767), 309-312.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. J Cogn Neurosci, 8(6), 551-565.

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

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. Clin Neurophysiol, 110(1), 24-30.

Deguchi, C., Chobert, J., Brunelliere, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. Brain Res, 1366, 149-161.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage, 23(3), 840-848.

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. Trends Cogn Sci, 9(10), 481-488.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci Biobehav Rev, 25(4), 355-373.

Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., & Braff, D. (1990). Effects of rare non-target stimuli on brain electrophysiological activity and performance. Int J Psychophysiol, 9(3), 257-267.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. Neuron, 65(6), 852-858.

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

Ilvonen, T., Kujala, T., Kozou, H., Kiesilainen, A., Salonen, O., Alku, P., et al. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). Neurosci Lett, 366(3), 235-240.

Katayama, J., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. Int J Psychophysiol, 23(1-2), 33-40.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. Psychol Sci, 14(3), 220-224.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage, 22(2), 948-955.

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

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. Psychophysiology, 40(2), 291-305.

Matuoka, T., Yabe, H., Shinozaki, N., Sato, Y., Hiruma, T., Ren, A., et al. (2006). The development of memory trace depending on the number of the standard stimuli. Clin EEG Neurosci, 37(3), 223-229.

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid Brain Discrimination of Sounds of Objects. The Journal of Neuroscience, 26(4), 1293-1302.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology, 24(4), 375-425.

Naatanen, R., Syssoeva, O., & Takegata, R. (2004). Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. Psychophysiology, 41(4), 660-663.

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. Early Hum Dev, 18(1), 27-36.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. Nat Neurosci, 11(3), 367-374.

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? Neuroscientist, 15(5), 419-429.

Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. Int J Psychophysiol, 75(1), 44-47.

Samson, Y., Belin, P., Thivard, L., Boddaert, N., Crozier, S., & Zilbovicius, M. (2001). [Auditory perception and language: functional imaging of speech sensitive auditory cortex]. Rev Neurol (Paris), 157(8-9 Pt 1), 837-846.

Schweinberger, S. R. (2001). Human brain potential correlates of voice priming and voice recognition. Neuropsychologia, 39(9), 921-936.

Tiitinen, H., Sivonen, P., Alku, P., Virtanen, J., & Naatanen, R. (1999). Electromagnetic recordings reveal latency differences in speech and tone processing in humans. Brain Res Cogn Brain Res, 8(3), 355-363.

Titova, N., & Naatanen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. Neurosci Lett, 308(1), 63-65.

Van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. Lang Speech, 33 (Pt 3), 259-272.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. Cortex, 24(2), 195-209.

van Zuijen, T. L., Sussman, E., Winkler, I., Naatanen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity--a mismatch negativity study comparing musicians and non-musicians. Brain Res Cogn Brain Res, 23(2-3), 270-276.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. Brain Res Cogn Brain Res, 17(1), 48-55.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. Neuroimage, 24(4), 1052-1057.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. Neuroimage, 31(3), 1389-1397.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Naatanen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. Brain Res, 1072(1), 175-185.

Zaske, R., Schweinberger, S. R., Kaufmann, J. M., & Kawahara, H. (2009). In the ear of the beholder: neural correlates of adaptation to voice gender. Eur J Neurosci, 30(3), 527-534.

DISCUSSION GÉNÉRALE

La présente thèse s'intéressait à l'ontogénie et à la spécificité de la réponse corticale à la voix humaine. Le premier objectif était de mettre sur pied un protocole électrophysiologique permettant de mesurer objectivement le traitement de la familiarité de la voix chez le sujet adulte. Nous souhaitions ensuite utiliser ce même protocole afin de déterminer si une telle mesure pouvait aussi objectiver chez le nouveau-né de 24 heures un traitement préférentiel d'une voix familière, notamment la voix de la mère. Finalement, nous nous sommes intéressés à savoir si la discrimination entre des stimuli vocaux et non-vocaux se faisait de façon pré-attentionnelle.

MARQUEURS ÉLECTROPHYSIOLOGIQUES DE LA FAMILIARITÉ D'UNE VOIX CHEZ LE SUJET ADULTE

Dans un premier temps, admettant que notre capacité à discriminer et reconnaître des voix est socialement et biologiquement importante et qu'elle figure parmi les fonctions les plus importantes du système auditif humain, nous avons voulu déterminer si des marqueurs électrophysiologiques pouvaient être utilisés comme mesures objectives du traitement de la familiarité d'une voix chez le sujet adulte. Nous avons par conséquent mesuré deux composantes électrophysiologiques (la MMN et la P3a) en réponse à une voix familière et à une voix inconnue, toutes deux présentées en tant que stimuli rares dans le cadre d'un paradigme *oddball*. De fait, les ERPs ont fréquemment été utilisés pour mettre en évidence des changements dans l'activité corticale lorsque de l'information auditive est présentée. Plus précisément, les paradigmes *oddball* auditifs, qui impliquent la présentation de stimuli rares dispersés parmi la présentation de stimuli fréquents, ont souvent été utilisés pour générer une composante ERP connue sous le nom de MMN (Naatanen, 1995; Naatanen & Alho, 1995; Naatanen, Gaillard, & Mantysalo, 1978). Selon la théorie de Naatanen et Alho (1995), la discrimination de deux stimuli sonores différant sur un seul paramètre reflète la participation de deux représentations neuronales différentes. En d'autres termes, un

stimulus présenté à répétition forme une trace neuronale dans la mémoire sensorielle échoïque, qui pourrait persister dans le temps jusqu'à 8 à 10 secondes selon certains auteurs (Bottcher-Gandor & Ullsperger, 1992). L'entrée sensorielle d'un nouveau stimulus dont les caractéristiques sonores ne coïncident pas avec la trace neuronale laissée par le stimulus fréquent résulte donc en une déviation négative que l'on nomme MMN. Ainsi, la MMN est évoquée par toute modification discriminable d'aspects répétitifs d'une stimulation auditive stockés en mémoire sensorielle échoïque (Cowan, Winkler, Teder, & Naatanen, 1993).

Les principaux résultats de cette première étude étaient, en fait, la différence significative entre les réponses électrophysiologiques (MMN et P3a) évoquées par une voix familière par rapport à celles évoquées en réponse à une voix inconnue, ces composantes étant beaucoup plus amples en réponse à la voix familière. Ces résultats suggèrent, de façon préliminaire, que les aires corticales dédiées au traitement de la voix seraient particulièrement sensibles aux voix familières.

Bien que dans sa théorie originale, Naatanen (1990) ait suggéré que la mémoire à long terme ne soit reliée à la mémoire à court terme que par des processus conscients, dépendants de l'attention, ce chercheur, entouré de son équipe, a démontré dans une étude ultérieure que les traces mnésiques provenant de la mémoire à long terme ont un impact considérable sur la mémoire auditive à court terme et conséquemment, sur l'évocation de la MMN (Naatanen et al., 1997). De fait, ils ont démontré par le biais de leur étude que l'amplitude de la MMN était fortement influencée par le contenu de la mémoire à long terme. Plus précisément, au sein de leur étude, l'amplitude de la MMN était plus grande lorsqu'un phonème répété fréquemment était remplacé par un phonème appartenant à une langue connue (soit la langue maternelle) par opposition à un phonème provenant d'une langue étrangère, ne faisant pas partie du registre de la mémoire à long terme et pour lequel

aucune trace mnésique n'est créée en mémoire à long terme. Ainsi, la mémoire à long terme semble avoir un impact significatif sur la mémoire à court terme auditive. Nos résultats appuient aussi cette interprétation hypothétique : à savoir que la détection d'une voix familière, qui exige sa récupération en mémoire à long terme, se ferait à l'aide d'une connexion directe reliant la mémoire à long terme et le système d'analyse des caractéristiques contenu dans la mémoire à court terme auditive et utilisé pour détecter les caractéristiques acoustiques des stimuli (Huotilainen, Kujala, & Alku, 2001). En effet, puisque l'activation corticale mesurée est plus importante en réponse à la voix familière rare qu'à la voix inconnue, équiprobablement rare, elle ne peut qu'être le reflet de la formation préalable d'une trace mnésique à long terme des caractéristiques de la voix familière.

Les résultats électrophysiologiques corroborent également les données comportementales, démontrant que les sujets arrivent aussi très bien et explicitement à distinguer et reconnaître une voix familière parmi des voix inconnues. Ces résultats ne peuvent pas non plus s'expliquer et être simplement dus à de subtiles différences acoustiques entre les paramètres de la voix familière et ceux de la voix inconnue. De fait, les réponses électrophysiologiques de sujets contrôles, pour qui aucune des trois voix présentées n'était familière, ne montrent pas de différences significatives en réponse à une voix en particulier.

Rappelons aussi que les composantes MMN et P3a ont été évoquées en dépit du fait que les sujets avaient pour consigne de ne pas prêter attention aux stimuli auditifs, suggérant ainsi qu'il existe un certain degré d'évaluation pré-attentionnelle de la familiarité de la voix. La présente étude identifiait par ailleurs une composante électrophysiologiquement positive culminant juste avant 300 ms suivant la présentation des stimuli. La polarité de cette composante, sa latence de même que son activation frontale indiquent la présence d'une P3a élicitée par la présentation de distracteurs

saillants et exceptionnels, se distinguant particulièrement des autres stimuli présentés (Comerchero & Polich, 1999; Grillon, et al., 1990; Katayama & Polich, 1996). Il est généralement aussi admis que la P3a reflète la réponse corticale non intentionnelle à des stimuli nouveaux. Dans le cadre de notre paradigme, la voix familière semble être un stimulus suffisamment saillant et exceptionnel, facilement distinguable des voix inconnues, qu'elles soient fréquentes ou rares, pour susciter une plus grande amplitude de la composante P3a. Une voix familière est donc un stimulus suffisamment singulier et convainquant pour distraire les auditeurs de leur activité en cours. D'ailleurs, de façon analogue, des résultats similaires ont été obtenus dans le cadre d'études portant sur la reconnaissance des visages, suggérant la présence d'un processus automatique permettant l'identification de visages connus et familiers (Bobes, Martin, Olivares, & Valdes-Sosa, 2000). De fait, le traitement des visages familiers et inconnus a souvent été utilisé pour comprendre le phénomène de la reconnaissance des visages. L'étude de Bobes et collaborateurs (2000) a révélé une composante positive tardive, qu'ils ont nommée LP pour *late positivity*, qui était de plus grande amplitude en réponse aux visages familiers lorsque comparée aux visages inconnus.

Les résultats électrophysiologiques émanant de cette première étude confirment donc que des mécanismes neuronaux sont particulièrement sensibles à la familiarité d'une voix. Nos résultats soulignent en outre le rôle éminent joué par la voix humaine au sein de notre environnement sonore, non seulement à titre de support médiant la communication, mais également comme indice de familiarité.

ÉTUDE ÉLECTROPHYSIOLOGIQUE DU TRAITEMENT DE LA FAMILIARITÉ DE LA VOIX CHEZ LE NOUVEAU-NÉ

Sachant que des régions corticales semblent dédiées au traitement de la voix chez l'adulte (Belin, et al., 2000), que des mesures électrophysiologiques se sont montrées sensibles au traitement préférentiel de la familiarité de la voix (Beauchemin et al., 2006) et que nous connaissons très peu sur la façon dont le cerveau du nouveau-né traite les informations vocales, nous avons voulu, par le biais de cette deuxième étude, étudier, au moyen de l'électrophysiologie et des analyses de sources, le traitement cortical dédié à la voix de la mère par rapport à celui réservé au traitement d'une voix étrangère chez le nouveau-né de 24 heures. De fait, des études comportementales ont montré que les fœtus et les nouveau-nés savent reconnaître et réagissent de façon différente et préférentielle à la voix de leur mère (DeCasper & Fifer, 1980; Kisilevsky, et al., 2003; Ockleford, et al., 1988; Querleu, et al., 1984). Toutefois, la neurophysiologie de l'interaction mère-enfant demeure mal comprise comme aucune étude n'a jamais examiné les réponses corticales des nouveau-nés à la voix de leur mère. Nous avons donc mené une étude qui se voulait exploratoire, mais qui stipulait tout de même que les composantes électrophysiologiques que nous comptions mesurer (MMN et P3a) seraient, tout comme elles se sont montrées chez l'adulte, de plus grandes amplitudes en réponse à une voix familière, dans le cas présent, à la voix de la mère, lorsque comparées à celles évoquées par une voix inconnue. Par ailleurs, afin de déterminer quelles sont les régions du cerveau qui produisent l'activité électrophysiologique que nous allions mesurer, nous avons procédé à des analyses de sources. Nous postulions que les générateurs de l'activité mesurée en électrophysiologie se situeraient probablement au sein des régions dédiées au traitement de la voix chez l'adulte, notamment le cortex temporal droit.

Les résultats émanant de cette deuxième étude suggèrent que, peu après la naissance, les nouveau-nés traitent la voix de leur mère de manière différente et plus active que celle d'une étrangère et ce, à la fois à un stade précoce du traitement de l'information sonore que vers la fin du processus de traitement. Le décours temporel des sources d'activation a, quant à lui, révélé des patrons d'activation corticaux propres à la voix de la mère. En effet, la voix de la mère serait d'abord préférentiellement traitée au sein du lobe temporal gauche pour ensuite activer les régions centrales droites. Étant donné que le stimulus utilisé était la voyelle /a/, l'activation du lobe temporal gauche en réponse à la voix de la mère pourrait de ce fait suggérer que l'exposition à la voix maternelle suscite un traitement analogue à celui utilisé pour traiter le langage. En revanche, la voix étrangère aurait d'abord été traitée de façon prédominante au sein du lobe temporal droit, région spécifique dédiée au traitement de l'information vocale (Belin, et al., 2002), avant de stimuler les régions centrales droites à leur tour. En plus de soutenir davantage que les aires corticales spécialisées dans le traitement de la voix sont particulièrement sensibles aux voix familières (Beauchemin, et al., 2006), ces résultats suggèrent également que cette sensibilité accrue pour les voix familières est fonctionnelle dès la naissance, ou tout au moins dans les premières 24 heures suivant la naissance.

Les aptitudes générales perceptives et cognitives du nouveau-né peuvent rendre compte du traitement distinct réservé à la voix maternelle. En effet, la première distinction que l'on retrouve entre le tracé électrophysiologique en réponse à la voix de la mère et celui évoqué par la voix étrangère se situe au niveau du processus de discordance (*mismatch*), qui fait communément référence à la détection d'un changement dans un flot sonore de stimuli répétitifs, détection qui prend forme au sein de la mémoire sensorielle pré-attentionnelle. Ce processus de discordance, mis davantage en valeur lorsque la voix de la mère est traitée tel qu'en témoigne l'amplitude de la MMN, fait valoir que les nouveau-nés forment bel et bien des représentations neuronales des

caractéristiques spécifiques des stimuli entendus et sont à même de distinguer deux stimuli se différenciant par leurs caractéristiques sonores et ce, même pendant leur sommeil (Alho, et al., 1998; Cheour, Alho, et al., 1998; Cheour, Ceponiene, et al., 1998; Cheour et al., 2002; Sambeth, Ruohio, Alku, Fellman, & Huotilainen, 2008).

Il est généralement admis que l'amplitude de la MMN augmente en fonction de l'écart croissant qui existe entre les caractéristiques acoustiques des stimuli sonores déviants et standards (Tiitinen, May, Reinikainen, & Naatanen, 1994). Toutefois, Cheour et collaborateurs (1998) ont su démontrer avec élégance que les traces mnésiques contenues en mémoire à long terme pour les sons de la parole arrivaient à surpasser une plus grande différence acoustique entre deux stimuli. En d'autres termes, dans le cadre de leur étude, l'amplitude de la MMN s'est montrée plus grande en réponse à une voyelle faisant partie du registre de la langue première/maternelle d'un nouveau-né qu'elle ne s'est montrée en réponse à une voyelle appartenant au registre d'une langue étrangère, même si cette voyelle « étrangère » se distinguait davantage du stimulus standard en termes de ses caractéristiques acoustiques. Cette hypothèse, de l'importante contribution des traces mnésiques existantes, est également soutenue par nos travaux antérieurs montrant une plus grande activité corticale en réponse à une voix familière lorsque comparée à celle d'une voix inconnue (Beauchemin, et al., 2006). Selon la théorie prônant que les traces en mémoire à long terme exercent un impact marqué sur la mémoire à court terme auditive, et par conséquent sur l'évocation de la MMN (Beauchemin, et al., 2006; Huotilainen, et al., 2001; Naatanen, et al., 1997), les présents résultats fournissent de surcroît des données préliminaires suggérant que la mémoire à long terme serait fonctionnelle dès la naissance.

En termes de polarité, la MMN obtenue en réponse à la voix de la mère est davantage positive que celle obtenue en réponse à la voix étrangère, qui elle tend à être davantage négative. Certains auteurs ont proposé que, chez le nouveau-né, une réponse de polarité positive s'apparenterait à la composante P3a retrouvée chez l'adulte, indexant un changement involontaire de l'attention vers une stimulation déviant du standard (Kushnerenko, Ceponiene, Balan, Fellman, & Naatanen, 2002). Selon ces auteurs, la MMN serait évoquée chez les nouveau-nés à une latence comparable à celle de la P3a (i.e. plus tardivement que chez l'adulte), et l'amplitude, souvent très grande, de la composante P3a masquerait ou viendrait empiéter sur l'amplitude de la MMN, la rendant ainsi davantage positive. D'ailleurs, il a été démontré que l'amplitude de la P3a augmentait en fonction de l'ampleur de la différence séparant deux stimuli (Yago, Corral, & Escera, 2001) et nous avions démontré lors de notre première étude que l'amplitude de la P3a était aussi modulée par la familiarité de la voix, du moins chez le sujet adulte (Beauchemin, et al., 2006). Nous suggérons donc que la composante MMN en réponse à la voix maternelle déviante aurait été en partie chevauchée par la positivité subséquente de la composante P3a (plus importante en réponse à la voix de la mère), la rendant ainsi de polarité positive. Ceci suggère par ailleurs que la voix de la mère suscite un déploiement involontaire de l'attention, ce que la voix étrangère ne parvient pas autant à faire, faisant ainsi, une fois de plus, valoir la spécificité de la réponse à la voix maternelle.

Bien que les voix utilisées dans cette étude proviennent toutes d'interlocuteurs francophones unilingues, il est bien connu que les nouveau-nés sont capables de distinguer presque tous les contrastes phonétiques (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Streeter, 1976; Werker, Gilbert, Humphrey, & Tees, 1981). Beaucoup de caractéristiques acoustiques peuvent différencier entre les voix utilisées comme stimuli standards et rares, même si ces voix prononcent toutes la même voyelle (/a/) : la fréquence fondamentale de la phonation notamment. Toutefois, même lorsque

la fréquence fondamentale est similaire entre les voix, d'autres indices acoustiques, tels que les fréquences des formants (liées à la voyelle prononcée, mais aussi à l'appareil vocal de l'orateur) ou les aspects de la qualité de la voix comme l'harmonicité, les irrégularités temps/fréquence contribuent aux différences perçues entre les voix. Ceci étant dit, en raison des nombreuses caractéristiques acoustiques potentiellement responsables de la distinction acoustique entre la voix de la mère et la voix inconnue, nous ne pouvons nous prononcer quant à celle(s) ayant contribué à l'effet que nous avons mesuré. Cette intéressante question demeure à être investiguée et répondue par le biais de futures études, à l'aide de manipulations acoustiques des différentes voix, les uniformisant sur une ou plusieurs caractéristiques sonores (par exemple, les rendant équivalentes quant à leur fréquence fondamentale et la fréquence du premier formant). Il nous sera alors possible de déterminer quels sont les différents indices acoustiques utilisés par le cerveau pour permettre l'identité vocale. Cette question, bien que très pertinente et enrichissante, va au-delà de la portée des résultats émanant de cette deuxième étude dont l'emphase portait sur la première démonstration de réponses corticales à l'identité vocale chez le nouveau-né, plus particulièrement le traitement cortical dédié à la voix de la mère.

Une deuxième distinction, plus tardive, a aussi été identifiée entre le tracé électrophysiologique en réponse à la voix de la mère et celui évoqué par la voix étrangère et qui, à notre avis, dévoile un processus « cognitif » de plus haut niveau. Nous savons que les nouveau-nés sont en mesure de traiter activement l'information externe, même en dormant (Fifer et al., 2010) et certains auteurs ont même démontré certaines capacités d'apprentissage chez les nouveau-nés durant leur sommeil (Cheour, et al., 2002; Fifer, et al., 2010). De plus, Sambeth et collaborateurs (2008) ont montré que les nouveau-nés traitent les aspects structurels du langage pendant leur sommeil. Dans le cadre de cette deuxième étude, les sources d'activation provenant de façon

distinctive du lobe temporal gauche en réponse à la voyelle /a/ prononcée par la voix de la mère comparativement aux sources d'activation émanant plutôt du lobe temporal droit en réponse à la voyelle /a/ prononcée par une voix inconnue prônent en faveur de l'unicité de la voix de la mère dans l'acquisition du langage chez le nouveau-né. Il est facile de constater que peu importe la culture, le langage et la parole sont acquis rapidement et apparemment sans effort par les enfants. L'interaction sociale aiderait aussi à l'apprentissage du langage comme l'ont révélé Kuhl et collaborateurs (2003) dans leur étude comparant l'effet des interactions sociales vives dans l'apprentissage d'une langue étrangère avec l'utilisation d'un média télévisé. L'interaction sociale joue donc un rôle beaucoup plus important dans le développement du langage qu'on ne le pensait (P. K. Kuhl, 2007). Autant dans la perception que dans la production de la parole, la présence d'une personne interagissant avec un enfant a une forte influence sur l'apprentissage (P. K. Kuhl, 2004). Pensons simplement aux cas d'enfants élevés dans l'isolement social et l'impact important et négatif que cette privation sociale a sur le développement de leur habiletés langagières, dans la mesure où certaines compétences linguistiques « normales » ne sont jamais acquises (P. K. Kuhl, 2004; P. K. Kuhl, Coffey-Corina, Padden, & Dawson, 2005).

Si le développement du langage et de la parole dépendent de l'apprentissage et de l'environnement, il semble raisonnable de penser qu'il doit y avoir un certain partage des systèmes neuronaux, un certain recoupement entre les systèmes qui contrôlent la perception et l'action, ou en d'autres termes, un système miroir (P. Kuhl & Rivera-Gaxiola, 2008; P. K. Kuhl & Meltzoff, 1996). Il est connu que l'exposition prénatale à la prosodie de la langue maternelle influence la perception des nouveau-nés (Mehler et al., 1988; Moon, Cooper, & Fifer, 1993). D'ailleurs, Mampe et collaborateurs (2009) ont récemment montré que les nourrissons ont tendance à adopter des contours mélodiques similaires à ceux auxquels ils ont été exposés pendant la période prénatale. Ces données suggèrent

que non seulement les nouveau-nés parviennent à mémoriser les principaux patrons d'intonation contenus dans la langue maternelle environnante, mais ils arrivent également à les reproduire. Par ailleurs, Chen et ses collaborateurs (2004) ont démontré que les nouveau-nés produisent les mouvements de la bouche correspondant à des modèles vocaux de voyelles et de consonnes. Ces auteurs trouvent plausible et parcimonieux de concevoir leurs résultats en termes de cartographie intermodale unifiée sous-jacente, surtout compte tenu qu'il n'y avait aucune différence dans les performances d'appariement de mouvements de la bouche chez les nourrissons ayant les yeux fermés et ouverts. En somme, leurs résultats suggèrent que les nouveau-nés arrivent à transposer les sons perçus en mouvements correspondants de la bouche, n'ayant même jamais vu ces mouvements chez autrui, ce qui plaide en faveur du fait qu'ils possèderaient une carte auditivo-articulatoire fonctionnelle dès la naissance (P. K. Kuhl & Meltzoff, 1996).

Notre constat d'une activation corticale centrale droite subséquente significative en réponse à la voix de la mère, et dans une moindre mesure à la voix étrangère, suppose l'implication des aires pré-motrices et motrices supplémentaires comme substrat neurobiologique sous-jacent à la composante de production motrice faisant partie de la carte auditivo-articulatoire innée déjà proposée par Kuhl et Meltzoff (1996). En somme, la singularité du patron d'activation corticale gauche à la voix de la mère ainsi que l'activation subséquente soutenue et de plus grande amplitude en central droit pourrait très bien fournir le premier indice neurophysiologique de l'acquisition du langage se produisant à travers l'imitation, un processus qui serait particulièrement lié à l'interaction spéciale se déroulant entre une mère et son enfant.

LE CERVEAU ADULTE DISCRIMINE-T-IL RAPIDEMENT LES STIMULI VOCAUX DES STIMULI NON-VOCAUX?

Notre capacité d'analyser et de catégoriser l'information contenue dans la voix joue un rôle clé dans nos interactions sociales puisque la voix permet la parole (Belin, et al., 2004). Cependant, les auteurs que nous venons de citer sont aussi d'avis qu'il y a plus à la voix que la parole. Ils vont jusqu'à dire que la voix serait un visage auditif transmettant une panoplie d'informations importantes, notamment identitaires et affectives (Belin, et al., 2002). Ainsi, les sons vocaux sont significatifs au-delà de leur valence phonétique. En effet, la capacité à traiter les caractéristiques paralinguistiques de la voix est importante entre autres pour l'identification du locuteur (Van Dommelen, 1990). Le timbre de la voix porte également des indications importantes sur le genre, le statut et l'état affectif du locuteur et nous sommes dotés de cette capacité à extraire ces informations paralinguistiques des voix (Belin, et al., 2004). Par contre, afin d'y arriver, nous devons d'abord distinguer si oui ou non une structure acoustique harmonique entendue a été produite par un appareil vocal humain.

Notre troisième étude s'intéressait donc toujours au traitement cortical réservé à la voix, mais cette fois-ci dans un sens plus large. Les études de neuroimagerie ont fourni des preuves pour le traitement spécifique et localisé de la voix (Belin, et al., 2004; Belin, et al., 2002; Belin, et al., 2000), mais le décours temporel de ce traitement demeure mal compris. Bien que la résolution spatiale de l'imagerie par résonance magnétique soit séduisante, cette technique ne fournit pas d'informations précises concernant le décours temporel du traitement de la voix. Par conséquent, il est important de complémenter les résultats de neuroimagerie avec des mesures permettant une meilleure résolution temporelle, tels que les potentiels évoqués, qui se sont montrés utiles pour mettre en évidence des changements dans l'activité corticale lorsque de l'information auditive est présentée. Très peu d'études électrophysiologiques sur le traitement de la voix ont été menées à ce jour. Nous savons

toutefois que l'utilisation de l'électrophysiologie a démontré de l'activité neuronale, dans une fenêtre temporelle entourant les 200 ms, sensible aux différences vocale/non-vocale : composante dénommée FTPV (Charest, et al., 2009). Par contre, nous nous questionnions toujours à savoir si cette activité corticale était dépendante de processus attentionnels ou de processus automatiques, de nature pré-attentionnelle. Nous penchions effectivement en faveur de la dernière hypothèse compte tenu que nos travaux antérieurs ont démontré que les adultes et même les nouveau-nés arrivent à discriminer et reconnaître une voix familière d'une inconnue (Beauchemin, et al., 2006; Beauchemin et al., 2010). Ces résultats, obtenus en utilisant des paradigmes *oddball* passifs, suggèrent un certain degré d'évaluation pré-attentionnelle de la familiarité de la voix. Si le traitement de la familiarité d'une voix, qui nécessite la détection d'une voix familière de même que sa récupération en mémoire à long terme, est un processus pré-attentionnel, et suivant les principes de base de logique et de hiérarchie, le traitement seul d'une voix, qui nécessite la discrimination de stimuli vocaux parmi d'autres stimuli sonores, devrait également être un processus pré-attentionnel. Rogier et collaborateurs (2010) ont mis en évidence à l'aide d'un paradigme *oddball* passif un effet de voix, supposant une spécialisation précoce et pré-attentionnelle des mécanismes cérébraux impliqués dans le traitement de la voix. Rappelons que ces auteurs ont utilisé deux séquences de stimuli auditifs : (1) des stimuli vocaux à travers lesquels étaient insérés des stimuli non-vocaux, et (2) des stimuli non-vocaux à travers lesquels étaient insérés des stimuli vocaux; et qu'ils ont obtenu leur effet en analysant les stimuli standards fréquents (vocaux et non-vocaux). Nous voulions donc par le biais de cette troisième étude évaluer la robustesse de la réponse électrophysiologique. En utilisant le paradigme électrophysiologique élaboré et utilisé au sein des deux études antérieures, cette troisième étude expérimentale cherchait donc, d'une part, à confirmer le fait que la discrimination entre des stimuli vocaux et non-vocaux se fait de façon pré-attentionnelle et, d'autre part, à vérifier si une réponse électrophysiologique ne pouvait pas être évoquée en analysant

seulement les stimuli rares vocaux et non-vocaux présentés équiprobablement dans un paradigme *oddball* passif ayant comme standards une séquence de tons purs identiques. De cette façon, nous mettions l'accent sur la sélectivité et la sensibilité de la réponse corticale réservée au traitement de la voix en obtenant un corrélat de traitement cortical différentiel à un nombre plus limité de stimuli sans nécessiter que les sujets répondent aux stimuli.

D'une part les données découlant de cette troisième étude reproduisent les résultats déjà connus dans la littérature d'une plus grande positivité fronto-temporale en réponse aux stimuli vocaux (FTPV) (Charest, et al., 2009; Rogier, et al., 2010). En effet, en analysant uniquement les stimuli rares vocaux et non-vocaux, nous avons été à même d'observer des amplitudes significativement plus grandes aux électrodes temporales, entre 182-188 ms, en réponses aux voix par rapport aux sons de piano, prônant ainsi en faveur d'une réponse corticale robuste (sélective et sensible) à la voix qui n'a pas besoin d'être évoquée par la présentation répétitive de stimuli vocaux fréquents. Puisque la composante FTPV est fortement évoquée par la présentation rare et limitée de stimuli déviants vocaux insérés dans une séquence de stimuli standards non-vocaux, et que plusieurs exemplaires de voix, de sons de pianos et de tons purs ont été utilisés, tout porte à croire que cette composante FTPV est très sensible et induite de façon sélective par la voix.

D'autre part, contrairement à Charest et collaborateurs (2009), nous avons aussi enregistré des réponses différentielles à la voix au niveau des électrodes de la ligne médiane à des latences aussi précoces que 80 ms. La distinction entre nos résultats et ceux de travaux antérieurs s'explique probablement par des différences dans les paradigmes expérimentaux. En effet, afin d'explorer la facette pré-attentionnelle putative, nous devions induire un processus de discordance, d'inadéquation corticale habituellement obtenu à partir d'un paradigme *oddball* passif, dont la topographie est

généralement distribuée dans les régions fronto-centrales (Beauchemin, et al., 2006; Ilvonen et al., 2004; Naatanen, Syssoeva, & Takegata, 2004; van Zuijen, Sussman, Winkler, Naatanen, & Tervaniemi, 2005; Ylinen, Shestakova, Huotilainen, Alku, & Naatanen, 2006). Ce processus de discordance corticale déclenché, extrait de l'analyse des stimuli déviants rares vocaux et non-vocaux présentés dans une séquence de tons purs par ailleurs identiques, confirme donc que la spécialisation corticale à la voix est pré-attentionnelle. Nous savons donc maintenant que les aires temporales de la voix (TVA) (Belin, et al., 2000) sont non seulement sensibles et hautement sélectives à la voix, mais nous avons maintenant identifié que le traitement cortical réservé à la voix se produit de façon pré-attentionnelle. Nos résultats démontrent une fois de plus clairement que le cortex auditif nécessite beaucoup moins que les 320 ms déjà proposés dans la littérature (composante VSR) (Levy, et al., 2001, 2003) pour montrer une réponse distincte à la voix, par opposition à d'autres catégories sonores. De fait, nous avons non seulement identifié des morphologies différentes pour les composantes P1 et N1, mais ces composantes culminaient également plus tôt en réponse à la voix lorsque comparées à celles en réponse aux sons de piano, suggérant ainsi une discrimination sensorielle rapide entre les deux catégories sonores (Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006). Nos résultats secondent donc ceux de Rogier et collaborateurs (2010) stipulant que le traitement cortical réservé à la voix et à d'autres catégories sonores (dans le cas présent, les sons de piano) serait sous-tendu par des réseaux neuronaux activés différemment dès les premières étapes du traitement sonore.

Bien que certaines études aient montré un effet d'attention sur l'amplitude de la composante N1 (Hillyard, Woldorff, Mangun, & Hansen, 1987; Naatanen & Picton, 1987) de sorte que les stimuli auxquels on portait attention généraient de plus amples N1 par rapport aux stimuli ignorés, les deux types de stimuli déviants rares (voix et sons de piano) utilisés dans notre étude étaient ignorés et

présentés dans un paradigme *oddball* passif. Par conséquent, la différence d'amplitude que nous avons mesurée entre les N1 de nos deux conditions déviantes rares, où la N1 en réponse à la voix était de plus petite amplitude que la N1 en réponse au son de piano, ne peut s'expliquer par un tel effet attentionnel. Cependant, nous sommes d'avis que nos deux types de stimuli déviants rares (voix et sons de piano) ont bel et bien provoqué et suscité un changement involontaire de l'attention vers la stimulation auditive, tel qu'en témoigne la composante positive manifeste et visible suivant la N1. Cette composante positive, reflétant le processus de changement attentionnel, est plus importante en réponse à la voix et a donc pu influencer l'amplitude de la N1 la précédant, la rendant ainsi de plus petite amplitude en réponse à la voix. Une autre explication plausible pouvant expliquer cette différence d'amplitude de la N1 est que l'amplitude de cette composante est connue pour varier en fonction de la distance acoustique entre les stimuli rares déviants et les stimuli standards (Deguchi et al., 2010). En effet, même si toutes les triades de stimuli présentés ont été composées à partir de stimuli appariés pour leur hauteur tonale, normalisés pour leur niveau d'énergie et édités pour leur durée, les stimuli contenus dans une même triade différaient encore en termes de timbre, variable d'intérêt dans le cadre de notre étude. La plus petite amplitude de N1 en réponse à la voix pourrait donc aussi être expliquée par cette différence en timbre, reflétant sans doute une distance acoustique ou une distinction voix/non-voix plus grossière. On pourrait aussi supposer que le timbre des sons de piano est plus complexe que celui des voix, quoi que cette hypothèse nécessite davantage de preuves expérimentales.

Nous avons également identifié une composante positive culminant tout juste avant 200 ms qui se démarque significativement en termes d'amplitude lorsque suscité par la voix par rapport aux stimuli non-vocaux. La polarité de cette composante, sa latence ainsi que son activation frontale reflètent la présence d'une composante P3a, élicitée par des stimuli distracteurs dans un paradigme

oddball. Cette composante est généralement considérée comme reflétant une réponse corticale involontaire à des stimuli saillants, nouveaux, rares ou faciles à distinguer (Comerchero & Polich, 1999; Grillon, et al., 1990; Katayama & Polich, 1996). La voix semble donc être un stimulus à la fois saillant et plus facile à distinguer des autres stimuli auditifs, ayant suscité une P3a de plus grande amplitude. Nos résultats portent à croire que la voix est un stimulus suffisamment convainquant pour distraire les auditeurs de leur activité en cours. Nous croyons d'ailleurs, que la composante P3a identifiée dans notre étude serait semblable et analogue à la composante VSR (Levy, et al., 2001, 2003), soulignant la capture de l'attention et reflétant le déplacement de l'attention vers la catégorie de stimuli en question, à savoir les voix. En termes plus concrets, cette réponse d'orientation serait profitable et bénéfique pour le traitement phonologique ultérieur, tout comme pour des processus parallèles nécessaires à l'identification du locuteur notamment.

Le traitement différentiel réservé à la voix semble donc se produire dès 80 ms après la présentation du stimulus et se prolonge également à des latences plus tardives. Outre une plus grande activation induite par la voix, il n'existe aucune preuve comme quoi les topographies des champs électriques au scalp seraient différentes et, par extension, la configuration des aires corticales réellement activées, bien que des analyses de localisation de sources seraient nécessaires pour soutenir pleinement cette hypothèse.

Nos résultats mettent en évidence la raison pour laquelle les auditeurs normaux peuvent facilement extraire, évaluer et classer de grandes quantités d'informations contenues dans la voix (Belin, et al., 2004; Belin & Zatorre, 2003; Belin, et al., 2002; Fecteau, et al., 2004; Van Dommelen, 1990; Warren, et al., 2005; Warren, et al., 2006). Sachant que les voix sont traitées à un niveau pré-attentionnel semble rationnel étant donné que la voix est un des stimuli les plus biologiquement

importants, jouant un rôle clé dans les interactions sociales, qui sont au cœur de notre vie. La voix est notre signature acoustique, transportant notre patron unique de caractéristiques sonores. D'un point de vue phylogénétique, le traitement cortical de la voix a une longue histoire évolutive et n'est pas spécifique à l'Homme (Petkov et al., 2008; Petkov, Logothetis, & Obleser, 2009). Les humains ont vraisemblablement été victimes de pressions évolutionnaires similaires à celles des primates non humains pour développer des mécanismes spécialisés dans l'extraction précise d'informations contenues dans la voix (Belin, 2006). En effet, avec les évidences de déficits acquis limités à la perception de la voix et ne touchant guère la perception du langage sont venus les soupçons de mécanismes cérébraux sélectifs dédiés à l'extraction d'informations paralinguistiques de la voix (Belin, et al., 2004; Van Lancker, et al., 1988). Par ailleurs, un nombre croissant de résultats militent en faveur que les capacités perceptives et cognitives touchant la voix apparaissent tôt au cours du développement humain (Beauchemin, et al., 2010; Friederici, 2005; Kisilevsky, et al., 2003; Ockleford, et al., 1988). Ontogénétiquement parlant, la voix est perçue aussitôt que durant la vie fœtale (Kisilevsky, et al., 2003) et on sait que les nouveau-nés discriminent clairement entre la voix de leur mère de celles d'étrangères (Beauchemin, et al., 2010). Par ailleurs, Grossmann et collaborateurs (2010) ont récemment démontré que les régions temporales se spécialisent dans le traitement de la voix et ce, très tôt dans le développement, à partir de l'âge de 7 mois. Nous devenons donc, en quelque sorte, « experts » dans le traitement de la voix, possédant des parties de notre cerveau consacrées au traitement de la voix. L'expertise que nous avons développée pour traiter rapidement des stimuli à haute signification sociale, tels les voix, est secondée par nos résultats.

Les travaux futurs devraient entre autres tenter de déterminer si l'approche que nous avons employée pourrait être utilisée pour évaluer la discrimination voix/non-voix chez les nouveau-nés. Un

paradigme similaire s'est en fait révélé utile dans l'évaluation du traitement de la familiarité d'une voix chez les nouveau-nés, tel qu'en témoigne notre deuxième étude (Beauchemin, et al., 2010). Étudier les capacités de perception de la voix chez les nourrissons et les jeunes enfants pourrait éventuellement contribuer au diagnostic précoce et au traitement des troubles de la communication vocale, tel l'autisme.

CONCLUSION

L'être humain est un être social remarquable, et les mécanismes neuronaux sous-jacents qui permettent cette socialisation commencent à être étudiés et un peu mieux connus. Il existe un traitement cortical spécialisé et dédié à la voix humaine, entre autres en raison de sa prédominance dans notre environnement sonore, de son rôle général au sein des interactions humaines, et du fait qu'elle permet le langage. Notre capacité d'analyser et de catégoriser l'information contenue dans la voix joue d'ailleurs un rôle clé dans les interactions sociales humaines. La présente thèse s'intéressait à l'ontogénie et à la spécificité de la réponse corticale à la voix humaine. Elle a permis (1) de mettre sur pied un protocole électrophysiologique permettant de mesurer objectivement et de confirmer que des mécanises neuronaux sont particulièrement sensibles au traitement de la familiarité d'une voix chez le sujet adulte; (2) de mettre en lumière un patron d'activation corticale singulier à la voix de la mère chez le nouveau-né, fournissant le premier indice neurophysiologique de l'acquisition du langage, processus particulièrement lié à l'interaction mère-enfant; et (3) de confirmer l'aspect pré-attentionnel de la distinction entre une voix et un stimulus non-vocal tout en accentuant la sélectivité et la sensibilité de la réponse corticale réservée au traitement de la voix.

BIBLIOGRAPHIE

(pour l'introduction et la discussion générales)

Aitkin, L. M., Merzenich, M. M., Irvine, D. R., Clarey, J. C., & Nelson, J. E. (1986). Frequency representation in auditory cortex of the common marmoset (Callithrix jacchus jacchus). *J Comp Neurol, 252*(2), 175-185.

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc Natl Acad Sci U S A, 98*(21), 12301-12306.

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., et al. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neurosci Lett, 258*(1), 9-12.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb Cortex, 9*(5), 415-430.

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

Assal, G., Aubert, C., & Buttet, J. (1981). Asymétrie cérébrale et reconnaissance de la voix. *Rev Neurol (Paris), 137*(4), 255-268.

Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition, 94*(2), B45-53.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., et al. (2006). Electrophysiological markers of voice familiarity. *Eur J Neurosci, 23*(11), 3081-3086.

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., et al. (2010). Mother and Stranger: An Electrophysiological Study of Voice Processing in Newborns. *Cereb Cortex*.

Belin, P. (2006). Voice processing in human and non-human primates. *Philos Trans R Soc Lond B Biol Sci, 361*(1476), 2091-2107.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci, 8*(3), 129-135.

Belin, P., & Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nat Neurosci, 3*(10), 965-966.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport, 14*(16), 2105-2109.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res, 13*(1), 17-26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci, 8*(6), 551-565.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex, 10*(5), 512-528.

Bobes, M. A., Martin, M., Olivares, E., & Valdes-Sosa, M. (2000). Different scalp topography of brain potentials related to expression and identity matching of faces. *Brain Res Cogn Brain Res, 9*(3), 249-260.

Bottcher-Gandor, C., & Ullsperger, P. (1992). Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology, 29*(5), 546-550.

Bruce, V., & Young, A. (1986). Understanding face recognition. *Br J Psychol, 77 (Pt 3)*, 305-327.

Campanella, S., & Belin, P. (2007). Integrating face and voice in person perception. *Trends Cogn Sci, 11*(12), 535-543.

Caplan, D., Gow, D., & Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology, 45*(2), 293-298.

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

Chartrand, J. P., & Belin, P. (2006). Superior voice timbre processing in musicians. *Neurosci Lett, 405*(3), 164-167.

Chartrand, J. P., Filion-Bilodeau, S., & Belin, P. (2007). Brain response to birdsongs in bird experts. *Neuroreport, 18*(4), 335-340.

Chartrand, J. P., Peretz, I., & Belin, P. (2008). Auditory recognition expertise and domain specificity. *Brain Res*.

Chen, X., Striano, T., & Rakoczy, H. (2004). Auditory-oral matching behavior in newborns. *Dev Sci, 7*(1), 42-47.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., et al. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hear Res, 82*(1), 53-58.

Cheour, M., Alho, K., Ceponiene, R., Reinikainen, K., Sainio, K., Pohjavuori, M., et al. (1998). Maturation of mismatch negativity in infants. *Int J Psychophysiol, 29*(2), 217-226.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nat Neurosci, 1*(5), 351-353.

Cheour, M., Martynova, O., Naatanen, R., Erkkola, R., Sillanpaa, M., Kero, P., et al. (2002). Speech sounds learned by sleeping newborns. *Nature, 415*(6872), 599-600.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clin Neurophysiol, 110*(1), 24-30.

Cowan, N., Winkler, I., Teder, W., & Naatanen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J Exp Psychol Learn Mem Cogn, 19*(4), 909-921.

Creutzfeldt, O., Hellweg, F. C., & Schreiner, C. (1980). Thalamocortical transformation of responses to complex auditory stimuli. *Exp Brain Res, 39*(1), 87-104.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

DeCasper, A. J., Lecanuet, J. P., Busnel, M. C., Granier-Deferre, C., & Maugeais, R. (1994). Fetal reactions to recurrent maternal speech. *Infant Behav Dev, 17*, 159-164.

Deguchi, C., Chobert, J., Brunelliere, A., Nguyen, N., Colombo, L., & Besson, M. (2010). Pre-attentive and attentive processing of French vowels. *Brain Res, 1366*, 149-161.

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *J Cogn Neurosci, 12*(3), 449-460.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*(5600), 2013-2015.

Dehaene-Lambertz, G., Hertz-Pannier, L., & Dubois, J. (2006). Nature and nurture in language acquisition: anatomical and functional brain-imaging studies in infants. *Trends Neurosci, 29*(7), 367-373.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., et al. (2006). Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc Natl Acad Sci U S A, 103*(38), 14240-14245.

Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., et al. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain Lang, 114*(2), 53-65.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(968), 303-306.

Escera, C., Alho, K., Schroger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol Neurootol, 5*(3-4), 151-166.

Escera, C., Alho, K., Winkler, I., & Naatanen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *J Cogn Neurosci, 10*(5), 590-604.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage, 30*(2), 580-587.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage, 23*(3), 840-848.

Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Percept Psychophys, 59*(6), 839-849.

Fifer, W. P., Byrd, D. L., Kaku, M., Eigsti, I. M., Isler, J. R., Grose-Fifer, J., et al. (2010). Newborn infants learn during sleep. *Proc Natl Acad Sci U S A, 107*(22), 10320-10323.

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

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J Acoust Soc Am, 102*(2 Pt 1), 1213-1222.

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends Cogn Sci, 9*(10), 481-488.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev, 25*(4), 355-373.

Grillon, C., Courchesne, E., Ameli, R., Elmasian, R., & Braff, D. (1990). Effects of rare non-target stimuli on brain electrophysiological activity and performance. *Int J Psychophysiol, 9*(3), 257-267.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron, 65*(6), 852-858.

Gunji, A., Koyama, S., Ishii, R., Levy, D., Okamoto, H., Kakigi, R., et al. (2003). Magnetoencephalographic study of the cortical activity elicited by human voice. *Neurosci Lett, 348*(1), 13-16.

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

Holeckova, I., Fischer, C., Giard, M. H., Delpuech, C., & Morlet, D. (2006). Brain responses to a subject's own name uttered by a familiar voice. *Brain Res, 1082*(1), 142-152.

Huotilainen, M., Kujala, A., & Alku, P. (2001). Long-term memory traces facilitate short-term memory trace formation in audition in humans. *Neurosci Lett, 310*(2-3), 133-136.

Ilvonen, T., Kujala, T., Kozou, H., Kiesilainen, A., Salonen, O., Alku, P., et al. (2004). The processing of speech and non-speech sounds in aphasic patients as reflected by the mismatch negativity (MMN). *Neurosci Lett, 366*(3), 235-240.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport, 8*(12), 2809-2812.

Kaas, J. H., & Hackett, T. A. (1999). 'What' and 'where' processing in auditory cortex. *Nat Neurosci, 2*(12), 1045-1047.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Katayama, J., & Polich, J. (1996). P300 from one-, two-, and three-stimulus auditory paradigms. *Int J Psychophysiol, 23*(1-2), 33-40.

Kisilevsky, B. S., Hains, S. M., Lee, K., Xie, X., Huang, H., Ye, H. H., et al. (2003). Effects of experience on fetal voice recognition. *Psychol Sci, 14*(3), 220-224.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am, 87*(2), 820-857.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage, 22*(2), 948-955.

Kuhl, P., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu Rev Neurosci, 31*, 511-534.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat Rev Neurosci, 5*(11), 831-843.

Kuhl, P. K. (2007). Is speech learning 'gated' by the social brain? *Dev Sci, 10*(1), 110-120.

Kuhl, P. K., Coffey-Corina, S., Padden, D., & Dawson, G. (2005). Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Dev Sci, 8*(1), F1-F12.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: vocal imitation and developmental change. *J Acoust Soc Am, 100*(4 Pt 1), 2425-2438.

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci U S A, 100*(15), 9096-9101.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., & Naatanen, R. (2002). Maturation of the auditory change detection response in infants: a longitudinal ERP study. *Neuroreport, 13*(15), 1843-1848.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp, 24*(1), 11-20.

Lavner, Y., Gath, I., & Rosenhouse, J. (2000). The effects of acoustic modifications on the identification of familiar voices speaking isolated vowels. *Speech Communication, 30*(1), 9-26.

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

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40*(2), 291-305.

Linville, S. E. (1996). The sound of senescence. *J Voice, 10*(2), 190-200.

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Dev Sci, 6*(3), F1-F10.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage, 14*(4), 802-816.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Curr Biol, 19*(23), 1994-1997.

McAdams, S., & Cunible, J. C. (1992). Perception of timbral analogies. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 383-389.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143-178.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17*(4), 1742-1754.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495-500.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *J Acoust Soc Am, 98*(6), 3080-3095.

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid Brain Discrimination of Sounds of Objects. *The Journal of Neuroscience, 26*(4), 1293-1302.

Naatanen, R. (1990). The Role of Attention in Auditory Information-Processing as Revealed by Event-Related Potentials and Other Brain Measures of Cognitive Function. *Behavioral and Brain Sciences, 13*(2), 201-232.

Naatanen, R. (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear, 16*(1), 6-18.

Naatanen, R., & Alho, K. (1995). Mismatch negativity--a unique measure of sensory processing in audition. *Int J Neurosci, 80*(1-4), 317-337.

Naatanen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst), 42*(4), 313-329.

Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432-434.

Naatanen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Naatanen, R., Syssoeva, O., & Takegata, R. (2004). Automatic time perception in the human brain for intervals ranging from milliseconds to seconds. *Psychophysiology, 41*(4), 660-663.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., et al. (2001). Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia, 39*(10), 1047-1054.

Newman, J. D., & Wollberg, Z. (1973a). Multiple coding of species-specific vocalizations in the auditory cortex of squirrel monkeys. *Brain Res, 54*, 287-304.

Newman, J. D., & Wollberg, Z. (1973b). Responses of single neurons in the auditory cortex of squirrel monkeys to variants of a single call type. *Exp Neurol, 40*(3), 821-824.

Ockleford, E. M., Vince, M. A., Layton, C., & Reader, M. R. (1988). Responses of neonates to parents' and others' voices. *Early Hum Dev, 18*(1), 27-36.

Owren, M. J., Hopp, S. L., Sinnott, J. M., & Petersen, M. R. (1988). Absolute auditory thresholds in three Old World monkey species (Cercopithecus aethiops, C. neglectus, Macaca fuscata) and humans (Homo sapiens). *J Comp Psychol, 102*(2), 99-107.

Pena, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc Natl Acad Sci U S A, 100*(20), 11702-11705.

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nat Neurosci, 11*(3), 367-374.

Petkov, C. I., Logothetis, N. K., & Obleser, J. (2009). Where are the human speech and voice regions, and do other animals have anything like them? *Neuroscientist, 15*(5), 419-429.

Poremba, A., Malloy, M., Saunders, R. C., Carson, R. E., Herscovitch, P., & Mishkin, M. (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature, 427*(6973), 448-451.

Purhonen, M., Kilpelainen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *Int J Psychophysiol, 52*(3), 257-266.

Purhonen, M., Kilpelainen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2005). Four-month-old infants process own mother's voice faster than unfamiliar voices--electrical signs of sensitization in infant brain. *Brain Res Cogn Brain Res, 24*(3), 627-633.

Querleu, D., Lefebvre, C., Titran, M., Renard, X., Morillion, M., & Crepin, G. (1984). [Reaction of the newborn infant less than 2 hours after birth to the maternal voice]. *J Gynecol Obstet Biol Reprod (Paris), 13*(2), 125-134.

Querleu, D., Renard, X., & Crepin, G. (1981a). [Auditory perception and fetal reaction to react to sound stimulation (author's transl)]. *J Gynecol Obstet Biol Reprod (Paris), 10*(4), 307-314.

Querleu, D., Renard, X., & Crepin, G. (1981b). [Intra-uterine sound and fetal auditory perception]. *Bull Acad Natl Med, 165*(5), 581-588.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349-351.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A, 97*(22), 11800-11806.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *J Exp Psychol Hum Percept Perform, 23*(3), 651-666.

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (Macaca mulatta) vocalizations. *J Acoust Soc Am, 103*(1), 602-614.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 51*(5), 1007-1015.

Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. *Int J Psychophysiol, 75*(1), 44-47.

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci, 2*(12), 1131-1136.

Romanski, L. M., Tian, B., Fritz, J. B., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (2000). Reply to "What', 'where' and 'how' in auditory cortex'. *Nat Neurosci, 3*(10), 966.

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., & Huotilainen, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clin Neurophysiol, 119*(2), 332-341.

Scherer, K. R. (1995). Expression of emotion in voice and music. *J Voice, 9*(3), 235-248.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature, 259*(5538), 39-41.

Suga, N. (1992). Philosophy and stimulus design for neuroethology of complex-sound processing. *Philos Trans R Soc Lond B Biol Sci, 336*(1278), 423-428.

Tiitinen, H., May, P., Reinikainen, K., & Naatanen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature, 372*(6501), 90-92.

Titova, N., & Naatanen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett, 308*(1), 63-65.

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *J Acoust Soc Am, 85*(4), 1699-1707.

Van Dommelen, W. A. (1990). Acoustic parameters in human speaker recognition. *Lang Speech, 33 (Pt 3)*, 259-272.

Van Lancker, D. R., & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn, 1*(2), 185-195.

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: a dissociation between familiar and unfamiliar voices. *Cortex, 24*(2), 195-209.

Van Lancker, D. R., & Kreiman, J. (1987). Voice discrimination and recognition are separate abilities. *Neuropsychologia, 25*(5), 829-834.

Van Lancker, D. R., Kreiman, J., & Cummings, J. (1989). Voice perception deficits: neuroanatomical correlates of phonagnosia. *J Clin Exp Neuropsychol, 11*(5), 665-674.

van Zuijen, T. L., Sussman, E., Winkler, I., Naatanen, R., & Tervaniemi, M. (2005). Auditory organization of sound sequences by a temporal or numerical regularity--a mismatch negativity study comparing musicians and non-musicians. *Brain Res Cogn Brain Res, 23*(2-3), 270-276.

von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res, 17*(1), 48-55.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev Sci, 10*(2), 159-164.

Wang, X. (2000). On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A, 97*(22), 11843-11849.

Wang, X., & Kadia, S. C. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol, 86*(5), 2616-2620.

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage, 24*(4), 1052-1057.

Warren, J. D., Scott, S. K., Price, C. J., & Griffiths, T. D. (2006). Human brain mechanisms for the early analysis of voices. *Neuroimage, 31*(3), 1389-1397.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Dev, 52*(1), 349-355.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog Brain Res, 156*, 249-268.

Winter, P., & Funkenstein, H. H. (1973). The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey. (Saimiri sciureus). *Exp Brain Res, 18*(5), 489-504.

Winter, P., Ploog, D., & Latta, J. (1966). Vocal repertoire of the squirrel monkey (Saimiri sciureus), its analysis and significance. *Exp Brain Res, 1*(4), 359-384.

Yago, E., Corral, M. J., & Escera, C. (2001). Activation of brain mechanisms of attention switching as a function of auditory frequency change. *Neuroreport, 12*(18), 4093-4097.

Ylinen, S., Shestakova, A., Huotilainen, M., Alku, P., & Naatanen, R. (2006). Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. *Brain Res, 1072*(1), 175-185.

Zaske, R., Schweinberger, S. R., Kaufmann, J. M., & Kawahara, H. (2009). In the ear of the beholder: neural correlates of adaptation to voice gender. *Eur J Neurosci, 30*(3), 527-534.