Musique et Langage :
Spécificités, Interactions et Associations spatiales

par

Pascale Lidji

Thèse de doctorat effectuée en cotutelle

Faculté des Sciences psychologiques et de l'Education,
Université Libre de Bruxelles (U.L.B.)
Et
Département de Psychologie
Faculté des Arts et des Sciences
Université de Montréal

Thèse présentée à la Faculté des Sciences psychologiques et de l'Education de l'Université Libre de Bruxelles en vue de l'obtention du grade de Docteur en Sciences psychologiques

et à

la Faculté des Etudes supérieures de l'Université de Montréal en vue de l'obtention du grade de Philosophiae Doctor (Ph. D.) en psychologie, option Sciences cognitives et Neuropsychologie

Mars 2008

Université Libre de Bruxelles
Faculté des Sciences psychologiques et de l'Education
et
Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Musique et Langage : Spécificités, Interactions et Associations spatiales

présentée par Pascale Lidji
et soutenue à l'Université Libre de Bruxelles

a été évaluée par un jury composé des personnes suivantes :

Présidente

Pr. Jacqueline Leybaert, Laboratoire de Psychologie expérimentale,
Université Libre de Bruxelles (ULB)

Directrice de recherche (ULB)

Pr. Régine Kolinsky, Maître de Recherche FNRS, co-directrice de l'UNESCOG
Faculté des Sciences psychologiques et de l'éducation, ULB

Directrice de recherche (UdeM)

Pr. Isabelle Peretz, co-directrice du Laboratoire BRAMS, Département de
Psychologie, Université de Montréal (UdeM)

Membre du jury

Pr. Renée Béland, Faculté de Médecine, Ecole d'orthophonie et d'audiologie, UdeM

Examinatrice externe et secrétaire du jury

Pr. Séverine Samson, Directrice du Laboratoire Neuropsychologie et Cognition
auditive, Université de Lille III

Examinateur externe

Pr. Emmanuel Bigand, Directeur du LEAD, Université de Bourgogne, Dijon

Liste des coauteurs

Mireille Besson
John F. Connolly
Pierre Jolicoeur
Régine Kolinsky
Hélène Lafontaine
Aliette Lochy
José Morais
Patricia Moreau
Isabelle Peretz

Résumé

L'objectif de ce travail était d'examiner la spécificité fonctionnelle du traitement et des représentations des hauteurs musicales. À cette fin, ce traitement a été comparé à celui des phonèmes de la parole, d'une part, et aux associations spatiales évoquées par des séquences ordonnées, d'autre part. Nos quatre études avaient pour point commun d'adapter à un nouvel objet de recherche des méthodes bien établies en psychologie cognitive. Ainsi, nous avons exploité la tâche de classification accélérée (Etude 1) de Garner (1974), l'analyse des conjonctions illusoires en mémoire (Etude 2), l'additivité de la composante mismatch negativity (MMN) des potentiels évoqués (Etude 3) et l'observation d'associations spatiales de codes de réponse (Etude 4).

Les trois premières études, menées chez des participants non-musiciens, portaient sur la spécificité de traitement des hauteurs par rapport à celui des phonèmes au sein de stimuli chantés. Les deux premières études ont mis en évidence un effet surprenant de la nature des phonèmes sur leurs interactions avec le traitement des mélodies : les voyelles apparaissaient plus intégrées à la mélodie que les consonnes. Ceci était vrai à la fois lors du traitement en temps réel de non-mots chantés (Etude 1) et au niveau des traces en mémoire de ces mêmes non-mots (Etude 2, utilisant une tâche de reconnaissance à choix forcé permettant la mise en évidence de conjonctions illusoires). Cette dissociation entre voyelles et consonnes quant à leur intégration avec les traitements mélodiques ne semblait pas causée par des caractéristiques acoustico-phonétiques telles que la sonorité. Les résultats de la troisième étude indiquaient que les MMNs en réponse à des déviations de hauteur et de voyelle n'étaient pas additives et que leur distribution topographique ne différait pas selon le type de déviation. Ceci suggère que, même au niveau pré-attentionnel, le traitement des voyelles n'est pas indépendant de celui des hauteurs.

Dans la quatrième étude, nous avons comparé le traitement des hauteurs musicales à un autre domaine : la cognition spatiale. Nous avons ainsi montré que les non-musiciens comme les musiciens associent les notes graves à la partie inférieure et les notes aiguës à la partie supérieure de l'espace. Les deux groupes liaient aussi les notes graves au côté gauche et les notes aiguës au côté droit, mais ce lien n'était automatique que chez les musiciens. Enfin, des stimuli musicaux plus complexes (intervalles mélodiques) n'évoquaient ces associations spatiales que chez les musiciens et ce, uniquement sur le plan horizontal.

Ces recherches contribuent de plusieurs manières à la compréhension de la cognition musicale. Premièrement, nous avons montré que les consonnes et les voyelles diffèrent dans leurs interactions avec la musique, une idée à mettre en perspective avec les rôles différents de ces phonèmes dans l'évolution du langage. Ensuite, les travaux sur les représentations spatiales des hauteurs musicales ouvrent la voie à un courant de recherche qui aidera à dévoiler les liens potentiels entre habiletés musicales et spatiales.

Mots-clés

phonologie - voyelles – consonnes – intervalles – hauteurs musicales – mémoire – chant – effet SPARC – effet SMARC – mismatch negativity

Abstract

The purpose of this work was to examine the functional specificity of musical pitch processing and representation. To this aim, we compared musical pitch processing to (1) the phonological processing of speech and (2) the spatial associations evoked by ordered sequences. The four studies described here all use classical methods of cognitive psychology, which have been adapted to our research question. We have employed Garner's (1974) speeded classification task (Study 1), the analysis of illusory conjunctions in memory (Study 2), the additivity of the mismatch negativity (MMN) component of event-related potentials (Study 3), as well as the observation of spatial associations of response codes (Study 4).

The three first studies examined, in non-musician participants, the specificity of pitch processing compared to phoneme processing in songs. Studies 1 and 2 revealed a surprising effect of phoneme category on their interactions with melodic processing: vowels were more integrated with melody than were consonants. This was true for both on-line processing of sung nonwords (Study 1) and for the memory traces of these nonwords (Study 2, using a forced-choice recognition task allowing the occurrence of illusory conjunctions). The difference between vowels and consonants was not due to acoustic-phonetic properties such as phoneme sonority. The results of the third study showed that the MMN in response to pitch and to vowel deviations was not additive and that its brain topography did not differ as a function of the kind of deviation. This suggests that vowel processing is not independent from pitch processing, even at the pre-attentive level.

In the fourth study, we compared pitch processing to another domain: spatial cognition. We showed that both musicians and non-musicians map pitch onto space, in that they associate low-pitched tones to the lower spatial field and high-pitched tones to the higher spatial field. Both groups of participants also associated low pitched-tones with the left and high-pitched tones with the right, but this association was automatic only in musicians. Finally, more complex musical stimuli such as melodic intervals evoked these spatial associations in the horizontal plane only in musicians.

This work contributes to the understanding of music cognition in several ways. First, we have shown that consonants and vowels differ in their interactions with music, an idea related to the contrasting roles of these phonemes in language

evolution. Second, the work on the spatial representation of pitch opens the path to research that will help uncover the potential links between musical and spatial abilities.

Keywords

phonology - vowels – consonants – intervals – musical pitch – memory – song – SPARC effect – SMARC effect – mismatch negativity

Table des matières

Liste des tableaux

2. Section expérimentale

2. 1. Processing interactions between lyrics and tunes: Vowels sing but consonants speak

2. 2. Integration effect and illusory conjunctions in memory for lyrics and tunes: Vowels sing but consonants swing

Liste des figures

1. Introduction

1. 1. Introduction générale

1. 2. Intégralité et séparabilité : Revue et application aux interactions entre paroles et mélodies dans le chant

Liste des sigles et abréviations

ANOVA	ANalysis Of Variance
CV	Consonant-Vowel
dRTs	Difference of reaction times
dErrors	Difference of error rate
GRT	General Recognition Theory
EEG	électroencéphalogramme
ERP	Event-related potentials
e.g.	exempli gratia / for example
F0	Fréquence fondamentale
F1	Premier formant
F2	Deuxième formant
FAs	False alarms
I	Interval
ISI	Inter-stimulus interval
i.e.	id est / that is
MEG	Magnétoencéphalographie
MMN	Mismatch Negativity
NW	NonWord
p. ex.	par exemple
QI	Quotient intellectuel
REM	Retrieving Effectively from Memory
RTs	Reaction Times
SMARC	Spatial Melodical Association of Response Codes
SNARC	Spatial Numerical Association of Response Codes
SOA	Stimulus Onset Asynchrony
SPARC	Spatial Pitch Association of Response Codes

À Wanda.

Remerciements

Une thèse de doctorat, c'est un voyage initiatique à la recherche de la connaissance. Cette thèse fut un voyage, certes, mais au propre comme au figuré, puisqu'elle m'a permis de traverser un océan... et ce qui était au départ vécu comme une épreuve est devenu une extraordinaire expérience de vie.

Je n'aurai jamais assez de mots pour remercier mes guides lors de ce voyage, c'est-à-dire mes promotrices, Régine Kolinsky et Isabelle Peretz. Elles sont pour moi deux magnifiques modèles.

Régine m'a accompagnée depuis mon mémoire de licence avec une disponibilité et un investissement hors pair. Merci de l'exemple de persévérance, de rigueur et d'honnêteté scientifique que tu m'as donné, merci pour ta patience et ton soutien, y compris dans la recherche constante de financements afin de m'aider à continuer ces travaux dans les meilleures conditions possibles. Merci aussi de m'avoir laissé partir à plusieurs reprises, tout en continuant à m'encadrer et m'encourager de manière assidue malgré la distance.

Isabelle est arrivée un peu plus tard dans cette aventure en m'offrant l'opportunité exceptionnelle de réaliser mon doctorat en cotutelle et de découvrir par ce biais une autre culture. Merci de m'avoir poussée à me dépasser, y compris en m'incitant à modifier mon projet de recherche pour y inclure des méthodes avec lesquelles j'étais moins familière, comme l'EEG. Merci de m'avoir permis de collaborer avec des chercheurs prestigieux. Enfin, merci d'avoir suivi mes travaux d'aussi près ces derniers mois, nos rencontres, courriels et coups de fil ont été précieux.

Je me dois aussi de remercier mon directeur « non officiel », José Morais. Bien que sa participation à ce doctorat ne soit mentionnée qu'en tant que directeur de mémoire, collaborateur et président de comité d'accompagnement, José a joué un rôle primordial dans cette thèse. Premièrement, en étant le premier professeur à m'avoir fait découvrir la psychologie cognitive. Ensuite, parce qu'il m'a poussée à m'orienter vers la recherche malgré mes réticences d'origine en me proposant de postuler comme assistante dans son unité de recherche. Enfin, parce qu'il a continué à suivre mes recherches avec autant d'attention que s'il était toujours mon co-directeur. Merci pour ton investissement.

Ma reconnaissance va ensuite aux membres de mon jury doctoral, de mon comité d'accompagnement à Bruxelles et de mon jury d'examen de synthèse à Montréal. Merci donc à Jacqueline Leybaert, Séverine Samson, Renée Béland, Emmanuel Bigand et Martin Arguin. Vos commentaires ont été extrêmement constructifs tout au long de ce travail. C'est un honneur pour moi de vous voir siéger dans ce jury ; j'espère que le texte que je vous soumets aujourd'hui répondra à vos attentes.

La réalisation de cette thèse n'aurait pas été possible sans le soutien financier de plusieurs organismes que je souhaite remercier ici. Ces travaux ont, en effet, été réalisés dans le cadre d'un mandat d'assistante à mi-temps au sein de la Faculté des Sciences psychologiques et de l'Education de l'ULB. En complément, j'ai eu la chance de bénéficier d'un financement HFSP de 2003 à 2005, puis d'un financement Mini-Arc de 2005 à 2008. Une partie des frais inhérents à ces recherches ont également été pris en charge par un financement FRFC du FNRS.

Une thèse de doctorat, en cotutelle de surcroît, c'est aussi beaucoup de déplacements. Merci au département des relations internationales de l'ULB de m'avoir attribué une bourse de mobilité pour mon séjour à Montréal en 2005 ; à la Communauté française de Belgique pour la bourse de voyage octroyée en 2007 ; et à l'Université de Montréal de m'avoir offert une bourse d'exemption des frais de scolarité dans le cadre de cette cotutelle. Enfin, merci à l'UNESCOG et au BRAMS d'avoir contribué financièrement à mes autres séjours, ainsi qu'au FNRS pour les multiples financements accordés pour assister à des conférences scientifiques à l'étranger. La cotutelle implique aussi une certaine lourdeur administrative, je tiens donc à adresser mes plus chaleureux remerciements aux membres des différents secrétariats et services qui ont pris en charge mon dossier, entre autres, Olivier Cornelis, Sabrina Dejongh, Pierre Quertenmont et Anita Tenenbaum à l'ULB ; Danielle Morand, Thérèse Lauzé, Pascale Vezina et Lise Perreault à l'UdeM. Merci aussi aux fées du labo, je pense à Virginie Borguet, Aude Fenaux et Monique Declercq dont les sourires ont, heureusement, accompagné tant de démarches administratives moins sympathiques. Merci aussi à Assaad Azzi et Cécile Van de Leemput, Doyens de la Faculté des Sciences psychologiques et de l'éducation pendant la réalisation de ma thèse, et un immense merci à Muriel Constas pour son aide à de multiples étapes de ce processus.

Ces quatre années de thèse ont également été l'occasion de magnifiques rencontres et collaborations. Merci donc à tous les experts qui m'ont fait l'honneur de participer à ces travaux : Mireille Besson, Corine Astesano, Aliette Lochy, Pierre Jolicoeur, John Connolly, Cécile Colin et Paul Deltenre. Merci également à Olivier Bélanger, Bernard Bouchard, Olivier Piché, Manon Robert et Mihaela Felezeu pour leur support technique. Je n'aurais pu réaliser tant d'expériences sans l'aide des jobistes qui sont souvent devenus des collègues et amies: Julie Bertels, Olivia Gosseries, Sandrine Lahou, Dyna Delle-Vigne, Marjolaine Gailly et Olivia de Vylder ; ni sans mes stagiaires de choc : Ana Franco, Hélène Lafontaine et Julien-Pierre Vanasse-Larochelle. Je suis également reconnaissante à Thierry Vallier et Florent Hubert d'avoir prêté leur voix magnifique à mes stimuli... disons peu harmonieux.

Que serait la recherche en psychologie sans participants ? Merci pour leur patience aux cinq générations d'étudiants de 1ère candi - BA1, à mes étudiants de psycholinguistique, à tous les amis et amis d'amis pleins de bonne volonté, aux participants rémunérés venus de tous horizons (et particulièrement aux musiciens) qui se sont prêtés au jeu de mes expériences. Certaines populations sont particulièrement difficiles à atteindre. Merci à Hilde et Nelson Delle-Vigne de m'avoir donné accès aux musiciens de leurs masterclasses, ainsi qu'à l'Académie d'été de Libramont qui m'a accueillie plusieurs semaines.

Mes activités d'enseignement ont été la bouffée d'air nécessaire pendant ces années de thèse, je voudrais donc saluer les étudiants dont j'ai croisé la route et qui m'ont parfois inspiré des idées d'études ou de titres d'articles ... merci pour votre enthousiasme.

Grâce à la cotutelle, j'ai eu la chance de partager la vie de deux laboratoires à l'ambiance irremplaçable. Merci donc à la joyeuse bande « du 10ème » de Bruxelles, les anciens comme les nouveaux, pour tous ces petits moments agréables : Maud, Katia, Perrine, Chotiga, Muriel, Zohra, Laure D., Damien, Barbara, Cédric, Julie B., Ingrid, Julie N., Bert, Laure L., Stéphanie, Dominique, Hichem, Antoine, Orly, Ana, Stéphane, Florence, Rémy, Charline,... À Montréal, j'ai eu *ben du fun* avec Alexandra, Guylaine, Jean-Pierre, Stéphanie, Sandrine, Suzanne, Mathieu, Marie-Andrée, Dominic, Philippe, Emilie, les Sébastien, ... Parfois, les études et le travail donnent naissance à de belles amitiés : Virginie, Sandrine, Sonja, Nathalie, Patricia, Jessica, Bernard, j'espère que vous savez à quel point je suis heureuse de vous avoir

rencontrés sur les bancs de l'unif ou dans les couloirs d'un labo. Un merci spécial à Vinciane pour ses relectures et corrections. Votre amitié est un trésor précieux, merci d'avoir été là pendant les hauts et les bas de ma période thésarde.

Ma famille a également été un support indescriptible. Je ne sais comment exprimer ma gratitude envers mes parents, qui ont éveillé en moi la curiosité nécessaire en recherche, mais aussi le désir d'exceller. Merci de m'avoir fait aimer l'école, de m'avoir inscrite à l'Académie et fait découvrir la musique. Merci pour votre soutien moral et (parfois) financier, pour votre écoute, pour m'avoir accueillie quelques temps lorsque je cherchais un lieu propice à la rédaction, et pour avoir relu des sections de ce texte ardu. Merci aussi à la famille Couture d'avoir été une merveilleuse famille de substitution au Québec.

Enfin, merci à celui qui est resté à mes côtés pendant ce voyage et a stoïquement supporté toutes les étapes de mes études et de ma vie de chercheuse. Felipe, merci d'avoir fait le clown quand j'en avais besoin, de m'avoir sermonnée parfois, et, surtout, merci de m'avoir laissée partir plusieurs mois, d'avoir tout fait pour me suivre et de m'avoir donné l'envie de revenir… pour repartir à deux, j'espère.

1. Introduction

Introduction générale

1.1.1. Avant-propos

Chaque matin, vous vous réveillez au son de la radio, déjeunez en musique, passez peut-être plusieurs minutes dans les embouteillages en écoutant un disque compact ou prenez le métro avec, comme presque tous vos voisins, un lecteur mp3 sur les oreilles. La musique occupe donc aujourd'hui une place de choix dans nos sociétés occidentales. L'association actuelle entre musique et technologie pourrait nous faire oublier que la musique est universelle, puisqu'elle se retrouve dans toutes les cultures (p. ex., Blacking, 1995)[1] et qu'elle transcende l'Histoire (p. ex., D'Errico et al., 2003). Cette omniprésence incite à penser que l'être humain est fondamentalement musical (Peretz, 2006). En effet, la compétence musicale a souvent été confondue avec sa pratique, qui n'est l'apanage que d'une élite, ce qui a conduit à sous-estimer les habiletés musicales partagées par la population générale. Toutefois, plusieurs études ont mis en évidence le fait que les non musiciens possèdent des connaissances et compétences musicales non négligeables (pour une revue, voir Bigand & Poulin-Charronnat, 2006). Ces capacités semblent issues de la simple exposition à la musique de leur culture (Tillmann, Bharucha, & Bigand, 2000); la pratique musicale ne ferait que rendre ces connaissances implicites plus accessibles à la conscience.

Il est surprenant de constater qu'alors que nous sommes tous des êtres musicaux (ou presque, si l'on exclut les cas d'amusie congénitale, Ayotte, Peretz, & Hyde, 2002; Peretz, Cummings, & Dubé, 2007; Peretz & Hyde, 2003), le traitement cognitif de la musique n'est étudié spécifiquement que depuis quelques années (pour des revues, voir par exemple Patel, 2008; Peretz & Zatorre, 2005; Zatorre & Peretz, 2001). Une question qui a fait couler beaucoup d'encre dans cette littérature récente est celle de savoir si des ressources cognitives spécifiques sont dédiées à la perception et à la production musicales (p. ex., Jackendoff & Lerdahl, 2006; Peretz & Coltheart, 2003; Peretz & Morais, 1989) ou si la musique n'st qu'un « produit dérivé » de nos habiletés cognitives générales (p. ex., Pinker, 1997).

---

[1] Toutes les références de ce texte, y compris celles des chapitres issus d'articles, sont reprises dans la section « 4. Bibliographie » en fin de document.

Cette question de recherche renvoie à la problématique très actuelle de la nature potentiellement biologique des habiletés musicales humaines (McDermott & Hauser, 2005; Peretz, 2006). Si nous sommes biologiquement musicaux, cela devrait effectivement laisser des traces au sein de notre système cognitif et, éventuellement, au niveau neuronal.

Un nombre croissant de recherches visent donc à démontrer ou, au contraire, à discréditer l'origine biologique de la musique au moyen de méthodes extrêmement variées. Si certains se sont placés dans une perspective phylogénétique en cherchant à mettre en évidence des traces historiques d'activité musicale chez nos lointains ancêtres (Mithen, 2005) ou en utilisant une approche comparative avec diverses espèces animales (Fitch, 2006; Hauser & McDermott, 2003), des chercheurs en psychologie et neurosciences ont centré leur attention sur le fonctionnement cognitif humain actuel.

Ces chercheurs ont exprimé des points de vue contradictoires (pour un débat similaire au niveau du traitement de la parole, voir par exemple Diehl, Lotto, & Holt, 2004). Selon les tenants d'un courant théorique bien représenté, des compétences cognitives générales permettent l'acquisition des différentes habiletés spécifiques, y compris les habiletés musicales. La nature de ces compétences cognitives générales varie toutefois selon les auteurs. Elles peuvent consister en des capacités d'apprentissage statistique qui s'appliquent aussi bien au langage qu'à la musique (p. ex., Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999), ou encore reposer sur des réseaux connexionnistes qui ne présentent pas de structure dissociée selon le type d'information traitée (p. ex., Plaut, 1995; Rumelhart, McClelland, & Group, 1986), ou, enfin, postuler l'existence de ressources cognitives communes pour les traitements linguistiques et musicaux, ces ressources partagées traitant toutefois des représentations spécifiques à chaque domaine (Patel, 2003; Patel, 2007).

Une perspective alternative s'inscrit dans le cadre de la théorie de la modularité (Fodor, 1983). Selon la version forte, et sans doute exagérée, de cette théorie, la cognition opèrerait sur base de systèmes de traitement innés, spécifiques à un domaine (Hirschfeld & Gelamn, 1994), encapsulés informationnellement, associés à des substrats neuronaux spécifiques, et dont les traitements sont automatiques et rapides. Un argument solide en faveur de la nature biologique de la musique consisterait donc à démontrer que certaines compétences musicales répondent à un traitement de type modulaire (Peretz & Morais, 1989; mais voir, par exemple, Plaut, 1995).

L'approche traditionnelle de la modularité a toutefois été fortement nuancée depuis les travaux initiaux de Fodor (p. ex., Coltheart, 1999; Fodor, 2000). Ainsi, il est généralement admis qu'un système puisse opérer de manière modulaire sans être inné. C'est, par exemple, le cas de la lecture (Fodor, 2000). De même, l'idée de réseaux neuronaux spécifiques et, surtout, restreints au niveau anatomiques ne semble pas indispensable à la modularité. Bien que, dans certaines situations de traitement modulaire, des substrats neuronaux localisés puissent être précisés (voir par exemple Belin, Fecteau, & Bédard, 2004 pour le traitement de la voix; Farah, Wilson, Drain, & Tanaka, 1998 pour le traitement des visages), l'idée que d'autres traitements modulaires puissent être distribués au niveau cérébral est de plus en plus acceptée (voir Dehaene, Piazza, Pinel, & Cohen, 2003 pour un exemple en cognition numérique; Hickok & Poeppel, 2004 pour un exemple au niveau du langage). En revanche, la spécificité à un domaine est décrite comme une condition nécessaire à l'idée de modularité (Coltheart, 1999; Peretz & Coltheart, 2003).

Dans cette thèse, nous aurons précisément pour objectif d'enrichir la compréhension de la spécificité ou, au contraire, de la non-spécificité du traitement musical. À cette fin, nous avons pris comme références deux domaines auxquels la musique a régulièrement été comparée : la parole, d'une part, et les représentations spatiales, d'autre part.

1. 1 . 2. Spécificité du traitement musical

L'idée selon laquelle la musique pourrait faire l'objet d'un traitement modulaire a été clairement exposée et justifiée par Peretz et Coltheart en 2003. Ces auteurs basent essentiellement leur argumentation sur l'observation de déficits sélectifs du traitement musical, qu'ils soient congénitaux (p. ex., Ayotte et al., 2002; Peretz, 2001) ou consécutifs à des lésions cérébrales (p. ex. Griffiths et al., 1997; Peretz, Belleville, & Fontaine, 1997; Peretz et al., 1994; Steinke, Cuddy, & Holden, 1997). Ces patrons de performance se complètent par des dissociations où le traitement musical est préservé mais le traitement linguistique altéré (p. ex., Hébert, Racette, Gagnon, & Peretz, 2003; Mendez, 2001; Peretz, Gagnon, Hébert, & Macoir, 2004; Warren, Warren, Fox, & Warrington, 2003). L'observation de doubles dissociations entre le traitement musical et d'autres traitements issus de la sphère auditive, tels que le traitement de la parole, constitue donc un premier argument en faveur de la spécificité fonctionnelle de la musique.

Comme nous l'avons vu, une deuxième propriété des systèmes modulaires est, selon Fodor (1983), de s'appuyer sur une architecture neuronale spécifique. De nombreuses recherches ont donc été orientées vers la découverte de substrats neuronaux spécifiquement consacrés au traitement musical (pour des revues récentes, voir Peretz & Zatorre, 2005; Samson, 1999; Stewart, von Kriegstein, Warren, & Griffiths, 2006; Zatorre & Gandour, sous presse). Ces études concordent généralement pour montrer que des aires du gyrus temporal supérieur et du lobe frontal droits sont particulièrement impliquées dans l'analyse des relations de hauteur (p. ex., Hyde, Peretz, & Zatorre, sous presse; Johnsrude, Penhune, & Zatorre, 2000; Liégeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998; Samson & Zatorre, 1991; Tervaniemi et al., 2000; Tramo, Shah, & Braida, 2002; Zatorre & Belin, 2001). Toutefois, aucune aire purement musicale n'a encore été découverte, peut-être en raison du manque de comparaisons inter-domaines.

Un autre argument en faveur de la nature biologiquement spécifique des processus musicaux consisterait à montrer que les compétences musicales sont présentes de manière innée chez les humains, un défi qu'a tenté de relever, entre autres, l'équipe de Sandra Trehub (pour une revue récente, voir Trehub & Hannon, 2006). Ces chercheurs ont, en effet, démontré que les bébés perçoivent, très précocement, la musique de manière similaire aux adultes. Par exemples, des enfants de quelques mois reconnaissent une mélodie même lorsqu'elle a été transposée (Trehub, Thorpe, & Morrongiello, 1987), préfèrent les gammes présentant des intervalles de notes inégaux à celles dont les intervalles sont égaux (Trehub, Schellenberg, & Kamenetsky, 1999) et les intervalles consonants aux intervalles dissonants (Schellenberg & Trehub, 1996) alors que d'autres espèces animales ne manifestent pas de biais similaire (McDermott & Hauser, 2004). De plus, les berceuses, un genre musical directement destiné aux bébés, sont présentes et appréciées des enfants dans toutes les cultures (Trehub & Trainor, 1998). Ceci suggère que les nourrissons du monde entier sont enclins à écouter et aimer la musique. Il n'est toutefois pas possible, dans l'état actuel des connaissances, de préciser si les compétences musicales des bébés sont issues de prédispositions spécifiquement orientées vers le traitement musical ou de capacités générales d'extraction de règles (Saffran et al., 1999).

1. 1 . 3. Non-spécificité du traitement musical

Cette idée d'un traitement musical issu de mécanismes généraux ou détournés de leur fonction d'origine (p. ex., Sperber & Hirschfeld, 2004) est en phase avec le point de vue des opposants à la spécificité de la musique (p. ex., Pinker, 1997). Pour ces chercheurs, la musique agirait comme un parasite, exploitant des capacités cognitives vouées à d'autres types de traitements. La fonction la plus fréquemment décrite comme partageant de tels composants de traitement avec la musique est le langage. Cette comparaison entre musique et langage semble évidente, tant ces deux moyens d'expression présentent de points communs apparents (Besson & Schön, 2001). Tous deux empruntent le canal auditif, répondent à un ensemble de règles de type syntaxique (p. ex., Patel, 2003), ont une structure temporelle précise, possèdent un contour mélodique (p. ex., Patel, Peretz, Tramo, & Labrecque, 1998) et peuvent transmettre des émotions (p. ex., Juslin & Laukka, 2003).

Une proposition plus surprenante d'interaction entre traitement musical et non musical a, par ailleurs, provoqué une nouvelle effervescence dans la communauté scientifique (Stewart & Walsh, 2007). Selon cette hypothèse, basée sur la découverte de déficits de traitement spatial chez des participants amusiques (Douglas & Bilkey, 2007), la cognition musicale ferait appel à des ressources impliquées dans les habiletés visuo-spatiales. Cette proposition n'est pas sans rappeler la controverse (p. ex. Dalla Bella, Dunlop, Dawe, Humphrey, & Peretz, 1999; Steele, Dalla Bella et al., 1999) ayant entouré l'Effet Mozart (Rauscher, Shaw, & Ky, 1993, 1995), sur laquelle nous reviendrons dans la suite de ce texte.

Que la musique soit mise en parallèle avec le langage ou avec la cognition spatiale, l'idée sous-jacente à ces études comparatives reste d'étayer l'existence de spécificités de traitement ou, au contraire, l'existence d'interactions entre ces activités cognitives. Dans la suite de cette introduction, nous passerons donc en revue les données sur les relations entre musique et langage, d'une part, et entre musique et traitements spatiaux, d'autre part. Mais avant d'aborder ces questions, il convient d'examiner la structure du traitement musical et de préciser à quels aspects de la musique nous nous sommes intéressés.

1. 1. 4. Perspective de la thèse

Comme évoqué précédemment, la question à la base de cette thèse est la spécificité fonctionnelle du traitement musical, ce qui constituerait un indice de son caractère modulaire et biologiquement déterminé. Toutefois, il serait caricatural de vouloir montrer que le traitement musical, de manière globale, opère de manière modulaire (McDermott & Hauser, 2005). En effet, si de nombreuses études ont mis en évidence des similarités (p. ex., Koelsch et al., 2002; Koelsch et al., 2004; Levitin & Menon, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006) ou des transferts (p. ex., Anvari, Trainor, Woodside, & Levy, 2002; Besson, Schön, Moreno, Santos, & Magne, 2007; Slevc & Miyake, 2006; Thompson, Schellenberg, & Husain, 2004; Wong, Skoe, Russo, Dees, & Kraus, 2007) entre les traitements musicaux et non musicaux, un nombre tout aussi important de recherches apporte des arguments en faveur de la spécificité du traitement de la musique (p. ex., Besson, Faita, Peretz, Bonnel, & Requin, 1998; Bonnel, Faita, Peretz, & Besson, 2001; Peretz, 2001; Peretz et al., 1994; Tervaniemi et al., 1999; Tervaniemi et al., 2000).

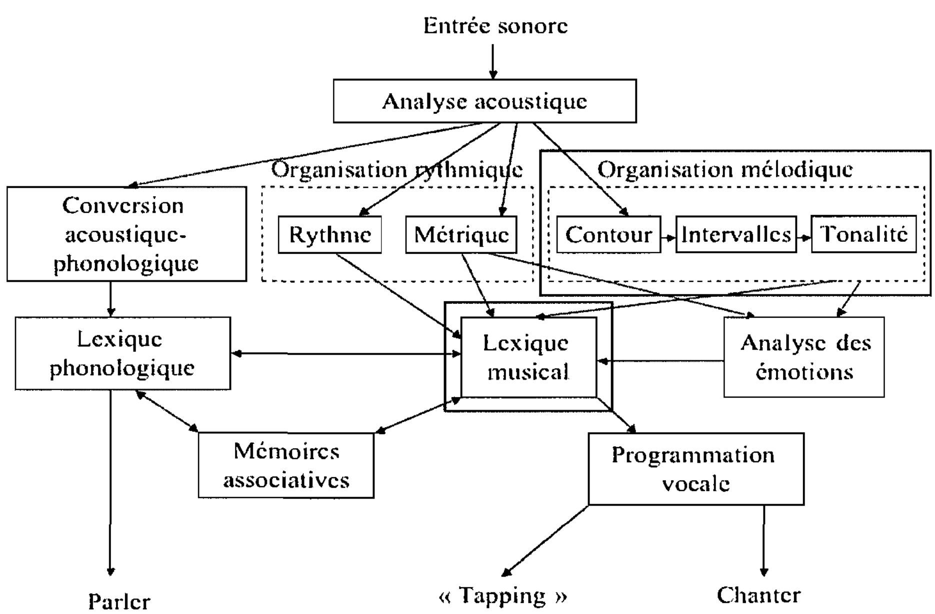
Ces données soutenant des points de vue théoriques apparemment contradictoires peuvent cependant être conciliées en adoptant une position plus nuancée. Selon la perspective que nous avons choisi d'adopter, certains aspects du traitement musical pourraient interagir avec des aspects particuliers du traitement linguistique (ou spatial), alors que d'autres s'en dissocieraient (Jackendoff & Lerdahl, 2006). Notre objectif devient donc de dépasser la dichotomie entre spécificité et non-spécificité, et de préciser à quels niveaux les éventuelles interactions ou dissociations pourraient prendre place. Nous serons ainsi à même de définir quels traitements sont spécifiquement musicaux et lesquels font appel à des ressources communes avec d'autres fonctions. Étant donné la multiplicité des objets musicaux et non-musicaux potentiellement comparables, il reste à choisir les composantes auxquelles nous allons consacrer notre attention.

1. 1. 5. Hauteurs musicales

Le traitement musical peut être subdivisé en sous-composantes dissociables. La Figure 1a, tirée de l'article de Peretz et Coltheart (2003), présente un modèle modulaire des processus musicaux. D'après ce modèle, la musique est constituée de deux dimensions principales, la dimension mélodique et la dimension rythmique. Sur base de données issues de la neuropsychologie cognitive indiquant que les patients amusiques présentent généralement un déficit circonscrit au traitement de la hauteur musicale (Hyde & Peretz, 2004; Peretz & Kolinsky, 1993; Peretz et al., 1994 ; mais voir Di Pietro, Laganaro, Leemann, & Schnider, 2004), Peretz (2006) affirme que l'encodage tonal de la hauteur, faisant partie des traitements mélodiques, constituerait une sous-composante spécifique à la musique. Dans la même veine, Patel (2008) propose que les sujets amusiques congénitaux présentent un déficit du traitement du contour, c'est-à-dire de la direction des hauteurs musicales. Malgré les nuances qui différencient ces deux visions, elles ont le point commun de placer le traitement des hauteurs au centre des compétences musicales.

Le traitement de la hauteur musicale, en raison de sa spécificité déclarée, constituera donc le fil rouge de cette thèse. La spécificité du traitement des hauteurs sera comparée (1) au traitement phonétique dans la parole et (2) au traitement des séquences ordonnées évoquant des associations spatiales.

Figure 1a. Modèle modulaire de traitement de la musique, d'après Peretz & Coltheart, 2003, p. 690. Les composants de traitements spécifiques à la musique sont encadrés en rouge.



1. 1. 6. Parole et musique

1. 1. 6. 1. Comparaisons langage - musique

La question de la spécificité fonctionnelle de la musique par rapport au traitement du langage peut être envisagée selon deux angles. Comme développé ci-dessus, des recherches récentes utilisent la comparaison langage-musique pour tenter de démontrer que la musique pourrait avoir des fondements biologiques (McDermott & Hauser, 2005; Peretz, 2006).

Par opposition à cette vision centrée sur la musique, la comparaison langage-musique a classiquement été utilisée dans le but de montrer que « la parole est spéciale », c'est-à-dire que des mécanismes spécifiques permettraient son traitement et son acquisition (Chomsky, 1957; Lieberman & Mattingly, 1985; Pinker, 1997; Pinker & Jackendoff, 2005). Cette position a, par exemple, été soutenue par les défenseurs de la théorie motrice de la parole (Galantucci, Fowler, & Turvey, 2006; Lieberman & Mattingly, 1985), selon lesquels la perception de la parole serait le résultat d'un « module phonétique », les autres informations auditives étant traitées par un système auditif général. D'autres auteurs (p. ex., Diehl et al., 2004; Tomasello, 1998) ont appuyé l'idée opposée, selon laquelle le langage émergerait de ressources cognitives globales et ne disposerait donc d'aucune spécificité fonctionnelle.

Les comparaisons entre musique et langage, quel que soit le niveau de traitement envisagé, ont fourni des arguments en faveur de ces deux positions théoriques (p. ex. Besson et al., 1998; Koelsch, Gunter, Wittgoth, & Sammler, 2005; Koelsch et al., 2004; Patel, Gibson et al., 1998; Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005; Schön, Gordon, & Besson, 2005).

Illustrons cet aspect en nous concentrant sur une analogie typique, celle qui compare les caractéristiques mélodiques de la musique et de la parole. La mélodie de la parole, ou prosodie, est constituée de variations de hauteur fournissant des informations sur l'état émotionnel du locuteur, l'intonation de la phrase (les indices de contour permettent par exemple de discriminer questions et affirmations) et l'accentuation. Ces variations de contour jouissent donc d'un statut fort différent dans la parole et la musique : alors qu'elles occupent une fonction essentiellement pragmatique dans la parole, elles sont une des clés de la structure musicale (Patel, 2008).

Une autre distinction entre mélodie et prosodie est qu'elles semblent donner lieu à des traitements dissociés. En effet, des sujets amusiques congénitaux éprouvent des difficultés à traiter des contours musicaux, y compris lorsque les contours mélodiques sont des analogues non-linguistiques de contours prosodiques (Ayotte et al., 2002; Patel, Foxton, & Griffiths, 2005; Patel, Peretz et al., 1998), alors qu'ils ne manifestent pas de déficit similaire face à des stimuli de parole. Enfin, il a été démontré à de multiples reprises (pour des revues, voir Wong, 2002; Zatorre & Gandour, sous presse) que les locuteurs de langues tonales, où certains contours intonationels ont des conséquences lexicales, présentent une activation supérieure des régions cérébrales gauches pour traiter les tons de leur langue (p. ex., Xu et al., 2006), alors que chez les locuteurs de langues non tonales, ces contours activent plutôt l'hémisphère droit (p. ex., Zatorre, Evans, Meyer, & Gjedde, 1992). Tous ces résultats semblent suggérer que le traitement de différences de hauteurs est modulé par le contexte dans lequel elles interviennent. Ces données apportent, de ce fait, un nouveau soutien à l'idée de processus spécifiquement linguistiques et / ou musicaux et confirment le rôle fondamental du traitement des hauteurs dans la musique.

Néanmoins, il existe plusieurs exemples de transferts entre compétences mélodiques musicales et linguistiques (pour une revue critique, voir Patel & Iversen, 2007), suggérant que ces traitements ne sont pas totalement indépendants. La pratique musicale semble augmenter la sensibilité à des différences prosodiques fines (Besson et al., 2007; Magne, Schön, & Besson, 2006; Marques, Moreno, Castro, & Besson, 2007; Schön, Magne, & Besson, 2004), y compris à un niveau précoce. Wong et collaborateurs (2007) ont récemment observé que, par rapport à des participants musicalement naïfs, des musiciens présentent une réponse plus robuste du tronc cérébral aux variations de hauteurs de tons linguistiques du Mandarin. Pourtant, ni les participants musiciens ni les non-musiciens n'étaient familiers avec les langues tonales. Il semble donc que l'expertise musicale puisse moduler la sensibilité aux variations mélodiques, y compris lorsque ces variations opèrent sur des stimuli linguistiques. Cette conclusion doit toutefois être nuancée par le fait que les individus enclins à développer une expertise musicale présentent peut-être une prédisposition au traitement de la hauteur, quel qu'en soit le contexte (Patel & Iversen, 2007). Des études longitudinales semblent toutefois suggérer que la pratique musicale a bien un impact direct sur le traitement mélodique et prosodique (Moreno & Besson, 2006).

En somme, il existe autant d'arguments en faveur de la spécificité de traitement des contours mélodiques linguistiques et musicaux qu'en faveur d'interactions entre ces dimensions. Cette variété des résultats expérimentaux se retrouve de manière similaire pour les autres dimensions linguistiques et musicales comparées dans la littérature (pour une revue, voir Patel, 2008).

1. 1. 6. 2. Phonèmes et hauteurs

Alors que les traitements mélodiques musicaux et linguistiques ont souvent été comparés en raison de leur similarité apparente, la spécificité de nombreux modules potentiels reste encore à examiner (Peretz, 2006). Ainsi, la comparaison entre le traitement des phonèmes et celui des hauteurs musicales a été relativement négligée jusqu'à présent, alors que ces dimensions constituent les éléments de base de la parole, d'une part, et de la musique, d'autre part (Patel, 2008).

De fait, la parole et la musique ont toutes deux un caractère discret (Besson & Schön, 2001; Hauser, Chomsky, & Fitch, 2002) : tout énoncé verbal et toute phrase musicale sont composés d'un ensemble limité d'unités, les phonèmes dans la parole et les notes dans la musique. Sur base de ces unités discrètes, un nombre illimité de productions sont possibles (Handel, 1989). Nous avons déjà discuté du statut privilégié des notes ou hauteurs dans la musique. Dans la parole, les unités discrètes sont les phonèmes. Ceux-ci ne diffèrent pas entre eux par la hauteur, mais plutôt par le timbre, attribut mutidimensionnel des sons complexes qui permet de distinguer des sons ayant la même hauteur, intensité sonore et durée (American Standard Institute, 1973). Le timbre est surtout fondamental pour la discrimination des voyelles car celles-ci sont définies par les relations entre la partie stable de leurs harmoniques les plus intenses ou formants (p. ex., Lieberman & Blumstein, 1988). Les consonnes sont, en plus de ces variations spectrales, caractérisées par des changements rapides d'amplitude et de fréquence fondamentale situés au niveau des transitions de formants (p. ex., Delattre, Liberman, & Cooper, 1955). Le timbre est également important en musique car il permet de discriminer les divers instruments. Cependant, il n'y joue pas un rôle aussi capital que dans la parole (mais voir Patel, 2008, pp. 34-37). Par exemple, le même morceau pourra être interprété et reconnu, qu'il soit joué au piano ou à la guitare, car ce qui en détermine l'identité est la tonalité, le choix des intervalles qui le composent, le contour mélodique et la structure rythmique.

Les notes et les phonèmes constituent donc les dimensions segmentales de la musique et de la parole, respectivement (McMullen & Saffran, 2004). Il apparaît ainsi nécessaire d'étudier directement les interactions entre ces dimensions constitutives de la musique et de la parole. Avant d'aborder cet aspect, passons en revue les connaissances sur les interactions entre timbre et hauteur dans un contexte plus général.

La majorité des études ayant examiné les interactions entre le timbre et la hauteur de notes isolées ont démontré que ces dimensions sont traitées de manière relativement intégrée. Par exemple, Melara et Marks (1990) ont montré, au moyen du paradigme de classification accélérée de Garner (1974, pour plus d'informations voir section 1. 2.) que des notes graves ou aiguës produites avec un timbre « nasillard » ou « creux » constituent des dimensions intégrées. Krumhansl et Iverson (Krumhansl & Iverson, 1992) ont reproduit ce patron d'interaction avec des timbres plus proches de ceux d'instruments existant.

Toutefois, cette interaction ne semble valide que dans le cas de notes isolées (Krumhansl & Iverson, 1992, Expériences 2 et 3) : lorsque les notes étaient placées dans un contexte plus long, les changements de hauteurs des notes constituant le contexte modifiaient la perception de la hauteur, mais pas celle du timbre, des notes cibles. Réciproquement, les modifications du timbre des notes du contexte influençaient la perception du timbre, mais pas celle de la hauteur, des notes cibles. Dans la lignée de ces résultats, Warrier et Zatorre (2002) ont observé que des modifications de timbre interféraient sur la capacité des sujets à décider si deux notes étaient identiques ou non, mais que cette interférence était réduite si le contexte tonal était accru, comme lors d'une tâche de jugement de congruence entre une note et la mélodie dont elle faisait partie. Si le contexte musical semble réduire les interactions entre le timbre et la hauteur d'une note cible, l'existence d'effets du timbre sur la taille perçue d'intervalles mélodiques a été récemment démontrée (Russo & Thompson, 2005). La transition d'un timbre sourd à un timbre plus brillant a tendance à augmenter la taille perçue d'un intervalle, tandis que la séquence inverse tend à réduire la distance ressentie entre les notes. Cette illusion de taille d'intervalle était observée même si l'influence du timbre sur la hauteur perçue des notes était neutralisée. En ce qui concerne l'influence de la fréquence fondamentale sur les jugements de timbre, les modifications de hauteur ne semblent moduler que de manière minime les distances perceptives entre les timbres, évaluées par la méthode du *multidimensional scaling* (Marozeau, de Cheveigné, McAdams, & Winsberg, 2003). Enfin, ces interactions entre timbre et hauteur semblent trouver leur source dans des substrats neuronaux communs (Warren, Jennings, & Griffiths, 2005).

En somme, les dimensions sonores de timbre et de hauteur semblent interagir dans le cas de stimuli non-linguistiques. Qu'en est-il des stimuli linguistiques ? Jusqu'à présent, les rares recherches qui se sont placées dans cette optique ont conduit à des résultats difficiles à interpréter. En effet, Bigand, Tillmann, Poulin, D'Adamo et Madurell (2001) semblaient avoir montré que la structure harmonique de mélodies influençait le traitement des voyelles sur lesquelles elles étaient chantées. Toutefois, ces données ont été récemment réinterprétées comme reflétant des processus attentionnels généraux, et non des interactions spécifiques entre traitement phonologique et mélodique (Escoffier & Tillmann, 2006).

D'autres recherches ont déclaré mettre en évidence un effet des habiletés musicales sur les compétences phonologiques, que ce soit dans l'acquisition de la lecture (Anvari et al., 2002) ou pour l'apprentissage d'une langue seconde (Slevc & Miyake, 2006). La portée de ces résultats est toutefois limitée par leur nature corrélationnelle qui empêche de tirer des conclusions sur le sens de la causalité (Patel & Iversen, 2007). Sur base de données comportementales, il n'est donc pas possible de conclure fermement qu'il existe des interactions entre traitements phonologiques et musicaux.

En revanche, la littérature sur l'origine neuronale de ces traitements indique l'existence d'une certaine spécialisation hémisphérique, les traitements phonétiques activant préférentiellement des aires de l'hémishère gauche (p. ex., Wong, 2002; Zatorre, Belin, & Penhune, 2002; Zatorre & Belin, 2001) alors que les hauteurs, intervalles et mélodies sont souvent associés à l'hémisphère droit (p. ex., Jamison, Watkins, Bishop, & Matthews, 2006; Johnsrude et al., 2000; Kimura, 1964; Liégeois-Chauvel et al., 1998; Tervaniemi et al., 2000; Tramo et al., 2002; Zatorre & Belin, 2001; Zatorre, Evans, & Meyer, 1994; Zatorre & Samson, 1991).

Cette littérature a été revue de manière extrêmement intéressante par Zatorre et Gandour (sous presse). Ces auteurs font, entre autres, le point sur une hypothèse prometteuse (Tervaniemi & Hugdahl, 2003; Zatorre et al., 2002) selon laquelle il n'y aurait pas de spécialisation cérébrale pour le traitement des hauteurs musicales, certaines régions de l'hémisphère droit étant plus aptes à traiter des différences fines de hauteur (Hyde et al., sous presse) et présentant, de ce fait, une supériorité pour le traitement musical. En revanche, des aires de l'hémisphère gauche seraient plus aptes à analyser les stimuli présentant des variations temporelles rapides (Ehrlé, Samson, & Baulac, 2001) caractéristiques de la parole (Price, Thierry, & Griffiths, 2006). Les caractéristiques acoustiques des stimuli linguistiques et musicaux pourraient donc jouer un rôle primordial dans les spécialisations hémisphériques décrites dans la littérature, bien que les données exposées précédemment nuancent ce point de vue. En effet, comme tout traitement linguistique y compris non auditif (Hickok, Bellugi, & Klima, 1998), le traitement des hauteurs tonales peut être associé à l'hémisphère gauche lorsque celles-ci ont une fonction linguistique. Cette approche est basée sur une vision physiologique mais peut être transposée en termes fonctionnels : la spécificité de traitement de la musique pourrait être directement liée à ses caractéristiques acoustiques fort différentes de celles du langage.

1. 1. 6. 3. Différences entre voyelles et consonnes

Si la dissociation entre traitements linguistiques et musicaux repose sur de telles propriétés acoustiques, l'on pourrait s'attendre à ce que les interactions entre mélodie et phonèmes soient modulées par la nature de ces derniers. Plus précisément, comme les consonnes, plus que les voyelles, dépendent d'une résolution temporelle fine, le traitement de ces phonèmes devrait être plus dissocié du traitement musical, basé sur la hauteur. Bien entendu, la distinction entre voyelles et consonnes n'est pas dichotomique puisque les phonèmes d'une langue diffèrent entre eux également par leur durée et leur sonorité (Clements, 1990, cité par Romani & Calabrese, 1998).

Le postulat selon lequel le traitement des consonnes serait fonctionnellement plus séparable de la mélodie que celui des voyelles repose sur deux ensembles d'arguments. Premièrement, les voyelles paraissent plus susceptibles que les consonnes de servir de support aux variations de hauteur en raison de leur durée et du fait qu'elles nécessitent une vibration des cordes vocales. Cette sonorité rend leur fréquence fondamentale facilement perceptible. Cette proposition est en accord avec des études sur la production du chant ayant montré que les voyelles sont le support du contour mélodique (Sundberg, 1982) alors que les consonnes auraient plutôt tendance à réduire la musicalité (Scotto di Carlo, 1993). Ces dernières peuvent difficilement transmettre de telles informations sur la hauteur sauf lorsqu'elles sont relativement sonores. D'ailleurs, il est intéressant de constater que des syllabes dont la consonne initiale est sourde provoquent une activation cérébrale gauche plus importante que des syllabes commençant par une consonne sonore (Jäncke, Wüstenberg, Scheich, & Heinze, 2002).

L'idée d'une origine commune des intervalles mélodiques et des voyelles a été suggérée par une étude récente (Ross, Choi, & Purves, 2007) selon laquelle les ratios de fréquence des intervalles de la gamme chromatique correspondent aux relations entre les fréquences des deux premiers formants des voyelles de plusieurs langues. Ces résultats sont interpellants car ils suggèrent que certains intervalles musicaux sont peut-être appréciés par les humains car ils font référence aux caractéristiques de leur langue. L'hypothèse inverse est également possible : l'évolution des voyelles d'une langue pourrait se baser sur sa musique (Mithen, 2005). Une interprétation plus parcimonieuse serait que le système auditif humain présente une préférence pour les intervalles aux ratios de fréquence entiers (Schellenberg & Trehub, 1994) et que ces intervalles aient été sélectionnés en parallèle dans le langage et la musique. Quoi qu'il en soit, il semble que la structure des voyelles et des intervalles musicaux présente une certaine parenté qui pourrait justifier des traitements cognitifs similaires, alors que les indices acoustiques spécifiant les consonnes sont plus différents de ceux qui caractérisent la musique.

Cette idée nous conduit au deuxième argument en faveur de la différence entre voyelles et consonnes dans leurs interactions avec la musique. Cet argument repose sur les fonctions contrastées de ces phonèmes dans la parole : les consonnes pourraient jouer un rôle plus important que les voyelles dans la perception et la production de la parole. Cette affirmation se base sur plusieurs types de données.

Le premier type est issu d'études comparatives entre les productions vocales humaines et celles de mammifères non-humains. Alors que différentes espèces animales produisent des cris dont la structure peut être assimilée à celle des voyelles, avec un conduit vocal ouvert et la présence de formants (Lieberman, 1968; Owren, Seyfarth, & Cheney, 1997; Rendall, Rodman, & Emond, 1996), la structure de la parole humaine est caractérisée par une alternance rythmique et très rapide d'ouvertures (voyelles) et de fermetures du conduit vocal (consonnes) (MacNeilage, 1998). La production de phonèmes de type consonantique semble donc typiquement humaine en raison des contraintes articulatoires qu'elle implique (MacNeilage & Davis, 2000 ; mais voir Fitch & Reby, 2001).

Un contraste similaire entre voyelles et consonnes a été observé dans des études d'apprentissage statistique basées sur les probabilités transitionnelles. Celles-ci correspondent à la probabilité de rencontrer certaines séquences de syllabes ou de mots en fonction de la structure d'une langue. Dans ces études, on présente aux participants un flux continu d'un langage artificiel où les probabilités transitionnelles des syllabes ou des segments phonémiques sont manipulées. Cette procédure permet d'examiner si ces variables statistiques interviennent dans l'extraction des propriétés lexicales et grammaticales de la langue. C'est le cas chez les adultes (Saffran, Newport, & Aslin, 1996; Schön et al., 2008) comme chez les bébés (Saffran, Aslin et al., 1996), y compris pour des segments non-adjacents (Newport & Aslin, 2004). Plus surprenant, ces capacités d'extraction de règles statistiques sont également présentes chez des espèces animales telles que les tamarins pinchés (Hauser, Newport, & Aslin, 2001) et les rats (Toro & Trobalòn, 2005).

Pour revenir à la question des voyelles et des consonnes, les humains adultes sont capables d'extraire les régularités statistiques de segments phonémiques non-adjacents, qu'il s'agisse de voyelles ou de consonnes (Newport & Aslin, 2004), mais présentent une performance supérieure pour les langages construits sur des racines consonantiques (Bonatti, Peña, Nespor, & Mehler, 2005; Mehler, Peña, Nespor, & Bonatti, 2006). Cette structure est d'ailleurs analogue à celle des langues sémitiques. Par contre, des tamarins étaient uniquement capables d'extraire les régularités statistiques de langues artificielles construites sur une base vocalique (Newport, Hauser, Spaepen, & Aslin, 2004) et ne présentaient aucun apprentissage pour les langues fondées sur des racines consonantiques. Ces résultats suggèrent, d'une part, que les consonnes bénéficieraient d'un traitement privilégié dans l'espèce humaine, ce qui les rendrait éventuellement spécifiques au langage. D'autre part, les consonnes et les voyelles pourraient occuper un statut différent au sein même de la parole, les consonnes jouant un rôle crucial dans le traitement lexical alors que les voyelles combineraient une fonction syntaxique et un rôle plus général de support prosodique et de transmission d'informations sur l'identité du locuteur (Bonatti et al., 2005; Bonatti, Peña, Nespor, & Mehler, 2007; Owren & Cardillo, 2006).

Si les consonnes et les voyelles bénéficient bien de statuts différents dans le langage humain, nous pourrions donc nous attendre à ce qu'elles présentent un degré de spécificité différent avec le traitement mélodique de la musique. L'objectif des trois études de cette thèse où le traitement phonétique et le traitement des hauteurs musicales ont été comparés était de tester cette hypothèse au moyen d'un matériel expérimental particulier : le chant.

1. 1. 7. Le chant

Le chant est certainement l'expression musicale la plus universelle : il est pratiqué dans toutes les cultures et par presque tous les individus, même si ce n'est parfois que sous la douche (Dalla Bella, Giguère, & Peretz, 2007). Le chant constitue, par ailleurs, un matériel expérimental privilégié pour comparer les traitements linguistiques et musicaux (Patel & Peretz, 1997; Schön et al., 2005).

En effet, la plupart des chercheurs qui se sont intéressés aux liens ou aux dissociations entre langage et musique se sont basés sur la comparaison entre les performances de sujets sur deux types de matériel distincts : un matériel musical et un matériel linguistique (p. ex., Koelsch et al., 2002; Tervaniemi et al., 1999; Tervaniemi et al., 2000). Or, si l'on se place dans l'optique d'une comparabilité optimale du matériel verbal et musical, l'intérêt du chant réside dans sa structure même. En effet, dans le chant, le langage et la musique sont combinés au sein d'un même stimulus et donc appariés au niveau structurel, ce qui limite les risques de biais liés à des différences de complexité des tâches musicales et langagières. C'est pourquoi nous avons utilisé du matériel chanté dans l'ensemble des études de cette thèse visant à comparer les traitements phonétiques et mélodiques.

1. 1. 8. Présentation des études sur la comparaison entre traitements phonétiques et mélodiques dans le chant

Un des objectifs de cette introduction sur la comparaison entre langage et musique, qui ne se veut pas exhaustive, était de souligner les incohérences des résultats expérimentaux publiés jusqu'à présent. Une manière d'interpréter ces contradictions est de considérer que certaines dimensions linguistiques et musicales pourraient faire l'objet de traitements communs alors que d'autres seraient pris en charge de manière spécifique. Notre choix théorique de comparer les traitements linguistiques et musicaux les plus fondamentaux, c'est-à-dire les traitements phonétiques et de hauteur se justifie, premièrement, par le fait qu'il s'agit des dimensions constitutives de la parole et de la musique (Patel, 2008). Par conséquent, montrer une spécificité de traitement à ce niveau fondamental constituerait un argument extrêmement puissant en faveur de la nature biologique de cette spécificité. Ce choix est aussi lié au nombre limité d'études centrées sur ces dimensions et par le fait que les recherches existantes utilisaient des stimuli physiquement différents et des méthodologies variables pour comparer les phonèmes et les hauteurs, ce qui ne permet pas d'en tirer des conclusions claires. L'utilisation du chant a pour objectif de pallier ces limitations. Enfin, une constante de ce travail de thèse a été le choix d'adapter des méthodes expérimentales classiques, et donc bien établies, aux nouveaux objets d'études que sont les interactions entre phonèmes et hauteurs musicales.

Par exemple, dans la **première étude** décrite dans la section expérimentale de ce travail (*Processing interactions between lyrics and tunes: Vowels sing but consonants speak*), nous avons utilisé le paradigme de classification accélérée de Garner (1974) pour examiner les interactions entre le traitement en temps réel des caractéristiques phonologiques de non-mots et les intervalles mélodiques sur lesquels ils étaient chantés. Dans ce but, des participants non-musiciens devaient classifier ces stimuli chantés soit sur base de leurs « paroles », soit sur base de leur contour mélodique. Afin d'examiner l'hypothèse selon laquelle les consonnes sont traitées de manière plus séparée du contour mélodique que les voyelles, nous avons comparé des non-mots variant au niveau consonantique et des non-mots variant au niveau vocalique. L'approche expérimentale utilisée et ses implications théoriques sont décrites de manière détaillée dans la deuxième partie de l'introduction (Lidji, 2007), qui fait le point sur les connaissances actuelles quant aux concepts d'intégralité et de séparabilité dimensionnelles. Cet article théorique vise également à démontrer l'intérêt de cette approche classique pour l'étude des interactions entre langage et musique.

La **deuxième étude** (*Integration effect and illusory conjunctions in memory for lyrics and tune: Vowels sing but consonants swing*) emploie elle aussi une méthode classique, à savoir une tâche de reconnaissance à choix forcé (Serafine, Crowder, & Repp, 1984; Thompson, Hall, & Pressing, 2001) de stimuli chantés où les distracteurs incluaient des stimuli *mismatch*, combinant les paroles et la mélodie de deux stimuli encodés séparément pendant la phase d'apprentissage. Nous avons exploité deux phénomènes bien connus, l'effet d'intégration (Morrongiello & Roes, 1990; Serafine et al., 1984; Serafine, Davidson, Crowder, & Repp, 1986) et l'occurrence de conjonctions illusoires (Treisman & Schmidt, 1982), pour explorer les interactions entre les traces mnésiques des phonèmes et intervalles. À nouveau, le caractère consonantique et vocalique de la dimension linguistique a été manipulé afin de spécifier si les consonnes sont plus séparables que les voyelles, mais cette fois en mémoire.

La **troisième étude** (*Early integration of vowel and pitch processing: A Mismatch Negativity study*) utilise la composante Mismatch Négativité (MMN) (Näätänen, 1992) des potentiels évoqués cérébraux afin d'établir si l'intégration entre voyelles et hauteurs musicales observée dans les études comportementales brièvement présentées ci-dessus opère de manière précoce ou à un stade de traitement plus tardif et éventuellement stratégique. Si les voyelles et les hauteurs musicales sont prises en charge par des substrats neuronaux communs, nous nous attendrions à ce que les MMN évoquées par des déviations vocaliques et des déviations de hauteur au sein de séquence de voyelles chantées ne soient pas additives mais, au contraire, interactives.

En résumé, dans les trois premières études de cette thèse, le traitement de la hauteur musicale, isolée ou au sein d'intervalle mélodiques, a été comparé au traitement des phonèmes, car ceux-ci sont les éléments de base de la parole comme la hauteur est l'élément de base de la musique. Parce que les consonnes semblent, pour des raisons acoustiques et fonctionnelles, plus spécifiques à la parole que les voyelles, nous prédisons l'apparition de différences systématiques entre voyelles et consonnes en ce qui concerne leurs interactions avec la mélodie.

1. 1. 9. Espace musical

1. 1. 9. 1. Transferts entre traitements musicaux et non-musicaux

La lecture de ce qui précède démontre clairement la place prépondérante occupée par la comparaison entre musique et langage dans le domaine, en plein essor, de la psychologie cognitive de la musique. Cependant, l'on ne peut ignorer que les processus musicaux pourraient établir des ponts avec d'autres habiletés cognitives. Ces connexions pourraient d'ailleurs être bi-directionnelles : s'il est possible que les compétences musicales se développent à partir de certaines compétences non-musicales, la pratique de la musique pourrait également avoir un effet positif sur ces dernières (pour des exemples au niveau du langage, voir Besson et al., 2007; Ho, Cheung, & Chan, 2003; Wong et al., 2007). Les traitements visuo-spatiaux constituent une des compétences non-musicales ayant fait l'objet d'investigations prometteuses dans ce sens.

L'étude des interactions entre traitements musicaux, spatiaux et numériques, s'est essentiellement développée dans la dernière décennie. L'engouement pour cette question est au moins partiellement lié à l'impact médiatique hors du commun des expériences ayant révélé l' « Effet Mozart » (Rauscher, Shaw, & Ky, 1993, 1995, Steele, Dalla Bella, Peretz, Dunlop, Dawe, Humphrey, et al., 1999) et à la légende scientifique qu'elles ont entraînée (Bangerter & Heath, 2004). Cet effet temporaire de l'écoute d'une sonate de Mozart sur la performance à des tâches de quotient intellectuel (QI) visuo-spatial a constitué l'un des premiers arguments en faveur de transferts entre traitements musicaux et visuo-spatiaux. Les auteurs (Rauscher et al., 1995) ont fourni une interprétation forte du phénomène : selon eux, l'écoute de la musique de Mozart amorcerait l'activation des neurones impliqués dans les traitements visuo-spatio-temporels. La portée de ces affirmations est toutefois limitée par divers facteurs, tels que les difficultés à répliquer l'effet (McKelvie & Low, 2002; Steele, Bass, & Crook, 1999) ou le fait qu'il puisse être expliqué par des facteurs non spécifiques à la musique tels qu'un degré d'activation (*arousal*) plus élevé suite à l'écoute musicale (Thompson, Schellenberg, & Husain, 2001) ou un effet de préférence (Nantais & Schellenberg, 1999).

Le débat sur l'effet Mozart est loin d'être clos, mais il n'est pas l'objet principal de notre discussion. Cet effet nous intéresse surtout parce qu'il a constitué un incitant exceptionnel pour étudier les interactions entre traitements musicaux et visuo-spatiaux et qu'il a souvent été confondu avec d'autres types de recherches examinant, cette fois, les liens entre une pratique musicale active et les compétences cognitives générales (Rauscher & Hinton, 2006).

Plusieurs études font état d'effets à long terme de la pratique musicale pendant l'enfance sur les compétences spatiales et numériques (Cheek & Smith, 1999; Costa-Giomi, 2004; Gardiner, Fox, Knowles, & Jeffrey, 1996; Graziano, Peterson, & Shaw, 1999). Certaines de ces études sont toutefois critiquables au niveau méthodologique car la pratique musicale est généralement confondue avec des variables scolaires et socio-économiques qui pourraient également favoriser la performance dans des tâches numérico-spatiales (Schellenberg, 2001). Cependant, Schellenberg (2004) a montré, dans une étude mieux contrôlée, une augmentation plus importante du QI général chez des enfants de 6 ans ayant suivi des cours de musique pendant un an par rapport à des groupes contrôles qui avaient suivi des cours de théâtre ou aucun cours complémentaire. De plus, les cours de musique avaient des effets prolongés à la fois sur le QI et sur la réussite scolaire, y compris quand les variables socio-économiques étaient contrôlées statistiquement (Schellenberg, 2006).

Des données récentes (Sluming, Brooks, Howard, Downes, & Roberts, 2007) indiquent, de plus, que la performance de musiciens dans des tâches de rotation mentales à trois dimensions est similaire à celle de participants ayant suivi un entraînement intensif dans de telles tâches et que leur score était corrélé au nombre d'années d'expérience musicale. Dans le même ordre d'idées, Brochard, Dufour et Després (2004) ont observé que des adultes musiciens présentent une performance supérieure à celle de non-musiciens dans des tâches évaluant la perception et l'imagerie visuo-spatiales. La rapidité de la discrimination visuelle des participants était évaluée dans des tâches où ils devaient indiquer de quel côté d'une ligne horizontale ou verticale se trouvait un point lumineux. Dans une autre condition, dite d'imagerie mentale, la ligne de référence devait être imaginée par les sujets. Les musiciens étaient plus rapides que les non-musiciens, et ce, particulièrement dans la condition d'imagerie mentale verticale. En bref, la pratique musicale intense semble bien avoir un effet bénéfique sur les habiletés visuo-spatiales, un effet qui serait, selon Brochard et al. (2004) conjointement dû à la stimulation des habiletés perceptivo-motrices par la pratique d'un instrument et à l'activation des traitements visuo-spatiaux verticaux par la lecture musicale.

Cette idée est à mettre en correspondance avec le patron de performance opposé : Douglas et Bilkey (2007) rapportent, parmi d'autres résultats qui seront présentés dans la discussion générale, que des sujets amusiques congénitaux présentent une performance significativement inférieure à celle de sujets contrôles musiciens et non musiciens dans une tâche de rotation mentale. Ils en concluent que le traitement mélodique pourrait être lié à des traitements de nature spatiale.

1.1. 9. 2. Associations spatiales de codes de réponse

Il semble donc bien que la pratique et, éventuellement, l'écoute musicale exercent des effets bénéfiques sur les compétences visuo-spatiales. Toutefois, à l'exception notable des études sur l'Effet Mozart et de celle de Brochard et ses collègues (2004), aucune explication n'a été fournie pour ces transferts. Une hypothèse séduisante pour expliquer l'effet de la musique sur le traitement spatial serait que les processus musicaux et visuo-spatiaux partagent des composants de traitement. Cette interprétation, si elle peut sembler similaire à celle de Rauscher et al. (1995), est cependant plus prudente car elle ne fait pas référence à l'idée de traitements spécifiques : il se peut que les traitements spatiaux interviennent dans plusieurs compétences dont la musique ferait partie au même titre que, par exemple, la cognition numérique.

Le lien entre traitements spatiaux et numériques repose essentiellement sur l'observation d'associations spatiale de codes de réponse ou effet SNARC (Spatial Numerical Association of Response Codes, Dehaene, Bossini et Giraux, 1993). Ainsi, dans des tâches comme le jugement de parité à partir de nombres présentés visuellement avec des réponses manuelles par pression de boutons disposés horizontalement, les réponses aux petits nombres sont plus rapides du côté gauche que du côté droit, et, inversement, les réponses aux grands nombres sont plus rapides du côté droit que du côté gauche (Dehaene, Bossini, & Giraux, 1993; Fias, 2001; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996; Nuerk, Wood, & Willmes, 2005; Saffran et al., 1999). Un effet SNARC vertical a également été observé à plusieurs reprises (Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2006; Ito & Hatta, 2004; Schwarz & Keus, 2004). L'observation de ces effets lors de tâches où aucun jugement explicite de magnitude n'est exigé (Fias et al., 1996; Fias, Lauwereyns, & Lammertyn, 2001) suggère que ces associations spatiales sont activées automatiquement (Gevers et al., 2006). L'effet SNARC se manifeste également par une orientation de l'attention vers le champ visuel droit ou gauche (Fischer, Castel, Dodd, & Pratt, 2003) et ne semble pas lié à l'effecteur utilisé pour répondre (Fischer, Warlop, Hill, & Fias, 2004; Schwarz & Keus, 2004). Cette dernière proposition a toutefois récemment été remise en question par des études montrant, dans une tâche avec réponse manuelle, que la position des mains interfère avec l'effet SNARC (Müller & Schwarz, 2007b). Enfin, tel qu'indiqué par des données comportementales

(Müller & Schwarz, 2007a) et électroencéphalographiques (Keus, Jenks, & Schwarz, 2005), l'effet semble prendre sa source à l'étape de sélection de la réponse.

Sur le plan théorique, l'effet SNARC est généralement interprété comme reflétant l'existence d'une représentation mentale de nature spatiale (Dehaene, 2001; Fias & Fischer, 2005), où les nombres sont conceptualisés comme des positions sur une «ligne mentale numérique» orientée de gauche à droite ou de bas en haut. Sur base de données neuropsychologiques et d'imagerie cérébrale, il semble que cette association entre représentations numériques et spatiales soit liée à des réseaux neuronaux communs situés dans le cortex pariétal (pour une revue récente, voir Hubbard et al., 2005). En effet, des patients héminégligents présentent un biais « vers la droite » lorsqu'ils doivent indiquer la médiane de deux chiffres, c'est-à-dire qu'ils répondent par un chiffre de plus grande magnitude que celui attendu (Zorzi, Priftis, & Umiltà, 2002). De plus, les aires cérébrales pariétales activées lors de la comparaison de nombres (Sandrini, Rossini, & Miniussi, 2004) correspondent de manière étonnante à celles dévolues aux activités visuo-spatiales (p. ex., Andersen, Snyder, Bradley, & Xing, 1997; Colby & Goldberg, 1999; Göbel, Walsh, & Rurshworth, 2001). Ceci nous conduit à un lien surprenant avec la musique : ces régions peuvent également être activées lors de la comparaison de hauteurs musicales (Schmithorst & Holland, 2003).

1. 1. 9. 3. Associations spatiales pour les hauteurs musicales

Il se pourrait donc que les hauteurs musicales évoquent des associations spatiales similaires à celles classiquement attribuées à la séquence des chiffres (Dehaene et al., 1993) mais aussi à d'autres séquences verbales comme les lettres, les jours de la semaine ou les mois de l'année (Gevers, Reynvoet, & Fias, 2003, 2004). L'objectif de la **quatrième étude** de cette thèse était de tester cette hypothèse au moyen de six expériences (Lidji, Kolinsky, Lochy, & Morais, 2007). À nouveau, nous avons adapté une méthode expérimentale classique, celle qui a mis en évidence l'effet SNARC, pour l'appliquer à un nouvel objet de recherche : le traitement des hauteurs musicales.

Nous avons confronté des participants non-musiciens et musiciens à des tâches de jugement de timbre ou de hauteur de notes de musique, avec un dispositif de réponse disposé horizontalement ou verticalement. Nous nous attendions, en accord avec les résultats obtenus par Rusconi, Kwan, Giordano, Umiltà et Butterworth (Rusconi et al., 2006) dans une expérience similaire dont nous n'avions pas connaissance à l'époque de la réalisation de cette étude, à observer un effet d'associations spatiales de codes de réponses pour ces hauteurs. Autrement dit, la séquence des notes de musique devrait activer des représentations de nature spatiale le long d'un axe horizontal ou vertical, avec les notes graves à gauche (ou en bas) et les notes aigues à droite (ou en haut) de cette ligne mentale hypothétique. Si la lecture musicale ou la pratique d'un instrument a pour effet de renforcer les traitements spatiaux, comme suggéré par Brochard et al. (2004), ces associations devraient être plus robustes chez les participants musiciens. C'est ce qui a effectivement été observé par Rusconi et ses collègues (2006) : les musiciens manifestaient un effet dit SMARC (*Spatial Musical Association of Response Codes*) horizontal plus important que les non-musiciens. Par contre, l'effet vertical était similaire dans les deux groupes, ce qui suggère que la représentation intuitive des hauteurs musicales serait orientée de bas en haut. Cet effet associé à la dimension de hauteur pouvait, toutefois, difficilement être qualifié de musical en raison de la simplicité des stimuli. Afin de vérifier si les associations spatiales découvertes pour les hauteurs isolées se généralisent à des stimuli musicaux plus complexes, nous avons appliqué ces mêmes tâches à des intervalles mélodiques ascendants et descendants.

La découverte de tels effets d'associations spatiales pour des stimuli musicaux apporterait des arguments supplémentaires en faveur des liens fréquemment décrits entre traitements musicaux et spatiaux. Dans le cadre de cette thèse, cette étude éclaire aussi d'un jour nouveau la question de la spécificité du traitement des hauteurs musicales en utilisant un point de comparaison différent de celui choisi dans les autres expériences.

1. 1. 10. Structure de la thèse

Le titre de cette thèse de doctorat met l'accent sur la comparaison entre musique et langage. En réalité, nous nous sommes placés à la fois dans une perspective plus spécifique et plus large que cet intitulé général ne le laisse entendre.

Notre approche est plus spécifique car nous avons focalisé notre attention sur le traitement de la hauteur pour la musique, et sur le traitement des phonèmes pour le langage. La comparaison entre ces dimensions musicales et linguistiques fait l'objet des **études 1, 2 et 3** de la section expérimentale de ce travail. Cette section expérimentale est précédée d'un complément à la présente introduction où nous passons en revue la littérature sur le paradigme de classification accélérée (Garner, 1974) dans ses applications à l'étude des interactions entre paroles et mélodies dans le chant.

Notre perspective est également plus large, puisque, dans **l'étude 4**, nous avons examiné si la séquence des hauteurs musicales donne lieu à des associations spatiales analogues à celles observées pour d'autres séquences ordonnées, telle que la séquence des chiffres. Cette dernière étude élargit donc notre point de vue en examinant la spécificité du traitement musical en sortant des sentiers battus de la comparaison classique entre musique et langage.

1. 2. Article 1

Intégralité et séparabilité : Revue et application aux interactions entre paroles et mélodies dans le chant [i]

[i] D'après Lidji, P. (2007). Intégralité et séparabilité : Revue et application aux interactions entre paroles et mélodies dans le chant. *L'Année Psychologique, 107* (4), 659-694.

Résumé

L'objectif de cet article est de montrer la pertinence de l'approche des interactions dimensionnelles (Garner, 1974) pour étudier les relations entre paroles et mélodies dans le traitement du chant. Une revue de la littérature sur l'application du paradigme de Garner aux dimensions auditives permet de conclure que cette approche concourt à dépasser les contradictions des études antérieures sur le chant et à préciser à quel niveau de traitement ces interactions ont lieu. Son avantage est de permettre d'examiner les interactions entre différentes caractéristiques des paroles et de la mélodie au moyen d'une seule méthode bien validée, contrairement aux études antérieures qui employaient des techniques variables. Enfin, cette approche appliquée à des participants sains évite les critiques émises envers les doubles dissociations.

Integrality and separability:
Review and application to the interactions between lyrics and tune in songs

The aim of this paper is to show the relevance of Garner's (1974) dimensional interactions paradigm to investigate the relations between language and music in song processing. A review of the literature on Garner's paradigm use with auditory dimensions suggests that the dimensional interactions approach goes beyond the contradictory results of former studies on song processing and enables one to specify on which processing level the interactions occur. Its advantage is to allow one studying the interactions between several dimensions of the lyrics and the tune with one, valid, method. On the contrary, former studies presented much methodological differences. Finally, this approach is applied to healthy subjects and thus avoids the critics against the double dissociation principle.

1. 2. 1. Introduction

Lorsque vous écoutez le dernier titre à la mode, traitez-vous les paroles et la mélodie de cette chanson comme deux dimensions distinctes ? Ou, au contraire, la chanson constitue-t-elle un objet global, dont les paroles et la musique sont indissociables ? Cette question peut sembler futile. Pourtant, la réponse qui y sera apportée a des implications fondamentales pour un problème au cœur de l'actualité en psychologie cognitive : celui de la spécificité respective du traitement de la musique et du langage.

En effet, depuis plusieurs décennies, une controverse fait rage entre les défenseurs de l'idée selon laquelle le *langage* fait l'objet d'un traitement spécifique (e.g., Chomsky, 1957) et les auteurs qui considèrent que le langage partage les mêmes principes organisationnels et fonctionnels que d'autres habiletés cognitives et ne serait qu'un épiphénomène de nos capacités cognitives globales (e.g., Tomasello, 1998). Plus récemment, le même problème a été abordé dans une autre perspective : la *musique* constitue-t-elle une fonction cognitive à part entière (Peretz, 2006, Mithen 2005) ou est-elle un « sous-produit » de nos capacités cognitives générales, un *auditory cheesecake* (Pinker, 1997) qui émergerait d'autres fonctions? Loin de se limiter, pour le langage, au cercle restreint des linguistes et psycholinguistes et, pour la musique, à celui des musicologues, musiciens et psychologues spécialisés, ce débat est susceptible d'interpeller tout scientifique qui s'intéresse à la modularité de l'esprit humain (Fodor, 1983 ; voir aussi Fodor, 2000).

Le domaine des sciences cognitives de la musique est aujourd'hui en plein essor, comme l'indique le nombre croissant de conférences (e.g., Avanzini, Koelsch, Lopez, & Majno, 2005 ; Zatorre & Peretz, 2001) et de publications sur ce thème depuis le début du millénaire. Ainsi, un numéro spécial de *Nature Neurosciences* consacré aux liens entre musique et cerveau a été publié en juillet 2003, tandis que la revue *Cognition* consacrait son numéro de mai 2006 à la cognition musicale. Une des raison de cet intérêt pour l'étude scientifique de la musique est sans nul doute son omniprésence dans la civilisation humaine qui pourrait suggérer que, contrairement à des croyances répandues, elle n'est pas un produit culturel mais une fonction fondamentale de l'être humain (Mithen, 2005 ; Peretz, 2006). Un courant important de ces recherches est basé sur l'idée que le traitement de la musique serait opéré par des modules cognitifs spécifiques (Peretz, 2002 ; Peretz & Coltheart, 2003 ; Peretz &

Zatorre, 2005) qui auraient une origine biologique et évolutive (Fitch, 2006 ; Peretz, 2006). En particulier, l'existence d'un module spécifique à l'encodage tonal de la hauteur (Peretz, 2006) constitue une hypothèse séduisante. Comme la musique présente de nombreux points communs avec le langage (Besson et Schön, 2001), à commencer par l'utilisation du canal auditif, la capacité à transmettre des émotions, et l'existence d'éléments de base (les notes et les phonèmes) permettant un nombre quasi illimité de productions, la comparaison de ces deux fonctions constitue une approche de choix pour tester la modularité de l'une comme de l'autre.

Dans la suite de ce texte, l'argument principal en faveur de l'idée de la spécificité du traitement de la musique et les critiques émises à son égard seront présentés. Nous verrons en quoi l'étude du chant constitue une alternative intéressante pour mettre en évidence la spécificité ou, au contraire, la non spécificité du traitement de diverses dimensions linguistiques et musicales. Enfin, l'utilisation d'une approche théorique et méthodologique bien connue en psychologie cognitive, à savoir la théorie des interactions dimensionnelles de Garner (1974), sera proposée pour étudier ces interactions. Afin de démontrer la pertinence de cette approche pour étudier la perception du chant, ses implications théoriques et pratiques seront examinées au cours d'une revue de la littérature.

1. 2. 2. La double dissociation prouve-t-elle la spécificité du traitement de la musique ?

Un des arguments les plus convaincants en faveur de l'idée de la spécificité du traitement de la musique provient d'études neuropsychologiques ayant mis en évidence une double dissociation entre deux troubles cognitifs acquis : l'aphasie, dans laquelle le traitement du langage est altéré et celui de la musique est préservé (Hébert, Racette, Gagnon, & Peretz, 2003; Peretz, Gagnon, Hébert, & Macoir, 2004 ; Racette, Bard, & Peretz, 2006), et l'amusie acquise, dans laquelle le traitement de la musique est altéré sans atteinte des compétences langagières (Hébert & Peretz, 2001 ; Peretz, 2002). Cependant, certains auteurs (Chater, 2003 ; Plaut, 1995 ; Van Orden, Pennington, & Stone, 2001) ont remis en question les conclusions de modularité généralement tirées des doubles dissociations. Ainsi, Van Orden et al. (2001) dénoncent le caractère circulaire du principe de double dissociation : selon eux, une double dissociation indique la modularité de sous-systèmes cognitifs à la seule condition que l'on se situe dans un cadre théorique de type modulaire. De plus, ces auteurs soulignent la tendance supposée des défenseurs de la double dissociation à rechercher des cas purs et à rejeter ceux ne correspondant pas à leurs conceptions théoriques. Enfin, Plaut (1995) a montré que des patrons comportementaux similaires à ceux d'une double dissociation peuvent être simulés par des réseaux connexionnistes, qui ne sont pourtant pas configurés sous forme de modules. Selon cet auteur, il serait donc hardi d'inférer un fonctionnement modulaire à partir de l'étude de doubles dissociations.

1. 2. 3. L'étude du chant : une solution miracle ?

Une manière de répondre aux critiques visant les doubles dissociations consisterait à confirmer l'indépendance des traitements musicaux et linguistiques autrement que via l'étude de patients. Plusieurs études ont, ainsi, comparé le traitement du langage et celui de la musique chez des sujets sains (par exemple, Knosche, Maess, Nakamura, & Friederici, 2005 ; Koelsch, Kaspers, Sammler, Schulze, Gunter, & Friederici, 2004 ; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). Cependant, malgré les contrôles effectués sur le matériel, les données obtenues sont toujours issues de stimuli linguistiques et musicaux par essence physiquement différents, ce qui pose des problèmes évidents de comparabilité des résultats.

Une alternative intéressante consiste à confronter des participants sains à des stimuli chantés afin d'évaluer s'ils en traitent les paroles et la mélodie de manière indépendante ou interactive. En effet, le chant peut être considéré comme un matériel écologique et multidimensionnel (Schön, Gordon, & Besson, 2005), constitué de deux dimensions principales : la dimension linguistique et la dimension musicale. Ces dimensions sont physiquement combinées dans le chant, et il est connu qu'elles influent l'une sur l'autre. Ainsi, la hauteur à laquelle des voyelles sont chantées tend à en modifier les caractéristiques spectrales (pour une revue, voir Astesano, Schön et Besson, 2004). Ceci implique que la mise en évidence de l'indépendance des paroles et de la mélodie dans le chant constituerait un argument bien plus fort que le même résultat sur de la musique et du langage physiquement distincts.

La dimension linguistique comme la dimension musicale du chant peuvent également être subdivisée (Besson et Schön, 2001). D'une part, les paroles de chansons comportent une sémantique, une syntaxe, mais elles peuvent également être étudiées à des niveaux inférieurs comme celui du mot, du morphème ou du phonème. D'autre part, la musique comporte des caractéristiques harmoniques, un contour mélodique pouvant être divisé en intervalles et en notes, et des caractéristiques rythmiques et de mesure (Peretz & Coltheart, 2003). Ces diverses dimensions de la musique et du langage ont souvent été comparées (Besson, 1998 ; Besson & Schön, 2001) mais leurs interactions dans le chant n'ont été que rarement étudiées.

Deux groupes d'études font exception à cette règle. Le premier groupe est constitué des études, citées plus haut, où la reconnaissance des paroles et des mélodies des mêmes chansons a été comparée chez des patients amusiques (Hébert & Peretz, 2001 ; Peretz, Belleville, & Fontaine, 1997). Elles mettent en évidence la dissociation attendue entre la reconnaissance des paroles (préservée) et celle de la mélodie (altérée). Mais ces études reposent toujours sur les doubles dissociations tant décriées par certains auteurs (Chater, 2003; Van Orden et al., 2001).

Le second groupe d'études a évité cette critique potentielle en étudiant le traitement du chant chez des sujets sains, au moyen de méthodes comportementales et électrophysiologiques (Besson, Faita, Peretz, Bonnel, & Requin, 1998 ; Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001 ; Bonnel, Faita, Peretz, & Besson, 2001 ; Crowder, Serafine, & Repp, 1990 ; Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005 ; Schön et al., 2005 ; Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). Ces études ont conduit à des résultats apparemment contradictoires selon le matériel et les techniques d'investigation utilisés. Pour une partie d'entre elles, les résultats suggéraient l'indépendance du traitement des paroles et des mélodies (Besson et al., 1998 ; Bonnel et al., 2001), tandis que d'autres soutenaient l'hypothèse selon laquelle ces dimensions ne seraient pas indépendantes mais, au contraire, interactives (Bigand et al., 2001 ; Poulin-Charronnat et al., 2005 ; Schön et al., 2005 ; Serafine et al., 1984, 1986). Selon ce dernier point de vue, le traitement de la mélodie serait influencé par les caractéristiques des paroles et réciproquement.

Les études suggérant l'indépendance des paroles et de la mélodie ont utilisé un matériel constitué d'extraits d'opéra pouvant se terminer par un mot congruent ou incongru. Ce mot était chanté sur une note congruente ou non par rapport à la mélodie qui le précédait. Les incongruités sémantiques et mélodiques conduisaient à des potentiels évoqués de sources et de latence différentes. De plus, en situation de double incongruité, c'est-à-dire lorsque le dernier mot et la dernière note étaient tous deux incongrus par rapport au contexte, les potentiels évoqués étaient additifs (Besson et al., 1998). Par ailleurs, la performance en attention divisée était aussi bonne qu'en attention sélective pour ces stimuli, et la performance sur les paroles n'était pas corrélée à la performance sur la mélodie (Bonnel et al., 2001). Tous ces éléments suggèrent l'indépendance des traitements sémantique et mélodique.

Bigand et al. (2001) et Poulin-Charronnat et al. (2005) ont obtenu des résultats très différents avec un paradigme d'autant plus puissant que la composante harmonique du matériel musical n'y est traitée que de manière implicite. Les auteurs présentaient aux sujets un matériel musical composé d'une séquence d'accords se terminant soit par l'accord le plus attendu (la tonique), soit par un accord consonant, mais moins attendu (la sous-dominante). Or, la tâche portait sur une tout autre dimension : la voyelle (Bigand et al., 2001) ou le mot (Poulin-Charronnat et al., 2005) sur lesquels l'accord final était chanté. Malgré la consigne d'ignorer la dimension musicale, les participants étaient plus performants quand la dimension linguistique cible était chantée sur l'accord le plus attendu. Ceci suggère que le traitement phonologique (Bigand et al., 2001) et le traitement lexical des paroles (Poulin-Charronnat et al., 2005) ne seraient pas indépendants du traitement harmonique de la musique.

En outre, Schön et al. (2005) ont étudié les interactions entre les paroles et la mélodie sur des paires de mots chantés. Ces auteurs ont adapté une méthode déjà utilisée avec succès par Peretz et Kolinsky (1993) pour mettre en évidence les interactions entre le rythme et la mélodie. Dans l'étude de Schön et al. (2005), les participants devaient décider si soit la mélodie, soit les paroles de deux mots trisyllabiques chantés étaient les mêmes. La dimension non pertinente [(1)] était également manipulée, menant à quatre types de paires possibles : paroles et mélodies identiques, paroles identiques et mélodies différentes, paroles différentes et mélodies identiques, ou paroles et mélodies différentes. Le traitement des paroles et celui de la mélodie ne semblaient pas indépendants puisque, d'une part, les participants jugeaient plus lentement la dimension pertinente lorsque la dimension non pertinente ne correspondait pas à la même réponse et que, d'autre part, la variation de la dimension non pertinente engendrait des composantes électrophysiologiques

[(1)] Par exemple, les paroles dans le cas où le participant doit porter attention à la mélodie.

caractéristiques de l'orientation attentionnelle. Ces données comportementales et électrophysiologiques suggèrent que cette variation n'a pu être ignorée.

Enfin, une série d'études (Crowder et al., 1990, Serafine et al., 1984, 1986) indique que les paroles et la mélodie de chansons seraient également associées en mémoire. Les résultats d'une tâche de reconnaissance à choix forcé suggèrent que la reconnaissance de la mélodie d'une chanson apprise récemment est facilitée lorsque cette mélodie est appariée aux mêmes paroles que pendant de la phase d'apprentissage. Cet effet a été intitulé « effet d'intégration » par les auteurs. Toutefois, Crowder et al. (1990) ont montré que la contiguïté temporelle des paroles et de la mélodie des chansons pourrait fortement contribuer à cet effet, puisqu'il a également été mis en évidence lorsque les paroles étaient lues et la mélodie fredonnée en musique de fond. Plus récemment, Peretz, Radeau, et Arguin (2004) ont constaté que la présentation de la mélodie d'une chanson connue amorce la reconnaissance de ses paroles et réciproquement. L'effet était présent même en situation d'amorçage inversé au niveau temporel, c'est-à-dire lorsque la cible provenait du début de la chanson et l'amorce de la suite de celle-ci. Ces résultats suggèrent, contrairement aux affirmations de Crowder et al. (1990), que les liens entre les paroles et la mélodie d'une chanson se situraient à un niveau plus abstrait que celui de simples contingences temporelles.

L'ensemble des données présentées ci-dessus dessine donc un tableau ambigu : dans des extraits d'opéra, les paroles et la mélodie semblent traitées de manière indépendante (Besson et al., 1998 ; Bonnel et al., 2001). Pourtant, un patron d'interaction a été observé avec un matériel constitué d'un seul mot chanté (Schön et al., 2005) et dans des situations où l'influence de la dimension musicale est étudiée de manière implicite (Bigand et al., 2001 ; Poulin-Charronnat et al., 2005), ainsi que lors de la mémorisation de chansons (Crowder et al., 1990, Peretz et al., 2004, Serafine et al., 1984, 1986).

L'on peut cependant se demander si les mêmes aspects du langage sont traités dans les différentes études portant sur le traitement du chant en temps réel. En effet, dans les expériences de Besson et al. (1998) et de Bonnel et al. (2001), un traitement sémantique est réalisé, alors que chez Bigand et al. (2001), Poulin-Charonnat et al. (2005) et Schön et al. (2005), la tâche peut être réalisée via un traitement phonologique ou lexical du mot. D'autre part, certaines études (Besson et al., 1998, Schön et al., 2005) utilisent une approche comportementale associée à des méthodes électrophysiologiques, méthodes qui ont l'avantage de nous informer sur le traitement précoce de l'information, tandis que d'autres (Bigand et al., 2001 ; Poulin-Charonnat et al., 2005) se basent uniquement sur des données comportementales, qui peuvent refléter des processus plus tardifs.

Dans l'état actuel des connaissances, il n'est donc pas possible de déterminer si les paroles et les mélodies de chansons sont traitées de manière indépendante ou non. En effet, l'incohérence de ces résultats pourrait s'expliquer soit par la dimension linguistique traitée par les participants, soit par la variété des méthodes expérimentales utilisées. Une manière d'éviter ce problème consisterait à étudier les interactions entre les caractéristiques phonologiques et sémantiques des paroles et la mélodie de chansons au moyen d'une seule méthode qui aurait déjà fait ses preuves dans d'autres domaines de recherche.

1. 2. 4. Une nouvelle approche des interactions entre paroles et mélodies dans le chant : les interactions dimensionnelles (Garner, 1974)

Curieusement, les auteurs étudiant les interactions entre les diverses dimensions linguistiques et musicales dans la perception chant ne se sont jamais référés à un courant influent en psychologie cognitive : l'approche des interactions dimensionnelles essentiellement développée par Garner (1974) et poursuivie par de nombreux auteurs jusqu'à nos jours.

Garner (1974) a développé une distinction entre les dimensions dites « intégrales » et les dimensions « séparables », distinction qui a été appliquée à de nombreuses dimensions visuelles (Ganel & Goshen-Gottstein, 2002 ; Gottwald & Garner, 1975; Macmillan & Ornstein, 1998), auditives (Melara & Marks, 1990b ; Pitt, 1994 ; Repp & Lin, 1990 ; Wood, 1975) et même audio-visuelles (Ben Artzi & Marks, 1995 ; Kaufmann & Schweinberger, 2005 ; Melara, 1989). Cette théorie a également donné lieu à une modélisation en réseau connexionniste (Tijsseling & Gluck, 2002). De plus, les méthodes expérimentales visant à mettre en évidence ces deux types d'interactions ont été validées (Garner, 1974 ; Maddox, 1992) et de nombreuses réflexions théoriques sur ces questions ont été publiées dans les trente dernières années (Ashby & Townsend, 1986 ; Kemler Nelson, 1993 ; Melara & Marks, 1990a ; Melara, Marks, & Potts, 1993). L'objectif de cet article est de démontrer que les concepts bien connus d'intégralité et de séparabilité pourraient s'appliquer avec succès à la question récente et de plus en plus discutée des interactions entre langage et musique dans le chant. Après avoir défini les concepts d'intégralité et de séparabilité dimensionnelles, différents points de vue théoriques sur ces idées seront présentés. Les données principales sur les interactions entre dimensions auditives seront examinées. En conclusion, nous commenterons leurs implications pour l'étude des interactions entre paroles et musique dans le chant.

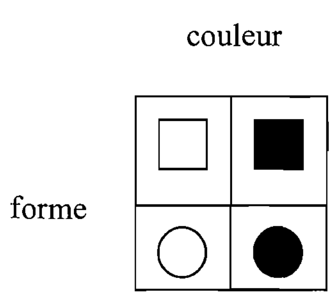
1. 2. 5. Intégralité et séparabilité dimensionnelles

1. 2. 5. 1. Les concepts d'intégralité et de séparabilité selon Garner

Selon Garner (1978), une dimension d'un stimulus est « tout attribut d'un stimulus qui possède au moins deux niveaux » (p. 98). Pour simplifier l'explication de ces concepts, prenons l'exemple de la couleur d'un objet : le même stimulus, par exemple un carré, peut exister dans plusieurs teintes. Mais des stimuli peuvent également varier conjointement sur plusieurs dimensions. Ainsi, un stimulus peut varier au niveau de la dimension « forme » (par exemple, être un cercle ou un carré) et de la dimension « couleur » (par exemple, être noir ou blanc, voir Figure 1).

Or, quand un objet est composé de dimensions multiples, celles-ci peuvent interagir d'au moins deux manières : elles peuvent être intégrales ou séparables.

Figure 1 : Illustration du concept de dimension.



Du point de vue théorique, Garner (1974) considère les dimensions d'un stimulus comme *intégrales* si elles n'ont pas de réalité psychologique pour l'observateur. Garner (1974, p. 119) estime même que le concept de « dimension » n'a pas de sens pour des stimuli faisant preuve d'intégralité : les dimensions sont psychologiquement fusionnées (Goldstone, 1999). Par conséquent, il est impossible de faire sélectivement attention à des dimensions intégrales. De plus, des dimensions intégrales ne peuvent exister l'une sans l'autre (Garner & Felfoldy, 1970). Par exemple, la hauteur et la sonie constituent des dimensions intégrales (Grau & Kemler Nelson, 1988) puisqu'un son doit posséder une hauteur pour que son intensité sonore soit perçue et réciproquement.

En revanche, les dimensions *séparables* ont une réalité psychologique pour l'observateur qui pourrait dès lors être capable d'y faire attention de manière sélective. Ces dimensions peuvent exister l'une sans l'autre (par exemple, une couleur peut exister sans forme).

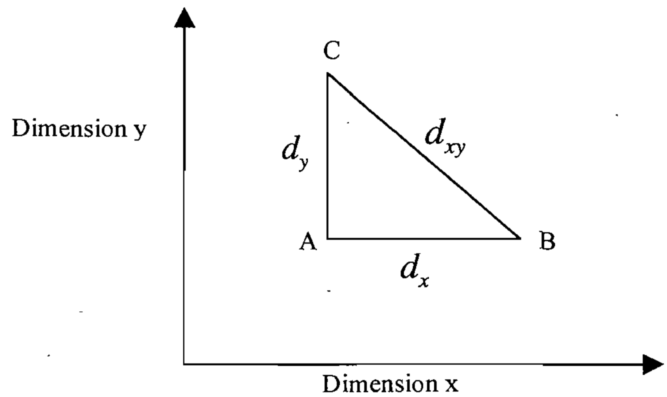
Il est important de souligner que ces interactions ne constituent pas des données du monde réel : le fait que des dimensions soient traitées comme *intégrales* ou *séparables* provient des traitements cognitifs effectués par l'observateur (Tijsseling & Gluck, 2002). Les concepts de séparabilité et d'intégralité ne sont donc pas des caractéristiques inhérentes au stimulus mais constituent des interactions *psychologiques*. La tâche des chercheurs est de déterminer quelles caractéristiques du matériel et des observateurs conduisent des dimensions à être traitées comme séparables ou intégrales.

1. 2. 5. 2. Les opérations convergentes

Les tâches ou « opérations convergentes » utilisées par Garner (1974) pour établir la distinction entre l'intégralité et la séparabilité peuvent être divisées en deux catégories : les tâches basées sur la métrique de l'espace dimensionnel et les tâches de classification.

En ce qui concerne la métrique, une tâche de jugement de similarité entre plusieurs stimuli est généralement utilisée (Hyman & Well, 1967 ; Shepard, 1964). Les participants doivent évaluer de manière numérique la similarité entre deux paires de stimuli variant sur une ou deux dimensions. Pour mettre en évidence la métrique utilisée, la similarité évaluée entre des paires de stimuli différant sur une dimension (par exemple, des cercles noirs et blancs) est comparée à celle de paires différant sur les deux dimensions (par exemple, un cercle noir et un carré blanc). Les observateurs peuvent comparer les stimuli sur base de deux métriques différentes : la métrique euclidienne et la métrique « city-block » (voir Figure 2).

Figure 2 : Illustration graphique des métriques euclidienne et city-block, d'après Garner (1974, p. 99).



$$d_{xy} = \sqrt{\left(d_x^2 + d_y^2\right)}$$

(métrique euclidienne)

$$d_{xy} = d_x + d_y$$

(métrique « city-block »)

Si l'on se trouve dans une métrique euclidienne, la distance évaluée entre des stimuli différant sur deux dimensions est liée à la distance évaluée quand les stimuli varient sur une seule dimension selon la même relation que celle qui lie l'hypoténuse à la longueur des côtés d'un triangle rectangle. En d'autres mots, l'observateur choisit le « chemin le plus court » pour comparer les stimuli B et C. Dans ce cas, il est impossible d'évaluer les dimensions séparément. La similarité de stimuli dont les dimensions sont intégrales est évaluée sur base de cette métrique euclidienne (Hyman & Well, 1967). Par contre, les dimensions séparables sont comparées sur base d'une métrique « city-block » (Shepard, 1964). Cela signifie que la distance évaluée entre des stimuli variant sur les deux dimensions est la somme des distances obtenues pour chaque dimension. En d'autres termes, la distance entre B et C sur la Figure 2 correspondrait à la somme des distances entre A et B et entre A et C.

La différence entre les dimensions intégrales et séparables est également mise en évidence par des tâches de classification. Il peut s'agir d'une tâche de classification libre où le participant effectue le classement selon ses propres critères. Ceux-ci permettent de définir les interactions dimensionnelles. En effet, Garner (1974) a constaté que les stimuli intégraux sont groupés en fonction de leur similarité globale alors que les stimuli aux dimensions séparables sont classés sur base des niveaux de chacune des dimensions.

Mais la plus célèbre tâche de classification développée par Garner (1974) est une tâche d'attention sélective, dite de classification accélérée, dans laquelle les sujets doivent classer des stimuli multidimensionnels sur base d'une seule dimension (Garner et Felfoldy, 1970, Garner, 1974). Dans ce cas, l'interaction est révélée par les temps de classification et les taux d'erreurs. Trois conditions de classification sont proposées sur un même ensemble de stimuli (voir Figure 1). Une première condition, dite de *contrôle*, consiste à ne faire varier que la dimension sur laquelle porte le classement (par exemple, la couleur dans la Figure 3), l'autre (la forme dans notre exemple) étant maintenue constante. Elle permet d'évaluer la performance des participants en l'absence de dimension « distractrice ».

Figure 3. Condition de CONTRÔLE : tâche sur la couleur.

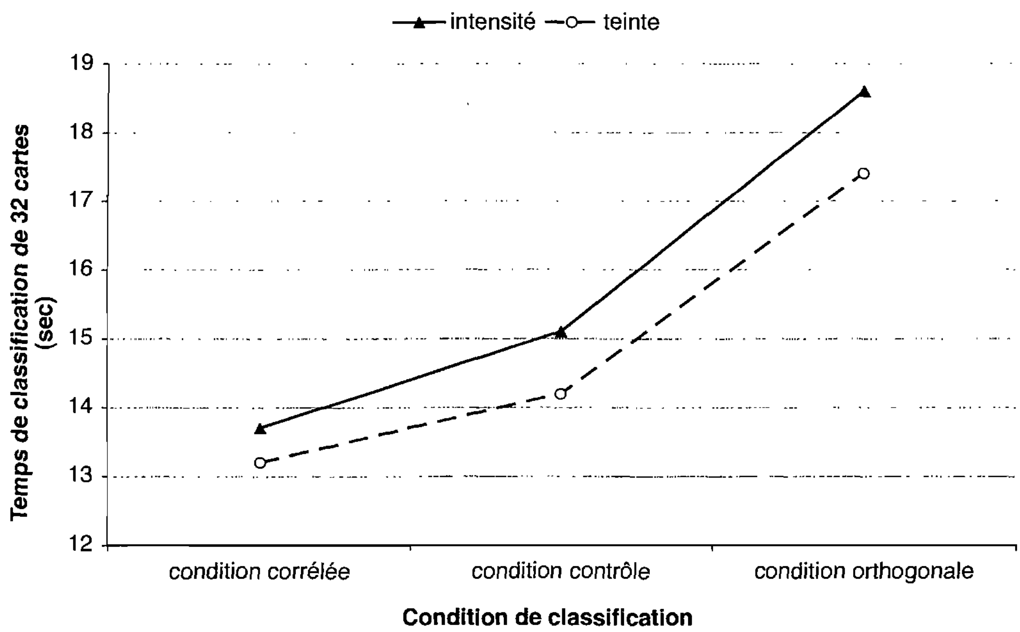
Dans une deuxième condition, dite *corrélée*, les deux niveaux de chaque dimension varient mais ils sont associés l'un à l'autre de manière systématique. Par exemple, la forme « cercle » sera toujours associée à la couleur « blanche » (Figure 4). Cette condition permet d'évaluer l'effet de la redondance d'information.

Figure 4. Condition CORRÉLÉE: tâche sur la couleur.

Enfin, dans la troisième condition, le set total ou *orthogonal* est présenté (Figure 1). Dans cette situation, les participants doivent ignorer les variations de la dimension non pertinente et grouper dans une même catégorie des stimuli différents (par exemple, les cercles et les carrés noirs). Cette situation permet de tester les capacités d'attention sélective pour une dimension donnée.

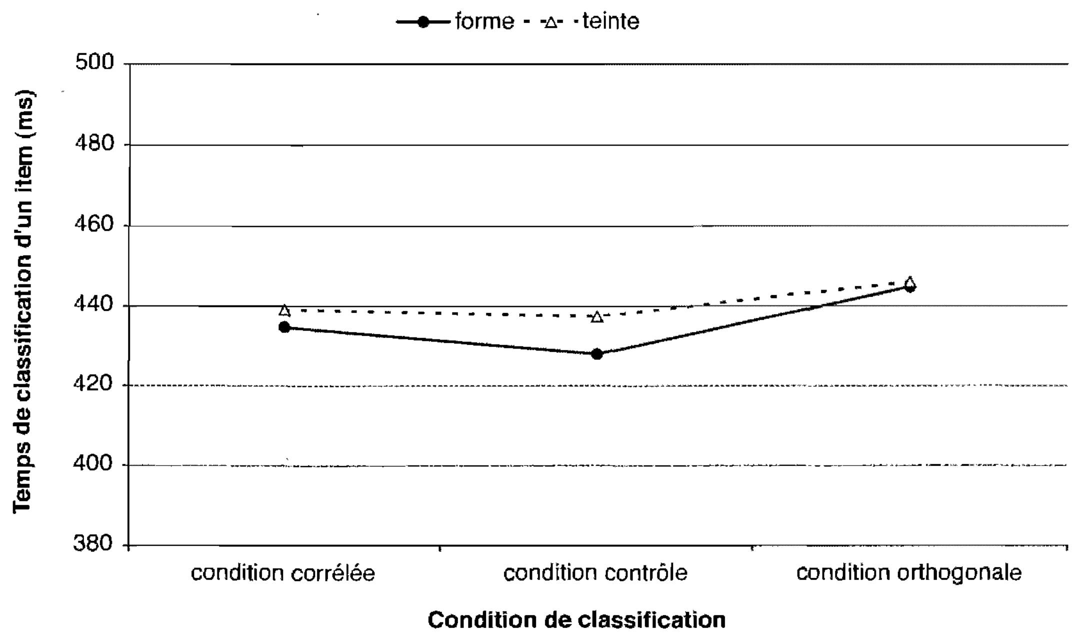
L'intégralité et la séparabilité dimensionnelles conduisent à des patrons de performance bien différenciés. Ainsi, dans le cas de dimensions intégrales, la redondance améliore la performance alors que des variations non pertinentes en condition orthogonale provoquent de l'interférence (Garner & Felfoldy, 1970). Un tel patron de résultat pour les dimensions de luminosité et d'intensité de couleurs est présenté sur la Figure 5). La facilitation en condition redondante est liée, selon Garner (1974) au fait que la corrélation augmente la dissimilarité globale des stimuli à classer. L'impossibilité de faire sélectivement attention à des dimensions intégrales expliquerait l'interférence en condition orthogonale.

Figure 5 : Temps moyen de classification accélérée de 32 cartes pour les dimensions intégrales d'intensité et de teinte de la couleur d'une pastille colorée, en sec. (d'après les données de Garner & Felfoldy, 1970).



En ce qui concerne les dimensions séparables, les auteurs n'observent ni facilitation en condition corrélée, ni interférence en condition orthogonale (Garner & Felfoldy, 1970 ; Maddox & Ashby, 1996, voir Figure 6). En effet, l'attention sélective envers une dimension ne pose pas de difficulté si les deux dimensions sont très facilement dissociables.

Figure 6 : Temps moyen de classification accélérée pour les dimensions séparables de forme et de teinte d'un rectangle, en ms (d'après les données de Maddox et Ashby, 1996).



En résumé, des dimensions sont dites *intégrales* si leur similarité est évaluée sur base d'une métrique euclidienne, si le critère choisi pour la classification libre repose sur la similarité globale des stimuli et si l'attention sélective pour une dimension n'est pas possible. Pour cette raison, la redondance facilite la performance alors que les variations orthogonales de la dimension non pertinente induisent de l'interférence. C'est, par exemple, le cas de la brillance et de la teinte d'une couleur (Hyman & Well, 1967).

Par contre, des dimensions sont *séparables* si elles sont comparées au moyen d'une métrique « city-block », classées librement sur une base dimensionnelle et si leur vitesse de classification n'est pas influencée par les variations de la dimension non pertinente. À première vue, ce patron de séparabilité semble rare car, comme nous le verrons, des dimensions peuvent interagir à différents niveaux (Melara et Marks, 1990a) et la force de cette interaction est variable. De plus, il peut arriver que des dimensions généralement considérées comme séparables telles que la taille d'un rectangle et la saturation de sa couleur (Maddox et Ashby, 1996) donnent lieu à une facilitation en condition corrélée chez des participants très entraînés (plus de 16 000 essais), sans interférence en condition orthogonale. Dans ce cas, cette facilitation est interprétée comme reflétant une stratégie des participants afin d'obtenir une performance optimale. Dans ce contexte, est-il possible d'obtenir un patron de séparabilité clair ? En d'autre termes, existe-t-il des dimensions réellement séparables ? Selon Garner (1974) et ses successeurs (par exemple, Lokhead et Pomerantz, 1991), la réponse est positive. En effet, un patron clair de séparabilité a été obtenu pour des dimensions visuelles telles que la taille d'un cercle et l'orientation de son rayon (Garner et Felfoldy, 1970) ou des bigrammes et leur position dans l'espace (Melara & Marks, 1990a).

Avant d'évoquer les développements associés aux concepts d'intégralité et de séparabilité, mentionnons qu'il ne s'agit pas des seules interactions dimensionnelles possibles. Ainsi, Pomerantz et Garner (1973) ont décrit une interaction dite « configurale ». L'interaction configurale reflète le fait que les composantes forment une configuration nouvelle dont les dimensions ont une certaine réalité psychologique (ne sont pas intégrales) mais auxquelles il est difficile de faire attention séparément, vraisemblablement à cause de l'apparition d'une "propriété émergente" nouvelle. Par exemple, les parenthèses ( ) possèdent la propriété émergente de symétrie, tandis que la configuration ((possède la propriété émergente de répétition (Pomerantz et Garner, 1973). Du point de vue de la tâche de classification accélérée, les dimensions configurales conduisent à une interférence en condition orthogonale sans facilitation en condition redondante (Garner, 1978). Cette interaction n'ayant, à notre connaissance, jamais été mise en évidence pour des dimensions auditives, nous ne la développerons pas plus avant.

Les concepts influents d'interactions dimensionnelles et les méthodes expérimentales décrites par Garner (1974) ont été appliqués dans de nombreux domaines. Cette approche a permis d'étudier des paires de dimensions aussi variées que la taille de cercles et l'angle de leur rayon (Potts, Melara, & Marks, 1998), la hauteur et l'intensité d'un son (Grau & Kemler Nelson, 1988 ; Melara & Marks, 1990b), mais aussi des stimuli complexes et dynamiques, comme l'identité d'un visage et ses mouvements labiaux (Kaufmann & Schweinberger, 2005). L'engouement pour cette méthode a également conduit à de multiples remises en question et raffinements qui seront passés en revue dans la suite de ce texte.

1. 2. 5. 3. Intégralité et traitement holistique

La définition donnée par Garner (1974) de l'intégralité, c'est-à-dire une perception holistique[(2)] des stimuli, a souvent été décriée. En particulier, le fait que des dimensions qui appartiennent à des modalités différentes, comme la couleur (noir ou blanc) et la hauteur d'une note (grave ou aiguë), mènent à un patron d'intégralité (Melara, 1989) s'oppose à l'idée selon laquelle les dimensions intégrales n'ont pas de réalité psychologique. Plus tard, Kemler-Nelson (1993) a nuancé ce propos en mentionnant que des dimensions intégrales, bien que préférentiellement traitées de façon holistique, peuvent être dissociées si nécessaire. Melara, Marks et Potts (1993) ont marqué leur désaccord par rapport à cette idée. Selon ces derniers, le traitement dimensionnel se déroule en plusieurs étapes : les dimensions sont toujours extraites dans un premier temps et elles sont ensuite réunies dans le cas de dimensions intégrales. Mais, selon eux, pour les dimensions intégrales l'information extraite d'une dimension est influencée par l'autre dimension, ce qui rendrait compte des patrons comportementaux de facilitation en condition redondante et d'interférence en condition orthogonale. Actuellement, au vu de la multiplicité de données indiquant des interactions entre des dimensions de modalités perceptives différentes (Ben Artzi & Marks, 1995 ; Kaufmann & Schweinberger, 2005 ; Melara, 1989 ; Melara & Marks, 1990d), l'hypothèse d'un traitement holistique précoce semble difficile à défendre. La méthodologie développée par Garner (1974) perd-elle pour autant son intérêt? Non, puisqu'elle permet néanmoins de définir si des dimensions perceptives sont indépendantes ou interagissent, même si cette interaction ne correspond pas à un traitement strictement « holistique ».

1. 2. 5. 4. Continuum et asymétrie

Il arrive parfois que les différentes opérations convergentes conduisent à des résultats contradictoires avec, par exemple, une facilitation en condition corrélée sans interférence en condition orthogonale (Biederman & Checkosky, 1970) ou, encore, que l'interférence en condition orthogonale soit moins importante pour certaines

---

(2) Au sens que l'objet constitue un tout dont les dimensions sont non seulement indissociables, mais n'ont pas de réalité psychologique pour l'observateur.

paires de dimensions que pour d'autres (Garner & Felfoldy, 1970). Ces observations ont incité certains auteurs (Garner, 1974 ; Grau & Kemler Nelson, 1988) à considérer l'intégralité et la séparabilité comme les deux extrêmes d'un continuum et non comme un phénomène de type « tout ou rien ». L'objectif des opérations convergentes devient donc de spécifier le degré d'intégralité des dimensions.

Un autre problème concerne les situations d'asymétrie entre interactions. En effet, certaines paires de dimensions semblent présenter une intégralité ou une séparabilité asymétriques, c'est-à-dire que les variations d'une des dimensions - par exemple, la fréquence fondamentale - influencent la classification de l'autre dimension - par exemple, la consonne (Wood, 1975) - mais que l'inverse n'est pas vrai. De telles asymétries sont fréquentes dans la littérature (par exemple Pitt, 1994 ; Potts et al., 1998) et peuvent être liées à différents phénomènes.

Tout d'abord, il est possible d'observer une asymétrie des résultats de la tâche de classification accélérée si l'une des dimensions testées est significativement plus discriminable ou saillante que l'autre. En effet, dans ce cas, la dimension dont les niveaux sont plus discriminables interférera sur la classification de la dimension moins discriminable, alors que la réciproque ne sera pas vraie. Cette situation se manifeste dans plusieurs études utilisant le paradigme de classification accélérée (par exemple, Patching et Quinlan, 2002) et les auteurs n'en sont pas toujours conscients dans l'interprétation des résultats. Une des contraintes fondamentales pour l'application du paradigme de Garner est, donc, que les niveaux de discriminabilité des deux dimensions soient équivalents. Ceci peut être évalué par le biais de la condition contrôle : si les vitesses de classification et les taux d'erreurs ne sont pas significativement différents pour les deux dimensions évaluées dans cette condition où seule la dimension cible varie, on peut en déduire que les niveaux de discriminabilité des deux dimensions sont équivalents (Wood, 1975).

En l'absence de différence de discriminabilité, les interactions asymétriques sont généralement interprétées comme reflétant des stades de traitement différents (Garner, 1983) : la dimension traitée au niveau le plus bas interfère avec le traitement de la dimension de niveau plus élevé. L'idée d'interpréter les interactions asymétriques sur base de hiérarchies entre les niveaux de traitement a été formalisée par Pomerantz, Pristach et Carson (1989) puis par Melara et Marks (1990a) via le concept de « levels of crosstalk ». Le terme *crosstalk,* difficile à traduire, exprime l'interférence d'un canal d'information sur un autre (Pomerantz et al., 1989). Des dimensions sont séparables si leurs traitements ont lieu dans deux canaux parallèles, sans connexion ou *crosstalk* entre eux (Figure 7, A).

Si des dimensions interagissent, leurs *crosstalks* peuvent être bidirectionnels ou unidirectionnels (Melara & Marks, 1990a). Un *crosstalk* bidirectionnel indique que les dimensions qui interagissent sont traitées au même niveau (Figure 7, B).

Un *crosstalk* unidirectionnel, se manifestant par une interférence asymétrique, reflète au contraire le fait que les dimensions sont traitées à des niveaux différents. Dans l'exemple cité plus haut (Wood, 1975), l'asymétrie de l'interférence entre le lieu d'articulation d'une consonne et sa fréquence fondamentale peut, ainsi, être expliquée par le fait que la fréquence, et donc la hauteur, est traitée à un niveau perceptif plus précoce que la consonne qui serait traitée à un niveau phonétique (Figure 7, C).

Melara et Marks (1990a, b, d) ont affiné la théorie de Garner (1974) en proposant de nouveaux types d'interactions dimensionnelles. Les concepts de séparabilité, d'intégralité et d'interaction configurale sont maintenus et correspondent aux définitions de Garner (1974). Toutefois, selon Melara et Marks (1990a), le patron d'intégralité reflète un *crosstalk* bidirectionnel à un niveau de traitement précoce, probablement perceptif (Figure 7, B). Une autre interaction définie par Melara et Marks (1990a) est l'interaction *correspondante.* Ce terme signifie que l'on peut établir une correspondance entre les valeurs des deux dimensions, par exemple entre la position spatiale d'un haut-parleur (haut ou bas) et le sens du mot prononcé (*high* ou *low*) (Melara & Marks, 1990d). Cette correspondance induit des effets de congruence (de type Stroop, 1935) qui s'ajoutent aux interférences de type Garner(3). Ainsi, en situation d'interaction correspondante,

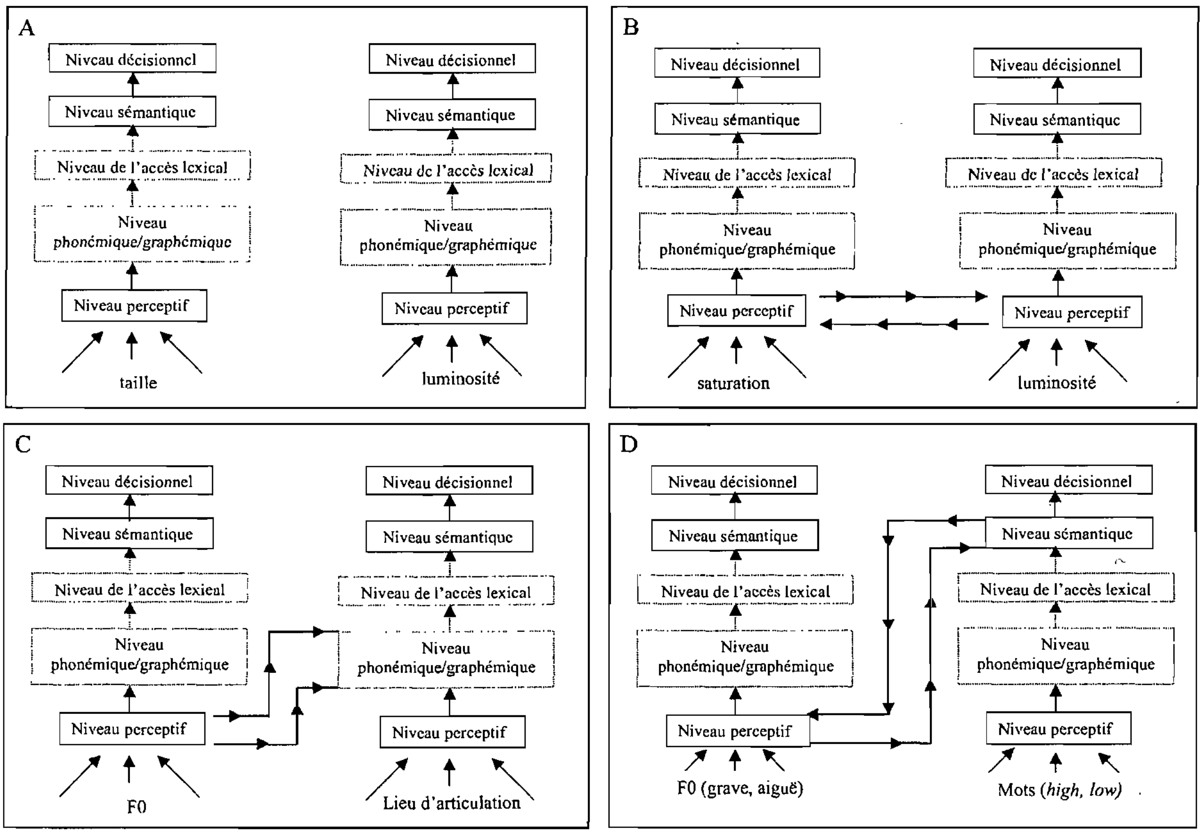
(3) Selon Pomerantz et al. (1989) les effets de congruence (de type Stroop) émergent du *contenu* de la dimension non pertinente alors que l'interférence Garner découle de la *variation* de la dimension non pertinente à travers les essais. Les effets de congruence interviennent donc lorsque l'observateur doit faire sélectivement attention aux valeurs d'une dimension et ignorer les valeurs d'une dimension donnant lieu à une réponse incompatible. En revanche, dans la condition orthogonale de la tâche de classification accélérée de Garner, l'observateur doit ignorer une dimension non pertinente qui varie de manière non prédictible d'un essai à l'autre.

on observera une facilitation en condition redondante uniquement lorsque deux dimensions congruentes sont associées (e.g. le mot *high* et le haut-parleur du haut). Par contre, lorsque les deux dimensions sont discongruentes (e.g. le mot *high* et le haut-parleur du bas), on n'observera pas la facilitation attendue. Ces interactions correspondantes, qui reflètent des interactions dimensionnelles autres que l'intégralité, permettent de rendre compte des interactions observées pour des dimensions multimodales qui trouveraient leur source au niveau sémantique (Ben Artzi & Marks, 1995 ; Kaufmann & Schweinberger, 2005 ; Melara, 1989 ; Melara & Marks, 1990d). Les interactions correspondantes sont donc des interactions liées à un *crosstalk* bidirectionnel au niveau sémantique (Melara & Marks, 1990d) – ce serait le cas des mots *high* et *low* prononcés sur des notes graves et aiguës (Figure 7, D).

En plus de ces interactions correspondantes, Melara et Marks (1990a) décrivent trois autres interactions, reflétant des relations asymétriques entre dimensions : (1) Un *crosstalk* bidirectionnel entre le niveau lexical et le niveau perceptif – l'interaction entre les mots *high* et *low* et l'intensité sonore à laquelle ils étaient prononcés (Melara & Marks, 1990a) – ; (2) un *crosstalk* unidirectionnel ascendant entre un traitement perceptif de bas niveau et un traitement phonétique de niveau supérieur (comme dans l'étude de Wood, 1975, Figure 7, C) ; (3) et, de manière plus surprenante, un *crosstalk* unidirectionnel descendant entre un traitement phonétique et un traitement perceptif – plus précisément entre des voyelles et leur sonie (Melara & Marks, 1990a). Les auteurs interprètent ce dernier effet comme reflétant une rétroaction du traitement phonétique vers le traitement perceptif.

L'approche de Melara et Marks (1990a) est intéressante car elle permet d'approfondir les interprétations déjà existantes des interactions dimensionnelles. Ainsi, ces interactions nous informent non seulement sur la séparabilité des dimensions mais, en cas d'interaction, elles permettent d'inférer les niveaux de traitement auxquels l'interaction a lieu. En effet, une interférence Garner symétrique couplée à des effets de congruence bidirectionnels plaiderait en faveur d'une interaction au niveau sémantique, une interférence Garner symétrique sans effet de congruence indiquerait une communication entre des niveaux plus précoces que le traitement sémantique, tandis qu'une interférence Garner associée à des effets de congruence asymétriques suggèrerait que seule une des dimensions est traitée au niveau sémantique.

Figure 7. Représentation graphique des traitements sous-tendant la séparabilité des dimensions de taille et de luminosité (Panneau A), l'intégralité symétrique des dimensions de saturation et de luminosité (Panneau B), le *crosstalk* unidirectionnel ou interaction asymétrique des dimensions de fréquence fondamentale (F0) et de lieu d'articulation d'une consonne (Panneau C), et le *crosstalk* bidirectionnel ou interaction correspondante des dimensions de fréquence fondamentale (F0) et de sémantique des mots *high* et *low* (Panneau D), d'après Melara et Marks, 1990a, p. 540-542.



1. 2. 5. 5. Modélisation théorique et nouvelles méthodes d'évaluation de la séparabilité dimensionnelle

Le caractère strictement opérationnel de la théorie de Garner (1974) a souvent été critiqué, en particulier par Ashby, Maddox et Townsend (Ashby & Maddox, 1994 ; Ashby & Townsend, 1986 ; Maddox, 1992 ; Maddox & Ashby, 1996). Ces auteurs regrettent le fait que les opérations convergentes de classification libre, de classification accélérée et d'évaluation de similarité ne soient pas suffisamment fondées sur le plan théorique. La Théorie Générale de Reconnaissance (*General Recognition Theory* ou GRT, Ashby & Townsend, 1986) a été développée pour combler cette lacune. La GRT décrit les interactions dimensionnelles sur base de théorèmes mathématiques. Sa validité a été testée à la fois sur les taux de réponses correctes (Ashby & Townsend, 1986) et sur les temps de réaction (Maddox & Ashby, 1996). L'intérêt principal de la GRT est de dissocier le niveau perceptif du niveau décisionnel. La séparabilité perceptive correspond à celle décrite par Garner (1974) : elle se manifeste si les effets perceptifs d'une dimension ne sont pas affectés par la variabilité de l'autre dimension (Maddox, 1992). Ashby et Maddox (1994) ont également proposé une nouvelle méthode d'évaluation de la séparabilité perceptive : l'invariance des temps de réaction marginaux. Les temps de réaction marginaux correspondent à la distribution des temps de classification de deux stimuli présentant le même niveau pour la dimension pertinente et des niveaux différents pour la dimension non pertinente. Revenons à l'exemple de la Figure 1 pour illustrer ce concept. En condition orthogonale, les temps de réaction marginaux sont invariants si le temps nécessaire pour décider qu'un carré noir est noir est égal au temps nécessaire pour décider qu'un cercle noir est noir. La séparabilité décisionnelle, quant à elle, correspond au fait que la décision, et donc la réponse, sur une dimension est indépendante de la décision sur l'autre dimension. Il s'agit donc d'interactions plus tardives, pouvant être liées aux stratégies décisionnelles des sujets.

Les travaux de Ashby et Maddox ont eu un impact essentiellement théorique, puisqu'ils ont pu démontrer de manière mathématique la validité des concepts d'intégralité et de séparabilité dimensionnelles, tout en conduisant à de nouveaux tests de séparabilité (Ashby & Maddox, 1994; Thomas, 1996).

1. 2. 6. Interactions entre dimensions auditives

Bien que l'approche de Garner (1974) ait été développée en modalité visuelle, elle a été appliquée avec succès à la modalité auditive. Les interactions entre les dimensions auditives non linguistiques seront présentées dans un premier temps. Ensuite, les interactions entre les dimensions auditives linguistiques et non linguistiques seront examinées.

1. 2. 6. 1. Interactions entre dimensions auditives non linguistiques

La plupart des études concernant les dimensions auditives non linguistiques suggèrent qu'elles ne sont pas séparables. Ainsi, via les trois opérations convergentes de classification accélérée, de jugement de similarité et d'évaluation de la métrique, Grau et Kemler-Nelson (1988) ont montré que la hauteur et la sonie constituent des dimensions intégrales au sens défini par Garner (1974). Ces résultats ont été confirmés par Melara et Marks (1990c), suggérant selon leur théorie que le *crosstalk* aurait lieu au niveau perceptif.

Les interactions entre le timbre et d'autres dimensions auditives comme la hauteur ou la sonie (Krumhansl & Iverson, 1992 ; Melara & Marks, 1990b) présentent un profil similaire. Dans une étude où les caractéristiques spectrales de timbres synthétiques étaient manipulées, Melara et Marks (1990b) ont observé une intégralité symétrique entre, d'une part, la sonie et le timbre et, d'autre part, la hauteur et le timbre. Krumhansl et Iverson (1992) ont également obtenu un patron d'intégralité pour des timbres dont l'enveloppe spectrale et les caractéristiques temporelles de l'attaque étaient manipulées de manière à reproduire le timbre d'instruments existants. Les interactions entre différentes dimensions constitutives du timbre (l'attaque, le centre de gravité spectral et l'atténuation des harmoniques pairs) ont également été étudiées au moyen du paradigme de Garner (Caclin, Giard, Smith, & McAdams, 2007). Les résultats suggèrent, à nouveau, que ces trois dimensions interagissent de manière bidirectionnelle, probablement au niveau perceptif.

Aucune des études décrites ci-dessus ne prenait en compte l'expertise musicale des participants. Or, si les interactions dimensionnelles émanent du traitement cognitif de l'auditeur, il est possible que la pratique musicale modifie ce traitement. Pitt (1994) a examiné cette hypothèse en comparant les interactions entre la hauteur et le timbre chez des participants d'expertise musicale variable. En accord avec les résultats de Melara et Marks (1990c) et de Krumhansl et Iverson (1992), ces dimensions interagissaient dans les deux groupes. Cependant, chez les non-musiciens, les variations non pertinentes du timbre influençaient plus le traitement de la hauteur que l'inverse, donnant lieu à une intégralité asymétrique. Cette asymétrie n'était pas observée chez les musiciens, qui présentaient un patron d'intégralité symétrique classique. Ces résultats ont été confirmés par une autre procédure (Crowder, 1991) : dans une tâche de jugement de hauteur sur des paires de notes (tâche même / différent), il était plus difficile de considérer deux notes comme identiques si leur timbre différait. Les musiciens étaient toutefois moins perturbés par les différences de timbre que les non-musiciens. Ces données suggèrent que les *crosstalks* entre dimensions peuvent être modifiés par les caractéristiques des sujets, telles qu'une expertise spécifique pour le traitement de la hauteur chez des musiciens.

En résumé, plusieurs dimensions auditives constitutives des stimuli musicaux – le timbre, la hauteur et la sonie – ne sont pas indépendantes. Cependant, il est envisageable que des dimensions auditives ne puissent tout simplement pas être séparables, ou encore que le paradigme de Garner (1974) ne soit pas suffisamment sensible pour mettre en évidence la séparabilité de ces dimensions. L'étude des interactions entre des dimensions auditives linguistiques et non linguistiques a permis de démentir ces deux propositions.

1. 2. 6. 2. Interactions entre dimensions auditives linguistiques et non linguistiques

Différents auteurs (Lee & Nusbaum, 1993 ; Miller, 1978 ; Repp & Lin, 1990 ; Wood, 1975) ont adapté le paradigme de classification accélérée à l'étude des relations entre dimensions segmentales et suprasegmentales[4] des stimuli linguistiques. Ces études sont particulièrement intéressantes pour notre propos puisqu'elles concernent les relations entre les phonèmes et la hauteur ou le contour tonal, question similaire à celle des relations entre les dimensions linguistiques et musicales dans le chant.

Wood (1975) s'est intéressé à l'interaction entre la hauteur et le lieu d'articulation de la consonne initiale de syllabes consonne-voyelle. Les participants devaient classer des syllabes soit sur base de leur fréquence fondamentale (aiguë, 140 Hz, ou grave, 104 Hz), soit sur base de leur consonne initiale (contraste /b/ - /g/ suivi de la voyelle /ae/). Wood a mis en évidence une interférence asymétrique en condition orthogonale : la fréquence fondamentale était séparable du lieu d'articulation de la consonne alors que le lieu d'articulation n'était pas séparable de

---

[4] Les informations segmentales concernent les phonèmes alors que les informations suprasegmentales concernent des propriétés acoustiques qui s'étendent sur plus d'un segment, comme le contour intonationel ou le patron d'accentuation (Lee & Nusbaum, 1993, p.157).

la fréquence fondamentale. Ces résultats indiquent, tout d'abord, que des dimensions auditives peuvent être séparables. De plus, cette étude a apporté le premier argument en faveur d'une interprétation des asymétries entre les interactions dimensionnelles en termes de niveaux de traitement, idée déjà évoquée au point 5d.

Cette interprétation a cependant été mise en cause par les résultats de Miller (1978), selon lesquels l'asymétrie pourrait être liée aux caractéristiques acoustiques des stimuli. Cet auteur a étudié les relations entre la qualité des voyelles (/a/ et /ae/) et leur hauteur (140 vs 104 Hz). Les résultats indiquaient une interférence mutuelle et symétrique qui contraste avec l'asymétrie obtenue par Wood (1975). Il semblerait donc qu'il n'y ait pas qu'un seul type d'interaction possible et, surtout, que le facteur pertinent ne serait pas la nature segmentale ou suprasegmentale des dimensions traitées, mais bien leurs caractéristiques acoustiques. En effet, les caractéristiques acoustiques des voyelles et des consonnes diffèrent, les consonnes étant caractérisées par des changements rapides d'amplitude et de fréquence fondamentale (Delattre, Liberman, & Cooper, 1955) alors que les informations acoustiques spécifiant l'identité des voyelles sont plus stables (Fry, Abramson, Eimas, & Liberman, 1962). Selon Miller (1978), les caractéristiques acoustiques des voyelles les rendraient, de ce fait, plus intégrées avec la hauteur tonale que ne le seraient les consonnes.

L'utilisation du paradigme de Garner (1974) avec des dimensions pouvant être considérées comme linguistiques ou non linguistiques selon la langue des sujets peut permettre de trancher entre l'interprétation « acoustique » et l'interprétation « linguistique » des interactions entre dimensions segmentales et suprasegmentales. Les tons lexicaux répondent à ce critère puisque ces variations de hauteur influencent la signification des mots dans les langues tonales mais qu'elles apportent essentiellement des informations syntaxiques, pragmatiques et affectives dans les langues non tonales. Or, si les différences de caractéristiques acoustiques sont à l'origine des résultats divergents obtenus par Wood (1975) et Miller (1978), les locuteurs de langues tonales et non tonales devraient présenter le même patron de résultats à une tâche de classification accélérée avec des tons et des dimensions segmentales comme cibles. Repp et Lin (1990) ont vérifié cette affirmation en confrontant des auditeurs de langue anglaise et chinoise à une tâche de classification accélérée sur base d'une dimension segmentale (voyelle ou consonne), d'une part, et suprasegmentale (tons), d'autre part. Un patron d'intégralité a été observé dans les deux groupes pour toutes les combinaisons de dimensions. Ces résultats ne soutiennent pas l'interprétation selon laquelle l'interférence observée par Wood (1975) serait liée aux niveaux de traitement des dimensions en présence. En effet, l'interaction était la même que les tons apportent ou non une information linguistique dans la langue des sujets.

Cependant, Lee et Nusbaum (1993) ont remarqué que les résultats de Wood (1975) et ceux de Repp et Lin (1990) n'étaient pas réellement comparables puisque Wood avait utilisé des notes stables alors que Repp et Lin avaient employé un contour tonal dynamique. Afin de clarifier les rôles respectifs des caractéristiques des stimuli et de la langue maternelle des participants, Lee et Nusbaum (1993) ont utilisé une tâche de classification accélérée sur base de dimensions segmentales (consonnes) et suprasegmentales (tons et fréquences fondamentales fixes) chez des participants de langue anglaise et chinoise. Les auditeurs anglophones présentaient une intégralité symétrique pour les tons et les consonnes (comme les participants de Repp et Lin, 1990), mais une intégralité asymétrique pour les hauteurs stables (comme les participants de Wood, 1975). En revanche, chez les sujets dont la langue maternelle était le chinois, il y avait une intégralité mutuelle à la fois pour les tons et les hauteurs stables. Lee et Nusbaum (1993) ont interprété ces résultats comme reflétant l'importance de l'expérience linguistique dans la perception de la parole. En somme, le caractère segmental ou suprasegmental, comme les caractéristiques acoustiques des stimuli, ne jouent qu'un rôle partiel sur le type d'interaction dimensionnelle. Les caractéristiques des auditeurs, comme la langue maternelle (Lee & Nusbaum, 1993) ou l'expertise musicale (Pitt, 1994) semblent également contribuer à ces interactions.

Un argument fort en faveur de cette idée a été apporté par Tomiak, Mullennix et Sawusch (1987). Ces auteurs ont montré que le fait de considérer les mêmes stimuli comme linguistiques ou non en modifie les interactions dimensionnelles. Ainsi, la « consonne » et la « voyelle » des syllabes synthétiques /fae/, /ʃae/, /fu/ et /ʃu/ présentaient une intégralité réciproque lorsque la consigne les décrivait comme de la parole. Le comportement des participants était identique à celui observé pour de la parole naturelle, alors que les stimuli de Tomiak et al. (1987) ne comportaient pas de transition de formant. Par contre, lorsque la consigne présentait les mêmes stimuli synthétiques comme un bruit suivi d'une note, on obtenait un patron de séparabilité. Ces résultats sont essentiels pour quatre raisons. Premièrement, ils confirment que des dimensions auditives peuvent être séparables. Deuxièmement, ils montrent que les mêmes dimensions peuvent être traitées comme intégrales ou séparables selon le contexte, ce qui renforce l'idée que les interactions dimensionnelles ont une origine psychologique. Troisièmement, le fait que les consonnes et les voyelles synthétiques soient perçues comme intégrales même en l'absence de transition de formant indique que, contrairement à l'idée de Miller (1978), les concepts d'intégralité et de séparabilité ne dépendent pas strictement des caractéristiques acoustiques des stimuli mais bien du traitement cognitif effectué par l'auditeur. Enfin, les données de Tomiak et al. (1987) impliquent que les résultats obtenus sur de la parole ne peuvent être généralisés au traitement du chant puisque les interactions dépendent de « l'étiquette » que les auditeurs attribuent aux stimuli. Les mêmes stimuli chantés pourraient donc conduire à des interactions dimensionnelles différentes selon qu'ils soient plutôt considérés comme linguistiques ou musicaux par les participants.

Les études décrites jusqu'à présent font référence aux interactions entre les dimensions phonologiques de la parole et les dimensions suprasegmentales. D'autres auteurs, dont Melara et Marks (1990d), se sont penchés sur les interactions entre la dimension sémantique du langage et des dimensions non linguistiques. Pour résumer leurs résultats, ces auteurs ont mis en évidence des interactions correspondantes entre les mots *high* et *low*, qu'ils soient présentés par écrit ou auditivement, et des dimensions pouvant être congruentes ou non au niveau sémantique. Ainsi, les syllabes écrites HI et LO interagissaient avec des sons graves et aigus, mais aussi avec la position verticale où elles étaient écrites. Les mêmes interactions étaient obtenues entre les mots parlés *high* et *low,* leur origine spatiale, la hauteur de la voix qui les prononçait (Melara & Marks, 1990d) et leur intensité sonore (Melara & Marks, 1990a). Ces études ont confirmé l'idée selon laquelle des interactions entre dimensions linguistiques et non linguistiques peuvent avoir lieu même à des niveaux de traitement plus élaborés qu'un simple traitement perceptif. De telles interactions ont également été observées dans des situations où la correspondance sémantique est moins directe : des enfants présentaient une interférence de type Garner et des effets de congruence lorsqu'ils devaient classer les mots *Daddy* et *Mommy* prononcés par une voix masculine ou féminine (Jerger, Elizondo, Dinh, Sanchez, & Chavira, 1994), suggérant que le genre du locuteur et le mot prononcé constituent des dimensions correspondantes.

En somme, les recherches passées en revue indiquent que le paradigme de Garner (1974) est bien adapté à l'étude des interactions entre des dimensions auditives linguistiques et non linguistiques, statiques (hauteur, voyelle) ou dynamiques (contour tonal, consonnes), simples (hauteur, sonie) ou plus complexes (sémantique).

1. 2. 7. Conclusion : Limites et apports de l'approche des interactions dimensionnelles pour l'étude des relations entre paroles et mélodies dans le chant

L'objectif de cet article était de montrer que les opérations convergentes développées par Garner (1974) constituent une méthode de choix pour étudier les interactions dimensionnelles entre les paroles et la mélodie dans le chant. Après avoir défini les concepts d'intégralité et de séparabilité, il a été précisé que d'autres types d'interactions dimensionnelles sont possibles, comme, par exemple, les interactions correspondantes (Melara & Marks, 1990a). La tâche de classification accélérée a également permis d'étudier les interactions entre de multiples dimensions auditives, qu'elles soient linguistiques ou non. De ce fait, l'on pourrait se demander s'il est réellement nécessaire de développer un nouvel axe de recherche dans lequel cette approche serait appliquée au chant. En effet, les études passées en revue apportent des informations intéressantes sur les relations entre les notes et les phonèmes (Miller, 1978 ; Wood, 1975), les contours tonaux et les phonèmes (Lee & Nusbaum, 1993) et entre la hauteur et la sémantique (Melara & Marks, 1990d ; Jerger et al., 1994). Cependant, Tomiak et al. (1987) ont montré que la manière dont les auditeurs considèrent les mêmes stimuli en modifie les interactions dimensionnelles. Par conséquent, les résultats des études décrites dans cet article ne peuvent simplement être transposés au chant car celui-ci pourrait être considéré par les auditeurs comme de la « musique vocale » ou comme du « langage chanté ». De nouvelles études utilisant les opérations convergentes de Garner sont donc nécessaires pour répondre de manière claire à la question des interactions entre paroles et mélodie dans le chant.

Le problème qui se pose à présent est celui de définir les limites et apports de cette technique de manière générale, d'une part, et pour l'étude des relations entre langage et musique, d'autre part.

1. 2. 7. 1. Limites de l'approche des interactions dimensionnelles

Une première restriction de l'approche des interactions dimensionnelles découle de ce qui vient d'être évoqué, c'est-à-dire le fait que le patron d'interactions dimensionnelles obtenu peut être modulé par les stratégies attentionnelles ou les caractéristiques des observateurs. Ainsi, une manipulation des consignes visant à focaliser l'attention de l'auditeur sur l'aspect plus musical ou, au contraire, plus linguistique des mêmes stimuli pourrait modifier le patron de performance des sujets et, par conséquent, les conclusions des expérimentateurs quant aux interactions entre paroles et mélodies. De tels phénomènes sont susceptibles d'altérer sévèrement la généralité des résultats obtenus. De manière similaire, le degré d'expertise musicale pourrait modifier le patron d'interactions entre les paroles et les mélodies du chant. Besson et Schön (2001) ont décrit une étude électrophysiologique répliquant celle de Besson et al. (1998) où les auteurs ont manipulé l'orientation attentionnelle des participants soit envers les paroles, soit envers la mélodie. Dans cette situation, les participants musiciens étaient capables de faire sélectivement attention à la mélodie ou aux paroles. Par contre, l'attention des non-musiciens était automatiquement attirée vers les incongruités de la dimension qu'ils devaient ignorer. Ces données laissent supposer que, dans la situation de classification accélérée (Garner, 1974) sur du chant, seuls les musiciens seraient susceptibles de présenter un patron de séparabilité.

Toutefois, loin de constituer une faiblesse, cette variabilité des résultats en fonction de variables attentionnelles ou de l'expertise des participants pourrait être considérée comme un atout du paradigme. En effet, l'objectif des recherches sur les interactions entre langage et musique n'est pas de démontrer qu'il s'agit de dimensions intégrées ou indépendantes dans l'environnement, mais de comprendre comment notre système cognitif les traite. Si les manipulations attentionnelles et les caractéristiques des observateurs peuvent conduire ces dimensions à être traitées comme plus ou moins intégrées ou séparables, cela suggérera clairement que les dimensions ne sont pas perçues comme fusionnées au point d'être indissociables. En résumé, de tels effets indiqueraient que les relations entre langage et musique, loin d'être une question de « tout ou rien », constituent une problématique complexe qui s'oppose à une vision purement modulaire ou intégrée de ces fonctions. Ainsi, il se pourrait que certaines composantes soient partagées par le langage et la musique, mais qu'un degré d'expertise musicale élevé contribue à la spécialisation de systèmes de traitement spécifiquement musicaux. Les études d'imagerie cérébrale indiquant la présence de différences anatomiques et fonctionnelles entre musiciens et non-musiciens vont dans ce sens (e.g. Ohnishi et al., 2001 ; Schlaug, 2001). L'approche des interactions dimensionnelles et les méthodes expérimentales associées pourront contribuer à préciser quelles dimensions linguistiques et musicales interagissent, et surtout, à examiner l'intensité et la robustesse de ces interactions.

Une autre contrainte de l'approche de Garner pouvant sembler difficilement compatible avec l'utilisation de matériel chanté réside dans l'obligation d'utiliser des stimuli brefs. Toutefois, des mots plurisyllabiques ont déjà été utilisés avec succès dans des tâches de classification accélérée (Jerger et al., 1994). De brèves phrases chantées pourraient donc constituer un matériel adéquat pour étudier les interactions entre paroles et mélodies au moyen du paradigme de classification accélérée. Il est évident qu'une telle approche sacrifie au caractère écologique des stimuli, mais le choix d'un matériel écologique ou d'un matériel bien contrôlé est une difficulté récurrente de l'approche expérimentale en psychologie. Une contrainte supplémentaire, à savoir la nécessité d'utiliser des dimensions de discriminabilité équivalente pour pouvoir conclure quant à leur intégralité ou séparabilité, est plus problématique dans le cas du chant. En effet, un dénominateur commun aux études faisant appel à du matériel chanté est la saillance plus importante des paroles par rapport à la mélodie (e.g. Serafine et al., 1984 ; Peretz et al., 2004). Toutefois, il semble possible, en choisissant des mélodies très discriminables, d'équilibrer la saillance des dimensions linguistiques et musicales (Kolinsky, Lidji, Peretz, Besson, & Morais, en révision).

La dernière limite, plus générale, du paradigme de Garner (1974) réside dans son aspect purement comportemental, à une époque où les méthodes d'imagerie cérébrale sont en plein essor. En effet, cette approche comportementale limite le contrôle sur l'étape de traitement à laquelle l'intégralité ou la séparabilité a lieu, ces interactions pouvant prendre place tardivement, au moment de la sélection ou de la production de la réponse (Holender, 1992). Pourtant, et c'est pourquoi nous l'avons développé précédemment, selon Melara et Marks (1990a), il serait possible de spécifier l'étape de traitement à laquelle l'interaction prend place sur base de ces données. Reconnaissons toutefois que cette approche comportementale ne dispose pas de la sensibilité de méthodes physiologiques comme les potentiels évoqués. Une manière d'examiner les interactions entre paroles et mélodies à un niveau précoce serait, justement, d'appliquer la technique de l'électroencéphalographie en situation de classification accélérée. De telles études ont été réalisées (Chen & Melara, 2003 ; Kaganovich, Francis, & Melara, 2006) et ont mis en évidence, pour les dimensions de timbre et de hauteur (Chen et Melara, 2003) et de voyelle et locuteur (Kaganovitch et al., 2006), l'apparition de composantes électrophysiologiques précoces (dès 100 ms) en réponse aux variations de la dimension non-pertinente pour la classification. Selon les auteurs, ces données suggèrent que l'interférence pour les dimensions intégrales provient d'une difficulté précoce à extraire la dimension cible, bien avant la sélection et l'exécution de la réponse. De manière similaire, des études exploitant la *mismatch negativity* (MMN, une composante électrophysiologique négative reflétant la détection précoce et non-intentionnelle d'une déviance parmi une série répétitive de stimuli auditifs) ont montré que les réponses physiologiques à des variations de hauteur et d'intensité sonore étaient indépendantes au niveau des générateurs précoces de la MMN, dans le cortex auditif (Paavilainen, Mikkonen, Kilpeläinen, Lehtinen, Saarela, & Tapola, 2003). Par contre, ces ondes interagissaient au niveau des générateurs frontaux (Wolff & Schröger, 2001), qui refléteraient une sous-composante plus tardive de la MMN. Ces données suggèrent que l'intégralité observée comportementalement par Grau et Kemler-Nelson (1988) serait issue de composantes indépendantes qui seraient ensuite intégrées, ces deux étapes se déroulant de façon précoce, après moins de 200 ms de traitement. L'on ne peut que plaider pour que cette utilisation combinée des techniques électrophysiologiques et des méthodes classiques de la psychologie expérimentale devienne la norme (voir aussi Caclin, McAdams, Smith, & Giard, 2008). En plus de cette question temporelle, la bonne résolution spatiale de la magnetoencéphalographie permettra, dans un avenir proche, de préciser si des dimensions considérées comme intégrales ou séparables sur base de données comportementales ont des sources neuronales communes ou séparées.

A l'issue de cette réflexion sur les limites de l'approche de Garner, il apparaît que, si les expérimentateurs sont conscients de leur existence, ils peuvent tenter de contourner ces limites voire de les exploiter comme des avantages. Examinons à présents les bénéfices que pourrait apporter la perspective des interactions dimensionnelles pour l'étude des relations entre langage et musique.

1. 2. 7. 2. Apports de l'approche des interactions dimensionnelles pour l'étude des relations entre paroles et mélodies dans le chant

Du point de vue théorique, l'approche des interactions dimensionnelles peut contribuer à dépasser la simple question d'indépendance ou d'interaction entre musique et langage (Schön et al., 2005) pour préciser, en cas d'interaction, à quel niveau et dans quelle direction ces interactions ou *crosstalks* ont lieu (Melara & Marks, 1990a). Ceci est d'autant plus vrai qu'il est aujourd'hui possible de combiner cette méthode comportementale avec l'acquisition de données électrophysiologiques (e.g. Kaganovitch et al., 2006) ou magnetoencéphalographiques pouvant apporter des informations importantes sur le décours temporel et les sources anatomiques des interactions.

Du point de vue pratique, l'approche de Garner (1974) autorise l'étude des interactions entre les dimensions linguistiques et musicales du chant chez des participants sains. Elle constitue ainsi un complément adéquat aux études de double dissociation chez des patients cérébro-lésés (Hébert & Peretz, 2001 ; Peretz, 2002 ; Peretz et al., 2004) et évite les critiques selon lesquelles la double dissociation ne constitue pas une approche valable pour mettre en évidence des spécificités de traitement (Van Orden et al., 2001, Plaut et al., 1995). De plus, les opérations convergentes décrites par Garner (1974) permettent d'étudier les interactions entre diverses dimensions linguistiques comme la dimension phonologique et la dimension sémantique, et diverses dimensions musicales comme la dimension mélodique et la dimension rythmique, au moyen d'une seule et même méthode bien validée par ailleurs. Cette approche aidera donc à comprendre si l'incohérence des résultats obtenus par le passé avec du matériel chanté (Besson et al., 1998 ; Bigand et al., 2001 ; Bonnel et al, 2001 ; Poulin-Charonnat et al., 2005 ; Schön et al., 2005) est liée aux différences méthodologiques entre ces études ou aux différences entre les dimensions linguistiques étudiées. En effet, ce paradigme peut permettre de comparer les interactions entre dimensions sémantique et mélodique, d'une part, et phonétique et mélodique, d'autre part, avec la même technique.

Ainsi, une tâche de classification accélérée de courtes phrases telles que « ça descend » et « ça remonte » chantées sur des mélodies ascendantes ou descendantes pourrait permettre de tester l'intégralité ou la séparabilité des dimensions sémantique et mélodique, et, en cas d'intégralité, d'examiner si l'interaction a lieu au niveau sémantique (interaction correspondante). Cette situation expérimentale est similaire à celle décrite par Melara et Marks (1990d) avec les mots *high* et *low* prononcés sur des notes graves et aiguës, à l'exception de l'ajout de la dimension mélodique. En cas d'interaction correspondante, l'on s'attendrait à observer un patron d'intégralité avec effet de congruence : les temps de classification seraient plus rapides en situation congruente (par exemple, la phrase *ça descend* chantée sur une mélodie *descendante*) qu'en situation discongruente (par exemple, la phrase *ça descend* chantée sur une mélodie *ascendante*). Ces résultats s'opposeraient donc à l'indépendance décrite par Besson et al. (1998) et Bonnel et al. (2001), mais correspondraient à l'interaction décrite par Schön et al. (2005). Au contraire, un patron de séparabilité serait en accord avec les résultats des études sur la perception de l'opéra (Besson et al., 1998 ; Bonnel et al., 2001).

En ce qui concerne les relations entre phonologie et mélodie, il serait possible d'examiner la classification accélérée de pseudo-mots chantés sur de courtes mélodies, les pseudo-mots variant sur le plan phonétique (par exemple, modification d'une voyelle ou d'une consonne) et les mélodies sur le plan du contour (par exemple, modification d'une note afin d'altérer le contour mélodique). Sur base des résultats de Bigand et al. (2001), nous nous attendrions à observer un patron d'intégralité. Toutefois, les résultats obtenus dans notre laboratoire (Kolinsky et al., en révision) suggèrent que seules les voyelles sont traitées de manière intégrale avec le contour mélodique, les consonnes semblant, au contraire, séparables de la mélodie. Ces données paraissent cohérentes avec les résultats indiquant l'intégralité des voyelles et de la hauteur dans le traitement du langage (Miller, 1978). Par contre, elles semblent contredire les résultats indiquant une interaction asymétrique entre consonne et hauteur (Wood, 1975 ; Lee & Nusbaum, 1993). Des recherches futures permettront certainement de dévoiler l'origine de ces contradictions. Quoi qu'il en soit, ces exemples démontrent qu'il serait possible d'examiner, avec la même méthode expérimentale, les interactions entre sémantique et mélodie, d'une part, et entre phonologie et mélodie, d'autre part.

Pour conclure, le but de cette revue critique de littérature était de démontrer que l'approche classique des interactions dimensionnelles peut contribuer à répondre à une question très actuelle en psychologie cognitive: celle de la spécificité fonctionnelle des traitements linguistiques et musicaux. Après avoir rappelé les concepts principaux de cette approche et passé en revue son évolution, nous avons souligné les apports de cette technique et ses limites. Ces limites devraient inciter les scientifiques à combiner cette méthode expérimentale classique aux techniques récentes d'imagerie cérébrale, construisant ainsi un pont entre le passé de la psychologie cognitive et son avenir.

2. Section expérimentale

2. 1. Article 2

Processing interactions between lyrics and tunes: Vowels sing but consonants speak [ii]

[ii] Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (en révision). Processing interactions between lyrics and tunes: Vowels sing but consonants speak. *Cognition.*

Abstract

The aim of the present study was to determine if two dimensions of songs, namely the phonological part of lyrics and the melodic part of tunes, are processed on-line as integral or separable dimensions. We submitted non-musician participants to Garner's (1974) filtering, redundancy and condensation tests, using materials consisting of bisyllabic nonsense words sung on two-tones intervals. The results of Experiment 1 showed that vowels and intervals are integral dimensions, whereas Experiment 2 suggested that stop consonants and intervals are more separable. To test whether the processing interactions between intervals and phonemes are related to the sonority of the latter, in Experiment 3, participants were presented with two new materials, one with a new vowel contrast and one with a contrast of sonorous, nasal consonants. This confirmed the integrality between vowels and intervals, whereas the pattern of processing interaction was less clear for nasal consonants. Nevertheless, the condensation task of Experiment 4 confirmed that consonants and melody were processed as separable dimensions, though selective attention to either one dimension was difficult. Finally, the use, in Experiment 5, of a synthesized vocalic material that controlled for the acoustical correlates between vowel quality and pitch height demonstrated that the integrality between vowels and intervals was not due to such a factor. Implications for the interactions between the musical and linguistic domains are discussed in light of the different evolutionary origins and linguisitic functions of vowels and consonants.

2. 1. 1. Introduction

A fundamental issue in human cognition is to determine how different stimulus dimensions combine and interact in processing complex materials. Speech and music are typical examples of such complex materials. Both language and music have been studied not only for their own sake, but also for comparing the cognitive processes involved in each of them. Indeed, some authors view music processing as a by-product of language processing (e.g., Pinker, 1997), while others argue that music involves specific computational processes (e.g., Peretz, 2006). Songs provide an ideal material to study the relations between language and music, since songs naturally combine a musical dimension, the tune, and a linguistic dimension, the lyrics (e.g., Patel & Peretz, 1997).

The aim of the present work was to examine whether the lyrics and the tunes are processed independently or in an integrated way in sung materials. More precisely, we examined the on-line processing independence or integration of the *phonological* and *melodic* dimensions of sung material.

Relatively little is known about how intimate speech and music are in song processing. Up to now, studies on songs have mainly investigated the processing relations between the semantics of the lyrics and the melody of the tune. Depending on the experimental approach and of the material used, these studies showed either independence (Besson, Faita, Peretz, Bonnel, & Requin, 1998; Bonnel, Faita, Peretz, & Besson, 2001), or interactions (Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005) between the semantic processing of the lyrics and the melodic / harmonic processing of the tune. Another study (Schön, Gordon, & Besson, 2005) focused on the relations between the processing of the melody and the lyrics, but this time at the word level. Again, it provided evidence for interacting processes.

As regards the interactions between the phonemic processing of the lyrics and the harmonic processing of the musical accompaniment, Bigand, Tillmann, Poulin, D'Adamo, & Madurell (2001) reported processing interactions between vowels and tunes. The structural relation between the last sung chord and the preceding musical context of an eight-chord sung sequence was manipulated in order to induce *harmonic priming,* an effect related to the musical function of the chord in the previous key context. Chords that do not belong to the musical key context (e.g., Bharucha & Stoeckig, 1986) or that are less stable in the key context (such as a *subdominant* chord compared to a *tonic* chord) are less primed by the context, resulting in slower processing even in musically naïve participants (e.g., Bigand & Pineau, 1997). In Bigand and collaborators' (2001) study, phoneme monitoring of the last sung vowel was faster when the target phoneme was sung on the tonic chord than when it was sung on the subdominant chord.

However, the effect of harmonic congruity on phoneme monitoring does not directly support the idea of specific processing interactions between language and music. According to Bigand et al. (2001) (see also Poulin-Charonnat et al., 2005), linguistic and musical domains share some attentional capacities, so that music can modulate linguistic processing by modifying the allocation of attentional resources necessary for linguistic computation. The facilitation in lyrics processing caused by the harmonic manipulation might thus arise from general attentional processes, rather than from specific music-language dependencies. This viewpoint has been further reinforced by Escoffier and Tillmann (2006) who demonstrated similar facilitation from harmonic relatedness with nonlinguistic stimuli such as geometric shapes. In addition, Bigand et al. (2001) only manipulated one category of phonemes, vowels. This does not allow generalizing the results to harmonic and phonemic processing at large, given that it has been well demonstrated that vocalic and consonantal processing do differ in both their acoustical properties and their linguistic function.

At the acoustical level, most consonants are defined by rapid transient acoustic cues typical of formant transitions, whereas vowels are characterized by more steady-state frequency informations (Delattre, Liberman, & Cooper, 1955; Fry, Abramson, Eimas, & Liberman, 1962; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). These acoustical differences have even been associated to specific cerebral lateralization, the processing of rapidly-changing acoustic information, like the one typical of consonants, being more left-lateralized than the processing of stable spectral information, like the one associated to vowels or music (for reviews, see Poeppel, 2003; Zatorre, Belin, & Penhune, 2002; Zatorre & Gandour, in press). Therefore, vowels might be more suitable to carry melodic information than the fast-changing consonants. This idea is in line with studies on opera singing suggesting that vowels are more intimately linked to the melodic variations of the tunes than consonants. The latter are located at the transition between notes, hence being sometimes reported as breaking the melodic line (e.g., Scotto di Carlo, 1993). For this reason, trained singers tend to shorten consonants and to reduce their articulation

(McCrean & Morris, 2005) while vowels are lengthened in singing compared to speech (Scotto di Carlo, 2007a, 2007b; Sundberg, 1982).

These acoustical differences between vowels and consonants are, however, not such straightforward. Speech segments may, indeed, be conceived as parts of a continuous stream of sound that varies in sonority rather than as two different categories of speech sounds. As a matter of fact, according to some researchers, in all languages, speech sounds would possess a scalar property that allows them to be ranked on a *sonority hierarchy*, ranging from the least sonorous stop consonants to the most sonorous glides and vowels, with fricatives, nasals, and liquids having an intermediate (and progressively more sonorous) status. Sonority is related to the degree of openness of the vocal apparatus during speech (e.g., Goldsmith, 1990; MacNeilage, 1998; Selkirk, 1982) and hence to relative loudness, perceptibility and acoustic intensity (but see Harris, 2006). Thus, vowels would be processed differently from consonants because the former are more sonorous than the latter. Such a view also holds that the sonority of consonants will affect how they are processed in relation to pitch in songs. In particular, the more sonorous consonants may be more apt to support pitch variations than the less sonorous ones, and hence may be more integrated with melody.

At the functional level, some authors have claimed that the difference between vowels and consonants is not limited to these distinct acoustical and articulatory properties, but that these differences indicate that vowels and consonants serve different roles in speech (Bonatti, Peña, Nespor, & Mehler, 2007). Grounded on comparative human-animal studies, it seems that consonants are more specific to human speech than vowels are. Nonhuman primates can produce harmonic sounds very similar to vowels (Owren, Seyfarth, & Cheney, 1997; Rendall, Rodman, & Emond, 1996) in order to provide indexical information (about sex, age, identity, emotion, etc.), while only humans have elaborated the supralaryngeal articulations that, by inserting consonants into the vocalic career (MacNeilage & Davis, 2000), allow the emergence of a rich set of meaningful contrasts. Consistent with this, primates can capture non-adjacent vocalic regularities in speech, but fail to capture non-adjacent consonantal regularities (Newport, Hauser, Spaepen, & Aslin, 2004), while human listeners display the opposite pattern of results (Bonatti, Peña, Nespor, & Mehler, 2005; Mehler, Peña, Nespor, & Bonatti, 2006).

The idea of a different linguistic status for vowels and consonants is further supported by the occurrence of double dissociations between the ability to produce vowels vs. consonants in aphasic patients, a pattern that cannot solely be explained by the differences of sonority between these phonemes (Caramazza, Chialant, Capasso, & Miceli, 2000; but see Monaghan & Shillcock, 2003). Moreover, consonants are more vulnerable than vowels to such impairments (Béland, Caplan, & Nespoulous, 1990; Canter, Trost, & Burns, 1985; for a review, see Monaghan & Shillcock, 2003; for contradictory evidence, see Semenza et al., 2007). Finally, the distinction between vowel and consonant processing seems to occur relatively early in life: 20 month-old infants can simultaneously learn two words that only differ by one consonant, being plosive or continous, but fail when the distinctive phoneme is a vowel (Nazzi, 2005; Nazzi & New, 2007).

In sum, comparative human-animal studies suggest that vowels may be less specific to speech than consonants, whereas experiments using statistical learning and developmental research show that vowels and consonants carry different kinds of information, the consonants being more tied to word identification and the vowels to speaker identification, prosody and grammar (Nespor et al., 2003, 2007).

Vowel processing may thus be more intricate than consonants with other, non-linguistic, auditory dimensions like melody. Surprisingly, few researchers have examined the consequences of such a difference between vowels and consonants for the issue of processing interactions between phonology and other domains. To our knowledge, this was done only in the framework of speech processing, using auditory adaptations of the *dimensional interactions approach* elaborated by Garner (1974), studied through *speeded classification* tasks (Garner, 1974; e.g., Garner, 1978a, 1978b; see Lidji, 2007, for a recent review in the song domain).

Such tasks aim at checking whether the presence of irrelevant, orthogonal, variations on one dimension (e.g., pitch: high or low) affects (or facilitates) processing of another, target, dimension (e.g., loudness: soft or loud) (Grau & Kemler Nelson, 1988). The irrelevant variations can be either redundant (e.g., when all presented soft tones are low and all loud tones are high) or orthogonal (when both soft and loud tones can be either low or high) to the variations that are relevant for the task at hand (e.g., sorting according to loudness; or, in another condition, according to pitch). Comparing sorting times and performance with a *baseline* control test (also called *standard* or *discrimination* task, e.g., Garner, 1981) in which only one dimension varies (e.g., just loudness, with only high soft and loud tones, or just pitch, with only soft high and low tones) permits to evaluate the participants' attentional filtering capacities. Indeed, if processing of the target dimension entailed processing of the non-target dimension, participants would be unable to filter out irrelevant variations on the latter, which would lead to poorer performance (mainly, slower reaction times, *RTs*) in the *filtering* task (following Posner's (1964) nomenclature, but also called *orthogonal classification*, e.g., Patching & Quinlan, 2002) than in either baseline tasks, an effect referred to as *Garner interference* (e.g., Pomerantz, 1983). Garner interference should thus be distinguished from Stroop-like congruency effects (Stroop, 1935) since, rather than arising from the content of an irrelevant component that is present on every trial and that is assigned an incompatible response, *Garner interference* arises from variations on the irrelevant (but not necessarily incongruent) dimension across trials.

Garner interference may still result from difficulties of selective attention to the underlying dimensions (Thibaut & Gelaes, 2002), sometimes because a new dimension emerges from the combination of specific values on these dimensions (e.g., Garner, 1978 b). According to Garner (1974), a demonstration of *integrality* of multidimensional stimuli, namely of integrated, "holistic" processing of the dimensions manipulated by the experimenter requires not only the occurrence of interference, but, in addition, that correlated variations on the non-target dimension lead to a benefit, namely a *redundancy gain*. Indeed, if the underlying dimensions were processed in a *unitary fashion* (Grau & Kemler Nelson, 1988), correlation between dimensions would turn the whole stimuli to be more different from one another. According to an Euclidean metric of (dis)similarity, the perceptual distance between integral dimensions would be enhanced. For separable dimensions, on the contrary, (dis)similarity is based on a "city-block" metric in which (dis)similarity between multidimensional stimuli is additive (Torgerson, 1958), and hence no gain is expected in the redundant situation.

As regards the study of the interactions between segmental (phonemes) and suprasegmental (pitch or pitch contour) dimensions in speech, participants heard syllables varying along two dimensions, e.g., pitch level (manipulated through the vowel fundamental frequency, F_0) and segmental constituency, and were asked to classify these syllables according to their values on a previously specified target dimension. For example, in Wood (1974; 1975), when pitch was the target dimension, listeners had to put together low vs. high-pitched syllables, irrespective of the orthogonal variations of their phonetic content (/bæ/ or /gæ/). Pitch classification was not affected by phonetic variations, while segmental classification was slowed down by variations in pitch. Asymmetrical interference may however be specific to consonantal information: when the segmental task concerned vowel quality (e.g., /b / vs. /bæ/) rather than consonants, mutual, symmetric, interference between the segmental dimension and either the pitch or the loudness dimension was reported (Miller, 1978). These data seem to support the idea of a different treatment of vowels and consonants relative to pitch. In addition, according to Melara and Marks (1990), they would suggest that vowel and pitch are processed by the same, general auditory mechanisms, while consonants are processed at a later, phonetic level.

However, these results cannot be easily generalized to the interactions between lyrics and tunes in songs. Indeed, the pitch levels used in these studies are far away from tunes: Wood (1974, 1975) and Miller (1978) used static pitch levels, namely synthetic syllables recorded at a constant F_0 (104 or 140 Hz). Both music (including songs) and, within the linguistic domain, intonation and tones (in languages in which these distinguish lexical items or grammatical functions), are characterized by dynamic contours. Consistent with this idea, Repp and Lin (1990) used dynamic tonal contours in speeded classification studies. They observed mutual interference between segmental (consonant or vowel) and tonal information in Mandarin Chinese. More crucially, Lee and Nusbaum (1993) showed, in a filtering task, that English listeners suffer mutual interference between consonantal and pitch information only for dynamic tonal contours like the ones involved in intonation (and like those used by Repp & Lin, 1990) but not for static, constant level pitches, for which they present asymmetrical interference. In short, contrary to what was observed with static pitch levels, both vowels and consonants interact with speech tonal contours.

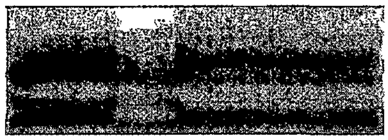
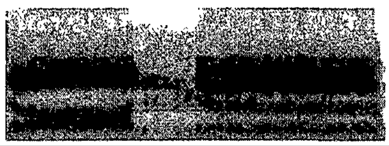
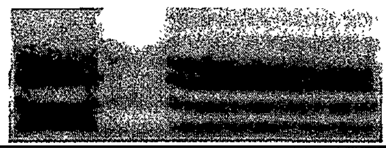
Yet, there are several shortcomings in these studies. The more important one is that, in most of them, processing interactions between dimensions were assessed only by examining the pattern of interference, which might at best reflect the listeners' inability to pay attention selectively to the target dimension, and certainly does not demonstrate integrality. To our knowledge, the only study that also used correlated stimuli to examine the interaction between linguistic tones and segmental information was Repp and Lin's (1990) one. Unfortunately, in this study there was a difference in the relative discriminability of the two dimensions, all participants discriminating tones more poorly than segments, and it is known that patterns of dimensional interaction are modulated by the relative discriminability of the involved dimensions (Garner, 1974; Garner & Felfoldy, 1970). In addition, it seems probable that even results obtained with speech tonal contours cannot be generalized to the processing interactions of other auditory dimensions such as the lyrics and tunes of songs. It has indeed been shown that the "consonant" and the "vowel" of synthesized C-V stimuli are processed as integral dimensions when the listeners consider them as purely linguistic, while they lead to a separability pattern when considered as a mixture of noise plus tone (Tomiak, Mullennix, & Sawusch, 1987).

In the present work, we adapted the speeded classification paradigm to the study of singing, and more specifically the study of the processing interactions between the phonological and melodic dimensions of consonant-vowel bisyllabic (CVCV) nonwords sung on a two-tone interval. In addition, we contrasted materials in which the CVCV nonwords differed by the middle consonant, either stop of nasal, and in which they differed by their final vowel. Naive, musically untrained participants were presented with a speeded classification set of tasks, using either natural (Experiments 1, 2, 3 and 4) or synthesized (Experiment 5) sung syllables. The speeded classification tasks of Experiment 1, 2, 3 and 5 included the filtering, redundant and baseline conditions designed by Garner (1974), an ensemble that constitutes the *filtering and redundancy tests* (Ashby & Maddox, 1994). In all these conditions, participants were required to respond according to the identity of either the "lyrics" (a bisyllabic pronounceable nonword varying on either the last vowel or the middle consonant) or the "tune" (an ascending or descending two-tones interval). The rationale for contrasting materials with either consonantal or vocalic contrasts was to confirm the hypothesis, based on the studies described above (Bonatti et al., 2005; Caramazza et al., 2000; Mehler et al., 2006; Nazzi, 2005), that different processes manage vowels and consonants and, henceforth, that these phonemes differ in their integration with melody.

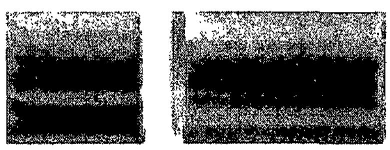
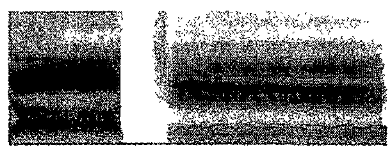
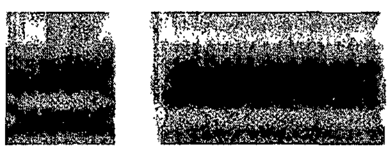
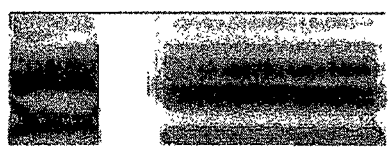
In Experiment 1, we examined the processing interactions between vowels and intervals. If, as found by Bigand and colleagues (2001), vowels and intervals are interactive dimensions, an integrality pattern reflected by both an interference cost and a redundancy gain was expected. In Experiment 2, the nonwords varied on the middle, voiceless stop, consonant. Given that stop consonants are claimed to be more specific to speech than vowels and constitute poor melodic supports, a separabiliy pattern was predicted. The purpose of Experiment 3 was to generalize the results beyond the case of stop consonants and to new vowel contrasts. If the difference of processing interactions between intervals and vowels vs. consonants is mainly due to their distinct acoustic features, one might expect nasal consonants to be more integrated with the tune than stop consonants, given that the former are more sonorous and more likely to provide melodic information than the latter. In contrast, if vowels and consonants are processed differently because they carry different functions in speech processing, we should not observe differences between nasal and stop consonants. The aim of Experiment 4 was to confirm, with a new task, the differences found between vowels and consonants. Participants were requested to perform a *condensation test* (cf. Posner, 1964) for which no single dimension can serve as the relevant basis of classification. Finally, in Experiment 5, we tested whether the integrality between vowels and intervals were related to the acoustical interactions between the spectral characteristics of the vowel and its pitch by replicating the filtering and redundancy test using a synthesized material in which these parameters ware carefully controlled. A summary of the stimuli used in all the experiments is presented in Table 1. Auditory examples of the stimuli can be heard on the website http://www.brams.umontreal.ca/plab/.

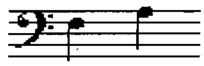
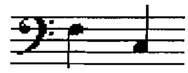
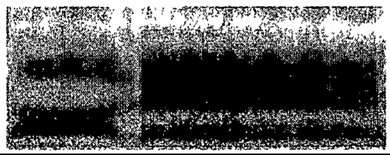
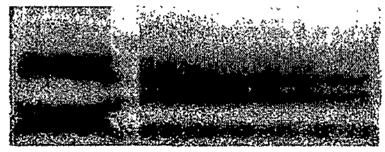
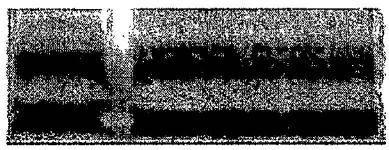
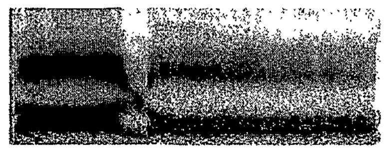
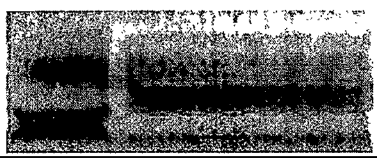
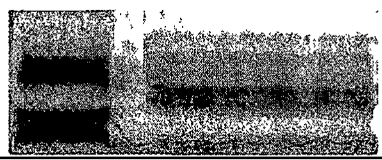
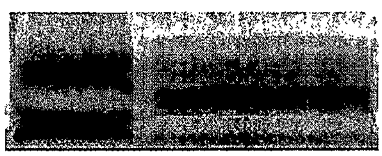
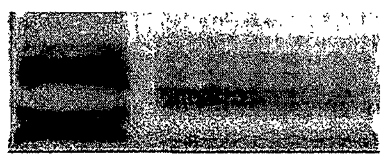
Table 1: The different combinations of nonword and interval, separately for Experiment 1 (part a, V1-material); Experiment 2 (part b, stop-C-material) and Experiments 3 and 4 (part c, V2- and nasal-C-matieral). The font styles of Table 1c should not be taken into account for Experiment 3. They refer to the condensation task of Experiment 4, in which the stimuli that had to be classified together are here presented in the same font style. The F_0 contour is marked in blue in the spectrograms.

a. Experiment 1

V1-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dalɔ̃/	I1NW1 	I2NW1 
Nonword 2 (NW2)	/dalø/	I1NW2 	I2NW2 

b. Experiment 2

Stop-C-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/daty/	I1NW1 	I2NW1 
Nonword 2 (NW2)	/daky/	I1NW2 	I2NW2 

c. Experiments 3 & 4			
V2-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dale/	**I1NW1** 	*I2NW1* 
Nonword 2 (NW2)	/dalɔ/	*I1NW2* 	**I2NW2** 
Nasal-C-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dany/	**I1NW1** 	*I2NW1* 
Nonword 2 (NW2)	/damy/	*I1NW2* 	**I2NW2** 

2. 1. 2. Experiment 1
On-line interactions between vowels and intervals in sung nonwords : Filtering and redundancy tests

In this experiment, participants had to classify bisyllabic nonwords sung on minor two-tones intervals on the basis of the identity of either the nonword (*phonological task*) or the interval (*melodic task*), in the three conditions defined by Garner (1974). The material is presented in Table 1a. The nonwords differed from each other by the identity of their final vowel, /ɔ̃/ vs. /ø/ and the intervals varied in their melodic contour, either ascending or descending.

Within each condition, the task remained formally the same: to associate each presented nonword (in the phonological task) or each interval (in the melodic task) to one of two response keys. In the baseline condition, only the dimension relevant for the task varied (e.g., for the phonological task, I1NW1 vs. I1NW2 in Table 1a). In the redundant condition, there was a systematic correlation between the interval and the nonword variations (e.g., I1NW1 vs. I2NW2). In the filtering, or orthogonal, condition, the four associations were presented and listeners had to ignore the irrelevant variations (e.g., for the phonological task, I1NW1 and I2NW1 vs. I1NW2 and I2NWP2).

Based on the idea that vowels constitute a natural melodic support and might thus be processed in an interactive way with melody, we predicted an integrality pattern (with participants showing facilitation in the redundant condition and interference in the orthogonal condition) to be observed.

2. 1. 2. 1. Method

2. 1. 2. 1. 1. Participants

Twenty-five undergraduate university students participated in return for credits toward an introductory psychology course requirement. One participant has been discarded from further analyses because his error rate in the baseline condition on intervals exceeded two-standard deviations from the average. The 24 remaining participants included 20 women and 4 men, average age: 20.8 years (range: 18-36). Nineteen participants had never learned music, and five had musical practice that did not exceed our inclusion criterion of four years of music experience stopped for at least five years. This inclusion criterion was identical in all following experiments.

2. 1. 2. 1. 2. Material

Stimuli were CVCV bisyllabic pronounceable nonwords sung on two-notes intervals. Each stimulus began with the first syllable /da/ sung on the note F3 and lasting around 500 ms. Thus, the differences between stimuli concerned only the longer (more than 1000 ms) second syllable. The slightly ascending F3-F3# and the descending F3-A2 minor intervals were combined with the nonwords /dalɔ̃/ and

/dalø/ varying by their last vowel (Table 1a). This material will henceforth be referred to as *V1-material.*

The stimuli, sung by a professional baritone, were recorded on mini-disc in an anaechoïc room. The baritone's range was chosen in order to avoid the major phoneme distortions linked to high frequencies that are generally observed in female opera singers (Scotto di Carlo, 2007a; Scotto di Carlo & Germain, 1985). During recording, the tempo was indicated by a silent luminous metronome. The chosen tempo was 120 beats/minute, with one pulsation for the first note and two for the second note.

To avoid response strategies based on obvious physical correlations between, for example, the pitch and the quality, or the length and intensity of the vowel, three physically different instances of each stimulus were used (i.e., three different recordings of I1NW1, I2NW1, I1NW2 and I2NW2). Selection of these three exemplars among five recordings of the stimuli was based on the following criteria: (i) Similar averaged F_0, vibrato and duration of the first syllable (/da/); (ii) pitch accuracy and duration of the second note / syllable: the recordings differing by more than one standard deviation from the averaged F_0 and duration were discarded; (iii) equivalent representation of relatively longer and shorter first and second syllables in each stimulus category, so that duration could not be a valid predictor of the interval or nonword category. The thorough description of the selected stimuli is provided in Appendix.

These selected stimuli were normalized for loudness and inserted in 2500 ms files. Within all files, silent intervals of different durations were inserted before and after the stimulus so that the transition between the two notes and syllables was centered at 1250 ms.

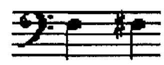
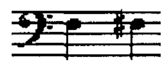
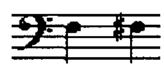
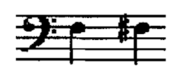
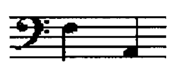
For each material, the same stimuli were presented in both a melodic task, which consisted of paying attention and respond to the intervals, and a phonological task, which consisted of paying attention and respond to the nonwords. Table 2 shows an example of the combinations of each task and condition.

A baseline block included only two values of the stimuli, which varied only on the target, attended dimension, the other dimension being held constant at one of its two possible values (see Table 2). There were thus two possible baselines for each task. For the phonological task, the /dalɔ̃/ and /dalø/ nonwords were both sung on the same interval, i.e., either both were sung with F3-F3# or both were sung with F3-A2.

Similarly, for the melodic task, stimuli were ascending F3-F3# and descending F3-A2 intervals sung on the same nonword, either /dalɔ̃/ or /dalø/. As illustrated in Table 2, the redundant blocks also included only two values of the stimuli, but these varied on both dimensions in a correlated way, with each value of the target dimension systematically associated with one of the two values of the other dimension. Again, there were two possible redundant pairs for each task: for both the melodic and phonological tasks, the redundant block included either the ascending F3-F3# interval sung on /dalɔ̃/ and the descending F3-A2 interval sung on /dalø/, or the reverse matching (F3-F3# sung on /dalø/ and F3-A2 on /dalɔ̃/). In the orthogonal condition, all four possible stimuli were presented.

For all conditions, each block included 72 trials, i.e. 12 presentations of each exemplar of each of the two different stimuli in the baseline and redundant conditions, and six presentations of each exemplar of each of the four different stimuli in the orthogonal condition. Within each block, trials were presented in random order.

Table 2: Example of response attribution for the different combinations of tasks and conditions in the speeded classification of Experiment 1 for the V1-Material. Each participant is presented with one of the baseline conditions, one of the redundant conditions and with the orthogonal condition, for each task (melodic and phonological).

Condition	**Task**			
	Melodic		**Phonological**	
	Response x	Response y	Response x	Response y
Baseline 1	 /dalɔ̃/	 /dalɔ̃/	 /dalɔ̃/	 /dalø/
or				
Baseline 2	 /dalø/	 /dalø/	 /dalɔ̃/	 /dalø/
Redundant 1	 /dalɔ̃/	 /dalø/	 /dalɔ̃/	 /dalø/
or				
Redundant 2	 /dalø/	 /dalɔ̃/	 /dalø/	 /dalɔ̃/
Orthogonal	 /dalɔ̃/	 /dalɔ̃/	 /dalɔ̃/	 /dalø/
	 /dalø/	 /dalø/	 /dalɔ̃/	 /dalø/

2. 1. 2. 1. 3. Procedure

Participants were tested individually in a quiet room, with stimuli presented through headphones. Stimuli presentation and response times were controlled via the *Psyscope* 1.2.5. PPC software (Cohen, Macwhinney, Flatt, & Provost, 1993) on a

Macintosh G3 computer. Errors and RTs were recorded via a button box, using only the left- and rightmost keys. Response time recording started at the beginning of the stimuli.

Each participant performed the speeded classification tasks on both dimensions (interval and nonword) in the three conditions (baseline, redundant and orthogonal) and thus completed 6 experimental blocks. They first performed the three experimental conditions (blocks) for one task (interval or nonword) before being presented with the other task. To avoid sequence effects, the presentation order of the tasks and conditions was counterbalanced between participants. Key assignment was also counterbalanced between-subjects, as was the choice of the specific baseline and redundant pairs. Between-subject assignment thus corresponded to a 2 (tasks) by 3 (conditions) by 2 (response key) by 2 (baseline and redundant pairs) design, so that at least 24 participants were needed for the experiment.

The instructions were presented simultaneously on the screen and orally by the experimenter, and emphasized both speed and accuracy. Participants were told that their task would be to classify nonsense sung words into two categories. For the task on intervals, it was explained that they ought to attend only to the melodies and not to the syllables, associating each melody to a response key. For the task on nonwords, instructions were identical but participants were told to attend only to the syllables and not to the melodies, associating each nonword to a response key. In order to familiarize participants with the task and the key assignment, a practice block of 12 trials preceded each experimental block. Auditory feedback was provided during the training: different beeps were triggered by correct, erroneous and timeout (longer than 2500 ms) responses. In the experimental blocks, no feedback of accuracy was given to participants: only the beep indicating a timeout response was maintained. Participants then completed the practice and the experimental blocks. The whole experiment lasted approximately 40 minutes.

2. 1. 2. 2. Results and discussion

The error rate averaged over task, condition and participants was 6.4%. Further analyses were performed on reaction times *(RTs)* to correct responses, except for the analyses aimed at addressing potential discriminability differences between tasks.[1]

Indeed, the relative discriminability of the involved dimensions is critical for interpreting the pattern of dimensional interactions (Garner, 1974; Garner & Felfoldy, 1970). We thus checked for potential discriminability differences between the intervals and nonwords on both accuracy and RT baseline values for each material. The analysis on accuracy actually revealed a discriminability difference, with less errors in the nonword than in the interval classification, $t(23) = 2.3, p < .05$. Such a difference was, however, not found on RTs, $t(23) = -.89, p > .10$.

The analysis of variance (ANOVA) on RTs to correct responses with task (melodic vs. phonological) and condition (redundant, baseline and orthogonal) as within-subject variables[2] showed no difference between the tasks, $F < 1$. The effect of experimental condition was, however, highly significant, $F(2, 46) = 36.58, p < .001$. In comparison to the baseline, there was significant interference in the orthogonal condition, $F(1, 23) = 37.86, p < .001$, and significant facilitation in the redundant condition, $F(1, 23) = 13.70, p < .001$. The interaction between task and condition was also significant, $F(2, 46) = 5.24$, $p < .01$. As illustrated in Figure 1, there was more interference in the orthogonal condition for the melodic than for the phonological task, $F(1, 23) = 8.93, p < .01$: the difference between baseline and orthogonal condition was highly significant for the melodic task, $F(1, 23) = 38.01, p < .001$ but only tended towards significance for the phonological task, $F(1, 23) = 3.087, < .10$. The redundancy gain was significant (and of similar amplitude, $F < 1$) in both tasks (nonword, $F(1, 23) = 7.51, p < .025$, interval, $F(1, 23) = 7.30$, $p < .025$).

The interference in the orthogonal condition was thus modulated by the task. This pattern of asymmetric interference can be explained by the slight difference of

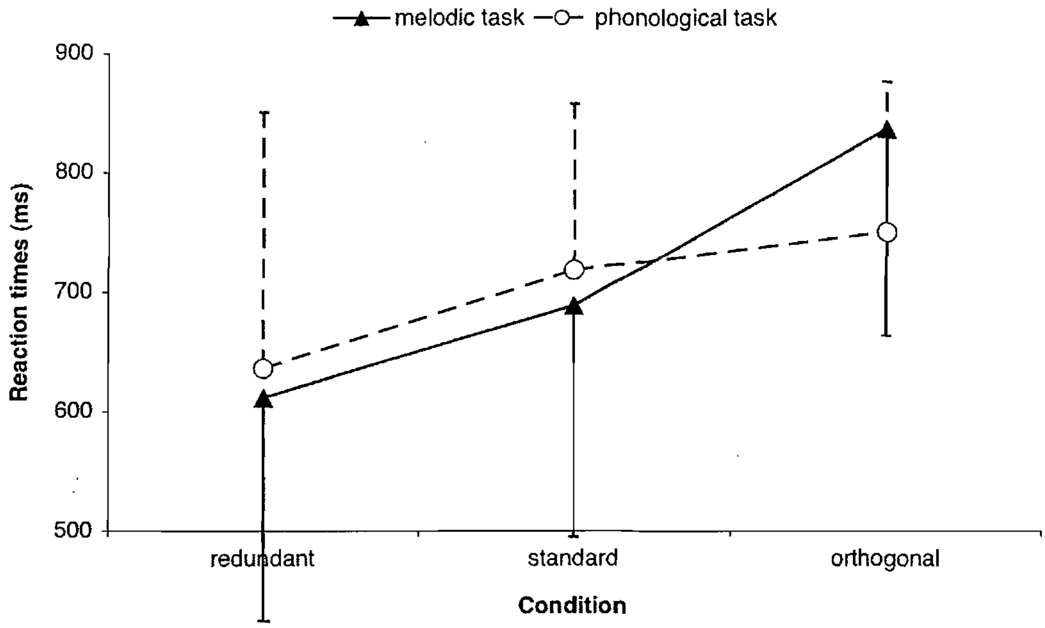
---

[1] In all these analyses, including those of the following experiments, RTs for correct responses were estimated by subtracting 1250 ms from the response times measured from the beginning of the stimuli, as the crucial information for the classification task (the transition between the two notes and syllables) was centred at 1250 ms.

[2] In a first step, the order of tasks and order of conditions were included as between-subjects variables. When the effects and interactions including these variables were not significant, they were removed from the reported analysis to allow focusing on the variables of main interest, i.e. task and condition.

discriminability between dimensions, at least in the error pattern. Indeed, if the values of one dimension are more salient than the values of the other dimension, the former will be easier to process, providing more opportunity for it to interfere on the processing of the latter (Ben Artzi & Marks, 1995; Garner & Felfoldy, 1970).

Figure 1: Average RTs for correct classification responses to V1-material as a function of target dimension (intervals: triangles and full lines, nonwords: circles and dotted lines) and condition. Error-bars represent one standard-deviation from the average.



In spite of this difference in the amount of interference, the results suggest that vowel and interval behave like integral dimensions. First, the difference in discriminability favoring nonwords cannot account for the almost significant interference effect observed in the phonological task: here, variations on the more difficult dimension (the interval) nevertheless tended to interfere with the processing of the easier dimension (the nonword). Neither can the difference in discriminability favoring nonwords account for the fact that in both tasks there was a significant facilitation effect in the redundant condition, without between-task difference in the size of the effect ($F < 1$). Indeed, a difference in discriminability between the two dimensions should induce an asymmetric redundancy gain, if participants were relying on the more salient dimension to perform the task (Ashby & Maddox, 1994).

2. 1. 3. Experiment 2

On-line interactions between stop consonants and intervals in sung nonwords : Filtering and redundancy tests

Our hypothesis that vowels and melodic intervals are integral dimensions was supported by the results of Experiment 1. In Experiment 2, we tested the additional hypothesis that consonants and intervals are less integrated, either because of their acoustic properties that prevent to carry melodic information or because of their higher linguistic specificity compared to vowels (Owren & Cardillo, 2006).

2. 1. 3. 1. Method

2. 1. 3. 1. 1. Participants

Among the 36 university students who initially took part in Experiment 2, two were discarded because of error rates exceeding two standard-deviations from the average in the baseline condition, and one because of musical experience above our inclusion criterion. The remaining group of 33 participants had an average age of 20.2 years (range: 18-23) and included three males. Most participants were non-musicians but nine of them had played a musical instrument during a maximum of four years, stopped for at least five years. They participated either in return for course credits (25 participants) or for financial compensation (8 participants).

2. 1. 3. 1. 2. Material and Procedure

The stimuli, recorded by the same baryton and in the same conditions as in Experiment 1, were the nonwords /daty/ and /daky/ varying on their middle, stop, consonant sung on the major ascending interval F3-A3 and the major descending interval F3-C3. As shown in the spectral representation of these stimuli in Table 1b, the consonants did not provide any F_0 information. In the following text, this material will be labeled *stop-C-material.* The procedure was identical to the one of Experiment 1.

2. 1. 3. 2. Results and discussion

The average error rate was 2.2%. On RTs, there was a significant discriminability difference between the phonological and melodic task in the baseline condition, $t(32) = 4.57$, $p < .001$, but the difference was not significant on accuracy, $t(32) = 1.38$, $p > .10$.

The ANOVA on correct RTs with task and condition as within-subjects variables[3] showed a highly significant effect of task, $F(1, 32) = 54.18$, $p < .001$ favoring nonwords. Both the effect of condition, $F(2, 64) = 15.53$, $p < .001$, and the interaction between condition and task, $F(2, 64) = 7.55$, $p < .001$, were significant. The effect of condition was due to significant facilitation in the redundant condition compared to the baseline condition, $F(1, 32) = 20.33$, $p < .001$, without significant interference in the orthogonal condition, $F < 1$. As depicted in Figure 2, the interaction arises from the effect of condition being restricted to the task on interval, with facilitation in redundant condition, $F(1, 32) = 34.68$, $p < .001$ but no Garner interference, $F(1, 32) = 2.84$, $p > .10$. On nonwords, there was no hint for between-condition differences, $F(2, 64) = 1.84$, $p > .10$.

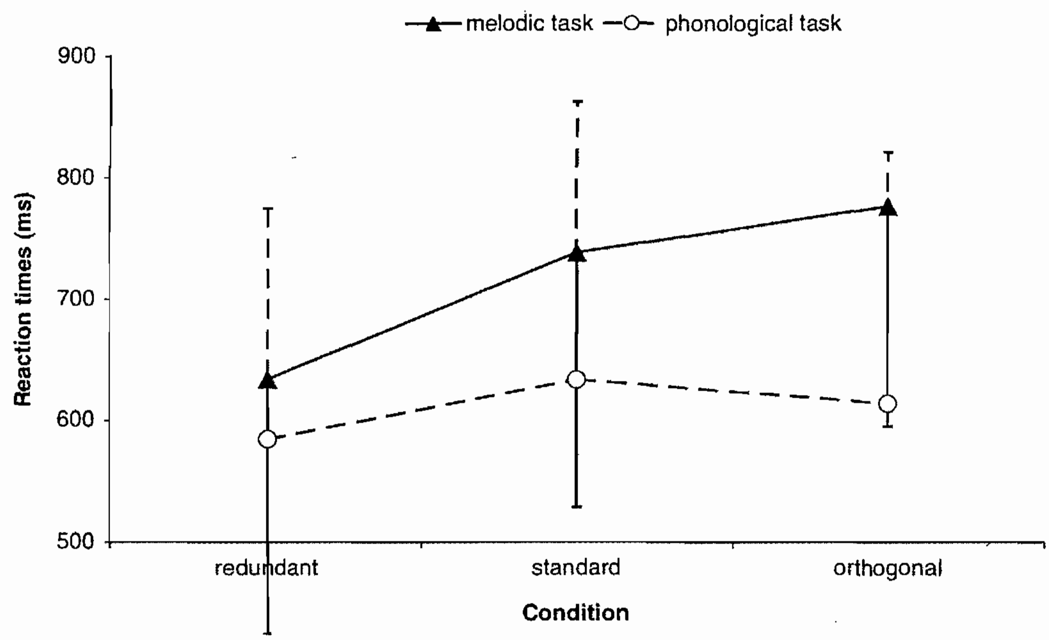
The redundancy gain restricted to the melodic task can easily be accounted for by the discriminability difference observed on reaction times. As a matter of fact, participants can, strategically, rely on the dimension that is easier to discriminate to carry on the redundancy test, hence leading to facilitation without significant interference in the orthogonal condition (Maddox & Ashby, 1996). Shepp, Barrett, & Kolbet (1987) proposed an insightful test to detect this « selective serial processing » strategy (Garner, 1984) : To compare the classification time in the redundant condition with the classification time in the faster baseline condition. If participants base their judgement on the identity of the nonword while ignoring the interval in the redundant condition, they should perform similarly as in the baseline condition on nonwords. Paired-samples t-test confirmed this view, $t(32) = .05$, $p > .10$.

---

[3] In a first analysis, the type of incentive for participants was involved as a between-subjects variable. Because this variable had no significant effect on the results and did not interact with either task and condition ($F < 1$), it was removed from further analyses.

Such a discriminability difference should also increase the interference of nonwords on the interval classification in the orthogonal condition, as discussed in Experiment 1. However, such interference was not significant. Altogheter, the evidence of selective serial processing strategy explaining the redundancy gain and the absence of significant interference in spite of discriminability differences that should favor it do not support the idea that stop consonants and intervals are integral dimensions. Instead, the results suggest that we encounter a separability pattern.

Figure 2: Average RTs for correct classification responses to stop-C-material as a function of target dimension (intervals: triangles and full lines, nonwords: circles and dotted lines) and condition. Error-bars represent one standard-deviation from the average.



2. 1. 3. 2. 1. Comparison between Experiment 1 (V1-material) and Experiment 2 (stop-C-material)

The results of Experiment 1 were compatible with an integrality between vowels and intervals, and the results of Experiment 2, with a separability between stop consonants and intervals. However, the occurrence of discriminability differences between dimensions weakens these conclusions. In order to directly compare the outcomes for vowels and stop consonants, we ran an ANOVA on the size of the effects. The amount of redundancy gain was computed as the difference in

RTs between the baseline and the redundant condition. The amount of Garner interference was computed as the difference in RTs between the baseline and the orthogonal condition. Material (Experiment) was a between-subjects variable and task was a within-subject variable. As expected, the vocalic or consonantal nature of the material significantly modulated the interference effect, $F(1, 55) = 15.82$, $p < .001$ with more interference for the V1-material (90 ms on the average) than for the stop-C-material (13 ms). However, there was no difference in redundancy gain between materials (80 ms on vowels, 73 ms on consonants), $F < 1$. The interaction between task and material were not significant on either the interference effect or the redundancy gain, $F(1, 55) = 1.95$ and $= 1.43$, $p > .10$, respectively.

In summary, the vocalic or consonantal nature of the material seems to influence the amount of integration with the melody but this difference was restricted to the Garner interference, probably because the participants used a selective serial processing strategy in the redundant condition on stop consonants.

2. 1. 4. Experiment 3
Generalization to new nonword-interval combinations, new vowels and nasal consonants

Experiment 3 had three objectives. First, in Experiments 1 and 2, there was a systematic association between, on the one hand, nonwords varying on vowels and minor intervals and, on the other hand, nonwords varying on consonants and major intervals. The size of these intervals also differed for the obvious reason that the baseline discriminability had to be as close as possible for the phonological and the melodic task in each material. This had the detrimental consequence of pairing the vocalic and consonantal nonwords with largely different interval pairs. Indeed, the V1-material was sung on either a slightly ascending or a largely descending interval, whereas the stop-C-material was sung on ascending and descending intervals of similar amplitudes. This confounding variable might have contributed to the observed integration difference between vowels and consonants. To control for the effect of interval, in Experiment 3, new nonwords varying on vowels were sung on the intervals that were associated to the stop-C-material of Experiment 2, that is, F3-A3 and F3-C3. Accordingly, new nonwords varying on consonants were sung on the intervals previously associated to the V1-material in Experiment 1 : F3-#F3 and F3-A2.

Second, to generalize the results of Experiment 1 on new vowels, we chose another vocalic contrast : the nonwords /dalɛ/ and /dal /. This material will henceforth be referred to as *V2-material,* see Table 1c. We expect to replicate the integrality pattern with these new vowel-interval asssociations.

Finally, to examine the contribution of the duration and sonority of the phonemes to their interaction with the melody, we used here a contrast of nasal consonants /n/ and /m/ (henceforth, *nasal-C-material,* Table 1c). These nasals, which are among the most sonorous consonants in French, are continuous consonants which allow air to flow through the nose. Therefore, they are more likely to carry melodic information than the voiceless, discontinuous stop consonants used in Experiments 2. A glance at the spectrogram of the stop-C-material (Table 1b) and the nasal-C-material (Table 1c) confirms this view: whereas no pitch information is carried by the stop consonants, which seem to break the melodic line, the nasal consonants support a continuous melodic contour. Furthermore, the transition between the first and the second tone are carried by these nasals, what makes them contribute in a fundamental way to melodic contour.

Two outcomes are possible with this nasal-C-material. If vowels are more integrated with the melody than consonants for acoustical reasons and given that integrality can be seen as a continuum (Grau & Kemler Nelson, 1988), the result pattern for the nasal consonants should be intermediate between integrality and a typical separability. If, on the contrary, the distinction between vowels and consonants rests on their different functions in speech processing, we should observe results similar to the ones of Experiment 2.

2. 1. 4. 1. Method

2. 1. 4. 1. 1. Participants

Nine of the 57 participants who initially took part in Experiment 3 were discarded: six because of poor performance in the baseline condition and three because of excessive musical experience. The 48 remaining participants were 22 men and 26 women aged 18 to 54 years (average: 22.6 years). Twenty-one had never learned music, and 27 had some musical practice corresponding to our criterion. Twenty were unpaid volunteers, 13 were undergraduate students in psychology who earned course credits and 15 were university students paid for their participation. All participants took part in the speeded classification task on both the V2- and the nasal-C-material.

2. 1. 4. 1. 2. Material

In the V2-material (Table 1c), the ascending F3-A3 and descending F3-C3 major intervals were combined with the nonwords /dalɛ/ and /dalɔ/. In the nasal-C-material (Table 1d), the F3-F3# and F3-A2 minor intervals were associated with the nonwords /dany/ and /damy/ varying by the place of articulation of the middle, nasal, consonant. Stimulus recording and editing was exactly as in the former Experiments, except that the stimuli were sung by another professional baryton.

2. 1. 4. 1. 3. Procedure

The procedure was the same as described above, except that each participant took part in the classification task on both the V2- and the nasal-C-material in one session. The order of presentation of the materials was thus counterbalanced between-subjects, each subject completing the whole classification on one material (two tasks, three conditions) before being presented with the other material. The order of materials, task, condition and the response key assignment were counterbalanced across participants. The whole experiment lasted about 80 minutes.

2. 1. 4. 2. Results and discussion

The speeded classification tasks were relatively easy: the error rates averaged over participants were 1.8% for the V2-material and 2.4% for the nasal-C-material.

In the baseline condition, accuracy was similar in the phonological and melodic tasks for the V2-material: $t(47) = -1.4$, $p > .10$. The trend to perform the phonological task slightly faster (644 ms) than the melodic task (664 ms) did not reach significance, $t(47) = 1.79$, $p < .08$. For the C-material, there was no discriminability difference between the phonological and the melodic dimension was observed, either on accuracy scores, or on RTs: $t(47) = -.017$ and $= 1.15$, respectively, $p >.10$ in both cases.

The ANOVA taking material (V2- vs. nasal-C-material), task (phonological vs. melodic) and condition (baseline, redundant, or orthogonal) as within-subject variables[4] revealed significant main effects of material, task and condition, $F(1, 47) = 14.29$, 20.94, and 105.74, all *p*s < .001. The reaction times were, on the average, 30 ms shorter for the V2- than for the nasal-C-material and the participants performed better in the phonological than in the melodic task. Overall, the effect of condition corresponded to an integrality pattern. However, the condition interacted significantly with material, $F(2, 94) = 12.7, p < .001$ and with task, $F(2, 94) = 18.68$, $p < .001$. The interaction between material and task was significant too, $F(1, 47) = 6.61, p < .025$, as was the second-order interaction between material, condition and task, $F(2, 94) = 17.74, p < .001$. This justified the more detailed follow-up analyses.

The decomposition of this second-order interaction shows, for the V2-material, significant effects of task, $F(1, 47) = 30.64, p < .001$ and of condition, $F(2, 94) = 107.48, p < .001$). In comparison to the baseline, RTs were faster in the redundant condition, $F(1, 47) = 20.81, p < .001$ and slower in the orthogonal condition ($F(1, 47) = 115.84, p < .001$).

For the V2-material, the interaction between task and condition was also highly significant, $F(2, 94) = 33.78, p < .001$. As can be seen in Figure 3a, although interference in the orthogonal condition was significant in both the melodic and phonological tasks, $F(1, 47) = 108.27, p < .001$ and $F(1, 47) = 10.26, p < .005$, respectively, this effect was more pronounced when participants processed the melodic contrasts, $F(1, 47) = 59.76, p < .001$. As in Experiment 1, this pattern of asymmetric interference could be explained by the slight difference in discriminability between the two dimensions.

For the nasal-C-material, there was a significant effect of task, $F(1, 47) = 5.20, p < .05$ but it was less prominent than in the V-material, what explains the

[4] In a first step, we introduced the type of incentive the participants received (credits, money or nothing) as an additional between-subjects variable. Given that this variable had no significant effect on the RTs and did not interact significantly with any of the within-subjects variables, it was removed from the analyses for sake of clarity.

interaction between task and material. The effect of condition was also significant: $F(2, 94) = 23.45$, $p < .001$. As illustrated in Figure 3b, in comparison to the baseline, there was significant interference in the orthogonal condition, $F(1, 47) = 28.32$, $p < .001$, but no significant facilitation in the redundant condition, $F(1, 47) = 1.82$, $p > .10$. Task and condition did not interact, $F < 1$.

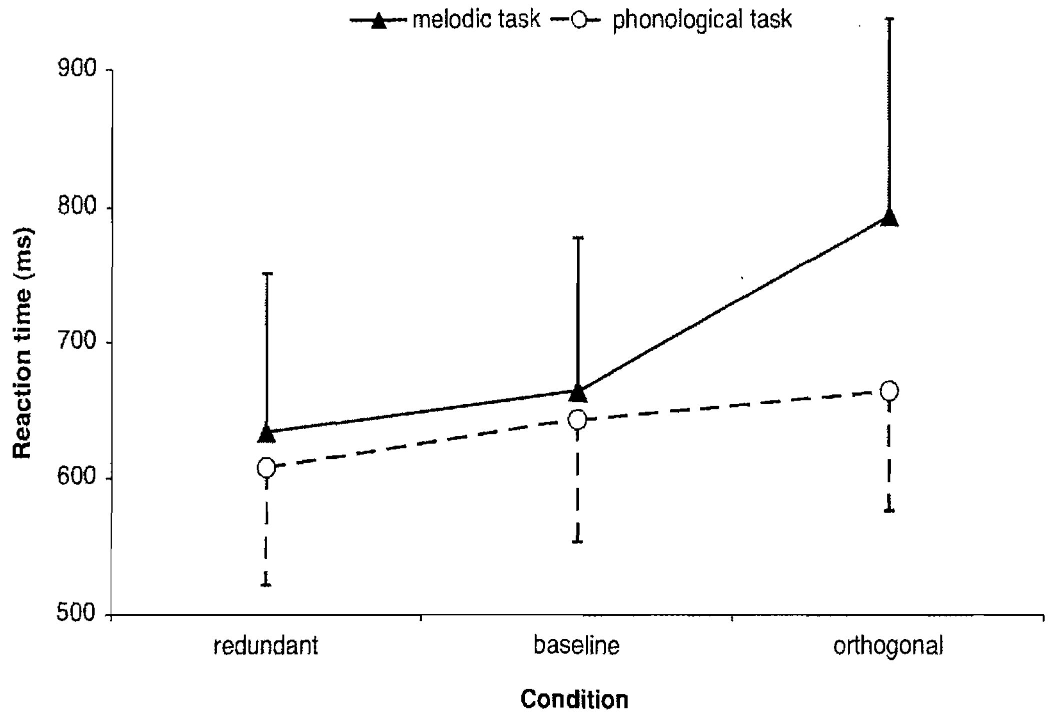
2. 1. 4. 2. 1. Comparison between V2- and nasal-C-materials

As described above, with the V-materials of Experiments 1 and 3 we observed not only mutual (although asymmetric) interference but also, and crucially, mutual and symmetric facilitation in the redundant condition. By contrast, for varying nasal consonants, we observed interference, but no significant redundancy gain. To further explore this processing difference between the V- and the nasal-C-materials, we ran additional analyses of variance on Garner interference and redundancy gain with task and material as within-subjects variables.

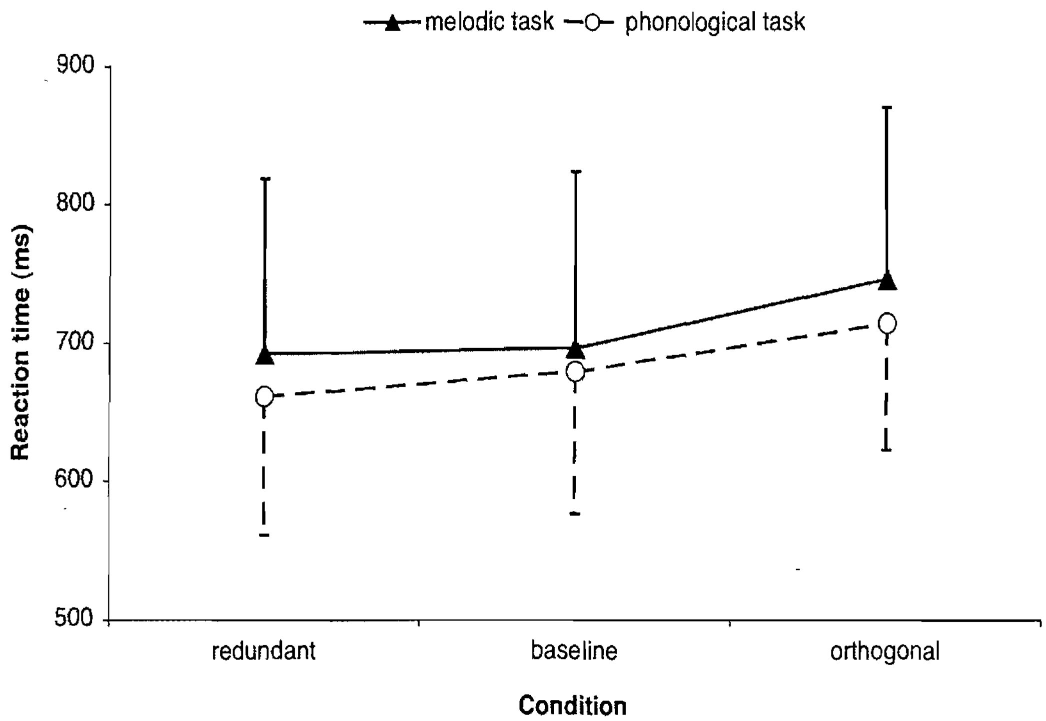
The material significantly modulates the amount of interference, $F(1, 47) = 10.03$, $p < .005$ with more interference for the V2- (76 ms) than the nasal-C-material (42 ms). The effect of task, $F(1, 47) = 32.39$, $p < .001$, and the interaction between material and task also reached significance, $F(1, 47) = 26.44$, $p < .001$. As depicted in Figure 3, this interaction reflects the fact that the interference differed between tasks for the V2-material only. Concerning redundancy gain, there was a non-significant trend towards larger facilitation for the V2- (32 ms) than for the nasal-C-material (12 ms), $F(1, 47) = 3.14$, $p < .09$. The effect of task and the interaction between task and material were not significant, both Fs < 1.

Figure 3: Average RTs for correct classification responses to V2-material (a) and nasal-C-material (b) as a function of target dimension (intervals: triangles and full lines, nonwords: circles and dotted lines) and condition. Error-bars represent one standard-deviation from the average.

a.



b.



Because the intervals of the V2-material of Experiment 3 were identical to the ones of the stop-C-material of Experiment 1, it seems that the choice of the intervals does not play a fundamental role in the dimensional interaction difference found for vowels and consonants. This experiment thus generalizes the conclusion that vowels and intervals are integral dimensions.

For the nasal-C-material, the lower Garner interference and slightly lower redundancy gain compared to the V-material could suggest that even nasal consonants and intervals are less integrated than vowels and intervals are. However, the picture is not so clear, given that significant interference without facilitation is not a typical separability pattern. It may either reflect the listeners' inability to pay attention selectively to the target dimension in spite of genuine dimensional separability (Thibaut & Gelaes, 2002) or the occurrence of an emerging dimension (Pomerantz & Garner, 1973). The aim of Experiment 4 will be to disentangle these interpretations but, before turning to this point, we will contrast the outcomes for the two kinds of consonants, i.e. stops and nasals.

2. 1. 4. 2. 2. Comparison between the stop-C- and the nasal-C-materials

To test whether the sonority of the phonemes contributes to their interactions with the melody, we compared the redundancy gain and the Garner interference between the stop-C-material of Experiment 2 and the nasal-C-material of Experiment 3. The task was a within-subject variable and the material a between-subjects variable. If the lack of sonority and the resulting inability to carry melodic information is the origin of the separability between stop consonants and intervals, we should observe more facilitation in the redundant condition and less interference in the orthogonal condition, compared to the baseline, for the nasal than the stop consonants.

The kind of consonantal contrast had actually a significant effect on the redundancy gain, $F(1, 79) = 12.10$, $p < .001$ but the direction of the effect was opposite to the one expected on the basis of sonority, with a larger facilitation for the material involving stop consonants (80 ms) than for the one involving nasals (12 ms). The task did not influence the results, $F(1, 79) = 1.24$, $p > .10$ but it interacted significantly with the material, $F(1, 79) = 4.81$, $p < .05$: the redundancy gain differed between the materials only for the task on intervals, $F(1, 79) = 20.84$, $p < .001$.

As regards Garner interference, the effect of task was significant, $F(1, 79) = 5.75, p < .025$ but neither the effect of material, nor the interaction between task and material reached the significance level, $F(1, 79) = 2.51$ and $= 1.845$, both $ps > .10$.

The present results provide no clue for the contribution of sonority to the integration between lyrics and tune : the more sonorous nasal consonants were not more integrated to the interval than the least sonorous, stop consonants. These surprising results can be related to two additional differences between materials, which might in fact be linked to each other : the occurrence of a discriminability difference in the stop-C-material resulting in a selective serial processing strategy, and the use of different melodic intervals in the stop- and the nasal-C-materials. An ideal study should compare consonants differing in sonority but sung on the same intervals and exhibiting no discriminability difference, a challenge that seems very difficult to take up.

2. 1. 5. Experiment 4
Condensation tests

The occurrence of an interference effect without redundancy gain in the nasal-C-material corresponds neither to integrality, nor to separability patterns. According to some authors, interference without facilitation merely arises from difficulties to pay attention selectively to separate dimensions, because of task difficulty and/or lack of discriminability between the values of the dimensions. This has been shown, for example, in developmental studies on attentional filtering capacities (e.g., Thibaut & Gelaes, 2002).

For other authors, including Garner (1974; see also Pomerantz & Garner, 1973), interference without facilitation may suggest *configural interaction* between dimensions, due to the presence of an *emergent feature*. In the visual domain, interference without facilitation has been observed, for example, when participants have to classify pairs of parentheses according to, say, the orientation of the left parenthesis, thus pooling ((and () on the one hand, and)) and)(on the other hand. According to the *configural interaction* interpretation, participants' difficulty to pay attention selectively to the orientation of the left parenthesis would be explained by the emergence of the more salient new features of symmetry and closure distinguishing ((and)) from () and)(. Independent evidence has been provided that the emergent features are more salient and discriminable from one another than are the underlying dimensions (the individual parentheses), leading to a *configural superiority effect* (Pomerantz, Sager, & Stoever, 1977) in discriminating () from)) in comparison to discriminating (and). In the filtering task, the emergent features are not useful since these are not mapped onto response categories in a one-to-one fashion (Garner, 1974, 1978b; Pomerantz & Garner, 1973; Pomerantz & Schwaitzberg, 1975), explaining a drop in performance.

Although in our view no obvious new dimension could have emerged from the association between a specific interval and a specific nonword of the kind we used in Experiment 3, we tested this alternative account of the nasal-C-material results by using a condensation task (also called *biconditional classification* by Garner, 1974). In addition, such a task is also appropriate for confirming the suggestion that vowels and interval processing is more integrated than consonants and interval processing.

Condensation is an attentional task in which no single dimension can serve as the relevant basis of classification: participants have to classify in the same category stimuli differing on the values of both dimensions. Considering the example of the pairs of parentheses, it would require to put ((and)) together on the one hand, and () and)(on the other hand. As illustrated by this example, when some emergent features distinguish these two categories of responses, as here parallelism and symmetry, the condensation task is carried out more easily than the filtering task, in which the emergent features are not mapped onto response categories. In contrast, in the absence of an emergent feature, condensation leads to poorer performances than filtering, in particular when dimensions are separable (e.g., Fitts & Biederman, 1965; Gottwald & Garner, 1972; Keele, 1970; Morin, Forrin, & Archer, 1961). Indeed, in the absence of an emergent dimension, the only way for participants to respond correctly is to pay attention to both dimensions at the same time, resulting in slower RTs than when filtering requires only consideration of a single dimension. More crucially for our purpose, the condensation task is easier on integral than on separable dimensions, even tough in both cases it is more difficult than the filtering task (e.g., Garner, 1974; Gottwald & Garner, 1975). This is because with separable dimensions the participant must attend to two discrete dimensions, while with integral dimensions the two classes of responses are more easily differentiated on the basis of the greater (Euclidean) perceptual distance between the whole stimuli (Garner, 1974).

Based on this logic, we presented the two materials used in Experiment 3 to a new group of participants, using two separate condensation tasks. Each required participants to group stimuli differing on the values of both dimensions (interval and nonword) in the same category. In other words, the two redundant subsets (of two stimuli each) had to be sorted into a single class. For example, as illustrated by the fonts in Table 1c, with the nasal-C-material participants had to classify /damy/ sung on F3-F3# (I1NW2) and /dany/ sung on F3-A2 (I2NW1) together, vs. /damy/ sung on F3-A2 (I2NW2) and /dany/ sung on F3-F3# (I1NW1). If in the nasal-C-material there were new features emerging from the association between consonantal and interval values, this task should be easier than the filtering task of Experiment 3. Furthermore, if the pattern of results observed in Experiment 3 corresponds to separability for the nasal-C-material and integrality for the V2-material, better condensation performance (less errors and faster RTs) should be observed with the latter than with the former.

2. 1. 5. 1. Method

2. 1. 5. 1. 1. Participants

Twenty-four undergraduate students took part in Experiment 4 for course credit. They were 4 men and 20 women aged between 17 and 31 year old (average: 20.7). Twelve participants had no musical experience, while the others had musical experience corresponding to our inclusion criterion.

2. 1. 5. 1. 2. Material and procedure

The two different materials of Experiment 3 were used, each including the four stimuli (with three physically different exemplars of each stimulus) formerly used in the orthogonal condition.

Each participant took part in two condensation tasks, one on the V2-material (variation of vowels, major intervals) and one on the nasal-C-material (variation of nasal consonants, minor intervals). As illustrated in Table 1c, each condensation task required the two redundant subsets (of two stimuli each) to be sorted into a single class. In other words, participants had to group stimuli differing on the values of both dimensions (interval and nonword) in the same category.

Each task included one experimental block of 72 trials (18 presentations of each stimulus, 6 presentations of each exemplar of one stimulus) preceded by a practice block of 12 trials (3 presentations of each stimulus corresponding to 3 different recordings). Order of trials within each block was random. For each material, response assignment to keys and order of materials were counterbalanced across participants.

The procedure was the same as in the previous experiments with testing time reduced to about 20 minutes.

2. 1. 5. 2. Results and discussion

Confirming the notion that condensation is a fairly difficult task, delayed responses (i.e., longer than 2500 ms and hence discarded from further analyses) were numerous, reaching about 8% of the trials for each material. On the remaining trials, error rates were also quite high: 28% for the C-material and 24% for the V-material.

We first compared the classification times as well as the accuracy rates between the filtering tasks (i.e., the orthogonal conditions) of Experiment 3 and the condensation task, separately for each material and for each dimension.

As shown in the left part of Figures 4a and b (illustrating mean RTs to correct responses and error rates, respectively), for the nasal-C-material participants were both slower and less accurate in the condensation than in the filtering task. This holds true for the comparison of condensation with both the melodic filtering task, in which participants had to classify the orthogonal stimuli on the basis of their interval values: $t(70) = 9.81$ for correct RTs and $= 9.26$ for errors, respectively, $p < .001$ in both cases, and the phonological filtering task, in which they had to classify the same stimuli on the basis of the identity of the nonwords: $t(70) = 13.90$ for correct RTs and $= 11.04$ for errors, $p < .001$ in both cases.

For the V2-material (see the right part of Figures 4a and b) participants were also slower and less accurate in the condensation task than in both filtering tasks (melodic task: $t(70) = 5.82$ for correct RTs and $= 6.51$ for errors; phonological task: $t(70)=14.02$ for correct RTs and $= 8.78$ for errors, $p < .001$ in all cases).

These results suggest that no new configural feature emerged from the dimensions used in the nasal-C-material. Had it been the case, performance would have been better in the condensation than in the filtering task (Garner, 1978b). Since the reverse pattern was observed, the interference effect reported in Experiment 3 for this material was in all likelihood related to a lack of selective attention to otherwise separable dimensions.

Further evidence for the idea that we encountered a separability pattern for the nasal-C-material and an integrality pattern for the V-material was provided by a second set of analyses, in which we compared the condensation performance as a function of material. Separate ANOVAs, each with material as within-subjects variable and order of materials as between-subjects variable, were carried out on RTs and correct responses. The main effect of material was significant in the RT analysis, $F(1, 22) = 7.44$, $p < .025$. As expected and as can be seen in Figure 4a, participants performed the condensation task by 45 ms faster, on the average, on the V2-material than on the nasal-C-material. No further effect or interaction was observed on RTs (all $Fs < 1$).

In the accuracy analysis the effect of material was close to significance and favored by 4%, on the average, the V2-material, $F(1, 22) = 3.41$, $p < .08$. While there was no effect of material order, $F < 1$, the interaction between material and material order was significant, $F(1, 22) = 19.34$, $p < .001$. Indeed, the participants' tendency to be more accurate for the second material they were presented with was modulated by the relative difficulty of the materials: those who were presented with the nasal-C-material first showed a 13% advantage, on the average, for the subsequently presented V2-material, $F(1, 22) = 13.34$, $p < .005$, while those who began the task with the V2-material had only a 5% advantage for the subsequently presented nasal-C-material, $F(1, 22) = 5.97$, $p < .05$. In other words, once familiarized with the particularly difficult condensation task, the performance advantage for the second presented material was greater with the V2-material than with the nasal-C-material.

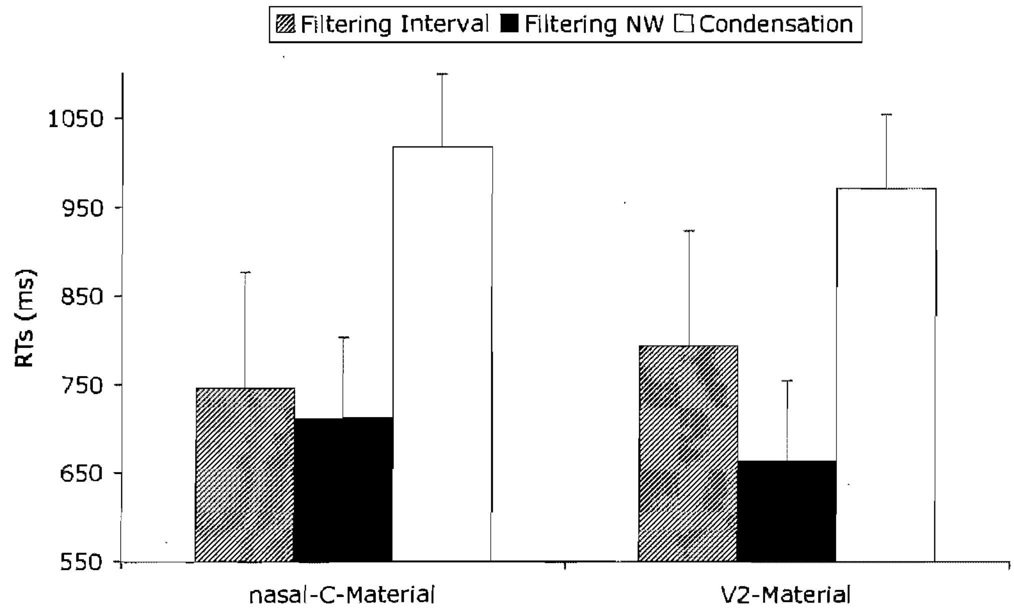
Both the RT and accuracy condensation performances thus support the idea that vowels and intervals are more integrated dimensions than consonants and intervals are, since the condensation task is relatively easier with the former than with the latter material (e.g., Garner, 1974; Gottwald & Garner, 1975).

The next experiment was aimed at better understanding the basis of the integrality pattern observed with the vocalic materials of Experiments 1 and 2.

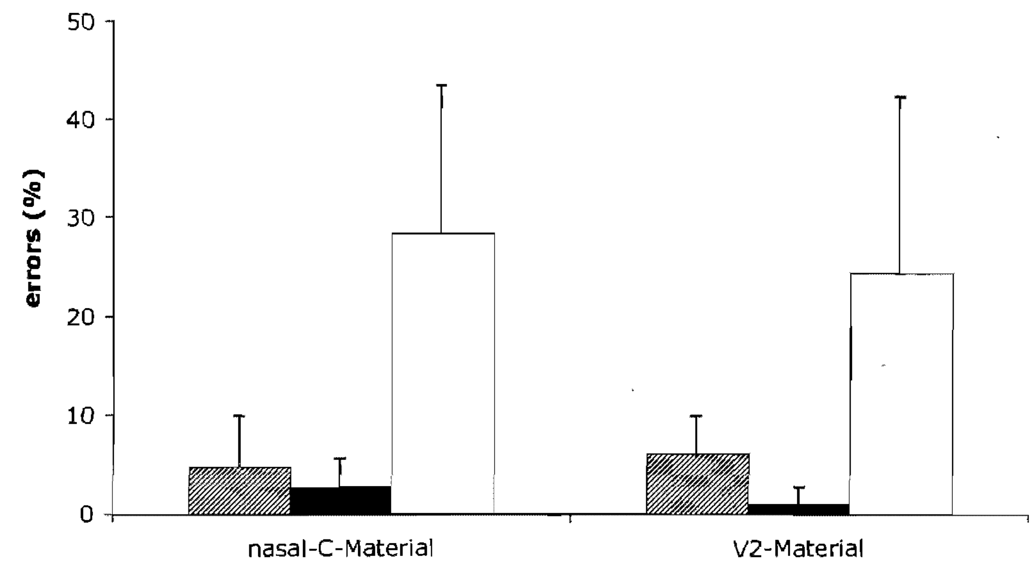
Figure 4. **a.** Average RTs for the nasal-C-material (left) and the V2-material (right) in the filtering (Experiment 3) and condensation (Experiment 4) tasks.

b. Average percentage of errors for the nasal-C-material (left) and the V2-material (right) in the filtering (Experiment 3) and condensation (Experiment 4) tasks. Error bars represent one standard-deviation from the average.

a.



b.



2. 1. 6. Experiment 5

Filtering and redundancy tests on synthesized singing with a vocalic material

In Experiments 1 and 3, we observed an integrality pattern for the vocalic materials, whereas it was not the case for the consonantal materials of Experiment 2 and 3. Such a difference may suggest that, in contrast to consonants, vowels and intervals are at least partly processed by common auditory mechanisms.

Alternatively, this response pattern may reflect physical interactions between the linguistic and musical dimensions. In singing, different tunes can alter the intelligibility of speech. This is well known by singers as well as by opera lovers, and studies on song production and intelligibility have confirmed how difficult it is to understand words that are sung (for a review, see Scotto di Carlo, 2007a, 2007b). This effect seems to depend mainly on the intelligibility of vowels, which is inversely proportional to their pitch (e.g., Gregg & Scherer, 2006; Scotto di Carlo & Germain, 1985; Sundberg, 1982). This is related to the fact that, in songs, the spectral characteristics of the vowels vary with pitch height (for a review, see Astesano, Schön, & Besson, 2004).

To control for such interactions in our V-materials, we examined the spectrum of the second vowel (Boersma & Weenink, 2007). The frequencies of the first (F1) and second (F2) formants of the vowels of the V-materials of Experiments 1 and 3, averaged over the three exemplars of each stimulus, are shown in Table 3 and the detailed values for each exemplar are provided in Appendix. The difference in pitch (A2 and #F3) did slightly influence the formant frequencies of the vowel /ø/ of V1-material (Experiment 1), especially by decreasing the frequency of F2. In the V2-material from Experiment 3, the effect of pitch change on the formant frequencies was more systematic, probably because the global pitch was higher than in the V1-material. The increase in pitch from C3 to A3 tends to slightly decrease F1 for /dal / and to increase it for /dalε/. On the contrary, a rise in pitch increases F2 for /dal / and decreases it for /dalε/. Thus, these vowels sung on high pitches should be slightly less discriminable, as is usually the case in song production (Astesano et al., 2004; Scotto Di Carlo, 2007 a, b).

This correlation between the spectral characteristics of the vowels and pitch height may have reinforced the integrality effect, at least, for the V2-material. This could, for example, explain that it was only for this material that the redundancy gain and the Garner interference were significant in both phonological and melodic tasks. Although variable from one item to the other (see Appendix), the spectral differences between, say, a high-pitched /ε/ and a low-pitched /ε/ might have induced the listeners to rely on these fine acoustical differences to perform the task. In order to check whether this could account for the results obtained with the V2-material, we synthesized and tested a new V-material that did not include any acoustical correlation between pitch height and vowel quality.

2. 1. 6. 1. Method

2. 1. 6. 1. 1. Participants

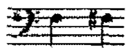
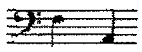
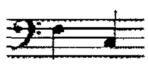
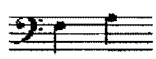
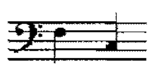
Twenty-four paid university students (16 men and 8 females aged between 18 and 37 years old, average: 25.8 yrs.) out of 26 met our musical experience criterion. Nine had never played music; the others had little practice.

2. 1. 6. 1. 2. Material

A source-filter synthesizer[5], simulating the vocal tract, was used to synthesize four stimuli that replicate V2-material. The nonwords /dalε/ and /dalɔ/ were sung on F3-A3 and F3-C3. For each vowel, five formants were modeled, the frequency of the first four formants determining the nature of the vowel, F5 being always 1000 Hz higher than F4. The central frequency values of the first four formants were fixed, so that they could not be influenced by pitch (Table 4, left panel, for F1 and F2 values). The plosive consonant /d/ and the liquid /l/ were modeled by manipulating the formant transitions. The model also fixed the fundamental frequency at 220 Hz for A3 and 130 Hz for C3. Syllable duration was 500 ms for the first syllable, and 1330 ms for the second. In order to obtain a singing-like result, a vibrato deviating from 2 % of the fundamental was added. Finally, slight random variations of pitch, harmonics, duration and vibrato were included to produce a more natural nonword. A cost of these random variations is that the measured formant frequency was not

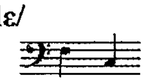
[5] This source-filter singing synthesizer is the work of Olivier Bélanger. We wish to thank him for his help in creating the material of Experiment 5.

Table 3: Averaged values of F1 and F2 in the final vowel of the natural stimuli of the V-materials used in Experiments 1 (V1-material) and 3 (V2-material), in Hz.

V1-Material / **Stimulus**	**Averaged F1 of the second vowel (Hz)**	**Averaged F2 of the second vowel (Hz)**
Averaged /daБ/ 	578	2139
Averaged /daБ/ 	583	2134
Difference	5	-5
Averaged /dalø/ 	403	1299
Averaged /dalø/ 	392	1246
Difference	-11	-53
V2-Material / **Stimulus**	**Averaged F1 of the second vowel (Hz)**	**Averaged F2 of the second vowel (Hz)**
Averaged /dalɛ/ 	497	1706
Averaged /dalɛ/ 	511	1636
Difference	14	-70
Averaged /dalɔ/ 	545	838
Averaged /dalɔ/ 	520	893
Difference	-25	55

identical across stimuli. However, a comparison of the left and right panels of Table 4 shows that these differences were minimal and, more crucially, that pitch differences did not modify F1 and F2 frequencies in a systematic way. As these random variations were also present in syllable 1 and could influence the participants' classification judgment, the two syllables of each of the four stimuli were cross-spliced just before the /l/ consonant, yielding four different exemplars of each stimulus.

Table 4: First and second formant frequency implemented in the source-filter synthesizer while programming (left) and measured afterwards in the synthesized stimuli (right) of Experiment 5.

Synthesized material / **Stimulus**	**Implemented F1 of the second vowel (Hz)**	**Implemented F2 of the second vowel (Hz)**	**Averaged measured F1 of the second vowel (Hz)**	**Averaged measured F2 of the second vowel (Hz)**
/dalɛ/ 	600	1770	609	1778
/dalɛ/ 	600	1770	607	1780
Difference	0	0	-2	2
/dalɔ/ 	500	900	507	897
/dalɔ/ 	500	900	503	896
Difference	0	0	-4	-1

2. 1. 6. 1. 3. Procedure

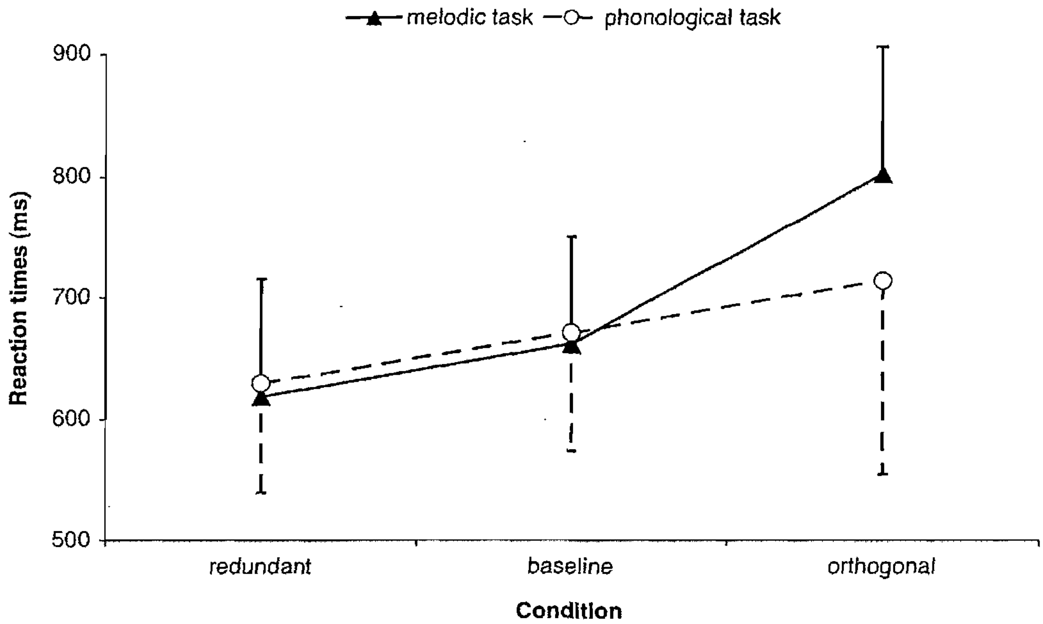
The procedural aspects were the same as in Experiments 1 and 2, except that the training blocks included 16 trials instead of the 12 used with natural speech, in order to include the 16 physically different stimuli (four exemplars of the four stimuli). Similarly, experimental blocks counted 80 trials instead of 72. The experiment lasted about 40 minutes.

2. 1. 6. 2. Results and discussion

The averaged error rate was 2.6%. There was no discriminability difference between the melodic and the phonological dimensions in the baseline condition, $t(47) = -.598$ on RTs, and $= .778$ on errors, both *ps* > .10.

The ANOVA run on correct RTs, with condition and task as within-subject variables, showed no significant effect of task, $F(1, 23) = 1.67$, $p > .10$. In contrast, both the effect of condition, $F(2, 46) = 43.89$, $p < .001$, and the interaction between task and condition, $F(2, 46) = 7.97$, $p < .001$, were significant. As apparent in Figure 5, the results replicated the ones obtained for the natural singing V2-material in Experiment 3. The effect of condition reveals highly significant interference in the orthogonal condition, $F(1, 23) = 40.99$, $p < .001$, and facilitation in the redundant condition, $F(1, 23) = 13.32$, $p < .001$, as compared to baseline. As with natural speech, there was more interference in the orthogonal condition for the melodic than for the phonological task, $F(1, 23) = 8.93$, $p < .01$, while the facilitation in redundant condition did not differ as a function of task ($F < 1$). In addition, the Garner interference and the redundancy gain were significant not only in the melodic task, $F(1, 23) = 29.70$, $p < .001$ and $= 7.36$, $p < .025$, respectively, but also in the phonological task, $F(1, 23) = 10.07$, $p < .005$ and $= 6.85$, $p < .025$. In brief, these results correspond to a typical integrality pattern except for the occurrence of asymmetric interference. Here, this asymmetry cannot be related to discriminability differences and might rather reflect difference in the processing levels of vocalic and pitch information (Garner, 1983; Melara & Marks, 1990).

Figure 5. Average RTs for correct classification responses with the synthesized V-material as a function of target dimension (interval: triangles and full lines; nonword: circles and dotted lines) and condition. Error-bars represent one standard-deviation from the average.



To check whether the natural vs. synthesized nature of the material influenced the pattern of interactions between vowel and pitch, a further ANOVA was run on the two sets of V-materials used in Experiment 3 and 5, with type of singing (natural vs. synthesized) as between-subject variable and task and condition as within-subject variables. Crucially for the present purpose, the effect of type of singing was not significant, $F < 1$, and did not interact with condition, $F(2, 140) = 1.574$, $p > .10$. Consistent with the fact that task had a significant effect for natural but not for synthesized singing, the interaction between task and type of singing tended to significance ($F(1, 70) = 3.324$, $p < .07$). However, the second-order interaction between task, condition and type of singing was far from significance ($F < 1$). In brief, the results observed with the natural and the synthesized material were very similar, hence suggesting that the occurrence of a redundancy gain in addition to an orthogonality cost reflect genuine psychological interactions and are not merely due to the acoustical correlation between pitch and vowel quality.

2. 1. 7. General Discussion

In the present study, we examined whether the phonological and melodic dimensions of sung material are processed independently or, instead, in an integrated way. To this aim, we adapted the filtering (Experiments 1, 2, 3 and 5) and condensation (Experiment 4) tasks designed by Garner (1974) to an auditory material made of CVCV nonsense sung syllables. More precisely, we compared materials with varying vowels (the V-materials, Experiments 1, 3, 4 and 5) and with varying consonants, either stops (the stop-C-material, Experiment 2) or nasals (the nasal-C-material, Experiments 3 and 4).

The underlying motivation for this separate manipulation of vowels and consonants is that both the phylogenesis and the informative values of these two classes of speech segments suggest that they pertain to distinct processing systems (e.g., Caramazza, et al., 2000). We therefore hypothesized that consonants and vowels may bear different relationships with other human auditory processing, involving music perception. Song is an obvious object for this inquiry, since it involves both music and speech.

The outcomes of the present series of Experiments are summarized in Table 5.

Overall, all Experiments involving materials varying on vowels, either naturally sung (Experiments 1 and 3) or synthesized (Experiment 5) led to an integrality pattern. Indeed, songs made out of nonwords that differed by their vowels did not only lead to interference in the orthogonal condition, but also to facilitation in the redundant condition. In other words, the classification times were shorter when, in the same block of trials, the value of the stimulus on one dimension was perfectly correlated with its value on the other dimension than when the latter dimension was kept constant.

It is worth noting that this interference-facilitation pattern occurred for each of the target dimensions and, in addition, for both natural and synthesized sung stimuli, as demonstrated in Experiment 5. All acoustic correlations between the two dimensions were eliminated from the latter stimuli, so that the observed results are likely related to the intrinsic characteristics of the perceptual processing of songs. For both the naturally sung and the synthesized V-materials, the participants' classification thus seem to reflect the use of an Euclidean metric of (dis)similarity,

Table 5. Summary of the main results of the five Experiments.

	Material(s)	Classification test	Results (RTs)	Interpretation
Experiment 1	V1-material	Filtering and redundancy tests	Melodic: Redundant < Baseline < Orthogonal Phonological : Redundant < Baseline ≤ Orthogonal	Integrality with asymetric interference
Experiment 2	Stop-C-material	Filtering and redundancy tests	Melodic: Redundant < Baseline = Orthogonal = Phonological: Redundant = Baseline = Orthogonal	Separability Redundancy gain compatible with serial processing strategy on separable dimensions
Experiment 3	V2-material	Filtering and redundancy tests	Melodic: Redundant < Baseline < Orthogonal Phonological : Redundant < Baseline < Orthogonal	Integrality with asymetric interference
	Nasal-C-material		Melodic: Redundant = Baseline < Orthogonal Phonological : Redundant = Baseline = Orthogonal	Emerging dimension or selective attention difficulties
Experiment 4	V2-material	Condensation test	Condensation < Filtering-Redundancy	
	Nasal-C-material		Condensation < Filtering-Redundancy Consensation : Nasal C < V2-material	No emerging dimension Nasal-C: more separable than V2-material
Experiment 5	Synthesized V2-material	Filtering and redundancy tests	Melodic: Redundant < Baseline < Orthogonal Phonological : Redundant < Baseline < Orthogonal	Integrality with asymetric interference

with perceptual distance between the stimuli enhanced in the redundant condition. Such a response pattern suggests that the underlying dimensions are processed in a somewhat integrated, unitary fashion. Finally, the integrality between vowels and intervals has been replicated with several intervals (Experiments 1 and 3) and, thus, appears to be independent of the identity of the intervals in presence.

However, the interference pattern was asymmetric in all experiments on vowels, with more pronounced interference of the vocalic variations onto interval processing than of the melodic variations onto vowel processing. We should however note that, except in the phonological task of Expertiment 1, the interference reached the significant level for both target dimensions. At first sight, this asymmetry could be interpreted as emerging from the higher discriminability of the linguistic compared to the musical information, a classical outcome in the literature on song processing (Peretz, Radeau, & Arguin, 2004; Serafine, Crowder, & Repp, 1984), especially with musically naïve participants. Still, this explanation cannot account for the asymmetry found with the synthesized material of Experiment 5, in which no discriminability difference between the target dimensions was found. Hence, it seems that this asymmetrical processing reflects deeper processing differences between vowels and intervals, such as the level at which the dimensions interact (Melara & Marks, 1990). As mentioned in the Introduction, asymmetries in Garner interference can reflect the fact that the least interfering dimension, here the interval, is processed at later level than the most interfering dimension, here the vowel. The finding that a phonetic processing occurs earlier than a supposedly more acoustical, pitch-related processing can seem counterintuive. This is particularly true if one remembers that the opposite asymmetry has been observed for stable pitch levels and consonants (Wood, 1974, 1975), leading to the idea that the phonetic processing of consonants occurred at an higher and, henceforth, later level than the acoustical processing of pitch. Furthermore, studies on the interactions between vowels and pitch in speech described reciprocal symmetric interference of these dimensions (Miller, 1978). How can we account for these puzzling results?

It may be that melodic intervals are much more complex musical dimensions than pitch height and that they, hence, call computations specific to music processing (Peretz & Coltheart, 2003). For example, it has been recently reported (Lidji, Kolinsky, Lochy, & Morais, 2007) that, whereas musical tones in isolation activate spatial mental representations similar to the ones associated to many ordered sequences, two-tones intervals do not evoke these general spatial representations. Similarly, patients suffering from acquired amusia are able to process pitch in isolated tones but exhibit deficits for tones embedded in a melody (Peretz & Kolinsky, 1993). Under that view, it is not surprising that the processing of melodic contour rests on more sophisticated musical processes, which could thus take place later than the phonetic processing of French vowel contrasts, which is natural to any native speaker of this language. Saying it briefly, it seems that vowel processing precedes interval processing in non-musicians. This asymmetry seems to be specific to a "musical mode of processing", given that vowels and tonal contours induce symmetric interference in tonal languages (Lee & Nusbaum, 1993; Repp & Lin, 1990).

In sharp contrast with the results on vocalic materials, for songs made out of nonwords that differed by the consonants, the pattern never corresponded to typical integrality. For stop consonants (Experiment 2), there was no interference in the orthogonal condition but a significant redundancy gain restricted to the task on interval. This facilitation could easily be explained by the differences in discriminability between the nonwords and the intervals leading the participants to, strategically, attend to the more salient dimension to perform the redundancy test. This phenomenon, labeled serial processing strategy (Garner, 1974), has been statistically confirmed in the present data. So, stop voiceless consonants and interval seem to be separable dimensions.

In sum, vowels seem to merge, or at least to interact with intervals during song processing, but consonants do not. This main finding from the present work is coherent with the research involving other types of tasks. As a matter of fact, studies of song production demonstrated that vowels act as a melodic skeleton whereas consonants interrupt the melodic line (McCrean & Morris, 2005; Scotto di Carlo, 1993; Sundberg, 1982). Our results show that the close relation between vocalic and melodic variations has processing consequences even when all acoustic correlations between vowel quality and pitch height are eliminated, as in the synthesized material used in Experiment 5. Thus, the pattern of results observed here must be interpreted at a higher processing level than the one of acoustic representation.

A likely reason of the tight link of vocalic and melodic variations in song processing is that, because of their temporal overlap and/or frequent acoustic correlation, human beings got used to process jointly the two types of variations.

With frequent exposure to sung materials, this joint processing would then apply to whatever situation, independently of the occurrence of acoustic correlations between the two dimensions. Under this view, what may be specific to the convergence of melody and phonology in songs is the rather unique overlap of spectral and temporal information supported by the vowels. This hypothesis should be studied through both developmental studies and experimental situations in which temporal overlap would be manipulated, e.g. by presenting separate verbal and melodic channels to the listeners as done by Crowder, Serafine and Repp (1990).

Another (compatible) interpretation is that intervals are linked to vowels because they have evolved together at the phylogenetic level. This idea is reinforced by the recent finding that the frequency ratio of the intervals composing the music of western and eastern cultures correspond to the ratio separating the formants of the languages of these same cultures (Ross, Choi, & Purves, 2007). Vowels and intervals may thus share the same auditory processing because they arise from the same root or, alternatively, because the development of one is based on the other. For example, the intervals of the musical scale may have been selected on the basis of the vowel spectrum because human beings were already used to process, and thus to privilege, these specific frequency ratios.

An additional explanation for the fact that processing of vowels and of musical intervals is integrated relies on the idea that vocalic distinctions depend on continuous quantity parameters (duration, pitch, intensity) that are much more similar to those used in tunes than are the consonantal distinctions, which mainly concern discrete quality categories, such as those of manner and place of articulation.

Under this view, the separability of stop consonants and melody, and, more generally, the detrimental effect of consonants on musicality, can also be due to obvious acoustical reasons. As can be seen in Table 1b, voiceless stop consonants consist in a silence followed by a noise-burst, so that such consonants are incompatible with melody continuity. In order to assess whether this separability was solely due to these acoustical constraints or, instead, to a more abstract dissociation of the processing of vowels vs. consonants as phonetic categories (Caramazza et al., 2000), we used nasal consonants in Experiment 3.

Indeed, patterns of interactivity may reflect a continuum of separable-to-integral processing rather than two categorically distinct processing modes (Garner, 1974; Grau & Kemler Nelson, 1988). Conversely, phonemes themselves vary continuously in sonority and the more sonorous consonants may be more apt to support pitch variations than the less sonorous ones, and hence may be more integrated with melody, either for acoustic reasons, or, as for vowels, because humans got used to process jointly these variations with melodic variations in songs. In Experiment 3, we examined through speeded classification tasks whether consonant sonority does modulate the interactivity pattern between intervals and nasal consonants. With this nasal material, there was no redundancy gain, but only interference in the orthogonal condition on interval. This interference could either reflect the occurrence of en emerging dimension (Pomerantz & Garner, 1973) or a failure of selective attention (Thibaut & Gelaes, 2002).

The notion that the interference effect observed on the nasal-C-material reflects limited capacity to selectively attend to the target dimension rather than the processing of a new, configural, property that would have emerged from interaction between the underlying dimensions (e.g., Garner, 1974) was confirmed in Experiment 4, which used a condensation task on both the nasal-C- and V2-materials. As expected, condensation was a far more difficult task than filtering for both materials. With respect to the nasal-C-material, in the absence of an emerging dimension, the only way for participants to perform correctly this task is to pay attention to both dimensions at the same time, while filtering requires only one single dimension to be processed at a time.

In addition, Experiment 4 showed that condensation was easier for the V2-material than for the nasal-C-material, which supports the notion that the nonwords varying on nasal, sonorous, consonants were more separable from the melody than the nowords varying on vowels. Indeed, while with separable dimensions, such as consonants and pitch, the participants must attend to two discrete dimensions in order to perform the condensation task, leading to poor performance, with integral dimensions the two classes of responses are more easily differentiated on the basis of the greater (Euclidean) perceptual distance between the whole stimuli (Garner, 1974).

This evidence of separability between nasal consonants and intervals, added to the fact that the pattern of interference did not differ significantly between the stop-C- and the nasal-C- materials, suggest that the sonority of the segments is not the key to their interactions with melody. Two alternative explanations, thought not mutually exclusive, are possible.

First, a glance at the pitch contour of the stimuli in Table 1 indicates that, whereas vowels are sung on relatively stable pitch levels (if one excludes vibrato), the nasal consonants are located at the transition between two tones. Consequently, pitch changes along the production of consonants. The kind of interval information provided by vowels and nasal consonants in the present materials is thus different (as is the case in natural singing), and this may at least partly explain the discrepancy in processing interactions with melody. Moreover, the consonants were far shorter in duration than the vowels, a phenomenon that is also typical of spontaneous singing (McCrean & Morris, 2005; Scotto di Carlo, 1993). A control for these two factors would be to present participants with filtering and redundancy tests on contrasts of nasal consonants sung on a stable pitch, and to the same contrasts sung on melodic intervals, a kind of singing called humming. The duration of these nasals could also be manipulated. If the nasals hummed on stable pitch, but not those hummed on intervals, lead to integrality, this would support the idea that the kind of musical information usually coupled to either vowels or consonants is crucial for these processing interactions.

Second, it might be that the linguistic function of vowels has anything more to do with music than the linguistic function of consonants, at large. According to recent studies (e.g., Bonatti et al., 2005; Mehler et al., 2006), this seems actually a plausible view. In addition to be responsible for prosody, the quantitative information carried by vowels gives important information about syntactic structure, as well as about some aspects of the morphological system. In contrast, while the role of consonants in signaling syntax is minor and limited to juncture phenomena that signal constituency (Nespor & Vogel, 1986; Selkirk, 1984), consonantal segments play a major role in making lexical distinctions. Semitic languages provide an extreme illustrative case of this role, since in these languages lexical roots are formed exclusively by consonants, whereas vowels are inserted to indicate morphological patterns (McCarthy, 1985). In line with our results, Nazzi (2005) and Nazzi and New (2007) even reported evidence that this dissociation of function is not related to the sonority of the phonemes. They indeed observed that infants exhibit difficulties in learning new words differing on one vowel, whereas they were able to perform this task when the distinctive phoneme was a consonant. In addition, their performance was identical when the consonant was a stop and when it was continuous, nasal, liquid or fricative consonant. Saying it briefly, vowels and consonants in words carry different kinds of information, the consonants being more tied to word identification and the vowels to grammar (Nespor et al., 2003).

More generally, some aspects of domain-specificity (or non-specificity), be it in the language or music domain, may be related to abstract properties, not to certain acoustic features. For example, in speech, it has been shown that left-hemisphere structures mediate abstract properties of language rather than acoustic features. Neuroimaging studies of linguistic pitch processing in tonal languages, such as Thai or Chinese, have indeed demonstrated that left-hemisphere structures are recruited for processing pitch contours specifically in speakers of such languages, whereas non-speakers process identical stimuli via right-hemisphere mechanisms (Gandour et al., 2000; Gandour, Wong, & Hutchins, 1998; Klein, Zatorre, Milner, & Zhao, 2001).

Although the notion that music has its own semantics is still under debate (e.g., Bharucha, Curtis, & Paroo, 2006; Koelsch et al., 2004), it is clear that both music and language display a generative structure that yields an infinite number of possible experiences (Bharucha et al., 2006). Thus, both have a grammar and a syntax. What vowels and pitches may share, instead of or in addition to temporal overlap and processing of similar acoustic parameters, is their important syntactic and grammatical role within the speech and the musical systems, respectively.

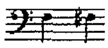
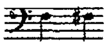
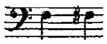
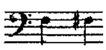
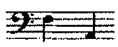
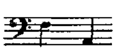
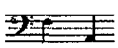
Whatever the correct explanation, temporal overlap, processing of similar acoustic parameters, or shared systemic functions, the present results imply that neither the speech nor the musical system are homogeneously modular. Indeed, while many data point to domain-specificity in both speech and music processing (for overviews, see Patel, 2008; Peretz, 2006), the observed interaction between vowel and pitch processing entails that speech and music processing are not totally independent. The difference of interactivity between pitch variations and vocalic vs. consonantal variations also clearly shows that whatever the functional domain, modularity and interactivity may best be evaluated, and are then likely to be both rejected, when one proceeds to a detailed level of empirical test (Peretz & Coltheart, 2003).

Future research is however needed to assess whether the processing interactions between the two systems depend on the actual function the speech and music parameters play in a specific experimental setting. For example, it has recently been suggested that learning a new language, especially in the first learning phase wherein one needs to segment new words, may largely benefit of the structuring properties of music in song: compared to speech sequences, a consistent mapping in sung sequences of linguistic and musical statistical dependencies (the location of dips in transitional probabilities between adjacent syllables and/or tones, cf. Saffran, Aslin, & Newport, 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, & Aslin, 1996) did enhance learning of an artificial language (Schön et al., 2008). Independent work has shown that, in speech, consonants are much more suitable than vowels to parse streams into words, using statistical dependencies, with "consonant words" significantly preferred over "vowel words" (Bonatti et al., 2005; Mehler et al., 2006). While the present results suggest stronger on-line processing interactions between vocalic and melodic variations than between consonantal and melodic variations, perhaps related to the similar syntactic functions of the former two sets of variations, it may be the case that, on the contrary, lexically parsing sung streams depends more on the statistical dependencies between consonants than on the statistical dependencies between vowels. In other words, processing sung sequences through a *lexically-oriented* (here, learning) mechanism might induce other interactions between language and music in sung material than the one observed here. This would also support the idea that studying the interactions between language and music at various and detailed levels of processing is the key for a better understanding of the similarities and specificities of each of these domains. In this research area, the fat lady has clearly not sung yet.

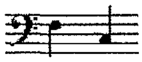
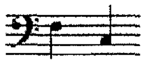
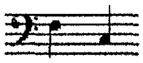
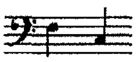
2. 1. 8. Appendix

F_0 and duration measures of the stimuli used in Experiments 1, 2, 3 and 4

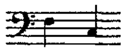
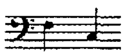
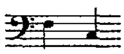
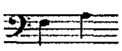
Experiment 1

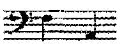
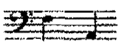
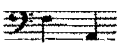
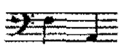
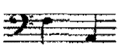
V1-material / Stimulus	Duration of first /da/ syllable (ms)	Averaged F0 of first /da/ syllable (Hz)	Duration of second syllable (ms)	Averaged F0 of second syllable (Hz)	Averaged F1 of the second vowel (Hz)	Averaged F2 of the second vowel (Hz)
1. /dalø/ 	421	167	1311	107	450	1345
2. /dalø/ 	509	172	1567	108	438	1302
3. /dalø/ 	447	170	1704	106	321	1250
1. /dalɔ/ 	617	167	1187	181	469	2148
2. /dalɔ/ 	443	166	1246	180	686	2137
3. /dalɔ/ 	557	170	1351	185	594	2117
1. /dalø/ 	598	169	1485	180	380	1109
2. /dalø/ 	614	166	1307	179	443	1289
3. /dalø/ 	448	166	1241	183	353	1340
1. /dalɔ/ 	457	169	1516	108	661	2173
2. /dalɔ/ 	701	170	1546	108	553	2097
3. /dalɔ/ 	408	168	1612	108	522	2148

Experiment 2

stop-C-material / Stimulus	Duration of first /da/ syllable (ms)	Averaged F0 of first /da/ syllable (Hz)	Duration of second syllable (ms)	Averaged F0 of second syllable (Hz)
1. /daky/ 	504	169	1621	125
2. /daky/ 	615	168	1603	126
3. /dakv/ 	451	171	1498	126
1. /daty/ 	618	169	1442	126
2. /daty/ 	611	171	1614	126
3. /daty/ 	538	170	1533	127
1. /daky/ 	605	168	1564	218
2. /dakv/ 	516	166	1148	217
3. /daky/ 	624	168	1487	217
1. /daty/ 	652	169	1249	218
2. /daty/ 	702	169	1524	217
3. /daty/ 	543	168	1143	220

Experiments 3 and 4

V2-material / Stimulus	Duration of first /da/ syllable (ms)	Averaged F0 of first /da/ syllable (Hz)	Duration of second syllable (ms)	Averaged F0 of second syllable (Hz)	Averaged F1 of the second vowel (Hz)	Averaged F2 of the second vowel (Hz)
1. /dalɛ/ 	479	170	1372	129	500	1725
2. /dalɛ/ 	584	170	1205	130	480	1702
3. /dalɛ/ 	539	171	1187	132	512	1691
1. /dalɔ/ 	496	170	1325	129	544	835
2. /dalɔ/ 	545	171	1210	131	540	827
3. /dalɔ/ 	513	172	1140	130	552	851
1. /dalɛ/ 	521	172	1332	219	522	1622
2. /dalɛ/ 	173	1312	219	599	499	1620
3. /dalɛ/ 	504	171	1298	222	511	1667
1. /dalɔ/ 	483	171	1375	220	550	878
2. /dalɔ/ 	437	171	1339	221	489	881
3. /dalɔ/ 	504	173	1259	224	520	920

nasal-C-material Stimulus	Duration of first /da/ syllable (ms)	Averaged F0 of first /da/ syllable (Hz)	Duration of second syllable (ms)	Averaged F0 of second syllable (Hz)	Duration of the middle consonant (ms)	Average F0 of the middle consonant (Hz)
1. /dany/ 	473	173	1314	182	94	173
2. /dany/ 	498	172	1286	184	111	175
3. /dany/ 	564	172	1089	183	79	174
1. /damy/ 	488	169	1378	183	128	174
2. /damy/ 	585	171	1173	184	102	173
3. /damy/ 	511	170	1139	185	114	172
1. /dany/ 	546	171	1267	110	108	129
2. /danv/ 	464	170	1263	108	160	125
3. /dany/ 	545	171	1205	112	98	141
1. /damy/ 	505	170	1390	110	125	116
2. /damy/ 	435	71	1324	111	119	111
3. /damy/ 	540	169	1170	112	121	117

2. 2. Article 3

Integration effect and illusory conjunctions in memory for lyrics and tunes: Vowels sing but consonants swing [iii]

[iii] Lidji, P., Kolinsky, R., Peretz, I., Lafontaine, H., & Morais, J. (en préparation). Integration effect and illusory conjunctions in memory for lyrics and tunes: Vowels sing but consonants swing.

Abstract

This study was aimed at determining how two components of songs, namely the phonological part of lyrics and the melodic part of tune, interact in memory. After an incidental learning of bisyllabic nonwords sung on two-tones intervals, the participants had to recognize the nonwords (Experiment 1), the intervals (Experiments 1 and 2) or the whole song (Experiments 1 and 2), that is, the exact association of nonword and interval. The old stimuli were mixed with four kinds of foils: completely new stimuli (new nonword and new interval), *partially new stimuli* made out of a new and an old dimension (new nonword and old interval or new interval and old nonwords) and *mismatch* stimuli in which both components were old but the original association had been disrupted by combining the nonword and the interval of separately studied songs. We observed an *integration effect*: the intervals of studied songs were better recognized when, in the test phase, they were presented with the same words as originally heard than with different words. This integration effect was coupled with *illusory conjunctions* of tunes and lyrics, which are revealed by the fact that participants give more erroneous "old song" responses to mismatch songs than to any other type of new song. These illusory conjunctions are generally interpreted as reflecting the separate encoding and storage of the dimensions. Hence, the combination of the integration effect with illusory conjunctions in Experiment 1 suggests that tunes and lyrics of songs are represented separately in memory, which allows illusory conjunctions to occur, but in an associative mnemonic network that leads to the integration effect. In Experiment 2, the role of vowels and consonants on the effects was examined. The results suggest that the separability of tunes and lyrics traces varies as a function of the involved phonemes. Consonants seem to be more separable from tunes than vowels, since they yield more illusory conjunctions. The results are discussed in the light of the function of vowels and consonants in song and speech.

2. 2. 1. Introduction

According to a popular belief, singing can help remembering verbal information (Dickson & Grant, 2003) because lyrics and tune are expected to leave some associated memory traces. However, the scientific evidence supporting this claim is surprisingly scarce. If some authors (Rainey & Larsen, 2002; Wallace, 1994) actually found a superiority of sung over spoken presentation for learning lyrics, the effect could be related to the slower presentation rate and the repetitive melodic pattern of songs (Wallace, 1994). Moreover, when the lyrics do not include any semantic content, the facilitative effect of song over speech is quite limited (Rainey & Larsen, 2002). There is also growing evidence that song lyrics are not better learned when encoded and produced as sung than as spoken (Calvert & Billingsley, 1998; Racette & Peretz, 2007). This is in agreement with the anecdote that occasional singers often fail to remember song lyrics after the first or second line and then try compulsively to retrieve the words while continuing to sing on "la, la, la" (Peretz, Radeau, & Arguin, 2004). This informal but common observation points to the fact that melody and song lyrics enjoy a high degree of autonomy in memory.

This claim is further supported by the finding that the melody and the text of familiar songs can be selectively impaired by brain damage. Aphasic patients are able to correctly sing the melody of well-known and newly learned songs, but not to pronounce their lyrics (Hébert, Racette, Gagnon, & Peretz, 2003; Racette, Bard, & Peretz, 2006), while patients suffering from amusia (e.g., Hébert & Peretz, 2001; Peretz, Belleville, & Fontaine, 1997; Peretz et al., 1994) can recognize and produce the lyrics, but not the melody, of familiar songs. Samson and Zatorre (1991) even demonstrated, in patients having undergone a surgical excision of left or right temporal brain regions, that retention of the lyrics of a song is related to left temporal lobe structures whereas memory of the melody depends on both temporal lobes. These results strongly suggest that the retention of the lyrics and of the melody of songs rests on different memory codes.

There is, still, experimental evidence of associated storing of lyrics and tune in healthy subjects (Serafine, Crowder, & Repp, 1984; Serafine, Davidson, Crowder, & Repp, 1986). In Serafine et al.'s (1984) Experiment 1, having studied 24 folk song excerpts, the participants were required to recognize them; if the test excerpt had not been presented in the learning phase, they had to state whether one of its components (the lyrics or the tune) was familiar. The "old songs" were mixed with four types of foils: (1) completely new songs (new tune and new lyrics), (2) a new tune combined with previously studied lyrics, (3) new lyrics combined with a previously studied tune, or (4) a new combination of familiar components, namely the lyrics and the tune of separately studied songs. These latter new songs, referred to as *mismatch* songs, elicited less correct "old tune" responses than the truly old songs, in which the original association between the tune and the lyrics was spared. The same effect was observed for the "old words" responses, but less reliably, probably because of a ceiling effect (Serafine et al., 1984). The authors labeled *integration effect* the fact that melodies of studied songs were better recognized when, in the test phase, they were presented with the same words as originally learned than with different words, even when the different words were equally familiar to the listeners (mismatch songs).

Whereas Serafine et al. (1984, 1986) initially interpreted this result pattern as reflecting the integrated nature of the lyrics and tunes representation in memory for songs, they eventually acknowledged that the integration effect merely comes from an association between separate components (Crowder, Serafine, & Repp, 1990). They actually observed that, with spoken texts accompanied by hummed melodies, participants again recognized the tune better when it was paired with the studied text (old song) than when it was paired with another old text (mismatch song). In this situation, the superior recognition for the old pairing over the mismatched one cannot be attributed to the retention of some physically integrated representation of melody and text, because they were encoded as physically separate. Thus, for songs, the integration effect of lyrics and tunes would result from the memorization of contingent links between separate memory traces. According to Crowder et al. (1990), this association was mainly caused by the temporal contiguity of the two features of songs. Peretz et al. (2004) have questioned this conclusion by showing evidence of backward priming for the lyrics and the tunes of familiar songs: the beginning of song lyrics like "happy birthday" was primed by the melody of later sections sung on "la, la, la" and vice-versa. These links thus seem to be reciprocal connections that cannot be reduced to simple time contingencies; they could be best understood in terms of more subtle perceptual rhythmic congruencies.

In short, though the idea that the components of songs leave separate but connected traces in memory seems quite well established, the nature of these links still needs to be clarified. The aim of the present study was to answer this question by uncovering factors that modulate the strength of these tune-lyrics connections. To this end, we combined two indexes of the mnemonic interactions between object dimensions: the now classical *integration effect* and the occurrence of *illusory conjunctions* (Kolinsky, 1989; Treisman & Schmidt, 1982). Illusory conjunctions are revealed by the erroneous recognition of mismatch stimuli that combine features from other, presented stimuli. These false recognitions of features combinations suggest that the component features of the stimuli have been initially extracted separately (Stefurak & Boynton, 1986; Treisman & Schmidt, 1982; Treisman, Sykes, & Gelade, 1977). While, in their original effort, Serafine et al. (1984; 1986) focused their attention on the integration effect, their initial data seem to include evidence for the independence of lyrics and tunes through the occurrence of these illusory conjunctions. Although no statistical test for illusory conjunctions was provided by the authors, a fine-grained analysis of the false recognition rates, or false alarms (henceforth, *FAs*) reported in the "old song" responses suggest that their participants have given more erroneous "old song" responses to mismatch songs than to any other type of new song, either completely new stimuli (new tune and new lyrics), or partially new stimuli made out of a new and an old feature, that is, of an old tune paired with new lyrics or of old lyrics paired with a new tune. This pattern has been replicated in both adults and children by Morrongiello and Roes (1990). In other words, this new look at Serafine et al.'s data supports the idea that lyrics and tunes of songs are extracted and stored separately in memory.

In the music domain, Thompson, Hall and Pressing (2001) already used illusory conjunctions to demonstrate the processing independence of melodic and temporal variations, a notion that emerged from the observation of impairments that selectively affect pitch (but not rhythm) processing (Hyde & Peretz, 2004; Peretz & Kolinsky, 1993; Peretz et al., 1994). Since the listeners of Thompson et al. (2001) failed to remember how the pitch and the duration of notes were combined in originally heard melodic sequences, they erroneously recognized mismatch stimuli combining the pitch of one note with the duration of another note from the melody. Thompson et al.'s conclusion that pitch and duration are processed independently has, however, been revisited later on in terms of familiarity of the mismatch stimuli (Jamieson, Thompson, Cuddy, & Mewhort, 2003). Using computer simulations based on the *REM* model (Retrieving Effectively from Memory, Shiffrin & Steyvers, 1997), Jamieson et al. (2003) demonstrated that the higher FA rates to mismatch foils, in which both components had been heard during the learning phase, can be explained by their superior familiarity compared to the other foils, in which none or only one component was familiar. This interesting interpretation is still not incompatible with the idea of dimensional separability since, in the model, the computed probability that an item is "old" is directly related to the familiarity of each of its component features.

In the present work, we investigated the memory interactions between lyrics and tune in the light of all these studies and examined new parameters that could modulate the separability or the association between these song features. These parameters are the phonetic properties of the lyrics. This idea is related to a hypothesis raised by Crowder et al. (1990): the "physical interaction hypothesis". It states that the phonetic properties of lyrics impose subtle physical changes on the melody, for example by transforming a staccato articulation for lyrics rich in plosive consonants, into a legato phrasing for lyrics rich in vowels and liquids. These *submelodic* changes, which keep the pitch, contour and rhythm of the melody unaffected, might result into the integration effect by slightly modifying the melody due to the lyrics changes. According to Crowder et al. (1990), this hypothesis was supported by the fact that the integration effect was maintained when the text was systematically replaced in the recognition test by nonwords phonetically similar to the learned lyrics. Although this shows that phonetically similar lyrics spare the recognition of the melody, a more direct way to test the physical integration hypothesis would be to examine the effect of phonetically dissimilar lyrics. If the integration effect relies on submelodic modulations due to the lyrics change in mismatch songs, we could predict a stronger integration effect for mismatch songs in which the lyrics differ from the original lyrics in more numerous phonemic features, hence favoring submelodic changes, than when the mixed lyrics are phonetically more similar.

In addition to this "physical interaction hypothesis", recent experimental evidence suggests that vowels and consonants might contrast as regards their interactions with music. Research on singing suggests that vowels are an exquisite melodic support (Sundberg, 1982), while consonants tend to break the melodic line (Scotto di Carlo, 1993). An intriguing finding by Ross and colleagues (Ross, Choi, & Purves, 2007) reinforces the idea that the processing of vowels and of musical intervals might be closely bound. These authors have observed that the frequency ratio of intervals from the chromatic scale correspond to the frequency ratio of the two first formants of English and Mandarin vowels. They conclude that the human preference for the intervals of the chromatic scale emerges from experience with the harmonics of vowels. Due to this link, vowels might be more integrated to the melodic line that consonants are. We have recently obtained results supporting that view (Kolinsky, Lidji, Peretz, Besson, & Morais, under review). In a speeded classification task (Garner, 1974), vowels and melodies have been shown to be perceptually integral dimensions while consonants and melodies are more separable, a result that has been labeled with humor: "vowels sing but consonants speak". If these on-line interactions also played a role in tune-lyrics memory associations, this asymmetry between vowels and consonants would yield less strong connections (as indexed by the occurrence of more illusory conjunctions) with the tune for lyrics varying on a consonantal basis than for lyrics varying on a vocalic basis.

To investigate the influence of these variables on lyrics-tune mnemonic interactions, we analyzed, in two experiments, both the integration effect and the occurrence of illusory conjunctions of lyrics and tunes in sung nonwords.

The goal of Experiment 1 was to test whether an integration effect could cohabit with illusory conjunctions in memory for sung stimuli and to assess the familiarity hypothesis brought up by Jamieson et al. (2003). An incidental encoding phase preceded the forced-choice recognition test, which was similar to the one used by Serafine and colleagues (Crowder et al., 1990; Serafine et al., 1984) with consonant-vowel bisyllabic (CVCV) nonwords sung on two-tones intervals. Three groups of participants[1], who had not been warned that their memory would be assessed, had to recognize the intervals (*I*), or the nonwords (*NW*), or the whole stimuli that had been presented during the incidental encoding phase. In addition to

[1] Contrary to Serafine et al. (1984), we used the recognition test as a between-subjects variable. In the first experiment of Serafine et al.'s original study, the participants had to perform the task sequentially: if they did not recognize the whole song, they had to assess whether one of its components was old. Hence, the "old tune", "old lyrics" and "old song" responses were inter-dependent: an error in the first, "old song recognition" step would yield to an error in the following "old tune" or "old lyrics" recognition steps. In our study, confronting different participants to the different recognition tests allowed them to really focus on the target dimension and avoided this sequential processing and these potential sequential errors. We are, however, aware that between-subjects designs have other limitations, such as a reduced statistical power.

truly old stimuli, there were four types of foils. The foils were completely new stimuli (henceforth, *New I/new NW*), or *partially old stimuli* made out of a new and an old feature. They could be composed by an interval that had been presented in the encoding phase paired with a new nonword (*old I/new NW*) or by a nonword presented in the speeded classification phase paired with a new interval (*old NW/new I*). The fourth type of foils was *mismatch* stimuli, created by pairing an interval from one of the materials heard in the learning phase with a nonword from the other material. Note that one material was made out of nonwords varying on their middle consonant, while the other material included nonwords varying on their final vowel. Creating the mismatch songs thus implied to switch nonwords differing on both vocalic and consonantal features.

The use of incidental encoding adds an ecological dimension to the experiment since, in everyday life, we usually learn songs through repeated passive exposure. It also allowed us to directly test the effect of feature familiarity on the conjunction errors (Jamieson et al., 2003). This learning phase consisted in a speeded classification task (Garner, 1974) of the sung CVCV nonwords, which results are reported in Kolinsky et al. (under review). In these tasks, the participants were required to classify items according to the identity of either their "lyrics" (i.e., the identity of the bisyllabic pronounceable nonword, that could take either one of two values) or their "tune" (i.e., the identity of the two-tones interval, that could also take either one of two values). Crucially for the familiarity hypothesis, in this encoding phase, the number of opportunities to encode the features was manipulated, so that some nonwords and some intervals were heard more often than others. If the expected higher FA rate to mismatch stimuli than to other foils were really due to the higher familiarity of their components, the participants should also give more erroneous "old" responses to the mismatch stimuli made out of the most frequently heard nonwords and intervals than to the other mismatch songs.

The predictions for Experiment 1 were thus the following: in both the interval and nonword recognition tests, we expected to replicate former results (Crowder et al., 1990; Serafine et al., 1984, 1986) as regards the occurrence of an integration effect. Thus, we predicted better recognition of completely old stimuli than of mismatch stimuli. However, in the whole stimuli recognition test, we predicted the occurrence of illusory conjunctions, indicated by a higher FA rate to mismatch stimuli than to other foils, namely, completely new stimuli and partially old stimuli, both old I/new NW and old NW/new I. If reflecting genuine illusory conjunctions, the number of FAs would not be modulated by the familiarity of the features, in other words, by the number of presentation of the nonword and the interval of the mismatch foils during the learning phase. The whole performance pattern would show that, while strongly connected to each other in memory, lyrics and tunes are nevertheless represented separately.

The frequency of these illusory conjunctions thus provides a new index of the strength of the connections between song features. In Experiment 2, we used this tool to investigate the nature of the connections between the separate traces of lyrics and tunes in memory for songs. Our hypothesis, based on our former data (Kolinsky et al., under review) showing that vowels and melodic intervals are more integrated in perception than are consonants and intervals, is that the phonological features of the text would influence its mnemonic interactions with the tune in the same way. We thus compared the integration effect and the illusory conjunction rates in two kinds of experimental materials made out of CVCV nonwords sung on the same two-tones intervals as in Experiment 1. By contrast with Experiment 1, the mismatch stimuli only combined lyrics varying by one consonant in one material, and by one vowel in the other material. If the phonetic properties of the lyrics play a role in the strength of association between lyrics and tune in memory, we expect to observe more numerous illusory conjunctions for lyrics varying on a consonantal basis than for lyrics varying on a vocalic basis.

Finally, the difference between the materials of Experiment 1 and of Experiment 2 provided an opportunity to evaluate Crowder et al.'s physical interaction hypothesis (1990). Indeed, the mismatch songs of Experiment 1 mixed stimuli differing by two phonemes, one consonant and one vowel, while in Experiment 2, the switched nonwords only differed by one phoneme. If submelodic changes were involved in the associations between lyrics and tune, the mismatch songs of Experiment 1 should thus elicit a stronger integration effect than the ones of Experiment 2, given that their melody articulation should be more altered by the lyrics change.

2. 2. 2. Experiment 1: Forced-choice recognition memory tests on a vocalic and a consonantal material

After performing a speeded classification task on both materials presented in Table 1, the participants of Experiment 1 were exposed to one of three recognition tests. They had to recognize the intervals, or the nonwords, or the whole stimuli that had been presented during the classification phase. We will present separately the methods of the speeded classification task, which results will not be reported here (see Kolinsky et al., under review), and the methods and results of the recognition tests. The description of the speeded classification task deserve some details because the number of presentations of each stimulus was modulated by the experimental condition, a specificity that was used to assess the familiarity effect in the recognition test (Jamieson et al., 2003). The reader looking for further information on the theoretical interpretation of this speeded classification task is referred to Garner (1974 ; see also Lidji, 2007).

2. 2. 2. 1. Method

2. 2. 2. 1. 1. General procedure

The participants first took part in an incidental encoding phase, which consisted in a speeded classification task on the lyrics and intervals of sung nonsense words. They were not warned that their memory of the sung stimuli would be assessed later on. This encoding phase lasted approximately 80 minutes and was followed by a 10-minute break during which participants filled in a questionnaire about their musical experience. They then performed the recognition test, which lasted about five minutes

2. 2. 2. 1. 2. Participants

The 48 participants were 22 men and 26 women aged from 18 to 54 years (average: 22.6 years), among whom 36 were right-handed. Twenty-one had never learned music, and 27 had some musical practice that did not exceed four years, stopped for at least five years. Twenty participants were unpaid volunteers, 13 were undergraduate students in psychology who earned course credits and 15 were university students paid for their participation. After the incidental encoding phase (speeded classification task), the participants were divided into three groups of 16 subjects each. One group was presented with the nonword recognition test, another with the interval recognition test and the last one with the whole stimulus recognition test.

2. 2. 2. 1. 3. Incidental encoding – speeded classification task

2. 2. 2. 1. 3. 1. Material.

The CVCV bisyllabic pronounceable nonwords sung on two-notes intervals were first encoded in a speeded classification task with two materials: one in which the sung nonwords varied on the middle consonant (C-material) and one in which the sung nonwords varied on the last vowel (V-material), see Table 1.

In the C-material (Table 1a), the slightly ascending F3-#F3 and the descending F3-A2 minor intervals were combined with the nonwords /dany/ and /damy/. In the V-material (Table 1b), the ascending F3-A3 and descending F3-C3 major intervals were combined with the nonwords /dalɛ/ and /dalɔ/. The stimuli, sung by a professional baritone, were recorded on mini-disc in an anaechoïc room and normalized for intensity with *Digidesign Protools*. The selected exemplars, in which the first syllable lasted around 500 ms and the second syllable around 1000 ms, were included in 2500 ms files. Within all files, silent intervals of different durations were inserted before and after the stimulus so that the transition between the two notes and syllables was centered at 1250 ms for all stimuli. To avoid response strategies based on obvious physical correlations between, for example, the pitch and the quality, or the length and intensity of the vowels, three physically different instances of each stimulus were used for each material (i.e., three different recordings of I1 NW1, I2 NW1, I1 NW2 and I2 NW2, see Table 1).

Table 1. The different combinations of nonwords and intervals used in the speeded classification phase of Experiment 1, separately for the C-material (part a) and the V-material (part b).

a

C-MATERIAL		Interval 1 (I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/dany/	I1 NW1	I2 NW1
Nonword 2 (NW2)	/damy/	I1 NW2	I2 NW2

b

V-MATERIAL		Interval 1(I1)	Interval 2 (I2)
			
Nonword 1 (N1)	/dalɛ/	I1 NW1	I2 NW1
Nonword 2 (N2)	/dalɔ/	I1 NW2	I2 NW2

2. 2. 2. 1. 3. 2. Task.

For each material, the stimuli showed in Table 1 were presented in both a melodic task, which consisted in classifying the stimuli according to their intervals, and a phonological task, which consisted in classifying the stimuli according to the nonwords. Within each of these tasks, three classification conditions were presented: *baseline*, *redundant* and *orthogonal* (Garner, 1974). The task remained formally the same: to associate as fast as possible each presented nonword (in the phonological task) or each interval (in the melodic task) to one of two response keys.

The baseline condition included no more than two values of the stimuli, which varied only on the target, attended dimension (e.g. the nonword), the other dimension (e.g. the interval) being held constant at one of its two possible values. For example, in the phonological task on the C-material, the baseline stimuli could be I1 NW1 and I1 NW2 (see Table 1a). The redundant condition also included only two values of the stimuli, but these varied on both dimensions in a correlated way, with each value of the target dimension systematically associated with one of the two values of the other dimension. For example, a redundant block in the phonological task on the C-material could consist in I1 NW1 and I2 NW2 in Table 1a. The orthogonal condition included all four stimuli, so that the classification involved selective attention to the target dimension in order to ignore the irrelevant variations of the other, non-target dimension.

Thus, in the standard and redundant conditions, the participants were only exposed to two out of four possible stimuli. The choice of these two stimuli was counterbalanced between-subjects, leading to two different possibilities (Standard 1 and Standard 2; see Appendix 1). Consequently, the participants were presented with some nonwords and intervals more often than with others. For example, a participant in the Standard-Redundant 1 group with the C-material heard the nonword /damy/ 252 times whereas /dany/ was heard 180 times. Similarly, the interval F3-#F3 was presented 252 times, while F3-A2 was only heard 180 times (see Appendix 1). This pattern was reversed for participants in the Standard-Redundant 2 group. Across tasks, conditions and materials, participants were presented with 12 experimental blocks of 72 trials each, conducting to a total of 864 trials.

2. 2. 2. 1. 3. 3. Procedure.

Participants were tested individually in a quiet room, with stimuli presented through two loudspeakers placed symmetrically on both sides. Stimuli presentation and response times were controlled via Psyscope 1.2.5. PPC (Cohen, Macwhinney, Flatt, & Provost, 1993) on a Macintosh G3 computer. Errors and response times were recorded with a Psyscope button box.

Each participant performed the speeded classification on both materials (C- and V-material) and tasks (interval and nonword) in the three conditions (baseline, redundant and orthogonal). They first completed all the conditions on one material before being presented with the other material. For one material, they first performed the three experimental conditions for one task (interval or nonword) before being tested with the other task. To avoid sequence effects, the presentation order of the materials, tasks and conditions was counterbalanced between participants.

2. 2. 2. 1. 4. Recognition tests

2. 2. 2. 1. 4. 1. Material.

Stimuli recording and editing were exactly as for the speeded classification task.

Five types of stimuli were presented to the participants: eight truly old stimuli, i.e. the stimuli heard in the incidental encoding phase (Table 1), and four kinds of foils, which will be described below. These foils were built on the basis of four new nonwords (/dabi/, /dagi/, based on the C-material; /dare/ and /daro/, based on the V-material) and four new intervals (unison: F3-F3, major ascending: F3-C4, descending octave: F3-F2, and minor ascending: F3-D4#). The new nonwords were adapted from the old stimuli by changing both their middle consonant and their final vowel. The intervals were chosen in order to get two small and two large intervals; typical exemplars such as unison and octave were chosen to get the foils sufficiently discriminable from the old intervals.

Based on the possible combinations of these components, 16 completely new stimuli (*New I/new NW*), eight stimuli made out of an old interval (presented in the speeded classification phase) paired with one of the four new nonword *(old I/new NW*), eight made out of an old nonword paired with one of the four new intervals (*old NW/new I*), and eight mismatch stimuli were formed. The mismatch stimuli were created by pairing an interval from one of the materials (either the C- or the V-material) of the speeded classification phase with a nonword from the other material (i.e., the V- or the C-material, respectively). Examples of these foils are provided in Table 2, and the complete list is available in Appendix 2.

All stimuli (foils and old stimuli) were presented twice, using two physically different recordings by the same singer in order to avoid judgments based on the acoustical identity of the stimuli. This resulted in 96 recognition trials, but the stimuli yielding old or new responses depended on the recognition test. In the nonword recognition test, the truly old, the old NW/new I and the mismatch stimuli requested an "old nonword" response. Each of these categories included 8 stimuli presented twice, leading to a maximum of 48 "old" responses out of 96. For the interval recognition test, the participants should respond "old interval" to truly old, old I/new P and mismatch stimuli, leading again to 48 "old" responses. Finally, in the whole stimulus recognition test, the participants should only recognize as "old" the truly old stimuli, corresponding to 16 trials out of 96.

Table 2. Examples of stimuli used in the recognition tests of Experiment 1 (I = interval; NW= nonword).

Completely old	 /damy/
Completely new	 /dagi/
Old I / New NW	 /dagi/
Old NW / New I	 /damy/
Mismatch	 /dalɛ/

2. 2. 2. 1. 4. 2. Procedure.

The conditions of presentation were similar to the ones of the encoding phase. The participants were instructed that they would have to recognize the stimuli they have been presented with during the classification task. Participants assigned to the nonword recognition group had to indicate if they had already heard the nonwords in the former phase, ignoring the interval. Participants in the interval recognition group had to decide if they had heard the interval in the former phase, ignoring the nonwords. Finally, participants included in the whole stimulus recognition group had to judge if they had heard the very same combination of a specific interval with a specific nonword in the former phase. The stimuli were presented in a pseudorandom order identical for all participants, with no consecutive presentations of the same stimulus. Positive, "old" responses were assigned to the green key, and negative, "new" responses were assigned to the red key of the response box. Correct responses and response times were recorded, but instructions emphasized accuracy. Therefore, only accuracy data were analyzed.

2. 2. 2. 2. Results and discussion

2. 2. 2. 2. 1. Comparison between recognition tests

For the recognition tests, the average percentages of correct old and new responses were 71% for the interval, 98% for the nonword, and 80% for the whole stimulus. One should note that despite the high number of presentations of each stimulus dimension during the encoding phase (at least 180) and the explicit attention allocation to the lyrics and the tune required in the speeded classification task, the accuracy rate was not at ceiling for the interval and the whole stimulus recognition tests. This suggests that two-note intervals and nonsense songs including such intervals are more difficult to encode and to recognize than longer songs like the ones used by Serafine et al. (1984).

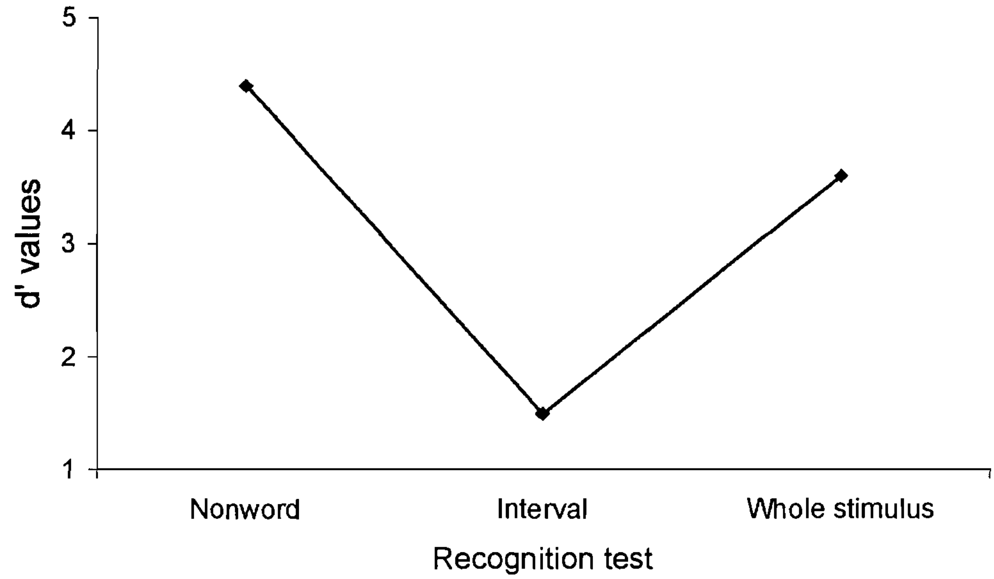
To compare the accuracy level and the error rate across these three recognition tests, two separate ANOVAs were performed on the arcsine transform of the proportions of hits and of false alarms with the recognition test (nonword, interval, whole stimulus) as a between-subjects variable. The correct "old" responses (hits) were computed on completely old stimuli and the incorrect "old" responses (FAs) on completely new stimuli. The remaining foils were not included in this analysis because the expected response to these depended on the task. For example, an Old I /New NW foil appeals a correct "old" answer in the interval recognition test but a correct "new" answer in both the whole stimulus and nonword recognition tests. Not surprisingly, this analysis revealed a significant effect of recognition test on both hits, F (2, 45) = 21.08, $p < .0001$, and false alarms, F (2, 45) = 123.68, $p < .0001$. The hit rate differed significantly between the three recognition tests as shown by Scheffe's tests, p at least $< .025$. By contrast, the FA rate did not differ between

[2] The use of the arcsine transform allows one to apply the analysis of variance on proportions, a variable that does not respect the rule of homogeneity of variance (Bishop, Fienberg, & Holland, 1975).

the nonword and the global recognition test, $p > .10$ but the participants made significantly more FAs in the interval recognition test, both *p*s < .0001.

Consistent with these results, a one-way ANOVA with task as between-subjects factor performed on the *Signal Detection Theory d'* scores (Green & Swets, 1966), that combine hits and FAs, also showed a highly significant effect of recognition test, $F(2,45) = 15.12$, $p < .0001$. According to Scheffe's tests, all between-tasks comparisons were significant, *p* at least < .005. As can be seen in Figure 1, interval recognition was thus the most difficult task, nonword recognition was the easiest, the whole stimulus recognition being intermediate in difficulty.

Figure 1. Average *d'* scores as a function of recognition test in Experiment 1.



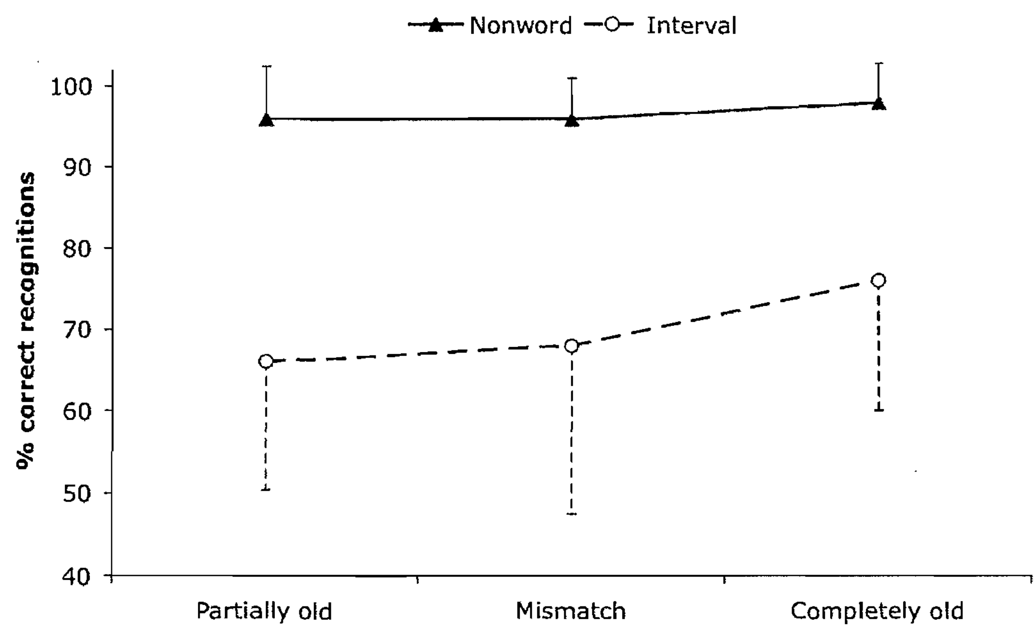
2. 2. 2. 2. 2. Integration effect

In order to check for the occurrence of an integration effect, only data from the nonword and interval recognition tests were taken into account. The hits, corresponding to correct detections of the relevant feature, either the nonword or the interval, were compared for completely old stimuli, partially old ones, and mismatch stimuli. The ANOVA, performed on the arcsine transform of the proportion of hits, included task (interval or nonword) as a between-subjects variable and stimulus type (completely old; mismatch; partially old, i.e. either old I/new NW or old NW/new I) as a within-subject variable. It revealed significant effects of task, $F(1, 30) = 66.86$, *p*

< .0001 and of stimulus type, $F(1, 30) = 3.99$, $p < .025$. The stimulus type by task interaction was not significant ($F < 1$).

Not surprisingly for musically untrained participants, the task effect reflects better recognition performance of nonwords than of intervals (on the average, 97 vs. 70%, respectively). The effect of stimulus type indicates that participants recognized better an old feature in a completely old stimulus (on the average, 87% correct) than in a partially old one (on the average, 81% correct, $F(1, 30) = 6.3$, $p < .025$), and, more crucially, than in a mismatch one (on the average, 82 % correct, $F(1, 30) = 7.71$, $p < .01$). This difference, illustrated in Figure 2, corresponds to Serafine et al.'s (1984) integration effect.

Figure 2. Average correct recognitions (hits) according to task and stimulus type observed in the interval and nonword recognition tests of Experiment 1. Error bars represent one standard-deviation from the average.

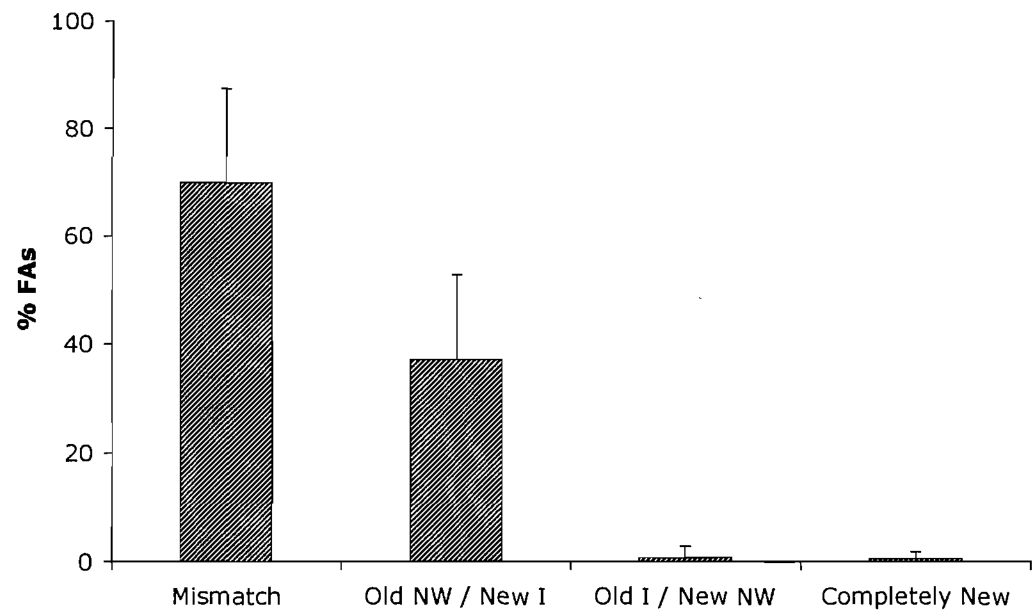


2. 2. 2. 2. 3. Illusory conjunctions

A separate ANOVA on FA rates in the whole stimuli recognition test aimed at examining the occurrence of illusory conjunctions of lyrics and tunes, which would be reflected by a higher FA rate for mismatch foils than for all other kinds of foils. The ANOVA run on the arcsine transform of FA rates took foil type as a within-subjects variable (completely new, old I/new NW, old NW/new I, mismatch).

It showed a highly significant effect of foil type, $F(3, 45) = 186.9$, $p < .0001$. As depicted in Figure 3, participants made much more FAs to mismatch stimuli than to all other kinds of foils, $F(1, 45) = 291.37$, $= 45.7$, and $= 222.66$, $p < .0001$ in all cases, for mismatch vs. old I/new NW, vs. old NW/new I, and vs. completely new foils, respectively. Reflecting the participants' tendency to rely on the identity of the nonwords, the old NW/new I foils also led to more FAs than both the old I/new NW and the completely new foils, $F(1, 45) = 292.31$ and $= 148.5$, $p < .0001$ in both cases. However, it should be noticed that the high FA rate for mismatch stimuli cannot be explained only by this tendency to rely on the presence of a very recognizable old nonword, since there were significantly more FAs to mismatch stimuli than to old NW/new I foils.

Figure 3. Average FA rates as a function of foil type observed in the whole stimulus recognition test of Experiment 1. Error bars represent one standard-deviation from the average.



This pattern of results may either reflect the occurrence of illusory conjunctions or be caused by the participants' higher familiarity with the two features (the nonword and the interval) presented in the mismatch stimuli (Jamieson et al., 2003). This latter possibility would be supported by an effect of the frequency of occurrence of one particular nonword or of one particular interval on the participants' FA rates.

As demonstrated in Appendix 1, participants who were presented with the Standard 1 materials were exposed to some nonwords and intervals 252 times but heard the remaining nonwords and intervals only 180 times, while the participants in the Standard 2 condition got the opposite exposition pattern. Thus, if frequency of occurrence were affecting the FA rate, we should observe an interaction between the type of Standard condition (Standard 1 vs. Standard 2) and the identity of the mismatch stimuli, with participants exposed to one of the standard conditions making more FAs to the mismatch stimuli combining the most frequently heard features in the classification task. Referring to Appendix 1, participants in the Standard 1 condition should thus have produced more FAs to /damy/ (C-material) sung on F3-A3 (V-material) and to /dal / (V-material) sung on F3-#F3 (C-material), because these features were most frequently heard in the speeded classification task, than to the mismatch stimuli made out of the least frequently heard features. Conversely, participants in the Standard 2 condition were expected to produce more FAs to the following mismatch stimuli: /dany/ sung on F3-A2 and /dalɛ/ sung on F3-C3. As illustrated in Table 3, which shows the average FA rates (in percentages) to these target mismatch stimuli, FAs did not seem to be modulated by familiarity. This was confirmed by the ANOVA performed on the arcsine transform of FA rates with the identity of the mismatch stimuli (/damy/ on F3-A3, /dal / on F3-F3#, /dany/ on F3-A2, and /dalɛ/ on F3-C3) as a within-subject variable and the type of Standard condition (1 vs. 2) as a between-subjects variable. It showed neither a significant effect of stimulus, $F(3, 42) = 1.82$, $p > .10$, nor a significant interaction between stimulus and material ($F < 1$). Although the sample was limited and we cannot be sure that 72 additional presentations of the features are enough to elicit a familiarity difference, the present results suggest that the higher FA rates to mismatch stimuli in comparison to other foil types does not, or not only, reflect the higher familiarity of participants' with the nonwords and intervals that are involved in these stimuli.

Table 3. Average FA rates in % in the global recognition test of Experiment 1 for the mismatch stimuli made out of the most and the least frequently heard features in each standard condition. Each cell corresponds to the average of two observations, that is, two presentations of each mismatch stimulus, on 8 participants in the Standard 1 and 8 participants in the Standard 2 condition. Most frequently presented stimuli in each group are in bold font.

	 /damy/	 /dal /	 /dalε/	 /dany/
Standard 1	**56.25**	**81.25**	68.75	50
Standard 2	56.25	81.25	**75**	**81.25**

In summary, nonwords seem to be easier to recognize than intervals and whole stimuli, but this difference, like the familiarity difference between mismatch stimuli and other foils, is not sufficient to explain the pattern of results of Experiment 1. These recognition results suggest that tunes (intervals) and lyrics (nonwords) are represented separately in memory, as indicated by the occurrence of illusory conjunctions. These separate representations are however tied in a mnemonic network that probably leads to the integration effect described by Serafine et al. (1984) and observed here. The origins of these connections will be further explored in Experiment 2.

3. 2. 3. Experiment 2:
Forced-choice recognition memory tests on either vocalic or consonantal materials

As lyrics and tune seem to be separately represented but nevertheless associated in memory, it is worth asking which features modulate the strength of these connections. Studies on song processing and production (Scotto di Carlo, 1993, Sundberg, 1982) and recent results (Kolinsky et al., under review) according to which vowels and intervals are processed on-line in an integrated way, whereas consonants and intervals are perceptually more separable, suggest that the phonetic characteristics of the lyrics could influence their interaction with the tune. The purpose of Experiment 2 was to test this hypothesis for song memory. This was not possible to do on the basis on Experiment 1, in which mismatch stimuli blend the C- and V-materials (namely, an interval of the C-material with a nonword of the V-material, or vice-versa).

In Experiment 2, after a speeded classification task acting as encoding phase, two groups of participants were exposed to memory tests similar to the ones of Experiment 1, but with different experimental materials. As depicted in Table 4, in the speeded classification task, one group was presented only with stimuli varying on a consonantal basis (henceforth, *C-materials*) and the other with stimuli varying on a vocalic basis (*V-materials*). The memory performance of these two groups could thus be compared: as illustrated in Table 5, for the participants presented with the two C-materials (*C1* and *C2*), the mismatch stimuli were created by pairing an interval of the C1-material with a nonword of the C2-material (or vice-versa). For those presented with the V-materials (*V1* and *V2*), the mismatch foils were created by pairing an interval of the V1-material with a nonword of the V2-material (or vice-versa). If in memory, like in perception, consonants and intervals were less integrated than vowels and intervals, we expected to observe more illusory conjunctions for the consonantal than for the vocalic materials.

2. 2. 3. 1. Method

2. 2. 3. 1. 1. Participants

The 49 participants (7 males, average age: 20 years, range: 18-36 years, 45 right handed) were undergraduates who participated in return for credits toward an introductory psychology course requirement. Most participants had never practiced music, but 15 of them had musical experience of maximum four years, stopped for at least seven years. Participants were randomly divided into four groups. One group of 25 participants was exposed to the two C-materials in the speeded classification phase. Among them, 13 were presented later on with the interval recognition test on the C-materials, the 12 others were presented with the whole stimulus recognition test on the same materials. The other 24 participants were exposed to the two V-materials in the speeded classification phase; finally, half of them were presented with the interval recognition test on the V-materials, and the others with the whole stimulus recognition test on the same materials. The material was thus a between-subjects variable, a choice justified by the length of the speeded classification task that did not allow submitting the same participants to the two materials in one session.

2. 2. 3. 1. 2. General material and procedure

Material preparation was the same as in Experiment 1, except that the stimuli were recorded by a different baryton. The procedure was also identical to the one of Experiment 1.

2. 2. 3. 1. 2. 1. Incidental encoding – speeded classification task.

In the speeded classification phase, one group of participants performed the classification task thoroughly described in Experiment 1, but on two different C-materials (see Table 4a), that is, /daby/-/dagy/ sung on the ascending F3-#F3 and the descending F3-A2 minor intervals (henceforth, *C1*-material), and /daty/-daky/ sung on the ascending F3-A3 and the descending F3-C3 major intervals (*C2*-material). The other group had to classify two different V-materials (see Table 4b): /dalɛ/ and /dalo/ sung on the ascending F3-A3 and the descending F3-C3 major intervals (henceforth *V1*-material) and /dalɔ̃/-/dalø/ sung on the slightly ascending F3-#F3 and the descending F3-A2 minor intervals (*V2*-material). Note that the same intervals were thus used in the C and the V materials, and that these intervals were identical to the ones of Experiment 1. As in Experiment 1, the number of occurrences of each stimulus varied between-subjects, depending on the assigned standard and redundant conditions (see Appendix 3).

Table 4. The different combinations of nonwords and intervals used in the speeded classification phase of Experiment 2, separately for the two C-materials (part a: C1- and C2-materials) and for the two V-materials (part b: V1- and V2-materials).

a.

C1-MATERIAL		Interval 1(I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/daby/	I1 NW1	I2 NW1
Nonword 2 (NW2)	/dagy/	I1 NW2	I2 NW2

C2- MATERIAL		Interval 1(I1)	Interval 2 (I2)
			
Nonword 1 (NW1)	/daty/	I1 NW1	I2 NW1
Nonword 2 (NW2)	/daky/	I1 NW2	I2 NW2

b.

V1- MATERIAL		Interval 1(I1)	Interval 2(I2)
			
Nonword 1 (NW1)	/dalɛ/	I1 NW1	I2 NW1
Nonword 2 (NW2)	/dalo/	I1 NW2	I2 NW2

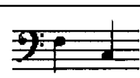
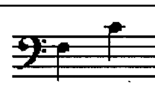
V2- MATERIAL		Interval 1(I1)	Interval 2(I2)
			
Nonword 1 (NW1)	/dalɔ̃/	I1 NW1	I2 NW1
Nonword 2 (NW2)	/dalø/	I1 NW2	I2 NW2

2. 2. 3. 1. 2. 2. Recognition tests.

This incidental encoding task was followed by one of two recognition tests: recognition of the intervals or of the whole stimuli. We did not include the nonword recognition because, in Experiment 1 as in Serafine et al.'s (1984) study, the performance in this task was at ceiling, so that these data were not much informative (Crowder et al., 1990; Serafine et al., 1986).

Like in Experiment 1, five types of stimuli, including eight old stimuli, 16 *New I/new NW,* eight *old I/new NW,* eight *old NW/new I,* and eight *mismatch stimuli,* were presented in the 96 trials of the recognition test (see Table 5 for examples). The foils were created from the same four new intervals as in Experiment 1 (unison: F3-F3, major ascending: F3-C4, descending octave: F3-F2, and minor ascending: F3-#D4). In addition, four new nonwords varying in their middle consonant (/damy/, /dany/, /dafy/ and /dapy/), but keeping the same final vowel as the targets, were chosen to create the foils of the C-materials. Four nonwords varying on the final vowel (/daly/, /dalɛ̃/, /dalu/ and /dalɑ̃/) but keeping the same consonants served as a basis for the foils of the V-materials. The complete list of these foils is provided in Appendix 4.

Table 5. Examples of stimuli used in the recognition tests of Experiment 2 (I = interval; NW = nonword), separately for the C-materials (middle column) and V-materials (right column).

	C-materials	V-materials
Completely old	 /daky/	 /dalø/
Completely new	 /dapy/	 /dalɛ̃/
Old I / New NW	 /dapy/	 /dalɛ̃/
Old NW / New I	 /daky/	 /dalø/
Mismatch	 /daby/	 /dalɛ/

2. 2. 3. 2. Results and discussion

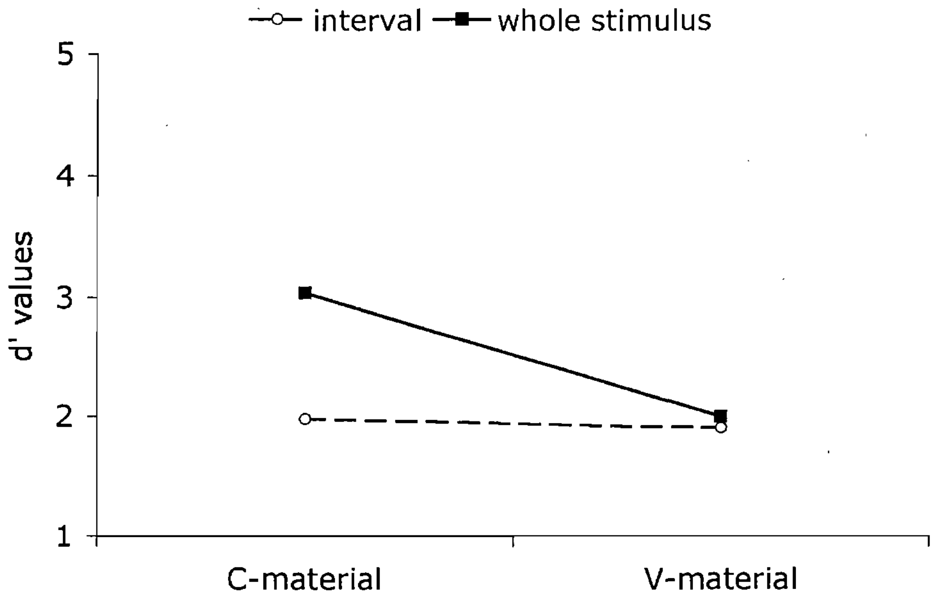
2. 2. 3. 2. 1. Comparison between recognition tests

The average percentages of correct responses were 63% for the interval recognition test and 73.5% for the whole stimulus recognition test. The performance is thus below the one of Experiment 1, maybe because the new nonwords were more similar to the old ones here. Indeed, only one phoneme was changed in Experiment 2 vs. two in Experiment 1. The pattern is, however, in the same direction.

In order to compare the accuracy level across the two tests and to check whether the performance was similar for the vocalic and consonantal materials, we ran ANOVAs with task (interval vs. whole stimuli) and material (V- vs. C-material) as between-subjects variables on the arcsine transform of the hit and FA rates. The recognition test and the material did not influence significantly the hit rate, both *F*s < 1. The interaction between these variables fell short of significance, F (1, 46) = 3.97, $p < .06$, with slightly more hits on the global recognition test for the consonant than for the vowel material. The false alarms were significantly more numerous for the interval than for the global recognition test, F (1, 46) = 4.61, $p < .05$, but their number did not differ as a function of the material F (1, 46) = 1.05, $p > .10$ and these variables did not interact, $F < 1$.

On the *d'* scores (Green & Swets, 1966), although there was a trend towards better performance in the whole stimuli recognition test with the C-material (see Figure 4), no effect or interaction did reach the significance level (test: $F(1, 45) = 2.47$; material: $F(1, 45) = 2.3$; material by test interaction: $F(1, 45) = 1.73$, $p >.10$ in all cases). These results suggest that the interval and the whole stimulus recognition tests were roughly equivalent in difficulty on the V- and the C- material, what allows us to compare the participants' performance between materials.

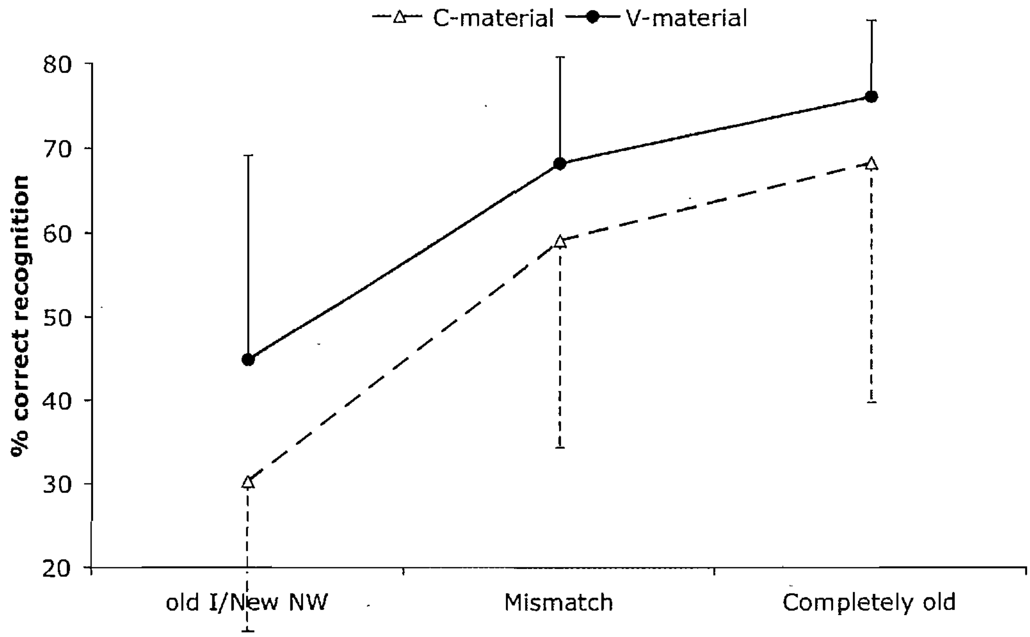
Figure 4. Average *d'* scores as a function of recognition test and material in Experiment 2.



2. 2. 3. 2. 2. Integration effect

In order to check for an integration effect, the arcsine transform of the proportion of hits (correct recognitions) in the interval recognition test were compared between completely old stimuli, partially old ones and mismatch stimuli. The ANOVA took into account material (C- vs. V-material) as a between-subject variable and stimulus type as a within-subject variable (completely old; mismatch; old I/new NW). This analysis revealed a significant effect of stimulus type, $F(2, 46) = 16.81$, $p < .0001$, but no significant effect of material, $F(1, 23) = 1.178$, $p > .10$, nor interaction with this variable, $F < 1$. The effect of stimulus type indicates that the participants recognized better an old feature in a completely old stimulus (on the average, 72% correct) than in a partially old one (on the average, 37.5%, $F(1, 23) = 18.49$, $p < .001$) and than in a mismatch one (on the average, 64% correct, $F(1, 23) = 8.81$, $p < .01$), a pattern that corresponds to the integration effect. This is illustrated in Figure 5. The nonsignificant interaction with material shows that the integration effect was not modulated by the vocalic or consonantal nature of the nonwords.

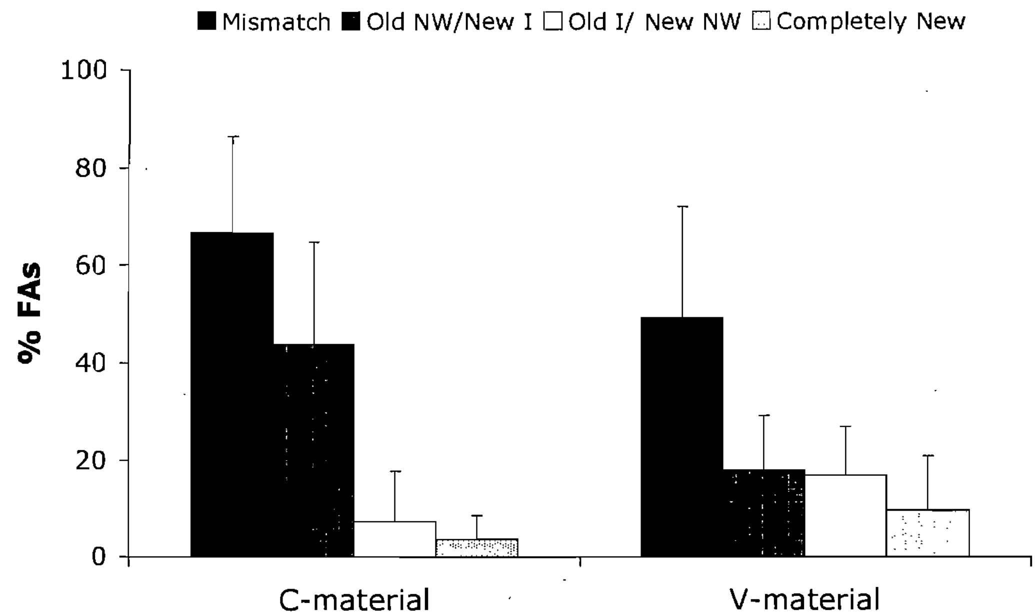
Figure 5. Average correct recognitions according to material and stimulus type observed in the interval recognition test of Experiment 2. Error bars represent one standard-deviation from the average.



2. 2. 3. 2. 3. Illusory conjunctions

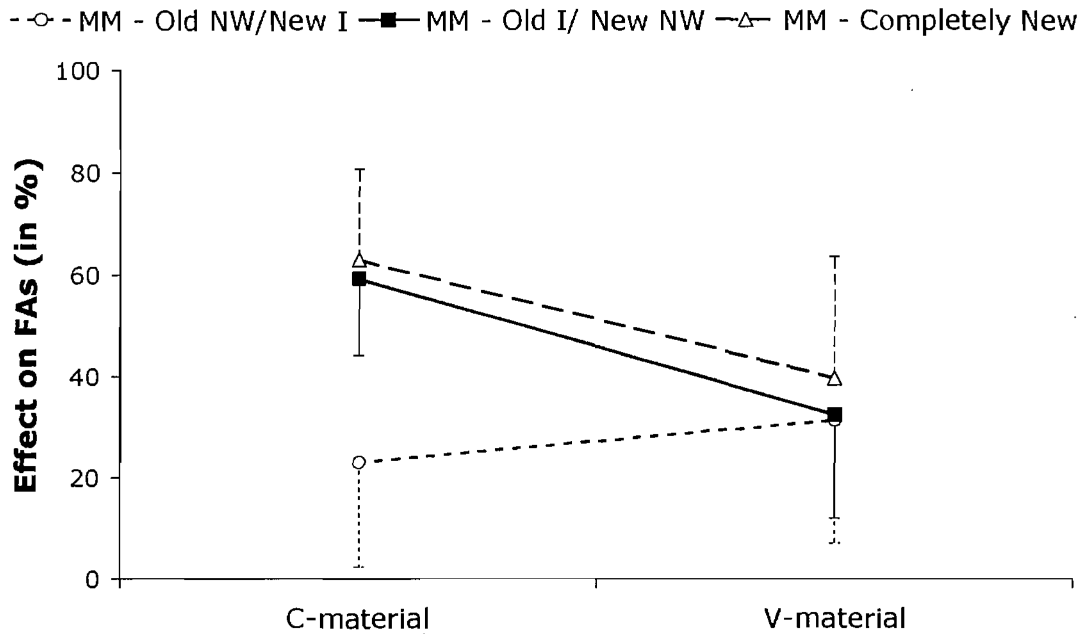
As in Experiment 1, a higher FA rate for mismatch foils than for all other kinds of foils in the whole stimuli recognition test will point to the occurrence of illusory conjunctions of lyrics and tunes. The ANOVA run on the arcsine transform of this score took foil type as a within-subjects variable (completely new, old I/new NW, old NW/new I; mismatch) and material (C- vs V- material) as a between-subjects variable. It showed a highly significant effect of foil type (F (3, 66) = 73.33, $p < .0001$). The effect of material was not significant, $F < 1$, but this variable interacted with foil type, F (3, 66) = 11.76, $p < .0001$. As illustrated in Figure 6, participants made much more FAs to mismatch stimuli than to all other kinds of foils, with both the C-material (F (1, 33) = 121.21; = 13.71; and = 143.11 for mismatch vs. old I/new NW, vs. old NW/new I, and vs. completely new foils, respectively, p always at least < .001) and the V-material (F(1, 33) = 29.49; = 25.57; and = 50.45 for mismatch vs. old I/new NW, vs. old NW/new I, and vs. completely new foils, respectively, $p < .0001$). However, the interaction suggests that the pattern of results differed between the consonantal and the vocalic material. Indeed, the FA rate to mismatch and to old NW / New I foils tended to be higher for the C-material than for the V-material (respectively, $p = .05$ and $p < .001$), but the reverse tendency was observed for old I/ New NW and completely new foils (respectively, $p < .05$ and $p < .08$), see Figure 6.

Figure 6. Average FA rates as a function of foil type and material observed in the whole stimulus recognition test of Experiment 2. Error bars correspond to one standard-deviation from the average.



To further explore this consonant-vowel dissimilarity, we compared the difference of FA rate to mismatch stimuli and to other foils between the C- and the V-material. This difference, computed by subtracting the FA rate to each kind of foils from the FAs to mismatch stimuli, will henceforth be referred as the *FA rate superiority for mismatch stimuli*. The results, depicted in Figure 7, show a significant interaction between the material and the size of this difference, $F(2, 44) = 15.15, p < .0001$. Indeed, the FA rate superiority for mismatch stimuli was larger in the C- than in the V-material when mismatch stimuli are compared to either Old I/New NW foils, $F(1, 44) = 8.12, p < .01$ or completely new items, $F(1, 44) = 7.47, p < .01$, but there was no difference when mismatch stimuli are compared to old NW/new I foils, $F < 1$. This difference between materials is certainly jointly due to the higher FA rate to mismatch foils for the C-material (see Figure 6) and to the more numerous FAs to Old I/New NW and completely new foils for the V-material. Altogether, the C-material looked to be more prone to illusory conjunctions than the V-material.

Figure 7. Whole stimulus recognition test of Experiment 2: average FA difference between the mismatch stimuli and other foil types, as a function of foil type and material. Error bars represent one standard-deviation from the average.

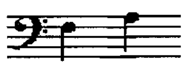


Finally, we checked whether this FA rate superiority for mismatch stimuli was indeed related to the occurrence of illusory conjunctions and not to the greater familiarity of their components. As in Experiment 1, we compared the FA rates to mismatch stimuli resulting from combinations of frequently presented features in the learning phase with the FA rate to mismatches of the least often heard nonwords and intervals. Based on Appendix 3, if the familiarity were sufficient to explain the higher FA rate to mismatch stimuli, we expected to see more FAs to /dagy/ sung on F3-A3 and to /daty/ on F3-#F3 than to other mismatch stimuli for the participants exposed to the Standard 1 condition of the C-materials. Conversely, subjects who had received the Standard 2 condition should have made more FAs to /daby/ on F3-C3 and to /daky/ on F3-A2. For the V-material, the highest FA rate was expected to /dalø/ on F3-A3 and /dalo/ on F3-#F3 for the Standard 1 and to /dalɛ/ on F3-A2 and /dalɔ̃/ on F3-C3 for the Standard 2 condition. Again, such a pattern was not found (see Table 6, which depicts the percentage of FAs to these mismatch stimuli): the ANOVAS with mismatch stimulus as a within-subject variable and standard condition as a between-subjects variable did not reveal effects of stimulus or of standard condition, neither an interaction between these variables, being on the C-material or on the V-material (stimulus and standard condition: $F < 1$ on both C- and

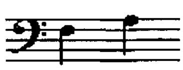
V-materials, interaction: $F(3, 30) = 1.25$ on the C- and $= 1.02$ on the V-Material, $p > .10$).

Table 6. Average FA rates in % in the global recognition test of Experiment 2 for the mismatch stimuli made out of the most and the least frequently heard features in each standard condition, for the C- (a) and the V-material (b). Each cell corresponds to the average of two observations, that is, two presentations of each mismatch stimulus, on 6 participants in the Standard 1 and 6 participants in the Standard 2 condition for each material. Most frequently presented stimuli in each group are in bold font.

a. C-material

	 /dagy/	 /daty/	 /daby/	 /daky/
Standard 1	**50**	**58.33**	58.33	83.33
Standard 2	66.66	58.33	**66.66**	**58.33**

b. V-material

	 /dalø/	 /dalo/	 /dalɛ/	 /dalɔ̃/
Standard 1	**57.14**	**64.28**	50	57.14
Standard 2	50	33.33	**58.33**	**41.66**

2. 2. 3. 2. 4. Comparison of Experiments 1 and 2: Integration effect

To evaluate Crowder et al.'s (1990) claim that the integration effect is, at least partly, due to physical interactions between lyrics and tune leading to submelodic changes in mismatch songs, we compared the integration effect between Experiments 1 and 2. Although the influence of these phoneme switches is difficult to objectivize, one might expect more obvious lyrics changes to have a stronger impact on melody articulation than more subtle lyrics changes. Thus, if submelodic variations were indeed involved in the integration effect, computed as the hit difference between mismatch and truly old stimuli, this effect should be stronger in

Experiment 1, in which the mismatch songs blend materials differing by two phonemes than in Experiment 2, in which the nonwords only differ by one phoneme.

The ANOVA on the interval recognition score with stimulus type (mismatch and truly old) as a within-subject variable and material (Experiment 1, Experiment 2 / V-material, Experiment 2 / C-material) as a between-subjects variable revealed a significant effect of stimulus type, $F(1, 38) = 9.94$, $p < .005$, corresponding to the integration effect. This integration effect was not modulated by the experiment and material, $F < 1$. Indeed, the size of the integration effect computed as the difference in interval recognition between the completely old and the mismatch stimuli was 8% in Experiment 1, 8% for the V-material of Experiment 2, and 9% in C-material of Experiment 2.

In summary, the results of Experiment 2 suggest that vowels and consonants interact differently with intervals in memory. Indeed, the participants made more illusory conjunctions for the C-material, as suggested by the more numerous FAs and the larger difference in FA rates for mismatch stimuli and other foils for this material. This phenomenon is not due to disparities in task difficulty given that the global performance was similar for the V- and the C-materials and that the same melodic intervals were used in both materials. As in perception (Kolinsky et al., under review), consonants thus seem more separable from intervals than vowels are. This dissimilarity does, however, not support the submelodic interpretation of the integration effect (Crowder et al., 1990). Indeed, the integration effect did not differ between Experiments 1 and 2 though the lyrics change between the exposure phase and the test phase was larger in Experiment 1.

2. 2. 4. General discussion

Although most studies that have addressed the interactions between lyrics and tunes in familiar and unfamiliar songs with various paradigms have concluded that these dimensions are separable but linked by strong connections in memory (Crowder et al., 1990; Peretz et al., 1994; Peretz et al., 2004; Samson & Zatorre, 1991; Serafine et al., 1984; Serafine et al., 1986; Steinke, Cuddy, & Jakobson, 2001), only a few of them provided convincing explanations of the nature of these associations. The present study was aimed at testing the following hypothesis: the phonetic properties of the lyrics, and more specifically the vocalic or consonantal characteristics of the text, modulate the lyrics-tunes interactions (Kolinsky et al., under review). This idea is related to Crowder et al.'s (1990) suggestion, according to which the connections between lyrics and tune are, at least partly, due to physical interactions between these features.

This goal was achieved by combining the integration effect (Serafine et al., 1984) and the analysis of illusory conjunctions of lyrics and tunes (Thompson et al., 2001; Treisman & Schmidt, 1982). Experiment 1 showed that the integration effect could co-occur with illusory conjunctions of lyrics and tune, indicating that the text and the melody of songs are extracted and stored as separated features but are, nevertheless, associated in memory. In Experiment 2, we supplied additional evidence for the claim that vowels and consonants vary as regards their interactions with melody (Kolinsky et al., under review) by comparing the integration effect and the illusory conjunctions between materials in which the mismatch songs were built by switching lyrics differing either on a consonants basis or on a vowel basis. However, one should note that the kind of phoneme change only modulated the rate of illusory conjunctions (lyrics with a consonant change lead to more illusory conjunctions than lyrics with a vowel change), but not the integration effect. Moreover, the comparison of the integration effect between Experiments 1 and 2 does not seem to support Crowder et al.'s (1990) physical interaction hypothesis. In addition to the observation of a spared, though weaker, integration effect when original lyrics are replaced by phonetically similar lyrics (Corwder et al., 1990), we proposed a new test of the physical interaction hypothesis. Mismatch songs in which the new lyrics differ from the original lyrics by more phonemes (Experiment 1, in which the new lyrics included two phonemic changes) should indeed elicit less correct recognitions of one dimension than mismatch songs mixing more similar lyrics (Experiment 2, in which the new lyrics included only one phonemic change). This was not the case.

Though these results must be interpreted with caution because the phoneme switches used here were maybe not sufficiently different to generate significant effects, the integration effect does not seem to be only related to submelodic modifications of the tune by the phonetic properties of the lyrics. Integration effects have, actually, been found for spoken lyrics and hummed melodies (Crowder et al., 1990) and phenomena comparable to the integration effect have been obtained on stimuli having nothing linguistic or musical, like faces. For example, Jones, Bartlett and Wade (2006) observed that the features of faces are less well recognized in mismatch faces than in old faces and state that this difference could simply be caused by a higher familiarity of old faces. For old faces, both the features and their *combination* (configural information) were familiar, while for mismatch faces the features were familiar but their conjunction was new. In addition, studies on the dissociation between recollection and familiarity (Jones & Jacoby, 2001; Yonelinas, 2002) suggest that old stimuli might lead to recollection, while the recognition of mismatch stimuli is probably only based on a feeling of familiarity without recollection. The integration effect, instead of being specific to songs, could thus emerge from the general functioning of recognition memory. That might explain why it is not affected by the vocalic vs. consonantal nature of the lyrics. This parsimonious interpretation is not at odds with alternative accounts of the integration effect for songs, like the physical interaction, the association-by-contiguity (Crowder et al., 1990) and the rhythmic congruencies accounts (Peretz et al., 2004). Indeed, the acoustical interaction between lyrics and tune, their temporal proximity, and the presence of rhythmic parameters are all typical properties of songs that might contribute in various extends to the recollection of old stimuli and, thus, to the well-known connections between lyrics and tune.

In addition to this integration effect, we have found that, when they had to state whether a stimulus had been heard in the same form in the learning phase, participants made significantly more false alarms to mismatch stimuli (made out of two old features from different songs) than to foils which had none or only one old feature. According to Treisman's theory (Treisman & Schmidt, 1982; Treisman & Gelade, 1980), these errors reflect illusory conjunctions and are diagnostic of dimensional separability, given that only separately extracted features can be erroneously recombined.

An alternative account for this outcome is that mismatch stimuli are overall more familiar that the other foils because both their lyrics and tune had been presented in the exposure phase, compared to only one or none feature for the remaining foils (Jamieson et al., 2003). The mismatch foils would thus be more likely to elicit an "old" response independent of the presence of illusory conjunctions. In the present study, we had the opportunity to compare the illusory conjunctions hypothesis (Treisman & Schmidt, 1982) and the familiarity hypothesis

(Jamieson et al., 2003). In the speeded-classification task, which acted as an incidental learning phase, some stimuli were heard slightly more often than others. The familiarity hypothesis will thus predict a higher FA rate for mismatch stimuli combining the most often heard lyrics and tunes, which allowed more opportunities of correct encoding, compared to mismatches of the less frequent features. However, we did not find any significant difference in the FA rates to these two kinds of mismatch stimuli either in Experiment 1 or in Experiment 2. Though these results have to be interpreted carefully because of the limited number of observations in each condition, they do not support the familiarity hypothesis. Even tough the familiarity effect were present but undetected due to this lack of reliability, this would not challenge the idea of independence of lyrics and tune in memory. Indeed, in the REM model (Shiffrin & Steyvers, 1997) on which Jamieson and colleagues' claim (2003) is based, the likelihood that an idem is recognized as old is related to its mean familiarity which, in turn, depends on the number of features in common between a foil and a studied item. This implicitly assumes that features are extracted separately. In sum, whatever the explanatory model chosen, either familiarity or illusory conjunctions, the observation of more numerous FAs to erroneous conjunctions of old lyrics and old tunes from different stimuli than to other foils suggest that lyrics and tune are separable dimensions.

The main finding of the present study was that the occurrence of these illusory conjunctions was modulated by the vocalic or consonantal nature of the material. Following Garner's (1974) theoretical framework, we had previously observed that vowels and melodies behave as integral dimensions but that consonants and melodies are more separable (Kolinsky et al., under review). In the light of the illusory conjunctions approach, since consonants have been found to be more perceptually separable from the tune than vowels are, they should also be more prone to be erroneously recombined with these tunes, and, as a consequence, yield more illusory conjunctions. Such an outcome would confirm that the phonetic properties of the lyrics modulate their interaction with the tune in memory.

The results of Experiment 2 support this hypothesis: both the number of FAs and the FA rate superiority to mismatch stimuli, i.e. the difference in FAs between mismatch stimuli and other foils, were higher for the C- than for the V-material. These two indexes suggest that, in memory too, consonants are more separable from the melody than vowels are. In other words, if in perception *vowels sing but* *consonants speak*, in memory, *vowels sing but consonants swing*, hence leading to more numerous illusory conjunctions. But why are, in both perception and memory, consonants less strongly associated to the melody than vowels ?

The origin of this difference can be discussed in the light of recent studies on the evolution of speech and music (for reviews, see Fitch, 2000, 2006; Hauser & McDermott, 2003). Based on data from paleontology, archaeology and human-animal comparisons, Mithen (2005) elaborates the idea that music and language may have evolved together as a "musical protolanguage". The respective status of vowel-like and consonant-like sounds in this precursor of language is an exciting matter for speculations. A central aspect of human speech is the extensive use of formants, which function as bandpass filters. The three first formants of a vocalization define the vowels of a language (Lieberman & Blumstein, 1988). However, formants production and perception is not specific to humans (Fitch, 2002): the vocalizations of non-human primates (Lieberman, 1968; Owren, Seyfarth, & Cheney, 1997; Rendall, Rodman, & Emond, 1996) and of birds (Suthers, 1994) also possess formants. On the contrary, there is no evidence, to our knowledge, of consonant-like production in nonhuman species, probably because the articulatory speed and accuracy needed to produce stopped and fricatives requires a vocal tract and tongue shape that is specific to humans thanks to their lowered larynx (Lieberman, 1984; but see Fitch & Reby, 2001). In brief, consonants production may emanate from the biomechanical properties of the humans' articulators (MacNeilage, 1998; MacNeilage & Davis, 2000). On the perceptual side, some studies have found that non-human primates have a (steady-state) formant perception comparable to the one of humans (Sommers, Moody, Prosen, & Stebbins, 1992) and that macaques can learn to discriminate the manner of articulation of consonants (Sinnott & Williamson, 1999). However, because the monkeys experimented problems in learning place of articulation contrasts, the authors conclude that their way of processing formant transitions differs from humans' (Sinnott & Gilmore, 2004). Similarly, in the statistical learning of artificial languages, cotton-top tamarins are able to capture non-adjacent regularities based on vowels but not on consonants (Newport, Hauser, Spaepen, & Aslin, 2004), while human adults exhibit the opposite pattern (Bonatti, Peña, Nespor, & Mehler, 2005; Mehler, Peña, Nespor, & Bonatti, 2006). The function of vowels and consonants in human speech also diverge: while vowels are fundamental for talker identification, consonants play a central role in meaning extraction (Bonatti et al., 2005; Owren & Cardillo, 2006).

In sum, vowel-like sounds defined by their formant-pattern do not appear to be specific to human speech, while consonant-like sounds require a vocal-tract virtuosity that is typically human and, therefore, probably specific to speech. As vocalizations based on formants, like vowels, do not seem to be diagnostic of speech and can be handled by several animal species, their processing may rest on auditory mechanisms that are shared with music, or at least melodic processing. The repeated finding of tight interactions between vowel and pitch (Kolinsky et al., under review; Ross et al., 2007; Sundberg, 1982) might thus be explained by these common processes

To conclude, consistent with previous research (Crowder et al., 1990; Peretz et al., 2004), our results support the claim that the lyrics and the tune of songs are separable in memory but are, however, tied by bidirectional connections. The origin of these connections may rest on temporal (Crowder et al., 1990) or rhythmic contingencies (Peretz et al., 2004), but also, as demonstrated here, on the phonological structure of the lyrics (see also Crowder et al., 1990). However, these phonological differences do not influence the so-called integration effect (Serafine et al., 1984) that probably rests on more general memory processes such as recollection. By contrast, illusory conjunctions seem to be specifically influenced by the processing differences between vowels and consonants at the memory level, showing that vowels and melodies are closely linked in song memory while consonants and melodies can be more easily disjoined. This conclusion points to the potential different evolutionary function of these phonemes in speech and song.

2. 2. 5. Appendix

Apprendix 1. Number of occurrences of each stimulus feature as a function of the between-subjects standard and redundant conditions for the C- and the V-materials of Experiment 1. The top panel (a) depicts the Standard 1 conditions; the bottom panel (b) shows the Standard 2 conditions. The nonword-interval pairs that were associated in the Redundant condition are in the same font style.

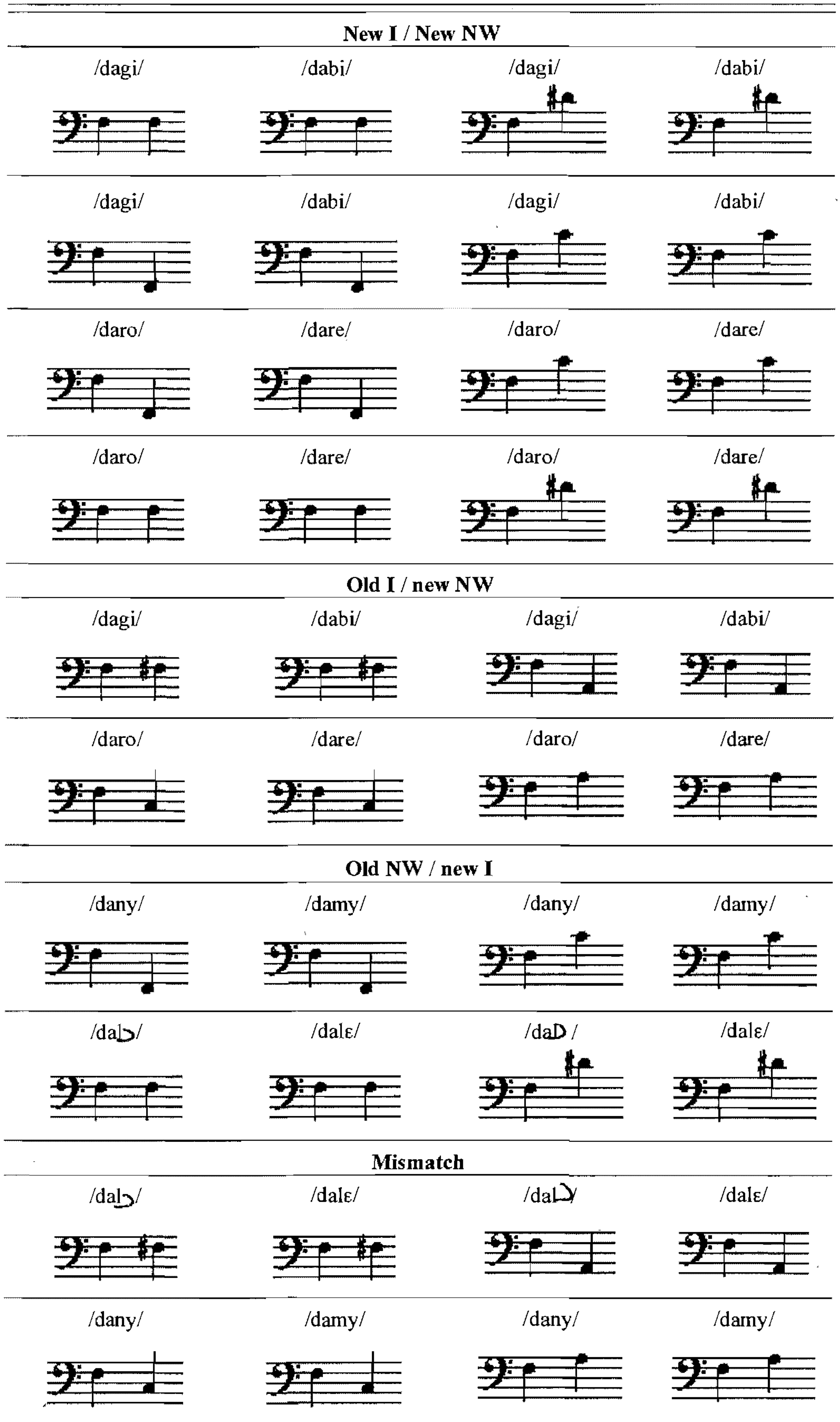
a. Standard and Redundant 1

Task	Condition	C-material				V-material			
		Nonword		Interval		Nonword		Interval	
		/damy/	/dany/	F3-F3#	F3-A2	/dalɛ/	/dalɔ/	F3-A3	F3-C3
	Standard	72	0	36	36	0	72	36	36
Melodic	Redundant	**36**	*36*	**36**	*36*	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
	Standard	36	36	72	0	36	36	72	0
Phono-logical	Redundant	**36**	*36*	**36**	*36*	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Total		252	180	252	180	180	252	252	180

b. Standard and Redundant 2

Task	Condition	C-material				V-material			
		Nonword		Interval		Nonword		Interval	
		/damy/	/dany/	F3-F3#	F3-A2	/dalɛ/	/dalɔ/	F3-A3	F3-C3
	Standard	0	72	36	36	72	0	36	36
Melodic	Redundant	*36*	**36**	**36**	*36*	*36*	**36**	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
	Standard	36	36	0	72	36	36	0	72
Phono-logical	Redundant	*36*	**36**	**36**	*36*	*36*	**36**	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Total		180	252	180	252	252	180	180	252

Appendix 2: Foils used in the recognition tests of Experiment 1 (I = interval; NW = nonword).



Appendix 3. Number of occurrences of each stimulus feature as a function of the between-subjects standard and redundant conditions for the C1- and C2 materials (1) and the V1- and V2-materials (2) of Experiment 2. The top panel (a) of each material depicts the Standard 1 conditions; the bottom panel (b) shows the 2 conditions. The nonword-interval pairs that were associated in each Redundant condition are in the same font style.

1. C1 and C2-materials

a. Standard and Redundant 1

Task	**Condition**	**C1-material**				**C2-material**			
		Nonword		Interval		Nonword		Interval	
		/daby/	/dagy/	F3-F3#	F3-A2	/daty/	/daky/	F3-A3	F3-C3
	Standard	0	72	36	36	72	0	36	36
Melodic	Redundant	*36*	**36**	*36*	**36**	**36**	*36*	**36**	*36*
	Orthogonal	36	36	36	36	36	36	36	36
	Standard	36	36	72	0	36	36	72	0
Phono-logical	Redundant	*36*	**36**	*36*	**36**	**36**	*36*	**36**	*36*
	Orthogonal	36	36	36	36	36	36	36	36
Total		180	252	252	180	252	180	252	180

b. Standard and Redundant 2

Task	**Condition**	**C1-material**				**C2-material**			
		Nonword		Interval		Nonword		Interval	
		/daby/	/dagy/	F3-F3#	F3-A2	/daty/	/daky/	F3-A3	F3-C3
	Standard	72	0	36	36	0	72	36	36
Melodic	Redundant	*36*	**36**	**36**	*36*	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
	Standard	36	36	0	72	36	36	0	72
Phono-logical	Redundant	*36*	**36**	**36**	*36*	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Total		252	180	180	252	180	252	180	252

2. V1 and V2-materials

a. Standard and Redundant 1

Task	Condition	V1-material				V2-material			
		Nonword		Interval		Nonword		Interval	
		/dalɛ/	/dalo/	F3-A3	F3-C3	/dalɔ̃/	/dalø/	F3-F3#	F3-A2
Melodic	Standard	0	72	36	36	0	72	36	36
	Redundant	*36*	**36**	**36**	*36*	*36*	**36**	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Phono-logical	Standard	36	36	72	0	36	36	72	0
	Redundant	*36*	**36**	**36**	*36*	*36*	**36**	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Total		180	252	252	180	180	252	252	180

b. Standard and Redundant 2

Task	Condition	V1-material				V2-material			
		Nonword		Interval		Nonword		Interval	
		/dalɛ/	/dalo/	F3-A3	F3-C3	/dalɔ̃/	/dalø/	F3-F3#	F3-A2
Melodic	Standard	72	0	36	36	72	0	36	36
	Redundant	*36*	**36**	*36*	**36**	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Phono-logical	Standard	36	36	0	72	36	36	0	72
	Redundant	*36*	**36**	*36*	**36**	**36**	*36*	*36*	**36**
	Orthogonal	36	36	36	36	36	36	36	36
Total		252	180	180	252	252	180	180	252

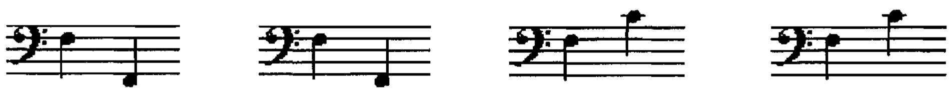
Appendix 4: Foils used in the recognition tests of Experiment 2 (I = interval; NW= nonword) for the C- and the V- materials.

a. C-Materials

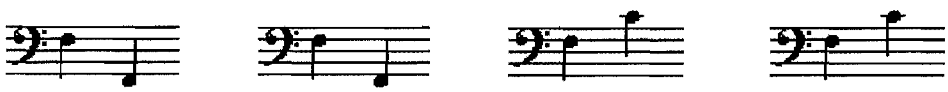
New I / New NW

/damy/	/dany/	/damy/	/dany/

/damy/	/dany/	/damy/	/dany/



/dafy/	/dapy/	/dafy/	/dapy/



/dafy/	/dapy/	/dafy/	/dapy/

Old I / new NW

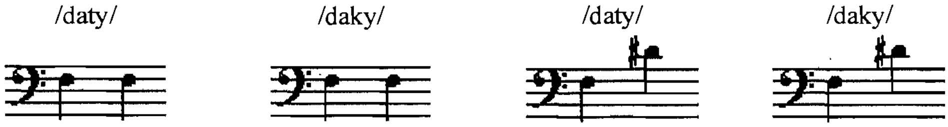
/damy/	/dany/	/damy/	/dany/

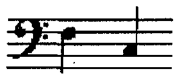
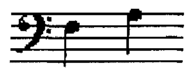
/dafy/	/dapy/	/dafy/	/dapy/

Old NW / new I

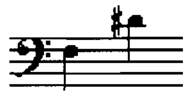
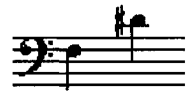
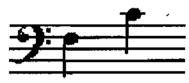
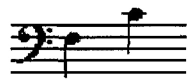
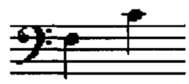
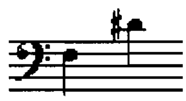
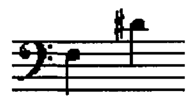
/daby/	/dagy/	/daby/	/dagy/

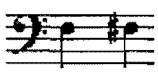
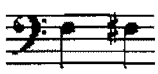
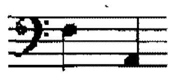
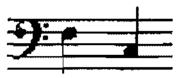
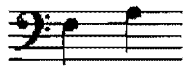
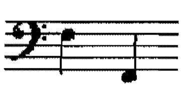
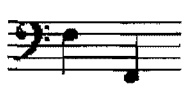
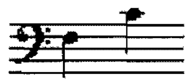
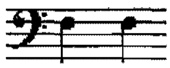
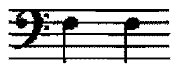
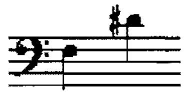
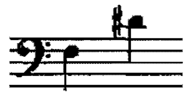
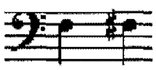
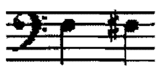
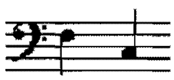
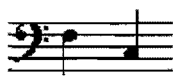
/daty/	/daky/	/daty/	/daky/



Mismatch			
/daty/	/daky/	/daty/	/daky/
			
/daby/	/dagy/	/dagy/	/dagy/
			

b. V-Materials

New I / New NW			
/daly/	/dalɛ̃/	/daly/	/dalɛ̃/
			
/daly/	/dalɛ̃/	/daly/	/dalɛ̃/
			
/dalu/	/dalɑ̃/	/dalu/	/dalɑ̃/
			
/dalu/	/dalɑ̃/	/dalu/	/dalɑ̃/
			

Old I / new NW			
/daly/	/dal$\tilde{\varepsilon}$/	/daly/	/dal$\tilde{\varepsilon}$/
			
/dalu/	/dalɑ̃/	/dalu/	/dalɑ̃/
			
Old NW / new I			
/dalɔ̃/	/dalø/	/dalɔ̃/	/dalø/
			
/dalo/	/dalɛ/	/dalo/	/dalɛ/
			
Mismatch			
/dalo/	/dalɛ/	/dalo/	/dalɛ/
			
/dalɔ̃/	/dalø/	/dalɔ̃/	/dalø/
			

2. 3. Article 4

Early integration of vowel and pitch processing: A mismatch negativity study [iv]

[iv] Lidji, P., Kolinsky, R., Jolicœur, P., Moreau, P., Connolly, J.F., & Peretz, I. (in preparation). Early integration of vowel and pitch processing: A mismatch negativity study.

Abstract

Objective: Several behavioral studies have explored the processing specificity of music and speech, but only a few have addressed the interactions between their building components, pitches and phonemes. Moreover, there is a discrepancy between behavioral studies showing that vowel and pitch are handled as integrated in singing, and electrophysiological experiments suggesting that separate neural networks underlie their processing. To disentangle these results, this study examines the additivity of the Mismatch Negativity (MMN) as an index of the early interactions between vowels and their pitch when sung.

Methods: Event-related potentials (ERPs) were recorded while participants were presented with standard sung vowels and stimuli deviating in pitch only, in vowel only or in both pitch and vowel. The MMN and P3a to these three kinds of deviants were compared.

Results: All three kinds of deviants elicited both a MMN and a P3a. The MMN was of similar amplitude for the three kinds of deviants but the P3a was larger for double deviants. The MMNs to vowel and pitch deviants were not additive. Finally, the MMN to pitch deviants occurred later than the MMN to vowel deviants.

Conclusions: The underadditivity of the MMN implies that vowel and pitch are processed by shared neural substrates, but the latency difference suggests that vowels elicit earlier responses, probably because of their greater significance in human speech compared to pitch, at least in non-tonal languages.

Significance: The present results indicate that vowel and pitch are processed as integrated, even at the pre-attentive level. Music-processing specificity thus rests on more complex dimensions of music and speech.

2. 3. 1. Introduction

The issue of the domain-specificity of music and speech has been a matter for debate for years in the growing community of scientists interested in music processing (e.g. Patel, 2008; Pinker, 1997; Zatorre & Peretz, 2001). Mounting experimental results from neuropsychology (Peretz & Coltheart, 2003; Peretz et al., 1994; Steinke, Cuddy, & Holden, 1997) support the hypothesis that specific processes are devoted to music perception and production. In contrast, brain imaging data show, for example, that music activates brain areas classically considered as specialized for language (Koelsch et al., 2002; Levitin & Menon, 2003) or that musical and linguistic syntax depend on shared neural resources (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). These data suggest that music and speech cannot be seen as totally independent from each other. Therefore, the most reasonable view in the present state of the art should be to consider that some aspects of music engage domain-specific processes (Peretz, 2006; Peretz & Zatorre, 2005) while other aspects engage processes that are shared between music and language (Patel, 2003; Price, Thierry, & Griffiths, 2006).

The identity of these aspects of music and speech that are handled by shared processes is, however, still unclear. Pitch and phonemes, which are the building blocks of music and speech, respectively, might be excellent candidates. More specifically, vowels and musical tones are both characterized by the dimensions of timbre and pitch. However, the relative importance of these dimensions differs between vowels, which are defined by their spectrum, and tones, which are defined by their pitch (Patel, 2008). However, recent data (Ross, Choi, & Purves, 2007) brought a new bridge between vowels and music by showing that the intervals from the music of most cultures have the same pitch ratio as the formants shaping the vowels of their languages. This suggests that the relationship between music and speech rest on more fundamental interactions than most authors have considered so far. While relatively complex parameters of music and speech, such as syntax, semantics and expectations (Besson, Faita, Peretz, Bonnel, & Requin, 1998; Koelsch et al., 2004; Patel, Gibson et al., 1998; Poulin-Charronnat, Bigand, Madurell, & Peereman, 2005; Schön, Gordon, & Besson, 2005) have been the focus of several comparative studies, little research has been dedicated to the interaction between the basic components that are vowels and pitches.

In one of the rare studies that compared phoneme identity and musical pitch, Bigand, Tillmann, Poulin, D'Adamo, and Madurell (2001) observed that the monitoring of the last vowel in a sung nonsense sentence is facilitated when the melody ends with the most expected chord, that is on the tonic, compared to the situation in which the last chord is the less expected subdominant. These results can be interpreted in two different ways. One the one hand, they could suggest the existence of interactions between the melodic and the phonological processing of sung vowels. On the other hand, recent evidence of similar interactions between visual forms and these melodies (Escoffier & Tillmann, 2006) support the idea that this effect simply reflects a general attention capture by the less expected subdominant, hence suggesting that the effect is not specific to vowels and melodies.

In a recent study on vowel and pitch processing interactions, Kolinsky, Lidji, Peretz, Besson, and Morais (under review) observed that these dimensions cannot be attended selectively in a speeded classification task (see Garner, 1974) in which the participants had to classify nonsense bisyllabic words sung on two-note intervals. When the nonword classification was based on vowel identity, irrelevant variations in pitch interfered with the classification process. A similar interference of irrelevant vowel variations was observed when the classification was based on pitch. In short, classification performance suggested that vowel identity and pitch were integral rather than separable dimensions (Garner, 1974).

Studies using similar behavioral methods to explore the interactions between pitch (or pitch contour) and vowels or consonants in the linguistic context of tone languages also found interactions between these dimensions (Lee & Nusbaum, 1993; Repp & Lin, 1990). However, the linguistic status of the pitch changes clearly modulated the strength of these interactions: the integration was stronger in speakers of Mandarin than in English-native participants, especially when the presented tones corresponded to linguistic tones of Madarin (Repp & Lin, 1990). This suggests that the lexical status of the tones plays a role in their interactions with the segmental information of speech (but see Taft & Chen, 1992 for contradictory results).

The limits of these speeded-classificaion studies (Kolinsky et al., under review; Lee & Nusbaum, 1993; Repp & Lin, 1990) rest on the various possible loci of the dimensional interaction, which can be as early as pre-attentive sound detection, but also as late as response selection, preparation or even response execution (Holender, 1992). In order to reveal that vowel and pitch processing are truly based on shared resources or components in speakers of non-tonal languages (as suggested by the results of Kolinsky et al., under review), one should rather focus on the early, pre-attentive processing of these dimensions, without the interference of task-related processes and participants' strategies. A useful tool to achieve this goal is the mismatch negativity (MMN) component of event-related potentials (ERPs), which reflects pre-attentive detection of auditory changes (Näätänen, 1992; Näätänen, Paavilainen, Rinne, & Alho, in press; Schröger, 1998). This negative deflection occurring between 100 and 250 ms after stimulus presentation is generally elicited in non-attentive conditions: the MMN cerebral generators are activated when the sensory trace of a deviant sound is detected as different from the sensory memory trace of the frequent sound, often called the standard (Näätänen, 1995). The MMN mainly originates from generators in the auditory cortex, described as related to pre-attentive detection, and from secondary frontal generators associated to later and probably conscious analysis of the change (Näätänen & Alho, 1995).

The localization of the sources of this physiological response has often been used to investigate the processing independence of different sound features (e.g. Giard et al., 1995; Levänen, Hari, McEvoy, & Sams, 1993). In a magnetoencephalographic (MEG) study, Tervianemi and colleagues (1999) used the magnetic counterpart of the MMN (MMNm) to compare the strength and location of the brain activation obtained in response to chords and to vowels. The results revealed a MMNm of greater amplitude in response to chords than to vowels in the right hemisphere but no difference in MMNm amplitude in the left hemisphere, both MMNm loci being posterior to the primary auditory cortex. These data thus support the idea of relative inter-hemispheric specialization for linguistic and musical dimensions (see also Eulitz, Diesch, Pantev, Hampson, & Elbert, 1995; Tervaniemi et al., 2000; Tervianemi & Hugdal, 2003; Zatorre, Evans, Meyer, & Gjedde, 1992). More specifically, these results suggest that the right hemisphere is most strongly involved than the left hemisphere in fine-grained pitch processing. In addition to this inter-hemispheric specialization, Tervaniemi et al. (1999) found evidence for intra-hemispheric specialization: in both hemispheres, the MMNm source location for the phoneme change was superior to that of the chord change. In short, these results suggest that spatially distinct neural populations are involved in the treatment of phonetic and musical stimuli. However, the musical (chords) and the phonetic (vowels) changes were presented in separate stimuli and separate blocks in

Tervaniemi et al.'s (1999) study. It might thus be that, although the material was controlled carefully, the independence of processing found by the authors was at least partly due to the intrinsic physical differences between the musical and phonetic material. The use of sung stimuli would constitute an important additional control because, in this case, the linguistic- and musical-related changes occur in exactly the same stimuli. Indeed, songs, as a natural combination of music and speech, are a great tool to directly compare music and language processing directly with the same experimental material for both dimensions (Lidji, 2007).

In the present study, motivated by the above considerations, we used sung vowels as stimuli, which enabled us to vary pitch and vowel identity in a tightly controlled context. In addition, we used the MMN component of the auditory ERPs as a way to measure early electrophysiological responses to these stimuli. Furthermore, we designed the experiment so as to be able to determine whether the MMN responses to a stimulus deviating from the standard along two stimulus attributes could be predicted from the simple additivity of the MMN response to two stimuli, each differing from the standard in a single dimension (e.g. Caclin et al., 2006; Paavilainen, Valppu, & Näätänen, 2001; Schröger, 1996b; Takegata, Paavilainen, Näätänen, & Winkler, 1999; Takegata, Syssoeva, Winkler, Paavilainen, & Näätänen, 2001; Wolff & Schröger, 2001). This test of the additivity of electrophysiological responses provides a way to probe the independence of processing, at the neural level, of vowel and pitch deviances in sung vowels.

The logic behind this approach is as follows: if the changes of two sound dimensions are processed by separate and independent neural generators, then the amplitude of the MMN response to a simultaneous change in both dimensions will be predicted by summing up the amplitudes of the MMN to changes in each single dimension. This idea has been corroborated by source localization studies (Caclin et al., 2006; Takegata, Huotilainen, Rinne, Näätänen, & Winkler, 2001), which have revealed that such additivity reflects the activity of at least partially distinct MMN neural generators, and thus separate memory traces, for each deviant feature. Indeed, for double deviants, the separate neural generators will fire for each change, the measured brain activity related to each single change summing up for the double change. This additivity approach has led to the discovery of several cases of apparently separate processing of auditory dimensions: Schröger (1996) found that the MMNs to interaural time and sound pressure were additive, as were the MMNs to duration and frequency (Levänen et al., 1993), duration and intensity (Wolff & Schröger, 2001), frequency and stimulus onset asynchrony (SOA) (Levänen et al., 1993; Paavilainen et al., 2001), frequency and location (Schröger, 1995), component dimensions of timbre such as spectral centroid and attack time (Caclin et al., 2006), and phoneme spectral quality and duration (Ylinen, Huotilainen, & Näätänen, 2005).

Conversely, if the sound dimensions have integrated or common neural representations (Paavilainen et al., 2001), a simultaneous change in both features will elicit an ERP response that will differ from the sum of the responses observed for a change in each of the two features, measured separately. This difference can be either an underadditivity, that is, an amplitude of the MMN to double deviants smaller than the one predicted by the additive model, or an overadditivity, corresponding to a larger observed MMN than predicted by the additive model. Both patterns may suggest that common or interacting neural populations are involved. The underadditivity can be explained if the same neurons respond to the two kinds of auditory changes. For double changes, they could fire as they did for single changes, or perhaps slightly more intensely if the double change leads to a greater saliency of the deviant, given that the MMN amplitude is proportional to the perceptual distance between standard and deviant stimuli (Näätänen et al., in press; Savela et al., 2003; Schröger, 1996a). Underadditivity has, for example, been observed for frequency and intensity (Wolff & Schröger, 2001). The interpretation of the overadditivity pattern is less clear, but we can speculate that the dimensions interact closely enough to yield a new kind of processing, different from the ones related to each single dimension, hence eliciting a new, larger, MMN. Such a result has been found for inter-stimulus interval and SOA (Takegata, Syssoeva et al., 2001).

Whereas the additivity of the MMN has been used to investigate the relations between various acoustical dimensions of sound (Caclin et al., 2006; Paavilainen et al., 2001; Schröger, 1995; Takegata, Huotilainen et al., 2001; Takegata, Syssoeva et al., 2001; Wolff & Schröger, 2001) or of speech (Ylinen et al., 2005), this paradigm has, to our knowledge, never been used to study whether the early pre-attentive brain responses to fundamental elements of music and speech, that is, pitch and phonemes, are elicited by independent or interactive neural processes. The aim of the present study was to fill in this gap by examining whether the MMNs to vowel and pitch deviants are additive, a result that would be consistent with independent neural generators for the extraction of vowel identity and pitch, at least in speakers of non-tonal languages. Additivity would suggest that the evidence for interaction found by Kolinsky et al. (under review) arises from relatively late cognitive processes, the processing of vowel and pitch at the stage producing the MMN being independent (Tervaniemi, 1999). Conversely, a non-additive pattern would suggest that vowel and pitch processing in songs is subserved by shared neural substrates.

2. 3. 2. Methods

2. 3. 2. 1. Participants

Twelve healthy right-handed French native speakers (five men, mean age 25 years, range: 18–38) volunteered to participate in this study. Five subjects had received musical training for 2 to 5 years but they all had stopped practicing for at least 6 years at the time of testing. All participants gave informed written consent after the procedures were explained to them. The study had received approval from the ethical committee of University of Montreal.

2. 3. 2. 2. Stimuli and procedure

The stimuli were the synthesized French vowels /ε/ and / /, sung at two different pitches separated by one semitone (C3 = 130 Hz and C3# = 138 Hz) with a duration of 300 ms, including 10 ms rise and fall time. The choice of this range was based on the results of calibration pilot studies that revealed that the MMN amplitude for the one semitone difference matched the amplitude of the MMN to the /ε/–/ɔ/ contrast. The stimuli were synthesized by a source-filter voice synthesizer simulating the vocal tract at a frequency of 130 Hz, and shifted up by one semitone with *Adobe Audition* software to create the 138 Hz stimuli. For each vowel, five formants were modeled. The frequency of the first four formants determined the nature of the vowel and F5 was always 1000 Hz higher than F4. In order to obtain a more natural singing feeling, a vibrato deviating by 2 % of the fundamental was added. Spectral analyses (PRAAT, Boersma & Weenink, 2007) showed that the pitch manipulation did not alter the formant structure of the vowels, as can be seen in Table 1.

Table 1. Fundamental frequency (F0) and frequency of the three first formants for the sung vowels in Hz.

	Vowel			
	/ɛ/		/ /	
	Pitch			
	C3	C3#	C3	C3#
F0	130	138	130	138
F1	601	602	513	540
F2	1795	1880	899	950
F3	2615	2760	2475	2626

The auditory stimuli were presented at a fixed inter-stimulus interval of 400 ms in a sequence including frequent (standard) and infrequent (deviant) stimuli, according to a passive oddball task. The presentation was pseudo-randomized so that two deviant sounds were separated by at least three standards. The deviants could differ from the standards in three different ways: pitch only, vowel identity only, and both pitch and vowel identity. A mixed design was used so that these three categories of deviants all occurred in one oddball block (Ylinen et al., 2005). Furthermore, each stimulus served as the standard, and the other three as deviants, across different blocks of stimulus presentations, as shown in Table 2. For example, in one block, /ɛ/ at C3 was used as the standard (1050 occurrences, P = .85), /ɔ/ at C3 was a deviant consisting of a vowel change (*vowel deviant*, 75 occurrences, P = .05), /ɛ/ at C3# was a deviant with a pitch change (*pitch deviant*; 75 occurrences, P = .05), and /ɔ/ at C3# was a deviant with simultaneous pitch and vowel changes (*double deviant*, 75 occurrences, P = .05). In sum, there were four blocks each lasting 15 minutes and the order of presentation was counterbalanced across participants. A total of 300 occurrences of each deviant category and 4200 standards were presented to each participant, summing up to 5100 sound presentations in the session. During the experiment, the participants were sitting in an electrically and acoustically isolated room. They were instructed to ignore the auditory stimulation presented binaurally through headphones at an intensity level of 70 dB SPL, while they watched a silent self-selected subtitled 3D-animation movie.

Table 2. Illustration of the experimental design : Example of sequence for one participant.

Stimulus category	Block 1	Block 2	Block 3	Block 4
Standard (P = .85)	/ɛ/ C3 N = 1050	/ɔ/ C3	/ɛ/ C3#	/ɔ/ C3#
Vowel deviant (P = .05)	/ɔ/ C3 N = 75	/ɛ/ C3	/ɔ/ C3#	/ɛ/ C3#
Pitch deviant (P = .05)	/ɛ/ C3# N = 75	/ɔ/ C3#	/ɛ/ C3	/ɔ/ C3
Double deviant (P = .05)	/ɔ/ C3# N = 75	/ɛ/ C3#	/ɔ/ C3	/ɛ/ C3

2. 3. 2. 3. Electroencephalogram recording and processing

The continuous electroencephalogram (EEG) was recorded (bandpass 0.1–70 Hz, sampling rate 256Hz, impedance < 10KΩ) via Synamp2 amplifiers (Neuroscan, Compumedics, El Paso, TX) from 64 Ag-AgCl electrodes at the standard 10-10 scalp sites. An electrode at the tip of the nose served as the reference. The electrooculogram (EOG) was monitored for horizontal and vertical eye movements using two bipolar electrode pairs placed at the outer canthi of left and right eyes and above and below the left eye. Because of a recording problem at C4 in several participants, this electrode and its homologous C3 were discarded from further statistical analyses.

The data were analyzed offline with Neuroscan 4.3.1. The EEG data were corrected for eye movements (Semlitsch, Anderer, Schuster, & Presslich, 1986), and filtered further with a 0.05–30 Hz band pass filter (24 dB/octave). Artifacts exceeding ± 100 μV were rejected, but the number of these trials did not exceed 22 put of 75 (in one subject and for one type of deviant). Prior to epoching, the standards sounds following a deviant sound were discarded from further analyses. These could indeed have evoked a MMN-like response because they were preceded by a physically different sound. Epochs of 800 ms, including a 100 ms pre-stimulus interval for baseline correction, were averaged separately for each oddball block (depending on the identity of the standard) and stimulus category (standard, pitch deviant, vowel deviant, double deviant). In a second step, epochs of the same stimulus category were pooled across blocks, hence controlling for stimulus-specific ERP variations.

2. 3. 2. 4. Data analysis

The grand average ERPs elicited by the standard and deviant sounds were examined on the non-subtracted waveforms (see Figure 1, Panels a and b). The MMN was delineated in each subject by subtracting the waveform to the standard from the waveform to each deviant category, hence leading to three difference MMN waves: pitch deviant, vowel deviant, and double deviant (see Figure 1, panel c). The MMN was quantified by computing the mean voltage within a 40 ms window centered at the each individual's peak detected within a time window ranging from 100 to 250 ms after tone onset at the fronto-central electrode (Fz), were the MMN amplitude was maximal, and at LM, in order to detect a potential mastoid inversion. The peak latency was defined as the time point of the maximum negativity at Fz in the same time window. The mean peak amplitude and peak latency of the P3a in the difference waves were measured in the same way, that is, averaged over a 40 ms around the peak amplitude, but in a window ranging from 200 to 350 ms. For analyses of variance, the Greenhouse-Geisser correction for non-sphericity was applied when required and the corrected p is reported along with the original degrees of freedom.

2. 3. 3. Results

2. 3. 3. 1. Standard and deviant ERPs

Figure 1 shows the grand-averaged standard and deviant waves at Fz (a) and LM (b), in each experimental condition. These electrodes were chosen because they exhibited the larger MMN amplitude. At Fz (Figure 1a), the standard waveform consisted of a P1 peaking at 70 ms, a N1 with a peak at 118 ms, and a P2 with a peak at 162 ms. These components were followed by a negative deflection peaking around 300 ms and a residual sustained negativity. Most of these components inverted polarity at the mastoids (Figure 1b).

A similar long-lasting negativity was observed in the three grand-average waves to deviants. Before this late potential, a P1, a negative peak in the MMN latency range and a positive deflection peaking between 260 (double) and 288 ms (pitch) were observed. Given its fronto-central topography (see Figure 1e) and relatively early latency, this positivity in the P300 latency range can be interpreted as a P3a, a component often assumed to reveal the orientation of attention even in tasks that do not require an overt response to the deviants (Escera, Alho, Schröger, & Winkler, 2000; Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007).

2. 3. 3. 2. Difference (deviant – standard) waveforms

2. 3. 3. 2. 1. MMN

The grand average difference waves (deviant – standard) obtained for each type of deviance at Fz and LM are depicted in Figure 1c. To confirm the presence of the MMN and of its polarity inversion for each type of deviant, *t*-tests against zero were performed on the mean amplitude of the difference wave at Fz and at LM. The double deviant, -1.6 ± 1 μV, $t(11) = -5.43, p < .0001$, as well as the vowel, -1.3 ± 0.6 μV, $t(11) = -8.29, p < .0001$, and the pitch deviants, -1.1 ± 0.8 μV, $t(11) = -5.22, p < .0001$, elicited a significant MMN at Fz, with a significant polarity inversion observed at the LM for the double, 1.4 ± 0.7 μV, $t(11) = 6.59, p < .0001$, the vowel, 1 ± 0.6 μV, $t(11) = 5.61, p < .0001$, and the pitch deviants, 1.4 ± 0.9 μV, $t(11) = 5.49, p < .0001$. The effects of condition (pitch deviant, vowel deviant, or double deviant) on the mean amplitude of the MMN at Fz and LM and on the latency of the MMN at Fz were analyzed in separate one-way within-subjects ANOVAs. The experimental condition did not significantly modulate the amplitude of the MMN at Fz or LM, $F(2, 22) = 1.34$, and $F(2, 22) = 2.093$, respectively, both *ps* > .1. However, it influenced the peak latency of the wave at Fz, $F(2, 22) = 17.75, p < .0001$ (see Figure 2a, which compares the difference waves across conditions). Post-hoc comparisons with Bonferroni's adjustment for multiple comparisons showed that the pitch deviants elicited a later MMN peak (average latency: 189 ± 27 ms) than did vowel (154 ± 23 ms) and double deviants (155 ± 21 ms), $F(2, 10) = 24.82, p < .001$.

Figure 1. Grand-average (N=12) ERPs elicited by the standard (black) and deviant (blue) stimuli recorded at Fz (panel a) and LM (panel b) for each category of deviants (Vowel, Pitch, Double), with negative up. Panel c shows the corresponding deviant-minus standard difference waverforms at Fz (red) and LM (green). Panel d depicts the average scalp topographies of MMN over a 40 ms window centered at the peak
for each condition ; and panel e shows the average scalp topographies of the P3a over a 40 ms window centered at the peak. (For the topographies, activity at C4 was computed as an average of the activity recorded at C2, C6, FC4 and CP4).

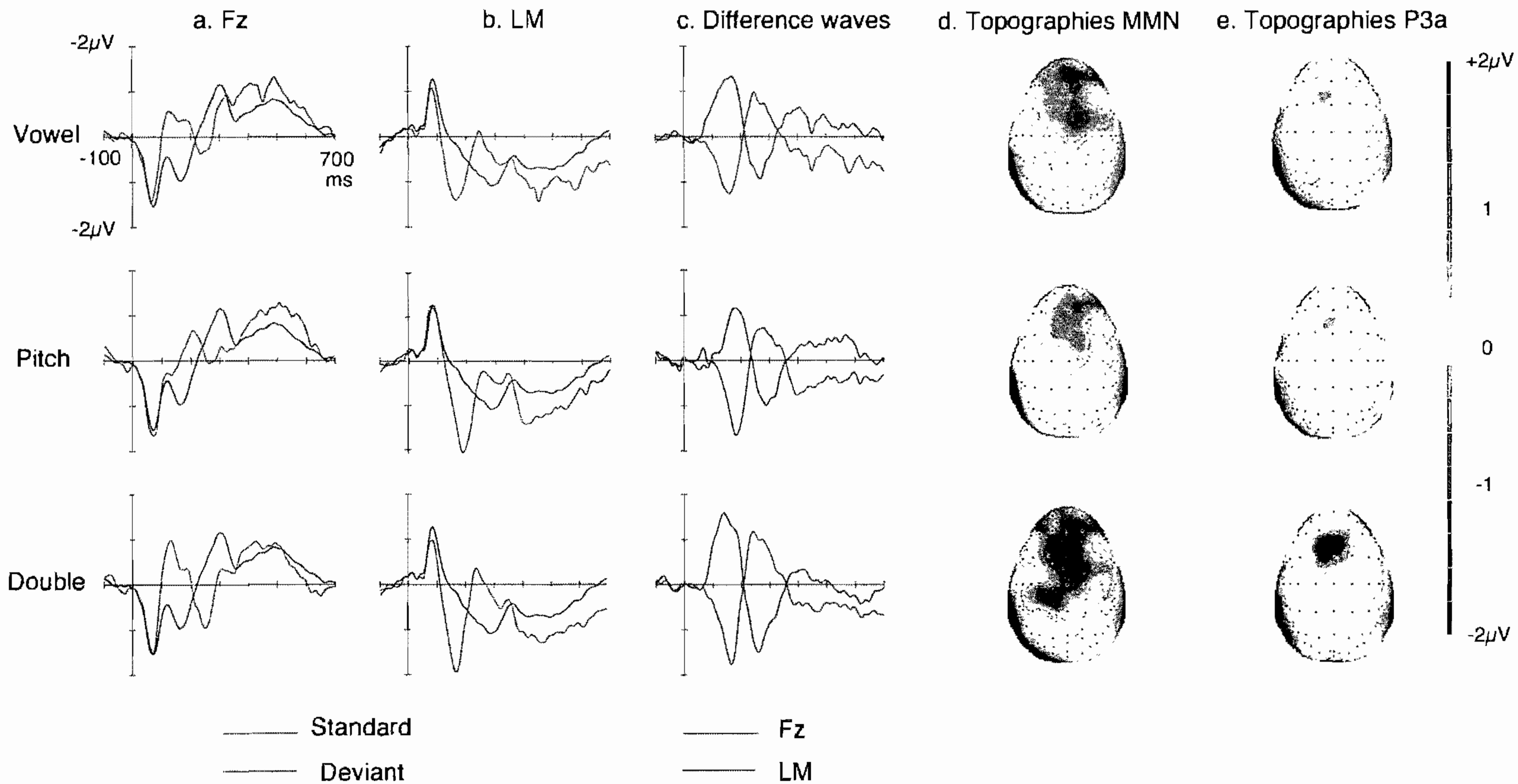
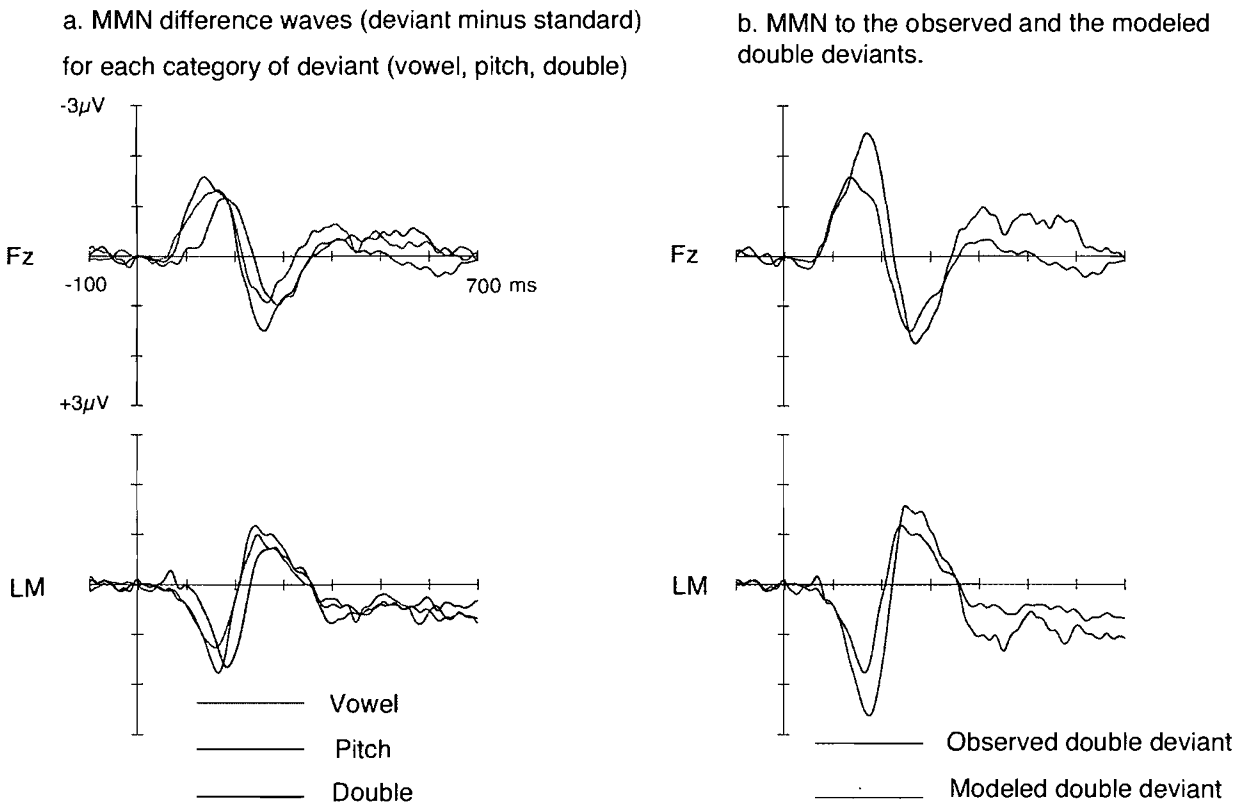


Figure 2. Panel a depicts the MMN difference waves (deviant minus standard) for each category of deviant (vowel, pitch, double) at Fz (top) and LM (bottom). Panel b shows the MMN to the observed (black) and the modeled (green) double deviant at Fz and LM.



In order to test whether the topographies of the brain responses to vowel, pitch, and double deviants differ, a two-way repeated measures analysis of variance was conducted on the MMN mean amplitudes, normalized according to McCarthy and Wood's procedure (McCarthy & Wood, 1985), with condition (3 levels) and electrode (62 levels) as within-subject variables. As expected, the analysis revealed a main effect of electrode, $F(61, 671) = 39.57, p < .0001$, but no interaction between electrode and condition, $F < 1$, providing no support for the hypothesis that the scalp distribution of the MMN might be different for vowel, pitch, and double deviants.

We also performed another test to examine whether there were effects of laterality on the MMN distributions. The scalp was divided in four sections according to laterality (left-right, midline excluded) and anteriority (from AF to C: anterior, from CP to O: posterior). The mean, normalized, MMN amplitude of the electrodes from each scalp region were thus pooled as follows: right-anterior (FP2, AF4, F8, F6, F4, F2, FT8, FC6, FC4, FC2, T8, C6, C2), left-anterior (FP1, AF3, F7,

F5, F3, F1, FT7, FC5, FC3, FC1, T7, C5, C1), right-posterior (M2, TP8, CP6, CP4, CP2, P8, P6, P4, P2, PO8, PO6, PO4, CB2, O2) and left-posterior (M1, TP7, CP5, CP3, CP1, P7, P5, P3, P1, PO7, PO5, PO3, CB1, O1). A three-way analysis of variance on these data, using laterality, anteriority and condition as within-subject variables, revealed that all MMNs were of greater amplitude for the anterior pooling, $F(1, 11) = 143.08, p < .0001$. However, there was not laterality difference, $F(1, 11) = 1.578, p < .10$ and no interaction involving the condition reached significance, all Fs < 1, converging with the previous omnibus analysis of the scalp distributions in showing that the topography of the MMN was not significantly different across the three types of deviants.

To test the additivity hypothesis, *a modeled double deviant* MMN was computed for each subject as the sum of the pitch deviant and the vowel deviant difference waves at Fz and LM. The empirical and modeled double deviant difference waves are depicted in Figure 2b. The mean amplitudes of the observed and of the modeled double deviants were measured in an 80 ms window between 120 and 200 ms post-stimulus. This window was chosen because it includes the peak latencies of both the double and the modeled double MMNs for each subject. The mean amplitude of the observed double deviant MMN was compared that of the modeled double deviant MMN by paired-samples t-tests. The observed double deviant MMN was significantly smaller than the modeled MMN at both Fz, $t(11) = 2.89, p < .02$, and LM, $t(11) = -3.58, p < .005$. In other words, the MMNs to vowel changes and to pitch changes are underadditive.

2. 3. 3. 2. 2. P3a

To confirm the presence of a P3a in each experimental condition, t-tests against zero were run on the mean 40 ms amplitude of the P3a centered at each individual's peak at Fz. The positive component was significantly different from zero for the double, $t(11) = 8.96, p < .0001$, and for the pitch deviants, $t(11) = 3.27, p < .01$, but only tended towards significance for the vowel deviant, $t(11) = 2.09, p = .06$. The effect of condition (vowel, pitch and double) on the amplitude of the P3a was computed in a repeated-measure ANOVA which revealed a significant difference, $F(2, 22) = 6.70, p < .05$. Post-hoc tests demonstrated that the P3a amplitude was larger for the double than for the vowel deviant, $p < .01$, but the difference did not reach the significance level for the double and pitch deviant comparison, $p < .09$. No difference was found for the pitch and vowel deviant comparison, either, $p > .10$. In summary, the double deviant stimuli produced a larger P3a.

2. 3. 4. Discussion

In this study, rare, deviant sung vowels were inserted in a repetitive sequence of more frequent sung vowels. The ERPs to deviants varying in vowel identity only, in pitch only or in both vowel identity and pitch were compared in order to assess the separability of the pre-attentive short term memory traces of these sound attributes, as indexed by the MMN. All three kinds of deviant stimuli elicited a reliable MMN, which was followed by a fronto-centrally distributed positive deflection than can be interpreted as a P3a (Friedman et al., 2001; Polich, 2007). The latency and amplitude of the MMNs were comparable to the ones obtained in studies with similar vowel identity contrasts (Jacobsen, 2004; Jacobsen, Schröger, & Alter, 2004; Näätänen et al., 1997; Savela et al., 2003). In the grand-averaged waves to standard stimuli, a sustained negativity was also observed. This negative deflection in the standard curve may be related to the fact that vowels, like other auditory stimuli (Picton, Woods, & Proulx, 1978), often evoke a sustained negativity that lasts until the offset of the stimulus (Eulitz et al., 1995; Hewson-Stoate, Schönwiesner, & Krumbholtz, 2006; Jacobsen et al., 2004). In the present case, this sustained negativity seems to arise around 200 ms and to decrease rapidly at 300 ms, corresponding to the offset of the stimulus. It does, however, not come back to zero and a remaining sustained negativity is present until 600 ms post-stimulus.

2. 3. 4. 1. MMN

More importantly, the amplitude of the MMN was similar for vowel, pitch, and double deviants. Since the MMN to double deviants was clearly not twice as large as the ones to single deviants, the results were underadditive. This was confirmed by the explicit test of the additive model: the observed double deviant MMN was significantly smaller than the one predicted by summing the MMNs to vowel and pitch deviants at both Fz and LM. The finding of similar patterns at the frontal and mastoid recording sites is interesting because the MMN recorded at Fz may be partially overlapped by an N2b component (Näätänen, Simpson, & Loveless, 1982) and includes contributions of both temporal and frontal generators (Alho, 1995), while the MMN polarity reversal recorded at the mastoid electrodes provides an estimate of the supratemporal component without N2b contamination. It implies that, contrasting with former reports of differential MMN additivity at frontal and mastoids sites (Paavilainen et al., 2003), both the temporally and the frontally generated MMNs to vowel deviants and pitch deviants were underadditive. These results strongly suggest that common neural networks handle the pre-attentive memory traces leading to the early response to vowel and pitch changes (Wolff & Schröger, 2001). This conclusion is further supported by the fact that the scalp distribution of the MMN did not differ for vowel, pitch, and double deviants.

At first glance, this finding contradicts Tervianemi et al.'s (1999) claim that vowel and chord changes are processed by anatomically different sources and that musical-pitch processing shows superiority over phonetic processing in right brain regions. How can we reconcile our underadditive pattern with the generator independence observed by these authors? A first possibility is that the sources of the change-detection for pitch and vowel identity are indeed separate and that the MMNs should have been additive in our study, but that some factors have prevented this pattern to occur. Such a factor might be that the MMN was already at its maximum amplitude for each single deviant, so that a more ample MMN could not be observed for the double deviants. Yet, several published data do not support this saturation hypothesis. In the present study, the amplitude of the MMN to the double deviants was -1.6 μV, which is below the amplitude of the MMN in experiments showing additivity for changes in vowel identity (Ylinen et al., 2005), with a double MMN of -2 μV. Moreover, in most studies of the MMN to vowel changes, the MMN amplitude even exceeds -2 μV (e.g. Ikeda, Hashimoto, Otomo, & Kanno, 2002; Jacobsen et al., 2004; Näätänen et al., 1997; Winkler et al., 1999), and can reach -4 μV when referenced to the mastoids (Savela et al., 2003). Additional evidence against the saturation hypothesis stems from a pilot, calibration study in which we used the same vocalic contrast as in the present experiment, but with a larger pitch difference of 9 semitones. With these stimuli, 7 pilot participants produced an average MMN of -2.9 μV for double deviants, hence showing that large MMN amplitudes can be obtained with sung vowels that were essentially the same as the stimuli used in the present study. These data were not reported here because the MMN elicited by the pitch difference was not matched in amplitude and latency

(Schröger, 1996b) to the MMN elicited by the vowel contrast, what prevented testing the additivity model.

Thus, the absence of additivity does not originate from a ceiling effect but rather reflects the pre-attentive interactive processing of vowel and pitch *when sung*. The discrepancy between Tervianemi et al.'s (1999) study and ours may thus rest on this singing aspect. Tervianemi and colleagues used chords and vowels, in other words, stimuli differing in both physical structure and cognitive domain (music for chords, speech for vowels) in separate experimental blocks. Because MMN amplitude, latency, and source are sensitive to the acoustical features of the auditory stimuli (Näätänen, 1992), the generator difference observed by Tervianemi et al. (1999) might be related to the physical difference between vowels and chords as well as to their abstract linguistic and musical nature. By contrast, our use of carefully-matched sung stimuli introduced vowel and pitch differences in the same acoustical stream, hence avoiding the potential caveat of using different stimuli to compare "music" and "language" processing. In this better controlled design, vowel identity and pitch stimulus differences appear to activate common neural generators (at least by the logic of the additive model we presented in the Introduction). This, in turn, confirms that songs are a great tool to examine the music-language interaction in relatively ecological conditions (Patel & Peretz, 1997; Schön et al., 2005). We should however note that, in songs, phonemes and pitches are intrinsically tied, the pitch being even known to modify the formant structure of the vowels (Scotto di Carlo & Germain, 1985). Although the spectral analyses run on our material show that the semitone difference did not strongly influence the formantic structure of the vowels, the slight formant shift depicted in Table 1 might have contributed to the integration of vowel and pitch processing in double deviants.

Further confirmation of the existence of shared neural generators for the detection of vowel and pitch changes in the same acoustical stream is however required to firmly conclude that common neural networks underlie vowel and pitch processing. This evidence could be provided by source analyses. However, the source difference between chords and phonemes found by Tervianemi et al. (1999) was in a range of 6 mm, which is probably below the spatial resolution and margin of error of EEG source modeling techniques (Luck, 2005). Further research using MEG might be able to corroborate the present findings by showing shared MMNm sources for the detection of the vowel and pitch deviances while also having the spatial resolution to rule out distinct, but proximal, sources.

In spite of these cautionary notes, the underadditivity of vowel and pitch deviants in our study might suggest that the interaction found by Kolinsky et al. (under review) is not merely due to attentional processes or processing strategies but corresponds to a genuine and early integration of the processing of vowel identity and pitch for sung vowels. Hence, our results confirm the idea that at least some of the building blocks of music (pitch) and language (vowels) can be processed conjointly. Music processing specificity probably occurs at more complex levels, such as tonal melodic processing (Besson et al., 1998; Bonnel, Faita, Peretz, & Besson, 2001), an idea further supported by the evidence that some patients suffering from acquired amusia perform as well as controls in single pitch judgments but exhibit deficits in tasks involving interval and melodic processing (Peretz & Kolinsky, 1993). In sum, pitch processing by itself is not specific to music, but the tonal encoding of pitch might be (Peretz, 2006). Such a conclusion is consistent with the fact that pitch, without tonal structure, is an important part of speech prosody (Patel, Peretz, Tramo, & Labrecque, 1998). The finding that pitch and vowel identity are early interactive dimensions is not trivial, given that several simple sound features such as duration and frequency (Levänen et al., 1993), duration and intensity (Wolff & Schröger, 2001) and timbre dimensions (Caclin et al., 2006) have been found to be processed independently with the same method as in the present study.

2. 3. 4. 2. P3a

The primary goal of this study was to explore the additivity of the MMNs to vowel deviants and pitch deviants, but it also brought up surprising and interesting results that deserve additional comments. All three types of deviants elicited a positive wave, which latency and scalp distribution allow interpreting as a P3a (Escera et al., 2000; Friedman et al., 2001; Polich, 2007). The P3a has been claimed to reflect attention switching towards the deviants (Friedman et al., 2001), hence suggesting that the deviants were sufficiently dissimilar from the standard to be detected, event though the participants were asked to ignore the auditory stimuli. In other words, the P3a represents an index of the attention-catching power, or saliency, of the deviants. Since this ERP was significantly more ample for double deviants than for vowel deviants (and tended to be more ample than the MMN to pitch deviants), it implies that the double deviant was a stronger attentional magnet than were the single deviants.

It is not that surprising that stimuli varying on two features elicit a stronger orientation response than stimuli varying on a single feature. However, MMN amplitude and latency are also known to be proportional to the magnitude of the stimulus change (Näätänen et al., in press), but here, the MMN amplitude was similar in the three experimental conditions and the MMN peak latency was longer for *pitch* deviants than for both vowel and double deviants. This contrasts with significantly smaller amplitude of the P3a for *vowel* deviants. This inconsistency between the MMN and the P3a patterns confirms the idea that the MMN (Näätänen, 1992; Näätänen et al., in press) and the P3a (Donchin, 1981; Squires, Squires, & Hillyards, 1975) reflect different mental processes. As already mentioned, the MMN is often described as indexing a pre-attentive sensory-memory trace of the incoming stimulus, and its neural generators can slightly differ as a function of the nature of the auditory change (Caclin et al., 2006; Giard et al., 1995; Levänen et al., 1993; Tervaniemi et al., 1999). In contrast, the P3a reflects the involuntary orientation of attention towards the deviant, or some of the deviant, stimuli (Polich, 2007; Squires et al., 1975). In our results, the similar MMNs amplitudes suggest that the three kinds of deviants might leave similar sensory traces at the pre-attentive level, but the P3a indicates that double deviant are more attention-catching than, at least, the vowel deviants.

2. 3. 4. 3. Latency differences

The slower MMN to pitch deviants relative to the other conditions still remains to be explained. An appealing hypothesis would be that this latency difference is not due to differences in dimensional saliency (if it were the case, the effect would, at least, have been in the same direction as for the P3a) but, rather, that the vowel dimension tends to be processed earlier than the pitch dimension (see Tervianemi et al., 1999, for a similar, though non significant, trend for vowels and chords). The similar latency of the MMN peak to vowel and double deviants can thus be accounted for by the fact that the fastest processed feature, here the vowel, dominates the MMN to double deviants, in a “race-like” process. The origin of the faster vowel processing could rest on the relevance of vowel information in speech understanding, while pitch information, like speaker's voice, is less relevant for verbal communication, at least in non-tonal languages such as French. This conclusion is supported by findings that language-irrelevant variations of pitch are ignored and do not interfere with the MMN for vowel (Jacobsen et al., 2004) and consonant contrasts (Aulanko, Hari, Lounasmaa, Näätänen, & Sams, 1993). This race-model is also consistent with repeated observations of processing superiority (faster or more accurate responses) for the linguistic than for the musical dimensions of songs (Kolinsky et al., under review; Peretz, Radeau, & Arguin, 2004; Serafine, Crowder, & Repp, 1984).

To conclude, the MMNs elicited by vowel and pitch changes do not produce additive effects when manipulated simultaneously in a single stimulus (double deviant), hence suggesting that the early, pre-attentive processing of these dimensions is integrated. This conclusion is corroborated by the analyses of the voltage scalp distributions of the MMNs across all three types of deviants, which provide no support for the hypothesis that the distributions were generated by different generators. Our result bring a new light on behavioral studies of pitch and vowel processing in songs and confirm that the integration between these dimensions occurs as early as the MMN latency range and may be independent of conscious attention allocation and of participants' strategies.

2. 4. Article 5

Spatial Associations for Musical Stimuli: A Piano in the Head? [v]

---

[v] D'après Lidji, P., Kolinsky, R., Lochy' A., & Morais, J. (2007). Spatial Associations for Musical Stimuli: A Piano in the Head? *Journal of Experimental Psychology: Human Perception and Performance,* 33 (5), 1189-1207.

Abstract

This study was aimed at examining whether pitch height and pitch change are mentally represented along spatial axes. A series of experiments explored, for isolated tones and two-note intervals, the occurrence of effects analogous to the SNARC effect (*Spatial Numerical Association of Response Codes*). Response device orientation (horizontal vs. vertical), task, and musical expertise of the participants were manipulated. The pitch of isolated tones triggered the automatic activation of a vertical axis independently of musical expertise, but the contour of melodic intervals did not. By contrast, automatic associations with the horizontal axis seem linked to music training for pitch and, to a lower extent, for intervals. These results, discussed in the light of studies on number representation, provide a new example of the effects of musical expertise on music cognition.

Keywords:

SMARC effect, SNARC effect, pitch, interval, music cognition

Response selection is a crucial aspect of human behavior. In choice-reaction tasks, one of the most important factors that affect response selection is how the stimuli are mapped onto responses (e.g., Fitts & Seeger, 1953; Kornblum, Hasbroucq, & Osman, 1990). The effects of such mappings on response latency or accuracy are labeled *stimulus-response compatibility effects* (*SRC*). For choice-reaction tasks in which both the stimuli and response alternatives are spatial (e.g., left- and right-side visual stimuli and left and right response keys), performance is superior when the response assigned to each stimulus matches the stimulus location than when it does not. This spatial correspondence effect occurs both when the response is based on stimulus location and when it is based on a non-spatial attribute, for example the color of the stimuli, as in the *Simon effect* (e.g., Simon, Sly, & Vilapakkam, 1981; for a review, see Lu & Proctor, 1995). Such effects are observed both in the visual and in the auditory modality (e.g., Tagliabue, Zorzi, & Umiltà, 2002; Vu, Proctor, & Urcuioli, 2003).

Stimulus-response compatibility effects also occur when the stimuli are not intrinsically spatial but considered as activating a mental spatial representation (however, see Proctor and Cho, 2006, for an alternative account). This kind of effect is usually referred to as reflecting *spatial association of response codes.* The label "response codes" may however be discussed, since the spatial compatibility effect can only occur if the same spatial code is shared by both stimulus and response coding processes. However, for sake of clarity, we keep the conventional "response codes" terminology since it is broadly used in the literature describing such effects (e.g. Dehaene, Bossini, & Giraux, 1993; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006).

The aim of the present study was to explore further whether auditory stimuli, in particular the pitch of isolated tones and the contour of melodic intervals, elicit such spatial associations. Indeed, up to now these effects have been studied almost exclusively in the visual modality (but see Nuerk, Wood, & Willmes, 2005, for similar effects in the auditory modality), and mainly in numerical cognition.

The Spatial Numerical Association of Response Codes, or *SNARC effect*, reflects the fact that, irrespective of the responding hand, participants respond faster to numbers of small magnitude on the left than on the right response side and, conversely, respond faster to numbers of large magnitude on the right than on the left response side (see Gevers and Lammertyn, 2005, and Hubbard, Piazza, Pinel, &

Dehaene, 2005, for reviews). This effect is observed in a variety of tasks and number formats, the most commonly used being parity judgment on written Arabic digits (e.g., Dehaene et al., 1993; Fias et al., 1996). The SNARC effect is generally interpreted as revealing a mental representation of number magnitude on an analogical and logarithmic mental line oriented from left to right (Dehaene, 2001). On this mental line, numerically small numbers would be placed on the left side and numerically large numbers on the right side.

Spatial associations of response codes are actually not restricted to numbers but also occur with sequences such as letters of the alphabet, months of the year, and days of the week (Gevers, Reynvoet, & Fias, 2003, 2004). This implies that mapping sets of ordered elements onto spatial positions may take place in different domains. If this were the case, musical stimuli might induce spatial associations of response codes as well.

Such a hypothesis has been recently tested by Rusconi, Kwan, Giordano, Umiltà, and Butterworth (2006). In three "SNARC-like" experiments, these authors observed the so-called *SMARC effect* (Spatial Musical Association of Response Codes). The results of a timbre judgment task showed that both musically trained and naive participants associated "up" responses with high-pitched tones and "down" responses with low-pitched tones. However, when horizontal responses were required, only musicians showed a horizontal association of faster and more accurate responses to low-pitched tones on the left and to high-pitched tones on the right. Non-musician participants had a trend toward this left-right association, but only in a pitch comparison task in which pitch processing was mandatory. According to the authors, these data suggest that "our cognitive system maps pitch onto a mental representation of space" (p. 126). The horizontal effect found in musicians could be attributed to an orthogonal SRC effect, in other words a conversion of an association with the vertical axis into one with the horizontal axis, as had already been shown in the SRC literature (e.g. Cho & Proctor, 2003). This effect could, however, also come from the influence of an external referent familiar to musicians, such as the piano keyboard. Indeed, the spatial representation of musical pitch could be influenced by the structure of the instrument played (Mikumo, 1998), like the direction of the SNARC effect seems to be modulated by writing direction (Dehaene et al., 1993).

Rusconi et al. (2006) may have prematurely called this effect "musical" because isolated tones cannot be systematically considered as music, except for exceptional individuals (those who possess absolute pitch, e.g. Takeuchi & Hulse, 1993) or in special circumstances (e.g. tuning an instrument). Therefore, it is important to examine whether the "musical mental line" is activated by more complex musical material like melodic intervals. Indeed, music emerges from the dynamic organization of pitches, defining contour, intervals, scales, consonance or dissonance (e.g. Dowling & Harwood, 1986), in a rhythmic framework.

In the present study, which we conducted without awareness of Rusconi et al. (2006), we compared the spatial representations of both the pitch of isolated tones (Experiments 1, 2, and 3) and the contour and pitch of two-note melodic intervals (Experiments 4, 5, 6). We examined the mapping of these stimuli both horizontally (with left and right responses; Experiments 1, 2, 4 and 5) and vertically (with bottom and top responses; Experiments 3 and 6) in cognitive space. Indeed, as suggested by the results of Rusconi et al. (2006), both horizontal and vertical associations of pitch could coexist. The direction of the "mental pitch line" is difficult to predict since, in addition to linguistic congruency effects (the « low » vs. « high » words for low- vs. high-pitched tones, respectively), two different spatial external referents are possible for music, namely the horizontal mapping similar to a piano keyboard, which is observed for numerical spatial associations (e.g. Gevers & Lammertyn, 2005, Fias & Fischer, 2005), and the vertical mapping which is adopted, for example, in musical notation and on some instruments like the recorder.

Two further aims of the present study were to specify on the one hand the generality, and on the other hand the automaticity, of the spatial associations for both tones and intervals. First, we studied musically untrained and trained participants (henceforth, *non-musicians* and *musicians*, respectively). This allowed us to examine the effect of musical expertise on the association between pitch or interval contour and space. Second, we manipulated the nature of the task, with indirect vs. direct responding to the target dimension. Indeed, both for numbers and ordered linguistic sequences such as the name of months (Gevers et al., 2003), the association between items from the beginning vs. end of the sequence with the left vs. right side also occurs when the task does not require the explicit processing of magnitude and/or ordinal position. As an example, such effects occur in tasks like discrimination of the orientation of a triangle or a line superimposed on a digit (Fias, Lauwereyns, & Lammertyn, 2001), color judgments (Keuz & Schwarz, 2005), phoneme-monitoring on visually presented Arabic digits (Fias, 2001; Fias et al., 1996), or letter detection in the name of a month (Gevers et al., 2003). This seems to support the idea of an automatic activation of the spatial representations (see however Fias, 2001, for other evidence). In the present case, if the pitch of tones and the contour of intervals induce automatic activations of a spatial representation, then spatial associations of response codes should take place even in tasks in which these dimensions are irrelevant. To test this idea, we used an instrumental timbre judgment task (piano vs. violin ; Experiments 1, 3, 4 and 6), and compared it to tasks in which the processing of the target[i] dimension (either pitch height or the contour of musical intervals) is mandatory (Experiments 2 and 5, respectively).

2. 4. 1. Experiment 1: Instrumental Timbre Judgment on Isolated Tones with Left-right Responses in Non-musicians and Musicians

The purpose of the first experiment was to assess the occurrence of an automatic left-to-right spatial association for pitch height, or *horizontal SPARC* (Spatial Pitch Association of Response Codes) effect in non-musicians and musicians. We chose the label *SPARC* instead of *SMARC* (Rusconi et al., 2006), because the effects observed on the pitch of isolated tones cannot be a priori generalized to all musical dimensions.

We used an instrumental timbre (piano vs. violin) judgment task. The target dimension, pitch height, was thus irrelevant to the task, just as number magnitude is irrelevant in the parity judgment tasks in which a SNARC effect is observed (e.g., Dehaene et al., 1993; Fias et al., 1996).

We predicted an association between low-pitched tones and left-side responses and between high-pitched tones and right-side responses in both groups. Indeed, in a serial reaction time task, Stöcker, Sebald, and Hoffmann (2003) showed that non-musicians learned a sequence of key presses better and had shorter response times (RTs) when the tones were mapped onto the response keys in an ascending order from left to right than in the opposite mapping (descending from left to right) or in a mixed, random mapping.

However, we expected a stronger SPARC effect in musicians because musical training could increase the spatial associations. Indeed, in the same way as the SNARC effect is sometimes described as depending on writing direction (with a reversed, right-to-left, SNARC effect reported in a group of Iranians by Dehaene et al., 1993, Experiment 7), the left-right effect for pitches may be reinforced by most musicians' familiarity with the piano keyboard and the left-side, low tone vs. right-side, high tone mapping of this instrument. Evidence supporting this idea comes from studies on expert and beginning pianists. Using a Stroop-like design (Stroop, 1935) with musical notation, Stewart, Walsh, and Frith (2004) asked participants to ignore the note written on the staff and to attend to a number indicating the finger to use for answering. Pianists were unable to ignore the note, which caused interference: the notes written on the low part of the staff (corresponding to low pitches) were associated with the left side of the response apparatus and the notes written on the high part of the staff (corresponding to high pitches) were associated with the right side. Non-musicians who were illiterate in musical notation did not show such interference. Thus, while playing, pianists transform the vertical mapping of musical notation into the horizontal mapping of the piano keyboard. Magnetoencephalography data (Haueisen & Knösche, 2001) further suggest that piano practice promotes an association between heard pitches and hand movements. Compared to non-pianists, pianists showed an increased activation of the motor cortex while listening to piano pieces. Moreover, naive non-musician participants who learned to read and play music during 15 weeks presented changes in the bilateral superior parietal cortex (Stewart, Henson, Kampe, Walsh, Turner, & Frith, 2003). All these results suggest that music reading in pianists involves an automatic association between the spatial features of musical notation and a (spatial) sequence of key presses. Thus, musical expertise may favor the use of spatial representations for music through learned associations between left-right space and pitch height. This hypothesis is tested here by calculating the correlation between the amount of piano training of the musicians and the regression coefficient of pitch.

2. 4. 1. 1. Method

2. 4. 1. 1. 1. Participants.

In the present experiment as well as in the following ones, the inclusion criterion for non-musicians was to have no musical experience or a maximum of 3 years of musical training interrupted for at least 7 years. For the musicians, the inclusion criterion was to have been playing an instrument or singing at a high level of proficiency (as professional, music student or high level amateur) for a minimum of 8 years.

Sixteen non-musicians and 16 musicians took part in Experiment 1. The non-musicians (3 males, average age: 21.6 yrs, all right-handed) were undergraduate psychology students taking part in the experiment as a course requirement. Six of them had no prior musical experience, while 10 had had an average of 1.9 years of music training but had stopped since 9.6 years, on the average. The musicians (4 males; average age: 24.3 yrs; 11 right-handed) were 3 professional musicians, 6 undergraduate music students and 7 high level amateur musicians. Eleven of them were pianists, 2 played the violin, 2 were singers, and one was a flutist. They had, on the average, 13.2 years of musical training and practiced their instrument around 15.3 hours a week. They were paid for their participation.

2. 4. 1. 1. 2. Material and procedure.

The stimuli consisted of two low-pitched tones (C3 and G3, respectively 131 and 196 Hz) and two high-pitched tones (E5 and B5, respectively 659 and 988 Hz). These pitches were chosen in order to get the same semitone distance between the two low and the two high tones (the interval of a fifth: 7 semitones).

Each tone was synthesized with piano and violin timbre with the Cakewalk software, for a total of eight different stimuli. Stimuli were then transferred to a Macintosh G3 computer for testing. Stimulus presentation and response recording with RTs measured from the beginning of the stimulus were controlled by the *Psyscope* 1.2.5 PPC software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Participants were tested individually in a quiet room. Each trial consisted of a silent interval of 400 ms preceding the presentation of the auditory stimulus (an isolated 500 ms tone), followed by a silence of 600 ms, so that the participant had a total of 1100 ms to respond, including the stimulus duration. The total inter-stimulus interval was thus 1000 ms. On each trial, participants were asked to judge if the tone was played by a violin or by a piano by pressing one of the two extreme left and right keys of a *Psyscope* button box (separated by 10.3 cm) with their left or right hand, respectively. A black fixation cross followed immediately the participant's response. The instructions were presented simultaneously on the screen and orally by the experimenter, and emphasized both speed and accuracy.

Participants took part in two experimental blocks of 96 trials each (12 presentations of each stimulus), one with the piano response assigned to the left side and the violin response assigned to the right side, and one with the reversed assignment. The order of blocks was counterbalanced across participants. There was a short resting period between the two blocks, during which participants had to fill in a questionnaire about their musical background.

Each block was preceded by examples of piano and violin stimuli and by a training block of 16 trials (2 presentations of each stimulus), during which visual feedback was provided (correct vs. error message on the screen).

Within each block, trials were presented in a pseudorandom order so that consecutive presentations of the same stimulus (same tone played by the same instrument) were not allowed and more than three consecutive responses with the same hand never occurred.

The whole experiment lasted about 20 minutes.

2. 4. 1. 2. Data Analysis

Because we were specifically interested in the association between pitch and response side, all statistical analyses were performed on the differences in median RTs (*dRT*) and error rates (*dError*) between right and left responses to each stimulus (each tone in the present case). The choice of this dependent variable facilitates the interpretation of the results because the spatial association effect is represented by a main effect rather than by the interaction between response side and pitch. According to our predictions, the right minus left difference should be positive for low pitches (faster left responses and fewer errors on the left) and negative for high pitches (faster right responses and fewer errors on the right).

Two analyses were run on the data: a repeated measures analysis of variance and a regression analysis. The ANOVA allows the description of interactions between variables. In order to better capture the structure of the data, the pitch height variable was here divided into two sub-variables. Indeed, two low and two high tones were used, creating two pitch categories. In the following analyses, these low and high pitch categories are labeled "global pitch". The SPARC effect will thus be reflected by a "global pitch effect", i.e. facilitation on the left for globally low tones and facilitation on the right for high tones. However, in each global pitch category, one tone was relatively lower than the other. The relative pitch within a category is coded as a two-level variable, "local pitch".

Contrary to ANOVAs, the regression analyses evaluate the linear relation between the target variable and the difference of reaction time in each participant, hence avoiding a potential overestimation of the effect due to group averaging. These regression analyses suggested by Lorch and Meyers (1990) and first used by Fias et al. (1996) in SNARC studies are popular in this domain, either as main analysis (e.g. Fias et al., 2001, Ito and Hatta, 2004) or in addition to analyses of variance (e.g. Fias, 2001, Gevers et al., 2003, Keus & Schwarz, 2005, Keus, Jenk, & Schwarz, 2005, Nuerk, Wood, & Willmes, 2005). In the present study, these regression analyses were performed on the differences in median RTs (*dRTs*) and error rates (*dErrors*) between right and left responses to each pitch, for each individual participant. "Global pitch" and "local pitch" were thus not distinguished in this analysis. Both pitch and instrumental timbre were predictor variables. The pitch predictor value was the logarithm of the fundamental frequency (F_0) of each pitch[ii]. Timbre was coded as -0.5 for the piano and +0.5 for the violin. We then tested whether the individual regression weights for pitch and timbre deviated significantly from zero across participants, in each participants group. A negative regression slope would reflect a horizontal SPARC effect.

2. 4. 1. 3. Results and Discussion

The error rate averaged over the 32 participants was 6.7% (non-musicians: 7.5%; musicians: 5.9%, $p > .10$). There was no speed-accuracy trade-off, as indicated by the positive correlation, $r = .69$, $p < .005$, between RTs and error rate. Table 1 displays the averaged median reaction times on each response side as a function of pitch, timbre and group.

Table 1. Average Median RTs (in ms) in Experiment 1 for each Pitch, Instrumental Timbre, Group and Response Side.

		Low				High			
		C3		G3		E5		B5	
		piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians									
	Left	554	622	572	615	581	570	510	556
	Right	558	608	557	638	557	571	538	537
	dRT (right - left)	4	-14	-15	23	-24	1	28	-19
Musicians									
	Left	527	609	550	589	581	563	543	556
	Right	562	638	530	616	526	576	534	536
	dRT (right - left)	35	29	-20	27	-55	13	-9	-20
	Mean dRT	19	8	-17	25	-40	7	10	-19

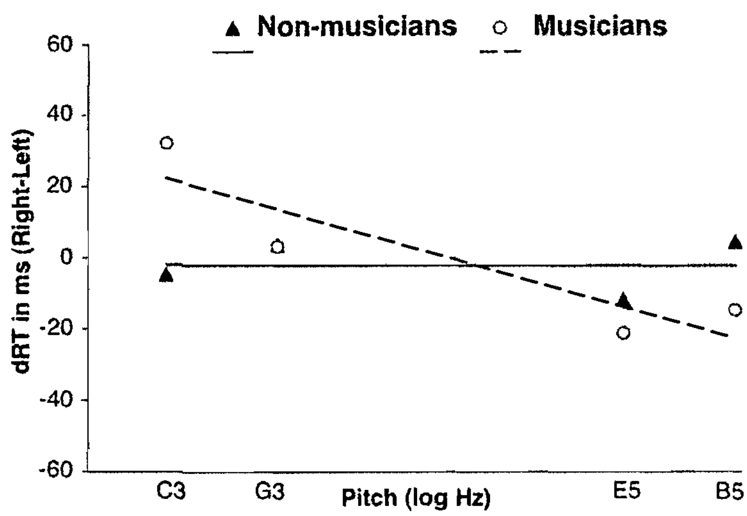
2. 4. 1. 3. 1. RT analyses.

The differences between median right and left reaction times (dRTs) for correct responses (see Table 1) were analyzed in a 2 X 2 X 2 X 2 analysis of variance. Global pitch (low or high pitch category), local pitch (relative pitch within one category) and instrument timbre (piano or violin) were included as within-subject variables. The participant group (non-musicians or musicians) was a between-subjects variable. The effect of global pitch was significant, $F(1, 30) = 7.63$, $p < .01$. As displayed in Table 1, the participants had shorter RTs on the left than on the right to low tones, and on the right than on the left to high tones, suggesting the occurrence of a SPARC effect. The effects of local pitch, of instrument and of group did not reach significance, all Fs < 1, $p > .10$. The significant interaction between group and global pitch indicated the presence of different patterns of results in musicians and non-musicians, $F(1, 30) = 5.27$, $p < .05$: while non-musicians did not display the SPARC effect, $F(1, 30) < 1$, $p > .10$, musicians did, $F(1, 30) = 19.07$, $p < .001$. Global pitch interacted significantly with local pitch and instrument timbre, $F(1, 30) = 20.32$, $p < .001$. As displayed in Table 1, when the tones were played on the violin, a SPARC effect was observed, but the dRTs was smaller for the lowest tone of a global pitch category than for the highest tone. For the piano tones, there was no evidence for a SPARC effect.

The results of the regression analyses[iii] on the non-musicians' dRT with the logarithm of each of the four pitches frequency as predictor showed that neither the regression coefficient of pitch, nor the one of timbre differed significantly from zero,

$t(15) = -.02, p > .10$; t $(15) = -.02, p > .10$, respectively. Thus, consistent with the results of the ANOVA and as shown in Figure 1, there was no evidence in favor of a horizontal SPARC effect in non-musicians. Conversely, as illustrated in Figure 1, the occurrence of a significant overall horizontal SPARC effect in musicians was confirmed by the fact that pitch had a negative regression coefficient that differed significantly from zero, $t(15) = -4.57, p < .001$, while timbre did not explain significantly the slope of the regression line, $t(15) = .78, p >.10$. Finally, the impact of piano training on the SPARC effect in musicians was explored through correlation analyses. The correlation between the number of years of piano instruction and the regression weight of pitch for each musician participant was calculated. If playing the piano tended to increase the SPARC effect, the amount of piano training would be negatively correlated with the regression coefficient. This was not the case, $r = -.02, p > .10$.

Figure 1. Observed data and regression line representing RT differences between right- and left-side responses as a function of pitch in the instrument timbre judgment task in Experiment 1. The full triangles and the full regression line represent the non-musicians' mean dRTs and corresponding regression line. The white circles and the dotted regression line correspond to the musicians' data.



2. 4. 1. 3. 2. Error analyses.

An ANOVA with the variables described above (global pitch, local pitch, instrument timbre and group) was performed on the differences between the percentage of errors committed with the right and left response keys. The effects of global pitch, local pitch and timbre were significant, $F(1, 30) = 4.68, p < .05$; $F(1,$

30) = 9.02, $p < .005$; $F(1, 30) = 5.51$, $p < .05$, respectively. The effect of timbre reflected an association between, on the one hand, piano and the right key and, on the other hand, violin and the left key. The effect of global pitch was consistent with the hypothesis of a SPARC effect: participants responded more accurately to low tones on the left and to high tones on the right. The direction of the local pitch effect was, however, unexpected since the low member of a tone group yielded fewer errors on the right than on the left and conversely for the high member. Both the interaction between local pitch and timbre, $F(1, 30) = 6.32$, $p < .025$, and the interaction between global pitch, local pitch and timbre, $F(1, 30) = 6.62$, $p < .025$, were significant. The interaction between local pitch and timbre was due to the effect of local pitch only occurring for the violin timbre; and the second order interaction reflected a more consistent SPARC effect for the piano than for the violin timbre. No interaction including the participants group reached significance.

Despite the observation of a SPARC effect in the ANOVA, the regression analysis on dErrors did not reach significance in either group. Indeed, the regression weights of pitch and of timbre did not differ significantly from zero in either non-musicians (pitch: $t(15) = -.75$, $p > .10$; timbre: $t(15) = -.99$, $p > .10$), or musicians (pitch: $t(15) = -1.72$; $p > .10$; timbre: $t(15) = -1.89$; $p < .08$).

According to the results of the ANOVAs and the regression analyses on dRTs and dErrors, it seems that non-musicians do not associate low pitches to left responses and high pitches to right responses. In contrast, musicians exhibited a significant SPARC effect on RTs both in the ANOVA and in the regression analysis.

These results suggest that the association between low-pitched tones and left responses and between high-pitched tones and right responses is specific to musicians. The presence of this horizontal SPARC effect in musicians is consistent with the results of Rusconi et al. (2006). Because piano training was not significantly correlated with the effect, the results may not be explained only by motor associations due to this external referent in musicians, but would rather seem to reflect the activation of an abstract mental representation.

Instead of being caused by the knowledge of the keyboard, the difference between groups varying in musical expertise observed in Experiment 1 and in Rusconi et al. (2006) may have been due to musicians and non-musicians differing in their ability to automatically activate a mental representation of pitch when processing timbre. Indeed, it remains possible that non-musicians associate pitch to a left-right axis, but that pitch and/or its spatial representation are not addressed automatically in this population when pitch processing is not mandatory, as was the case in the timbre judgment. Experiment 2 was aimed at examining whether non-musicians do explicitly associate pitch with a left-right space.

2. 4. 2. Experiment 2: Pitch Comparison with Left-right Responses in Non-musicians

To verify whether there is an explicit, though not implicit, left-to-right mental representation of pitch in non-musicians, the present experiment required them to process pitch intentionally. Non-musicians were asked to judge if isolated tones were higher or lower than a referent tone in a musical adaptation of Dehaene, Dupoux and Mehler's (1990) number comparison task. If musically naive participants associate pitch and response side when the task requires pitch processing, they should display a horizontal SPARC effect.

2. 4. 2. 1. Method

2. 4. 2. 1. 1. Participants.

Sixteen fresh non-musicians (1 male; average age: 18.6 yrs; 14 right-handed) participated in the experiment for psychology course credit. Eight of them had never learnt or played music, the remaining ones having on average one year of music training interrupted for 7.9 years. None of them had experience playing the piano.

2. 4. 2. 1. 2. Material and procedure.

Stimuli were the same as in Experiment 1. The procedure was also the same, except for the following points. On each trial, participants were asked to judge if the presented tone was higher or lower than a referent tone, F4 (349 Hz), located midway (in semitones) between the low-pitched (C3 and G3) and high-pitched (E5 and B5) tones. This referent tone was presented three times during the instructions and was not heard again during the experiment. The task was nevertheless easy because of the large distance between low- and high- pitched tones. The participants indicated their response by pressing a left or right key as fast and accurately as possible. Stimuli were played in a violin timbre for half of the participants, and in a piano timbre for the other half. All participants took part in two experimental blocks of 96 trials each, which included 24 presentations of each stimulus. The assignment of the "lower" and "higher" response to the left or right response key was reversed between the two blocks, the order of blocks being counterbalanced across participants. Within each block, trials were presented pseudo-randomly so that the same stimulus did not appear more than three consecutive times.

2. 4. 2. 2. Results and Discussion

The average error rate was 1%. There was no trade-off between RTs and error rate, r = .28, p > .10. Average median response times for each pitch, response side and timbre are shown in Table 2, timbre being a between-subjects variable in the present design.

Table 2. Average Median RTs (in ms) in Experiment 2 for each Pitch, Instrumental Timbre and Response Side.

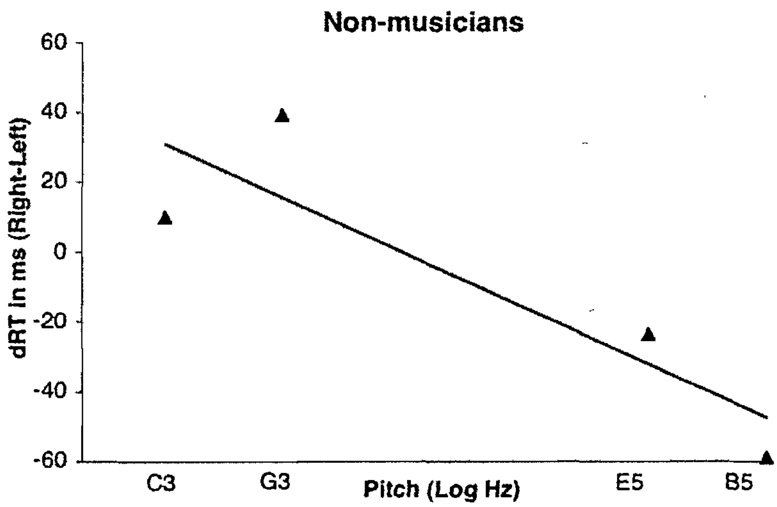
		Low				High			
		C3		G3		E5		B5	
		piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians	Left	390	426	388	465	408	422	411	447
	Right	400	437	423	508	373	411	354	487
	dRT (right - left)	10	11	35	43	-35	-11	-57	-60

2. 4. 2. 2. 1. RT analyses.

The observation of the regression slope in Figure 2 suggests the occurrence of a SPARC effect. The 2 × 2 × 2 analysis of variance run on the difference in median RTs between correct right and left responses with global and local pitch as within-subject variable and the instrument timbre (piano vs. violin) as between-subjects variable confirms this intuition. A significant main effect of global pitch was observed, $F(1, 14) = 14.38$, $p < .005$, with faster left responses to low-pitched tones and faster right responses to high-pitched tones (see Table 2). There was no significant effect of local pitch, $F(1, 14) < 1$, $p > .10$, but the significant interaction between global and local pitch, $F(1, 14) = 27.22$, $p < .001$, reflected the fact that, unexpectedly, the lowest of the low tones was less associated to the left key than the highest one. On the contrary, for the high tones, the relatively higher tone was, as expected, more associated to the right key than the lowest one. No other effect or interaction reached significance.

The results of the regression analysis confirmed those of the ANOVA: the regression coefficient of pitch height was negative and significantly different from zero, $t(15) = -4.03, p < .001$.

Figure 2: Observed data and regression line representing RT differences between right- and left-side responses as a function of pitch in the pitch judgment task presented in Experiment 2 to non-musicians.



2. 4. 2. 2. 2. Error analyses.

The ANOVA on the difference in the percentage of errors between right and left responses has only shown a significant interaction between local pitch and timbre, $F(1, 14) = 6.39$, $p < .025$: for the piano timbre, the lowest tone of each subgroup was associated to the left key and the highest to the right key, while the reverse happened with the violin timbre.

In line with the ANOVA, the regression analysis on dErrors failed to demonstrate a significant SPARC effect, $t(15) = -.26, p > .10$.

2. 4. 2. 2. 3. Comparison of Experiment 1 and Experiment 2 for non-musicians.

Contrary to what was observed in a timbre judgment task (Experiment 1), non-musicians showed a SPARC effect in a pitch judgment task (Experiment 2), at least on dRTs. To verify the influence of task, an additional ANOVA was run on non-musicians' dRTs with global and local pitch as within-subject variable, and task (timbre judgment, Experiment 1, or pitch judgment, Experiment 2) as between-subjects variable. It revealed a significant effect of global pitch, $F(1, 30) = 12.01, p < .005$, and, as expected, an interaction between global pitch and task, $F(1, 30) = 9.97$, $p < .005$, reflecting the fact that the SPARC effect only occurred in the pitch judgment task of Experiment 2. The interaction between global pitch, local pitch and task was also significant, $F(1, 30) = 6.07, p < .025$. It was caused by the fact that, in Experiment 1, there was an effect of local pitch without global SPARC effect, while in Experiment 2 there was a SPARC effect unmodulated by local pitch.

In summary, according to the RTs results of Experiment 2, it seems that when tones are judged as low or high, non-musicians, like musicians, associate pitch with lateralized responses. But, contrary to musicians, they do so only when explicitly asked to process pitch, as suggested by the interaction of global pitch and task (Experiment 1 vs. 2). With a similar pitch judgment task, Rusconi et al. (2006) only observed a trend toward this effect in non-musicians. The lack of significance in their study could be due to the difficulty of the task, as suggested by the high error rate they obtained. In our experiment, more distant low and high pitches were used, reducing task difficulty.

Since the non-musicians examined here never played the piano, this horizontal SPARC effect is unlikely to reflect overlearned motor patterns. Yet, the effect may have been caused by the theoretical knowledge of the keyboard structure: participants could have used the piano keyboard as a referent, thinking how a piano is arranged, hence causing the observed effect. More convincingly, the effect could be driven by the participants' knowledge of the usual order of singing or playing the notes of the musical scale, starting from low-pitched tones and progressing to high-pitched tones. Indeed, it is known that cultural biases like writing direction (Bertelson, 1972, Dehaene et al. 1993) and/or the western people's tendency to explore the world from left to right (Maass & Russo, 2003) may affect the spatial associations. Thus, musically untrained participants may tend to associate low-pitched tones with the beginning of a non-verbal scale-like sequence, and high-pitched tones with the end of it, and hence show a horizontal SPARC effect.

Whatever the correct interpretation of the non-musicians' results, the joint observation of a horizontal SPARC effect in the present experiment and of its absence in a timbre judgment task (Experiment 1) suggests that, in this population, pitch does not automatically evoke horizontal associations. Non-musicians might nevertheless automatically activate a representation of pitch in a different spatial direction, in particular along a vertically oriented axis. The next experiment was aimed at testing this hypothesis.

2. 4. 3. Experiment 3: Instrumental Timbre Judgment on Isolated Tones with Bottom-top Responses in Non-musicians and Musicians

Several arguments support the assumption that a vertical axis would constitute a more "intuitive" spatial representation of pitch than the horizontal one. First, in many languages, including French (our participants' language), a vertical mapping of pitch is linguistically congruous with the adjectives *high* and *low*, which are often used to refer to the pitch of auditory stimuli. Linguistic variables are known to affect spatial associations, at least in number processing, as illustrated by the *MARC* effect (linguistic markedness of response codes; Nuerk, Iversen, & Willmes, 2004), a presumably linguistically-based association of lexical entries sharing the same markedness with spatial locations. In this view, congruent spatial associations between linguistically marked or unmarked lexical items and response codes (odd-left and even-right) lead to faster responses than the reverse incongruent associations. One can thus expect the congruency between the vertical spatial position of the response key and pitch height to induce facilitation, in comparison to an incongruent assignment. Such congruency effects have already been reported in multidimensional speeded classification tasks involving a linguistic and a perceptual dimension, such as the spoken words "low" and "high" and their fundamental frequency (Melara & Marks, 1990; Ben-Artzi & Marks, 1999).

It is far from clear, however, whether these linguistically-mediated congruency effects occur when the situation does not require pitch labeling or spatial height judgment, like the timbre judgment task used in the present study. In addition, not all perceptual interactions can be reduced to linguistic congruency effects. For example, high-pitched tones have been shown to be located higher in space than low-pitched tones, even when the source location is constant and when measured in young children unfamiliar with the expressions "low pitch" and "high pitch" (Roffler & Butler, 1968). In other words, as suggested by Ben-Artzi and Marks (1999), the "lowness" and "highness" of pitches may not simply be an accidental linkage of labels but rather reflect a perceptual commonality of pitch and vertical space.

Whatever the value of the linguistic congruency account, other evidence supports the idea of an "intuitive" vertical mental line for music. Several models aimed at describing the mental representation of pitch have implicitly assumed that low pitches are placed low and high pitches are placed high on a representational axis (Shepard, 1982a, b; Ueda & Ohgushi, 1987). It is worth noting that this low-to-high representation of pitch is also used in music notation, low notes being written lower on the staff than higher notes.

Moreover, Rusconi et al.'s results (2006) suggest that the vertical association of pitch and space is more natural than the horizontal one. First, they found a vertical association of pitch and space in both non-musicians and musicians in a timbre judgment task, although, in non-musicians, the effect was only significant for the extreme pitches. Second, in a pitch comparison task, these authors observed a more consistent association with the vertical than with the horizontal key mapping in non-musicians.

In the following experiment, we used the timbre judgment task on isolated tones with bottom and top response keys instead of the usual left-right assignment, in both non-musicians and musicians. If there were indeed a *vertical SPARC effect*, low-pitched tones would be preferentially responded to with the bottom key and high-pitched tones with the top key, independently of the response hand assigned to each key.

2. 4. 3. 1. Method

2. 4. 3. 1. 1. Participants.

A fresh group of 32 participants, including 16 non-musicians and 16 musicians, took part in the experiment against course credit or money. The non-musicians were university students and university workers (2 males; average age: 23.6 yrs; 14 right handed). Nine of them had no prior musical training and 7 of them had 2 years of musical education, interrupted for 10 years, on the average. The musicians (9 males; average age: 23.2 yrs; 15 right-handed) were 3 professional musicians, 2 undergraduate music students and 11 high level amateur musicians. Nine were pianists, 4 were guitarists, one was a saxophonist, one a drummer and one a singer. They had played music for 12.9 years on the average and practiced 10 hours a week.

2. 4. 3. 1. 2. Material and procedure.

The tones were the same as in the previous experiments. Participants had to judge if they were played on a piano or on a violin by pressing a bottom or top response key. The procedure was the same as in Experiment 1, except that the button box was turned by 90° so that the response keys were placed vertically on the transverse plane[iv]. Participants responded with the index fingers of their two hands. In order to avoid a systematic association between hand and spatial position, half of the participants used the right hand to press the top key and the left hand to press the bottom key, and the other half responded with the reverse hand assignment (left hand top and right hand bottom).

2. 4. 3. 2. Results and Discussion

The average error rate was 6% (non musicians: 7%; musicians: 5%, $p > .10$). The significant positive correlation between RTs and error rate suggests there was no speed accuracy trade-off, $r = .737$; $p < .001$. The averaged median RTs for each pitch, timbre, response side, hand position and group are displayed in the Table 3.

Table 3. Average Median RTs (in ms) in Experiment 3 for each Pitch, Instrumental Timbre, Hand Position, Group and Response Side.

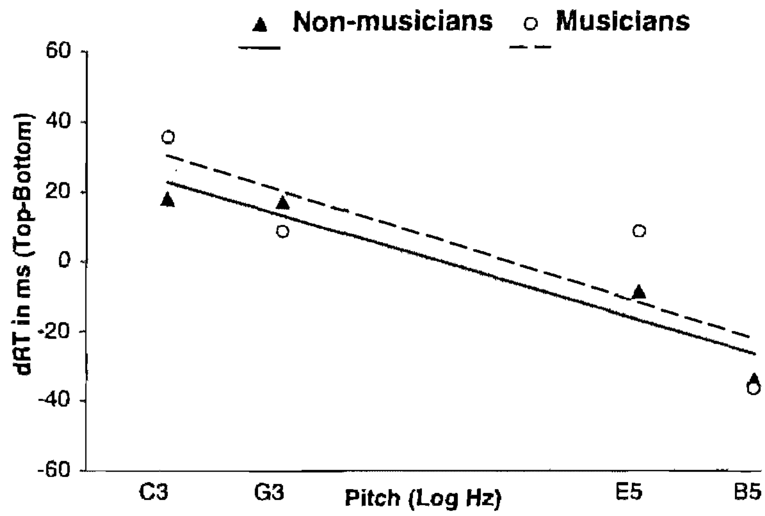
		Low				High			
		C3		G3		E5		B5	
		piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians	Bottom	558	615	563	620	551	591	571	527
	Top	566	653	556	648	564	580	546	521
Left hand top	dRT (top-bottom)	8	38	-7	28	13	-11	-25	-6
Right hand top	Bottom	541	602	490	605	535	584	533	570
	Top	539	630	550	593	556	526	468	532
	dRT (top-bottom)	-2	28	60	-12	21	-58	-65	-38
Mean dRT Non musicians		3	33	26	8	17	-35	-45	-22
Musicians	Bottom	440	536	450	486	454	487	463	437
	Top	477	542	453	508	494	495	484	434
Left hand top	dRT (top-bottom)	37	6	3	22	40	8	21	-3
	Bottom	446	490	450	523	486	476	513	460
Right hand top	Top	491	545	483	501	460	489	400	411
	dRT (top-bottom)	45	55	33	-22	-26	13	-113	-49
Mean dRT Musicians		41	31	18	0	7	11	-46	-26
Mean dRT		22	32	22	4	12	-12	-46	-24

2. 4. 3. 2. 1. RT analyses.

A 2 X 2 X 2 X 2 X 2 ANOVA was run with global pitch, local pitch and instrument timbre as within-subject variables, and musicianship and hand position (left or right hand on the top key) as between-subjects variables. The significant effect of global pitch indicated a SPARC effect, $F(1, 28) = 15.071$, $p < .001$. Participants responded faster to the lower pitches with the bottom key and to the higher pitches with the top key (see Table 3). The effect of local pitch was also significant, $F(1, 28) = 12.580$, $p < .001$: within a pitch group, participants tended to exhibit a stronger SPARC effect for the extreme exemplar. Interestingly, global pitch interacted with hand position, $F(1, 28) = 6.75$, $p < .025$. Indeed, the SPARC effect was more robust for participants having the right hand on the top key than for the ones having the left hand on the top key. Although the interaction between global pitch, hand position and group was not significant, $F(1, 28) = 1.15, p > .10$, a careful observation of Table 3 suggests that the influence of hand position on the SPARC effect was larger in musicians than in non-musicians. Indeed, in musicians, the difference of SPARC effect between participants having the right hand on the top key and the ones having the left hand on the top key was 67 ms, while it was 29 ms in non-musicians. This observation, added to the fact that the musicians' dRTs suggest the absence of a SPARC effect when the left hand was on the top key (mean dRT for low tones: 17 ms, mean dRT for high tones: 16 ms) but that non-musicians tended to show a SPARC effect even with this hand position (mean dRT for low tones: 17 ms, mean dRT for high tones: -7 ms), encouraged us to look at the interaction between hand position and the SPARC effect separately in each participants' group. The ANOVA on non-musicians showed a vertical SPARC effect, as reflected by the significant effect of global pitch, $F(1, 14) = 7.13, p < .025$. This SPARC effect did not interact with hand position, $F(1, 14) = 1.03, p > .10$. On the contrary, the musicians' SPARC effect, $F(1, 14) = 8.09, p < .025$, was modulated by hand position, $F(1, 14) = 7.80, p < .025$. As suggested by Table 3, the SPARC effect occurred only for musicians having the right hand on the top response key.

The regression analyses on the difference in median RTs between top and bottom response included as predictors pitch height (logarithmic function of F_0) and timbre (recoded as -0.5 and +0.5 for piano and violin, respectively). As shown in Figure 3, the significant contribution of pitch to the pattern of RT differences confirms the vertical SPARC effect in both non-musicians and musicians, $t(15) = -2.95, p < .01$ and $t(15) = -2.58, p < .025$, respectively. The regression weight of timbre did not differ significantly from zero in either group (non-musicians: $t(15) = -.09, p > .10$; musicians: $(15) = -.06, p > .10$). The regression coefficient for each musician was not significantly correlated with the amount of piano training, $r = -.07, p > .10$.

Figure 3. Observed data and regression line representing RT differences between top and bottom responses as a function of pitch in the instrument timbre judgment task used in Experiment 3. The full triangles and the full regression line represent the non-musicians' data. The white circles and the dotted regression line correspond to the musicians' data.



2. 4. 3. 2. 2. Error analyses.

The ANOVA on dErrors also showed a significant SPARC effect on global pitch, $F(1, 28) = 7.07, p < .025$. This vertical SPARC effect was however modulated by the group, $F(1, 28) = 6.05, p < .025$: only musicians exhibited it. The effect of timbre was also significant, $F(1, 28) = 15.02, p < .001$, with fewer errors on the top key for the piano and on the bottom key for the violin. Finally, hand position interacted with local pitch, $F(1, 28) = 8.65, p < .01$, the association between relatively lower tones and the bottom key and relatively higher tones and the top key being only significant for participants having the right hand on the top key.

Consistent with the results of the ANOVA, the regression analyses on dErrors showed that, in non-musicians, the regression weight of timbre differed significantly from zero, while the regression weight of pitch did not, $t(15) = -2.58, p < .025$ and $t(15) = .65, p > .10$, respectively. On the contrary, in musicians, the regression slopes of pitch and timbre were both negative and significantly different from zero, timbre: $t(15) = -2.65, p < .025$; pitch: $t(15) = -3.95, p < .001$.

2. 4. 3. 2. 3. Effect of instrument practice.

We thus obtained results similar to those of Rusconi et al. (2006): both musicians and non-musicians showed a vertical SPARC effect in timbre judgment, suggesting that people who do not play music associate pitch height and spatial height even when pitch processing is not required by the task. However, participants displayed a more reliable SPARC effect when their right hand was on the top key and their left hand on the bottom key. The SRC literature, and especially the up-right / down-left advantage (e.g. Cho & Proctor, 2003; Lippa & Adam, 2001; Weeks et al., 1995), provides an interesting account of this interaction. It may be that the SPARC effect is reinforced when the response hand and the response key are congruent with the natural correspondence of right and top.

However, as described above, a closer analysis of the data suggests that the modulation of the SPARC effect by hand position is stronger in musicians. We thus tested an additional hypothesis, according to which the interaction between pitch and hand position might also be modulated by the instrument played. Indeed, Mikumo (1998) showed in a visual tracking task that pianists possess a left-right image of pitch height, while, for example, cellists associate pitch with the top-down dimension. Our group included mainly pianists, which could partly explain the association between hand position and pitch. We tested this idea within the musician sample by introducing another variable in the ANOVA on dRTs: the instrument played (piano or other). As expected, the effect of global pitch and the interaction between pitch and hand position were significant, $F(1, 12) = 10.78, p < .01$; $F(1, 12) = 10.33, p < .01$, respectively. The crucial interaction for our present question, i.e. the interaction between global pitch, hand position, and instrument played, did also reach significance, $F(1, 12) = 14.07, p < .01$: the interaction between hand position and pitch was significant only in the pianist subgroup, (pianists: $F(1, 12) = 22.12, p < .01$; non-pianists: $F(1, 12) < 1, p > .10$).

Altogether, the present results confirm the idea that the representation of pitch along a vertical axis does not depend on musical expertise, since it is present in both musicians and non-musicians. Nevertheless, in agreement with the orthogonal SRC literature, the vertical association between pitch and space is reinforced when the left-right dimension of hand position is congruent with the bottom-top dimension of the response key. This phenomenon is stronger in musicians and especially in pianists, probably because the influence of the keyboard structure is added to the natural top-right / down-left advantage. Rusconi et al. (2006) did not report this influence, but they counterbalanced hand position within subjects, whereas we used a between-subjects design. Crossing hands in the same experimental session could have cancelled this effect of hand position.

2. 4. 3. 2. 4. Comparison of Experiments 1 and 3 for non-musicians.

Experiment 3 was a replication of Experiment 1, using a timbre judgment task on isolated tones, but with a different response mapping. Since non-musicians only showed a SPARC effect in the vertical dimension, an additional ANOVA was run on their dRTs to verify the influence of response mapping. Global pitch, local pitch and timbre were the within-subject variables, response mapping (horizontal or vertical) in Experiments 1 and 3 being the between-subjects variable. Only the effect of global pitch was significant, $F(1, 30) = 5.16, p < .05$. As indicated by the trend toward an interaction between global pitch and response mapping, $F(1, 30) = 3.68, p < .07$, the SPARC effect tended to be more reliable in the vertical mapping.

The three previous experiments, like the ones of Rusconi et al. (2006), had the purpose of studying the mental representation of musical stimuli. However, it can be argued that every listener may not consider isolated tones as music. Given that the relevance of the results obtained in Experiment 1 to 3 to music processing is matter for concern, the next experiments were aimed at investigating whether automatic horizontal and vertical spatial associations may be elicited by melodic intervals. Of course, two-note intervals are also a very simple musical material: these experiments are a first step to understand spatial associations for basic elements of the musical system.

2. 4. 4. Experiment 4: Instrumental Timbre Judgment on Melodic Intervals with Left-right Responses in Non-musicians and Musicians

The purpose of Experiment 4 was to assess whether ascending and descending melodic intervals could activate left-to-right associations automatically.

Another aim was to define the relative importance of pitch and of interval contour in these effects. Indeed, the pitch of the tones composing the melodic interval (e.g. a descending interval can be played on a low or high pitch range) could influence the spatial association.

By reference to our metaphor "a piano in the head", we supposed that the low-to-high pitch movement in ascending intervals would be translated into a left-to-right movement on the piano-like mental line, hence leading to faster right responses, and the reverse for descending intervals. Since this effect would be obtained with melodic materials, it could then be referred to as a *SMARC effect* (Spatial *Melodic* Association of Response Codes).

In Experiment 4, intervals beginning with tones from different octaves, either ascending or descending, were presented for a timbre decision task. We predicted an association between descending intervals and the left response side and between ascending intervals and the right response side. If the tones composing intervals also activate the mental line, participants would present a SPARC effect, too, with faster left responses to intervals in a low pitch range, and faster right responses to intervals in a high pitch range. In other words, the effect of interval (SMARC effect) would be modulated by the pitch range of the tones (SPARC effect).

2. 4. 4. 1. Method

2. 4. 4. 1. 1. Participants.

A fresh group of 17 non-musicians (6 males; average age: 25 yrs; 13 right-handed) and 16 musicians (13 males; average age: 23.6 yrs; 14 right-handed) participated in Experiment 4. The non-musicians were university students and university workers without prior musical training or, for 7 of them, with 2.9 years of music training, stopped for 10.6 years, on average. The musicians were 5 professional musicians, 6 undergraduate music students, and 5 high level amateur musicians. Eleven of them were pianists, 2 played the trumpet, one the violin, one the saxophone, and one was a composer who played mainly the piano. They had 14.6 years of formal musical training and practiced their instrument 22 hours a week, on the average. All participants were paid.

2. 4. 4. 1. 2. Material and procedure.

The stimuli were six melodic intervals consisting of two consecutive tones. All the intervals were fifths (7 semitones between the two pitches). In order to dissociate the effects of pitch height and of interval contour, both were manipulated. Ascending and descending intervals were constructed with three beginning tones: a low-pitched tone (F3, 175 Hz), a medium-pitched tone (G4, 392 Hz) and a high-pitched tone (A5, 784 Hz), and four possible ending tones (B2b, 117 Hz, C4, 262 Hz, D5, 587 Hz and E6, 1319 Hz). The ascending intervals were thus F3-C4, G4-D5, and A5-E6, while the descending intervals were F3-B2b, G4-C4, and A5-D5. The use of common first tones for ascending and descending intervals reduced interval predictability. Each stimulus was preceded by 400 ms of silence and followed by 800 ms of silence in order to leave 1300 ms to the participants to respond, including the stimulus duration of 500 ms (250 ms per tone). The inter-stimulus interval was thus 1200 ms. As in Experiment 1, in each trial participants were asked to judge if the melodic interval was played by a piano or by a violin by pressing a left or right response key. The key assignment was reversed between the experimental blocks. However, to avoid the timbre judgment task to be performed only on the basis of the first tone and to force participants to process the interval, the stimuli were constructed the following way. In 25% of the trials, both tones were played in a piano or a violin timbre; in another 25%, only the first tone was played by the piano or by the violin while the second tone was played by an instrument irrelevant to the task (the ocarina, a kind of flute); in a further 25%, only the second tone was played by one of the target instruments, the first tone being played by the ocarina, and, finally, in the remaining 25% of trials, the entire stimulus was played by the ocarina and required no response. The participants thus had to respond to 75% of the trials. Each block consisted of 192 trials (2 target timbres, 6 intervals, 4 positions of the target timbre, and 4 presentations of each stimulus) and was preceded by 24 practice trials. The instructions emphasized both speed and accuracy. The experiment lasted about 35 minutes.

2. 4. 4. 2. Results and Discussion

The average error rate in this timbre judgment task was 2.6%. (non-

musicians: 2.9 %, musicians: 2.4 %, $p > .10$). The negative correlation between RTs and percentage of errors did not reach significance, hence suggesting the absence of a speed-accuracy trade-off, $r =$
$-.22$, $p > .10$. Average median response times for each pitch, interval, response side, timbre and group are shown in Table 4.

2. 4. 4. 2. 1. RT analyses.

Data were analyzed through a 2 X 3 X 2 X 3 X 2 repeated measures ANOVA on median dRTs with the following within-subject variables: interval (descending vs. ascending), pitch of the first tone (F3, G4, A5), instrument timbre (piano vs. violin), and position of the target timbre (whole interval, first tone, second tone). Group was a between-subjects variable. The stimuli that did not yield a response, i.e. the intervals completely played by the ocarina, were not included in the analyses.

No significant effect of pitch or of interval contour emerged, $F(2, 62) < 1$, $p > .10$ and $F(1, 31) = 1.64$, $p > .10$, respectively. No other main effect or interaction reached significance.

Table 4. Average Median RTs (in ms) in Experiment for each Interval, Pitch, Instrumental Timbre, Group and Response Side.

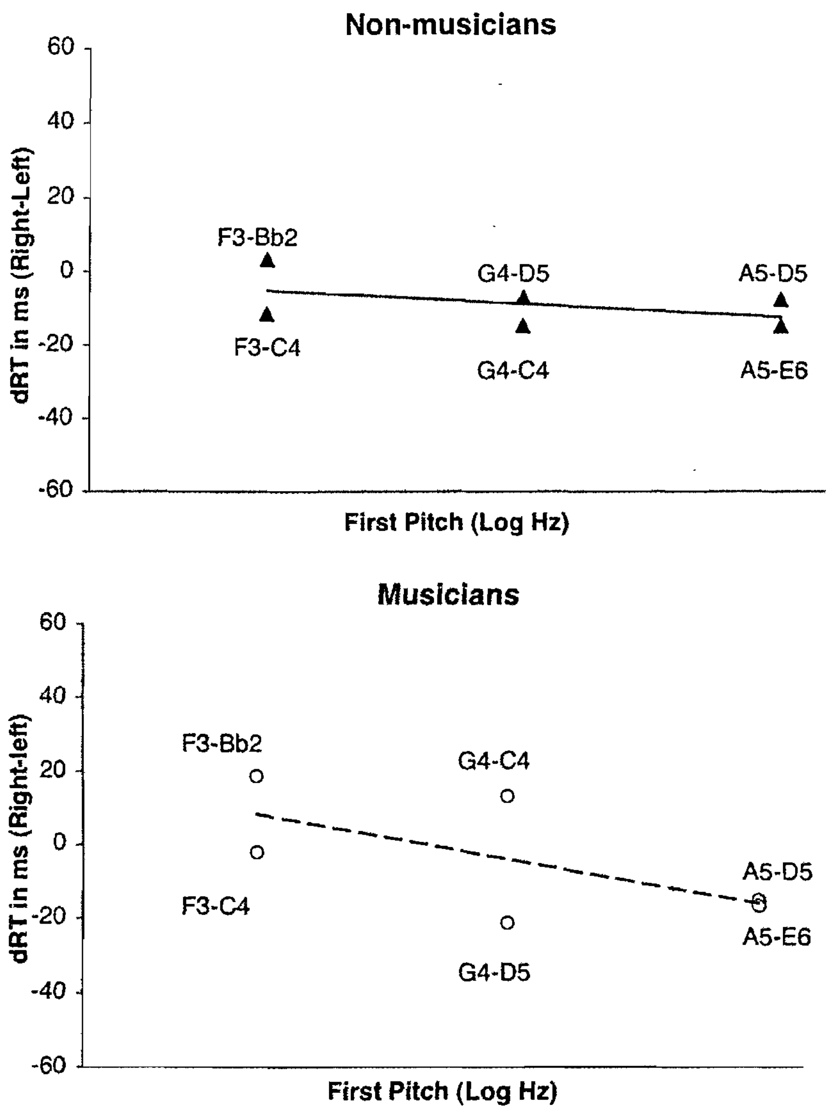
		Descending						Ascending					
		Low		Medium		High		Low		Medium		High	
		F3-Bb2		G4-C4		A5-D5		F3-C4		G4-D5		A5-E6	
		piano	violin	piano	violin	piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians	Left	733	777	726	777	725	757	716	815	724	752	760	757
	Right	742	775	727	747	719	747	732	776	727	736	739	749
	dRT (right-left)	9	-2	1	-30	-6	-10	16	-39	3	-16	-21	-8
Musicians	Left	693	720	689	727	749	733	705	750	708	757	732	719
	Right	714	737	700	742	708	744	696	755	687	735	721	697
	dRT (right-left)	21	17	11	15	-41	11	-9	5	-21	-22	-11	-22
Mean dRT		15	8	6	-8	-24	1	4	-17	-9	-19	-16	-15

Regression analyses were performed with interval, pitch and timbre as predictors. Timbre was coded -0.5 for the piano and +0.5 for the violin. The logarithm of the frequency of the first tone was used to code pitch height. Intervals were coded on the basis of their semitone distance from the departure pitch (7), with negative vs. positive values for descending vs. ascending intervals, respectively.

In agreement with the idea, suggested by the results above, that non-musicians do not associate intervals or the pitches composing them to a left-to-right axis, none of the predictors contributed significantly to the regression line (interval: $t(16) = -.06, p > .10$; pitch: $t(16) =$
$-.80$, $p > .10$; timbre: $t(16) = .71, p > .10$). These data are displayed in the top panel of Figure 4.

Surprisingly given the absence of an interaction between interval and group in the ANOVA, $F(1, 31) = 1.67$, $p > .10$, the regression analysis provided some evidence for an association between interval and response side in musicians, i.e. for a horizontal SMARC effect. As depicted by the distance between the descending and ascending intervals in the lower panel of Figure 4, in musicians, the contribution of interval contour to the regression slope was significantly different from zero, $t(15) = -2.18, p < .05$, whereas the regression weight of timbre was not, $t(15) = -.36, p > .10$. As suggested by the descending regression slope of musicians in Figure 4, the effect of pitch tended toward significance, $t(15) = -2.01, p < .07$.

Figure 4. Observed data and regression line representing RT differences between right- and left-side responses as a function of pitch in the instrument timbre judgment task on intervals used in Experiment 4 in non-musicians (top panel) and in musicians (bottom panel).



In order to explore further the disagreement between the lack of group by interval interaction and the occurrence of a significant SMARC effect in the regression, an additional ANOVA was run selectively on the musician group. Consistent with results of the regression analysis, interval contour had a significant impact on musicians' dRTs: participants showed facilitation on the left for descending intervals and on the right for ascending intervals, $F(1, 15) = 4.77$, $p < .05$. No other effect or interaction reached significance except position of the target timbre, $F(2, 30) = 3.31$, $p < .05$: the participants responded faster on the right to stimuli beginning with an ocarina tone and ending with a piano or a violin tone than to stimuli completely played by the piano or the violin.

Finally, the regression weight of pitch was here negatively correlated with the number of years of piano training in musicians, $r = -.57$, $p < .025$, though not in Experiment 1, suggesting that piano playing may favor the occurrence of a SPARC effect within intervals. On the contrary, the regression weight of interval contour was not correlated with piano training, $r = .13$, $p > .10$.

2. 4. 4. 2. 2. Error analyses.

The results of the ANOVA on dErrors did not show any significant SMARC or SPARC effect, $F(1, 31) < 1$, $p > .10$ and $F(2, 62) < 1$, $p > .10$, respectively. However, the interactions between pitch and timbre, pitch and position of the target timbre, pitch and group, and between pitch, timbre, position of the target timbre and group, were significant, in order $F(2, 62) = 4.29$, $p < .025$, $F(4, 124) = 2.78$, $p < .05$, $F(2, 62) = 4.04$, $p < .025$, $F(4, 124) = 2.60$, $p < .05$. The decomposition of the last interaction revealed that, in non-musicians, the SPARC effect was more consistent with the violin than with the piano timbre, $F(2, 32) = 3.62$, $p < .05$. In musicians, the position of the target timbre influenced the SPARC effect, $F(4, 60) = 2.69$, $p < .05$, which went in the expected direction only for the stimuli completely played on the piano or on the violin.

The regression analysis did not show any significant effect of interval, pitch or timbre in either non-musicians (interval: $t(16) = 1.14$, pitch: $t(16) = .83$, timbre: $t(16) = .58$; all $ps > .10$) or musicians (interval: $t(15) = -.75$, pitch: $t(15) = 1.69$, timbre: $t(15) = .74$ all $ps > .10$).

In summary, while interval contour did not induce a SMARC effect on dRTs in non-musicians, the regression analysis suggested the presence of this effect in musicians. This was confirmed by an ANOVA restricted to the musician group. It is worth noting that the absence of an interaction between interval contour and pitch of the first tone suggests that the SMARC and SPARC effects do not reinforce or cancel each other.

As non-musicians did not show a horizontal SPARC effect when attending to timbre but did exhibit it when attending to pitch, one may wonder whether a similar phenomenon happens for interval contour. The aim of the next experiment was to see if, in non-musicians, interval contour and pitch of the tones composing intervals activate a left-to-right representation under explicit attention to interval contour.

2. 4. 5. Experiment 5

Judgment of Interval Contour with Left-right Responses in Non-musicians

In the present experiment, non-musicians were required to judge the ascending or descending contour of the musical intervals used in Experiment 4. Since interval must be processed, an effect of interval contour was expected. We also examined the effect of pitch range, thus looking for the occurrence of a SPARC effect in a situation that differs from both the pitch (Experiment 2) and the timbre (Experiment 1) judgment tasks.

2. 4. 5. 1. Method

2. 4. 5. 1. 1. Participants.

A fresh group of 17 non-musician paid participants (9 males; average age: 25.9 yrs; 15 right-handed) took part in the experiment. They either had no musical training (n = 7), or had practiced an instrument during 2.3 years but had stopped this activity for 10.4 years on average (n = 10).

2. 4. 5. 1. 2. Material and procedure.

The task was to judge if the same melodic intervals as in Experiment 4 (F3-C4, G4-D5, A5-E6, F3-B2b, G4-C4, and A5-D5) were ascending or descending, by pressing a left or right key, the key assignment being reversed between blocks and its order counterbalanced across participants. Only intervals completely played on the violin and on the piano were presented to participants. Eight participants only heard piano intervals, while violin intervals were presented to the other nine. Each experimental block contained 144 trials including 24 presentations of each interval in a pseudorandom order avoiding more than three consecutive presentations of the same stimulus.

2. 4. 5. 2. Results and Discussion

The average error rate was 20%, suggesting that this interval contour judgment task was difficult to perform for non-musicians. Indeed, it seems that they had difficulties ignoring the pitch of the first tone to focus on interval contour. The absence of a significant correlation between response speed and error rate, $r = -.29$, $p > .10$, excluded the hypothesis of a speed-accuracy trade-off. Average median response times for each interval, pitch, response side and timbre are shown in Table 5, timbre being a between-subjects variable in the present design.

Table 5. Average Median RTs (in ms) in Experiment 5 for each Interval, Pitch, Instrumental Timbre and Response Side.

		Descending						Ascending					
		Low		Medium		High		Low		Medium		High	
		F3-Bb2		G4-C4		A5-D5		F3-C4		G4-D5		A5-E6	
		piano	violin	piano	violin	piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians	Left	805	844	886	958	930	1009	912	1037	835	893	807	796
	Right	863	841	910	974	899	1027	967	983	831	917	772	789
	dRT (right-left)	58	-3	24	16	-31	18	55	-54	-4	24	-35	-7

2. 4. 5. 2. 1. RT analyses.

A 2 × 3 × 2 analysis of variance with interval contour (ascending vs. descending) and pitch of the first tone (F3, G4 and A5) as within-subject variables and timbre as between-subjects variable was run on dRTs to correct responses. As can be seen in Figure 5, this analysis did not reveal any significant effect either of interval or of pitch range, both Fs < 1, $p > .10$. Only the interaction between pitch and timbre was significant, $F(2, 30) = 4.34$, $p < .025$. The expected effect of pitch, i.e. faster responses to intervals beginning with F3 on the left and to intervals beginning with A5 on the right, was only obtained in participants hearing the piano timbre, $F(2, 16) = 7.24$, $p < .01$.

The results of the regression analysis (see Figure 5) with the logarithm of pitch height and interval contour showed that none of these predictors did contribute significantly to the regression slope; pitch: $t(16) = -1.13$, $p > .10$; interval: $t(16) = -.612$, $p > .10$.

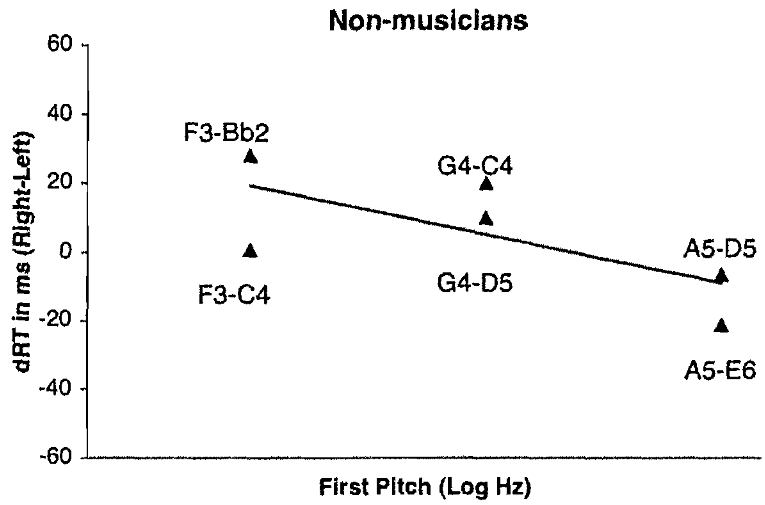
2. 4. 5. 2. 2. Error analyses.

In the ANOVA on dErrors, no significant effect or interaction was observed.

In line with these data, the regression analysis showed no effect of interval or pitch, $t(16) = -.26, p > .10$ and $t(16) = -1.38, p > .10$, respectively.

The results of this experiment suggest that interval contour does not activate a left-to-right mental line in non-musicians, even when explicit judgments of interval are required. The pitch of the tones composing the intervals only seems to activate the mental line when played on the piano.

Figure 5. Observed data and regression line representing RT differences between right- and left-side responses as a function of pitch in the interval contour judgment task on intervals presented in Experiment 5 to non-musicians.



2. 4. 6. Experiment 6:

Instrumental Timbre Judgment on Melodic Intervals with Bottom-top Responses in Non-musicians and Musicians

The previous experiment showed that non-musicians do not exhibit a horizontal SMARC effect in an interval judgment task. Because Experiment 3 suggested that a vertically oriented mental line could be automatically evoked in both non-musicians and musicians, the following experiment was aimed at examining whether such a vertical association could be obtained with melodic intervals instead of isolated tones.

2. 4. 6. 1. Method

2. 4. 6. 1. 1. Participants.

Thirty-two new paid participants, including 16 non-musicians (7 males; average age: 23.8 yrs; 15 right-handed) and 16 musicians (7 males; average age: 21.8 yrs; 15 right-handed), took part in this experiment. Most non-musicians had no musical training, but 7 of them had practiced music for 2.8 years and had stopped this activity for 10 years, on the average. The musician group included 4 professional musicians, 9 undergraduate music students and 3 amateur musicians, with 8 pianists, 4 violinists, 2 composers whose main instrument was the piano, one drummer and one flutist. They had practiced music for 15 years, playing or composing 22.4 hours a week, on the average.

2. 4. 6. 1. 2. Material and procedure.

Participants had to judge if the same intervals as those used in Experiments 4 and 5 (F3-C4, G4-D5, A5-E6, F3-B2b, G4-C4, and A5-D4) were played on a piano or on a violin by pressing a bottom or a top response key, the key assignment being reversed between blocks. Half of the participants responded with the left hand on the top key and the right hand on the bottom key, and the other half with the reverse hand position. As in Experiment 4, the target timbre could be present in both tones, only in the first or in the second tone, or absent.

2. 4. 6. 2. Results and Discussion

The average error rate was 2.4 % (non-musicians: 2%, musicians 2.8 %, $p >$ .10). There was no indication of a speed-accuracy trade-off, $r = -.09$, $p > .10$.

Average median RTs for each interval, pitch, timbre, hand position, response side and group are shown in Table 6.

2. 4. 6. 2. 1. RT analyses.

The 2 × 3 × 2 × 3 × 2 analysis of variance on dRTs with interval (ascending vs. descending), pitch of the first tone (F3, G4 and A5), timbre (piano vs. violin) and position of the target timbre as within-subject variables, and with hand position and group as between-subjects variables, did not reveal any significant effect or interaction.

In particular, participants did not associate interval contour or pitch of the first tone with a vertically oriented mental line, $F(1, 28) = 1.19, p > .10$ and $F(2, 56) = 2.38, p > .10$, respectively. The position of the hands did not interact with pitch or with interval, either, $F < 1, p > .10$ for both interactions.

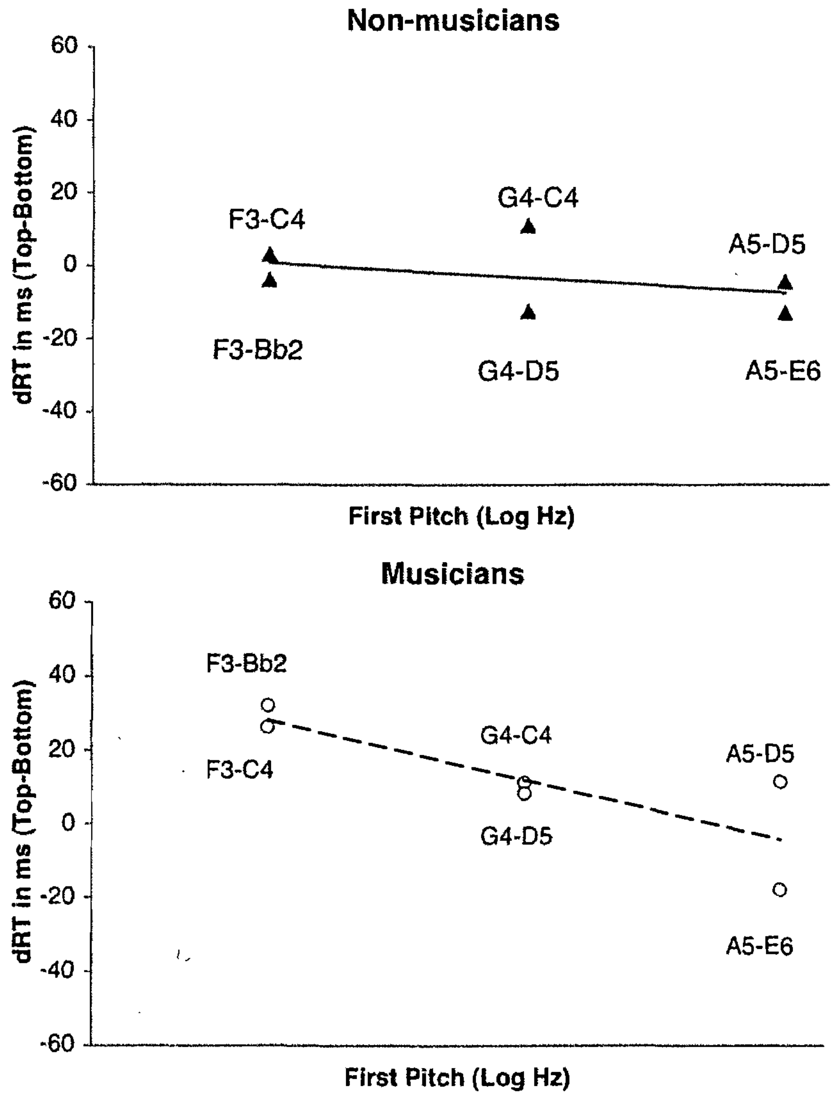
Table 6. Average Median RT (in ms) in Experiment 6 for each Interval, Pitch, Instrumental Timbre, Hand Position, Group and Response Side.

		Descending						Ascending					
		Low		Medium		High		Low		Medium		High	
		F3-Bb2		G4-C4		A5-D5		F3-C4		G4-D5		A5-E6	
		piano	violin	piano	violin	piano	violin	piano	violin	piano	violin	piano	violin
Non-musicians	Bottom	739	790	712	787	770	776	726	796	734	817	755	779
	Top	797	749	744	795	794	760	744	807	746	769	771	761
Left hand top	dRT (top-bottom)	58	-41	32	8	24	-16	18	11	12	-48	16	-18
	Bottom	783	828	732	790	800	821	776	835	754	826	798	792
	Top	764	813	716	806	789	808	754	840	734	831	740	800
Right hand top	dRT (top-bottom)	-19	-15	-16	16	-11	-13	-22	5	-20	5	-58	8
Mean dRT Non-musicians		20	-28	8	12	7	-15	-2	8	-4	-22	-21	-5
Musicians	Bottom	693	791	649	790	712	777	712	786	659	765	702	767
	Top	732	774	696	759	766	755	740	777	713	738	725	715
Left hand top	dRT (top-bottom)	39	-17	47	-31	54	-22	28	-9	54	-27	23	-52
	Bottom	699	680	654	719	687	751	678	705	701	756	723	730
Right hand top	Top	727	760	665	737	729	723	695	775	689	774	697	714
	dRT (top-bottom)	28	80	11	18	42	-28	17	70	-12	18	-26	-16
Mean dRT musicians		33	32	29	-6.5	48	-25	22	31	21	-4	-1	-34
Mean dRT		27	2	19	3	27	-20	10	19	9	-13	-11	-20

Consistent with the results of the ANOVA, the regression analysis showed no association between interval, pitch, timbre, and response side in non-musicians (interval: $t(15) = -.469$, $p > .10$; pitch: $t(15) = -.997, p > .10$; timbre: $t(15) = -.44$, $p > .10$). In musicians, the contribution of pitch to dRT only tended to differ from zero

(t (15) = -1.93, $p < .08$). Interval contour and timbre did not contribute significantly to the musicians' regression slope, $t(15) = -1.16$, $p > .10$; $t(15) = .74$, $p > .10$; respectively. Finally, the regression coefficients of pitch and of interval were not significantly correlated with piano training; $r = -.26$, $p > .10$ for pitch and $r = .40$, $p > .10$ for interval. The regression slope of pitch and the mean dRTs to each stimulus are displayed in Figure 6.

Figure 6. Observed data and regression line representing RT differences between top and bottom responses as a function of pitch in the instrument timbre judgment task on intervals used in Experiment 6 in non-musicians (top panel) and in musicians (bottom panel).



2. 4. 6. 2. 2. Error analyses.

Error rates mirrored the pattern of response times: There was no significant effect or interaction in the ANOVA, $F(1, 28) < 1$, $p > .10$ for interval and $F(2, 56) <$

1, p >.10 for pitch. Consistent with the ANOVA, the regression analyses did not reveal any significant effect in either non-musicians (interval: $t(15) = .64$, $p > .10$, pitch: $t(15) = .31$, $p > .10$; timbre: $t(15) = -.89$, $p > .10$) or musicians (interval: $t(15) = 1.44$, $p > .10$, pitch: $t(15) = .33$, $p > .10$; timbre: $t(15) = .95$, $p > .10$). In brief, interval contour does not seem to create vertical spatial associations of response codes in musicians or non-musicians, either on RTs and error rates.

2. 4. 6. 2. 3. Comparison of Experiments 4 and 6 for musicians.

Experiment 6 was identical to Experiment 4, except for the response mapping. Concerning musicians, in the regression analysis and in the separate ANOVAs on dRTs, the SMARC effect was significant in the horizontal direction but not in the vertical one. To compare directly the SMARC effects in the two response mappings, an additional ANOVA on the musicians' dRTs of Experiments 4 and 6 was run. Interval, pitch, timbre and position of the target timbre were included as within-subject variables, and response mapping as between-subjects variable. Unexpectedly, the interaction between interval contour and response mapping was not significant, $F(1, 30) < 1$, $p > .10$, but the main effect of interval was, $F(1, 30) = 5.05$, $p < .05$. Indeed, in both experiments the SMARC effect went in the same direction, even if it was only significant in the separate analysis for the horizontal response mapping. Given the presence of an effect of interval over Experiment 4 and 6 together, we looked for a potential effect of interval contour only in the musicians of Experiment 6. This analysis indicated that musicians taken alone did not exhibit a vertical SMARC effect, $F(1, 14) = 1.22$, $p > .10$. In the global ANOVA on Experiments 4 and 6 on musicians, the effect of pitch was significant, $F(2, 60) = 3.68$, $p < .05$, and in the expected direction of a SPARC effect. This SPARC effect in musicians was not modulated by response mapping, $F(2, 60) < 1$, $p > .10$. These results suggest than musicians might associate the pitch of melodic intervals to a spatial representation from low to high and, in a lower extent given the absence of a SPARC effect in the regression analysis of Experiment 4, from left to right. In general, Experiments 4 to 6 indicate that interval contour is generally not associated to horizontal or vertical mental axes, excepted in musicians for the left-right axis.

2. 4. 7. General Discussion

The present series of experiments was aimed at studying the relation between either the pitch of isolated tones (Experiments 1-3), or the contour of melodic intervals (Experiments 4-6), and a spatial response device oriented either horizontally or vertically. The idea underlying these experiments is that, if musical stimuli activate a spatial mental representation, this should be reflected in the responses through a spatial association of response codes (e.g. Dehaene, 2001).

The design of these studies allowed us to assess the role of musical expertise in the SPARC and SMARC effects by comparing non-musicians and musicians (Experiments 1, 3, 4 and 6). In addition, the representation of horizontal (Experiments 1, 2, 4, and 5) and vertical directions (Experiments 3 and 6) was distinguished for both pitches and intervals. In non-musicians, the role of task was also examined in the horizontal mapping; an indirect task such as timbre judgment (Experiments 1 on pitches and 4 on intervals) was compared to a direct one, either pitch or interval contour judgment (Experiment 2 and Experiment 5, respectively).

The data were analyzed through variance and regression analyses, on the differences of speed and accuracy between right (or top) and left (or bottom) responses. The analysis of variance highlights main effects and interactions, whereas the regression analysis evaluates the linear relation between the examined variables, a complementary approach necessary to conclude that a spatial (directional) representation exists. The combination of both analyses allows choosing a strict criterion to interpret our data: a SPARC or a SMARC effect is considered as reliable only if it is significant both in ANOVAs and regression analyses, in either speed or accuracy.

2. 4. 7. 1. SPARC Effects

According to our criterion, results for isolated tones (Experiments 1-3) demonstrate the existence of both horizontal and vertical SPARC effects. Pitch height is associated with space, more precisely low-pitched tones are associated with both left and bottom responses, and high-pitched tones with both right and top responses. However, these effects are modulated by at least two factors: task and musical expertise.

2. 4. 7. 1. 1. Horizontal SPARC effect.

In non-musicians, the horizontal SPARC effect seems to be influenced by task: it was present only when the task required pitch to be processed intentionally (Experiment 2). This effect, observed in response speed, was more systematic than in the study by Rusconi et al. (2006). However, in agreement with Rusconi et al., when the task explicitly required instrument discrimination and hence pitch was presumably processed only unintentionally (Experiment 1), non-musicians did not display the horizontal SPARC effect. On the contrary, musicians did show a significant effect when the task was to judge the timbre of isolated tones.

The difference between groups in the horizontal SPARC effect may be explained, at least partly, by the influence of musical expertise on music processing abilities (for a review, see Peretz & Zatorre, 2005). Musical expertise has been shown to enhance sensitivity to tiny pitch variations in both music (e.g., Kishon-Rabin, Amir, Vexler, & Zaltz, 2001) and spoken sentences (Schön, Magne, & Besson, 2004). At the cortical level, expert musicians exhibit enhanced cortical representations for the musical timbres associated with their usual instrument compared to another instrument (Pantev, Roberts, Schulz, Engelien, & Ross, 2000). Musicians also seem more prone to automatically process musical stimuli. Unlike non-musicians, they show a mismatch negativity to slightly impure chords even when instructed to ignore these auditory stimuli (Koelsch, Schröger, & Tervaniemi, 1999). This tendency to process some dimensions of music in an automatic way may, as suggested by the present results, lead to an unintentional activation of the horizontal spatial representations associated with them.

Since the left-to-right association was automatically evoked in musicians but not in non-musicians, one may argue that it is related to acquired knowledge in the same way as the orientation of the numerical SNARC effect is related to writing direction (Dehaene et al., 1993). Indeed, the piano keyboard may not be a referent specific to musicians, since even musically untrained people generally know that the piano keys corresponding to low pitches are on the left and those corresponding to high pitches on the right of the keyboard. Thus, when asked to respond to pitch by pressing left or right keys, the non-musician participants of Experiment 2 could have activated their internal representation of the piano, as the musicians of Experiment 1 may have done. Under this view, like the mental number line (Giaquinto, 2001), the "piano in the head" is a culturally supplied spatial representation, even if the absence of a correlation between piano practice and the horizontal SPARC effect in musicians suggests that this effect is not related in a simple way to experience with the piano structure. However, this cultural dependency does not necessarily imply that the "piano in the head" is only a "short-term representation resulting from task-specific resolution strategies" (Seron & Pesenti, 2001, p.84). Indeed, in addition to be unable to account for the fact that musicians did get a horizontal SPARC effect even when pitch was irrelevant to the task (Experiment 1), such an explanation implies that the participants who only heard violin tones would have called up the piano keyboard as a referent. This seems rather counterintuitive, especially in non-musicians.

The horizontal SPARC effect might rather be a general (although culture-dependent) order effect. In other words, it may reflect the fact that both musicians and non-musicians associate low-pitched tones with the beginning of a musical scale-like sequence, and high-pitched tones with the end of it. Akin to the effects observed with other ordered sequences like numbers (e.g. Dehaene et al., 1993), letters of the alphabet, months (Gevers et al., 2003), or days of the week (Gevers et al., 2004), the origin of the horizontal SPARC effect may thus be a directional bias from left to right in people from a culture in which writing direction follows this pattern. This bias may lead them to associate beginnings of actions (e.g. Maass & Russo, 2003) or sequences (e.g. Gevers et al., 2003) with left, and ends with right. This raises the question of whether the SPARC effect is simply caused by the ordered nature of tones, as for letters of the alphabet or months (Gevers et al., 2003) or, at a higher level, by their scalar structure, as may be the case for numbers (e.g. Dehaene , 2001). A scale may be defined as a graduated order. Like numbers, tones are ordered on a scale that does not only possess ordinal, but also quantitative information: the fundamental frequency. Further, one influential model of the mental number line describes it as logarithmic (Dehaene, 2001; for a linear view see Gallistel & Gelman, 1992), thus in a way that is appropriate to the frequency relations within the musical scale. The origin of the numerical SNARC effect as reflecting magnitude or ordinal information is still highly debated (e.g., Ito & Hatta, 2004). The same debate may arise concerning pitch, and the present data encourage further research in this direction.

2. 4. 7. 1. 2. Vertical SPARC effect.

In contrast to the influence of musical expertise on the horizontal SPARC effect, a vertical SPARC effect was observed in both musicians and non-musicians when pitch processing was unintentional (timbre judgment task, Experiment 3). Hence, there seems to be an automatic activation of associations between pitches and a vertical representation.

For the vertical SPARC effect, the data of Rusconi et al. (2006) together with the present evidence (Experiment 3) suggest that pitch representation along a vertical axis may be the "intuitive" one, shared by most Western people irrespective of their musical education. Indeed, even young children locate high-pitched tones higher in space than low-pitched tones (Roffler & Butler, 1968). The linguistic congruency between the words used to name pitch height and spatial height cannot fully account for this phenomenon since the children of Roffler and Butler's study (1968) were unaware of these labels. In addition, the participants of our Experiment 3 (vertically oriented responses) did not have to categorize the tones as high or low since they performed a timbre judgment task. The vertical SPARC effect would then reflect an intrinsic spatial characteristic of the pitch of isolated tones. This proposition is consistent with the fact that the dimensions of pitch and vertical position have been considered as the clearest examples of synaesthetically related dimensions (e.g. Melara & O'Brien, 1987, Roffler & Butler, 1968; Walker & Smith, 1984). Hence, the vertical SPARC effect seems to corroborate the description of pitch height as a rectilinear, bottom-to-top dimension that characterizes several representational models of musical pitch (Shepard, 1982a, b; Ueda & Ohgushi, 1987).

This idea that the vertical representation of pitch is the "natural one" may also explain the horizontal SPARC effect described above. Indeed, the low-to-high representation could have been turned into a left-to-right one because of orthogonal stimulus-response compatibility (Cho, & Proctor, 2003). Many studies have shown SRC effects when the stimuli vary along a vertical dimension and the response along a horizontal dimension (e.g. Weeks, Proctor, & Beyak, 1995). In this situation, an advantage is observed when the upper stimulus is mapped on the right response key and the lower stimulus is mapped on the left key. The so-called "up-right / down-left advantage" is classically explained through a saliency or polarity correspondence between asymmetric stimulus and response codes (e.g. Cho & Proctor, 2003; Proctor & Cho, 2006): the most salient dimension or the dimension sharing the same positive polarity, here top and right, tend to be associated. If pitch is mentally represented on a vertical axis, this up-right / down-left advantage could have caused the horizontal SPARC effect.

The vertical SPARC effect, even though it reflects an "intuitive" representation of pitch, seems nevertheless modulated by hand position and by musical expertise. Concerning hand position, the vertical SPARC effect appears to be stronger for participants having the right hand on the top key and the left hand on the bottom key than for participants in the reverse mapping. This modulation of the effect by hand position can be explained by the up-right / down-left advantage found in the SRC literature (Weeks et al., 1995; Lippa & Adam, 2001), too. Indeed, in our design, the association between the right hand and the top key may be considered as congruent, hence facilitating the vertical SPARC effect.

The fact that only musicians exhibited a vertical SPARC effect in both speed and accuracy introduces the question of the role played by musical expertise in the vertical SPARC effect. This could be related to musical notation, which would reinforce the association of pitch with a vertical axis. Moreover, it seems that the influence of hand position on the vertical SPARC effect is more pronounced in musicians, especially in pianists. This could be accounted for by the fact the natural correspondence between right and top was here reinforced by the repeated association between the right hand and high tones when practicing the piano.

2. 4. 7. 2. SMARC Effects

Our data on isolated tones confirm the findings of Rusconi et al. (2006) by showing that pitch height induces horizontal and vertical spatial associations of response codes. We went one step further and used intervals instead of isolated tones (in Experiments 4-6) to investigate whether the SPARC effect is restricted to isolated tones or whether it extends to melodic intervals. Our results revealed very limited evidence for an association between space and the contour of intervals (descending with left or bottom, and ascending with right or top).

Musicians showed a horizontal SMARC effect on speed when processing intervals unintentionally (timbre judgment task, Experiment 4), but only in the regression analysis and in the ANOVA restricted to this group. This horizontal SMARC effect could originate from the activation of associations between hand movements and interval contour, because ascending intervals are played with a left-to-right movement on the piano, and the reverse for descending intervals. This hypothesis is reinforced by the absence of a vertical SMARC effect in Experiment 6. One way to test the "hand movement" interpretation of the horizontal SMARC effect in musicians would be to see if this effect can be obtained with non-manual responses. In numeric cognition, some data already suggest that effectors other than the hand can lead to SNARC effects (Schwarz & Keus, 2004). Similarly, it would be interesting to explore the SMARC effect with non-manual responses such as ocular saccades.

Irrespective of task and response mapping, non-musicians never showed a SMARC effect. This suggests that non-musicians, contrary to musicians, do not associate interval contour (ascending / descending) with a spatial representation as they do for isolated tones.

In other words, in non-musicians, interval contour does not evoke a "music mental line", whereas isolated pitches do. The absence of spatial associations for intervals is likely to be related to the complex structure of these stimuli, which, contrary to what was expected, are probably not coded on the basis of their contour by non-musicians. This is surprising because, in Experiment 5, we explicitly asked participants to process interval contour by using the terms "ascending" and "descending", which may easily be associated with a vertical axis. The absence of a SMARC effect in this situation suggests that, even when the linguistic congruency is strong, non-musicians do not associate interval contour to space.

A parallel can be drawn between intervals and multi-digit numbers in SNARC experiments. In their seminal paper, Dehaene et al. (1993; Experiments 2 and 9) observed that the SNARC effect tended to disappear with numbers larger than 10. However, the effect of the decade digit, which is read first by Western participants, tended to significance in Dehaene et al.'s Experiment 2 (1993). In a similar way, although interval contour does not systematically induce a SMARC effect, the pitch of the first tone of these intervals tended to cause a SPARC effect in musicians when they responded to the timbre of melodic intervals (Experiment 4 in the regression analysis and Experiment 6 in both analyses on RTs) and in non-musicians when the tones were played on the piano (Experiment 5). However, these associations are far from strong and systematic. Further research is needed to understand the poor consistency of the effect of pitch when tones are presented in intervals.

2. 4. 7. 3. Conclusion and Perspectives

The present study provides new evidence on spatial associations to musical stimuli. Vertically oriented spatial representations seem to be automatically activated irrespective of musical expertise, but are triggered only by isolated pitches, not by musical intervals. These vertical associations probably constitute the "natural" association between pitch and space. In contrast, the automatic activation of horizontally oriented spatial representations seems to be linked to music knowledge for both pitch and intervals.

In the broad field of research comparing musical and non-musical dimensions (e.g., Besson, 1998; Schellenberg, 2001), the present work also suggests that musical pitches, numbers, and ordered linguistic sequences may share similar spatial mental representations. Whether these similarities arise from the ordinal or the scale properties of pitches remains an open question.

Recent behavioral evidence (Notebaert, Gevers, Verguts, & Fias, 2006) suggests that numbers and explicit spatial information are related to common spatial representations. Indeed, the SNARC (numerical) effect is modulated by spatial compatibility, just as the Simon (spatial) effect is modulated by numerical compatibility. Moreover, psychophysical measurements indicate that the Simon and

SNARC effects share the same electrophysiological signature (Keus et al., 2005). In the same way, some neuroimaging evidence indicates that close and probably similar neural structures in the parietal lobe are involved in the representation of both number and space. Indeed, number processing tasks considered as evoking the hypothetical number line, such as number comparison tasks, activate parietal regions (e.g. Dehaene, Piazza, Pinel, & Cohen, 2003; Sandrini, Rossini, & Miniussi, 2004) that are also involved in visuo-spatial processing (e.g., Göbel, Walsh, & Rushworth, 2001) and in visuo-spatial imagery (Suchan et al., 2002), hence suggesting that the mental representation of numbers could be spatial in nature (for a recent review, see Hubbard et al., 2005). Interestingly for our concern, an fMRI study on musicians also showed that music activates these "spatial areas". Schmithorst and Holland (2003) observed activation of the inferior parietal lobules during melodic and harmonic processing. Since these areas are also activated in number comparison tasks (Sandrini et al., 2004), these results suggest that shared areas are involved in number and in music processing. Added to our finding that musical pitch and numbers share similar spatial mental representations, these data may account for the controversial fact that music training increases spatial abilities (Brochard, Dufour & Despres, 2004) and plays a role in mathematics achievement (e.g., Cheek & Smith, 1999).

Endnotes

i In the present paper, by *target dimension*, we mean the one that is at the basis of the effect of spatial associations of response codes under study.

ii This was done to take into account the fact that, despite the larger absolute frequency difference between the high pitches E5 and B5 than between the low pitches C3 and G3, (respectively 328 Hz and 65 Hz), the frequency ratio (1.5) and the perceptual distance between these pitches were similar. The logarithmic scale is normally used in music research, and the human perceptual system seems to perceive pitch according to this non-linear scale (Bachem, 1950; Shepard, 1982a).

iii Similar analyses with either pitch names recoded 1-4 or pitch F_0 as predictor led to almost the same results.

iv In tasks with vertically aligned stimuli, spatial stimulus-response compatibility effects are found irrespective of whether the response keys are aligned along the actual, frontal, plane or along the transverse plane (Vu, Proctor & Pick, 2000). This holds true also for spatial associations, as shown by Ito and Hatta (2004) for the vertical SNARC effect.

3. Discussion générale

Le regretté Luciano Pavarotti avait déclaré : « Vous n'avez pas besoin d'un cerveau pour écouter de la musique ». Le ténor de génie aurait pu tenir des échanges animés avec les neuropsychologues et chercheurs en sciences cognitives qui, depuis quelques années, consacrent justement leur énergie à comprendre comment notre cerveau traite les informations musicales. Les recherches relatées dans ce travail s'intègrent à cette démarche, puisqu'elles ont pour objectif d'examiner si les hauteurs musicales font l'objet d'un traitement cognitif spécifique, c'est-à-dire distinct à la fois de celui des phonèmes de la parole et de celui de diverses séquences ordonnées, comme la séquence des chiffres. À cette fin, nous avons adopté deux perspectives.

Premièrement, nous avons examiné les interactions entre les phonèmes et les intervalles mélodiques ou les notes dans le traitement du chant lyrique (une approche qui, espérons-le, aurait convaincu Pavarotti !) et du chant synthétisé (là, le grand ténor aurait sans doute été moins enthousiaste...). Cette question a fait l'objet de trois études (Kolinsky, Lidji, Peretz, Besson, & Morais, en révision; Lidji, Kolinsky, Peretz, Lafontaine, & Morais, en préparation; Lidji, Peretz, Jolicoeur, Moreau, Connolly, & Kolinsky, en préparation).

La **première étude** (Kolinsky et al., en révision) utilisait le paradigme de classification accélérée de Garner (1974) afin de mettre en évidence les interactions dimensionnelles entre des non-mots et les intervalles mélodiques de deux notes sur lesquels ils étaient chantés. Cette étude sur le traitement en temps réel du chant a mis en évidence une dissociation intéressante entre voyelles et consonnes. En effet, dans les expériences où les non-mots variaient au niveau de la voyelle finale, un patron d'intégralité a été observé. En revanche, les consonnes et les intervalles se comportaient comme des dimensions plus séparables.

La **deuxième étude** (Lidji, Kolinsky et al., en préparation) a révélé une dissociation similaire, mais cette fois au niveau des traces en mémoire de ces stimuli chantés. Des distracteurs dits *mismatch*, issus de la combinaison de non-mots et d'intervalles entendus dans des stimuli distincts lors de la phase d'apprentissage, conduisaient à des conjonctions illusoires (Treisman et Schmidt, 1982). Celles-ci suggèrent que les paroles et mélodies sont représentées séparément en mémoire. Mais, surtout, les stimuli mismatch dont les non-mots intervertis variaient sur base de la consonne médiane conduisaient à plus de conjonctions illusoires que les stimuli mismatch dont les non-mots variaient au niveau de la voyelle finale. Les consonnes paraissent donc plus séparables de la mélodie que les voyelles également sur le plan mnésique.

La **troisième étude** (Lidji, Peretz et al., en préparation) utilisait la composante MMN des potentiels évoqués cérébraux (pour une revue, voir Näätänen, Paavilainen, Rinne, & Alho, sous presse) afin de préciser si l'intégration observée dans le traitement en ligne des voyelles et des hauteurs reposait sur des processus relativement tardifs, par exemple liés aux étapes attentionnelles de traitement nécessairement impliquées par la tâche de classification accélérée, ou, au contraire, si cette intégration était déjà présente lors de l'analyse auditive précoce de ces stimuli. Nous avons observé que les MMNs en réponse à des voyelles chantées différant du stimulus standard au niveau de l'identité de la voyelle ou de sa hauteur n'étaient pas additives et que la topographie de ces réponses cérébrales ne différait pas selon la dimension concernée, hauteur ou voyelle. Ces résultats confirment que les voyelles et les hauteurs sont traitées de manière interactive, y compris au niveau des traces mnésiques pré-attentionnelles reflétées par la MMN.

Ces études ont eu le mérite de démontrer que même des concepts aussi restreints que ceux de traitement phonologique et mélodique méritent d'être subdivisés en leurs sous-composantes si l'on souhaite investiguer de manière appropriée leur spécificité de traitement. Nous avons, en effet, constaté que la spécificité du traitement des intervalles mélodiques par rapport à celui des phonèmes varie en fonction des catégories de phonèmes sélectionnées. Si les méthodes choisies différaient dans nos trois premières études, la problématique examinée était similaire, c'est pourquoi nous les commenterons dans une même section.

Le deuxième axe de recherche comparait le traitement des hauteurs musicales isolées et des intervalles mélodiques à celui de séquences ordonnées donnant classiquement lieu à des associations spatiales de codes de réponses (Dehaene et al., 1993). Notre **quatrième étude** (Lidji, Kolinsky, Lochy, & Morais, 2007) a démontré que les notes de musiques isolées, comme les chiffres, évoquent des associations de nature spatiale. La robustesse et l'orientation de ces associations étaient toutefois modulées par l'expertise musicale. Par contraste avec les notes, des intervalles mélodiques n'étaient pas aussi systématiquement associées à des zones de l'espace : les seules associations mises en évidence pouvaient être directement liées à la pratique instrumentale. Ces résultats et leurs implications seront discutés dans une section séparée.

3. 1. La chanson des voyelles, la scansion des consonnes

Si les résultats des études 1, 2 et 3 se complètent de manière cohérente pour démontrer que les voyelles et les intervalles ou hauteurs sont perçus et mémorisés de manière plus intégrée que ne le sont les consonnes et les intervalles, il reste à comprendre l'origine de cette distinction entre catégories phonémiques. Nous examinerons séparément le cas des voyelles et des consonnes.

3. 1. 1. La chanson des voyelles

Plusieurs contrôles effectués dans l'étude 1 permettent de conclure que l'intégralité entre voyelles et intervalles n'était pas liée à un artefact tel que l'identité des voyelles ou des intervalles mélodiques choisis. Ainsi, le même patron de performance a été obtenu chez des sujets différents et avec des matériels expérimentaux distincts. De plus, cet effet ne peut être expliqué par la distorsion des caractéristiques spectrales des voyelles liée aux variations de hauteur, puisque les résultats étaient identiques pour des voyelles synthétisées dans lesquelles ces interactions acoustiques étaient soigneusement contrôlées. Enfin, les résultats de l'étude 3 suggèrent que l'intégralité observée sur le plan comportemental prend sa source dans des interactions relativement précoces, correspondant au moins à l'étape de traitement à l'origine de la composante MMN. En d'autres termes, les traces sensorielles de contrastes vocaliques ou de hauteur semblent prises en charge par les mêmes substrats neuronaux et ce, y compris lorsque les participants avaient pour instruction de ne pas porter attention aux stimuli auditifs présentés.

Toutefois, si les voyelles et leur hauteur donnent lieu à des traitements interactifs, ces traitements étaient asymétriques. Dans l'étude 1, l'identité de la voyelle interférait plus sur le traitement de sa hauteur que la hauteur n'interférait sur le traitement de la voyelle. Si l'on se place dans la perspective théorique de Melara et Marks (1990) (voir aussi Lidji, 2007), ces données suggèrent que ces deux dimensions sont traitées à des étapes différentes, la voyelle étant analysée de manière plus précoce que la hauteur. Ces résultats peuvent sembler surprenants si l'on prend en compte l'impression subjective de la plupart des chanteurs occasionnels d'oublier facilement les paroles de chansons dont la mélodie ne cesse de leur trotter en tête. Pourtant, à notre connaissance, aucune donnée expérimentale n'appuie ce sentiment.

En effet, dans l'étude 2, comme dans d'autres études évaluant les traces en mémoire des paroles et mélodies de chansons (Peretz, Radeau, & Arguin, 2004; Racette & Peretz, 2007; Serafine, Crowder, & Repp, 1984), la reconnaissance des paroles était bien supérieure à celle des mélodies.

L'hypothèse d'un traitement des voyelles plus précoce que le traitement de la hauteur est d'ailleurs corroborée par le décours temporel de la MMN dans l'étude 3. La MMN en réponse aux déviations de voyelles était significativement plus précoce que celle évoquée par des déviations de hauteur, alors que ces réponses cérébrales présentaient des amplitudes similaires. Il semble donc qu'en français, l'analyse de l'identité de la voyelle précède le traitement de la hauteur à laquelle cette voyelle est prononcée ou celui de l'intervalle sur lequel le non-mot dont fait partie cette voyelle est chanté.

Quelle pourrait être l'origine de cette succession apparente des traitements ? Tout d'abord, cet effet pourrait simplement être lié aux choix des stimuli et ne pas refléter les caractéristiques générales du traitement des voyelles et de leur hauteur. Plusieurs arguments expérimentaux sont, toutefois, en mesure d'infirmer cette hypothèse. Premièrement, le caractère asymétrique de l'interférence dans la tâche de classification accélérée de Garner (étude 1 ; Kolinsky et al., en révision) se manifestait avec une amplitude équivalente quels que soient la voyelle et l'intervalle sélectionnés, y compris lorsque les intervalles mélodiques étaient très contrastés (par exemple, une tierce majeure ascendante et une quarte descendante). En outre, la différence de latence obtenue dans l'étude de potentiels évoqués n'était pas couplée à une différence significative d'amplitude, alors que des différences de saillance des dimensions se manifestent généralement à la fois au niveau de la latence et de l'amplitude de la MMN (Nätäänen et al., sous presse). L'amplitude et la latence de la composante P3a ne variaient pas non plus en fonction de la dimension concernée par la déviation. La P3a est pourtant connue pour varier selon l'attractivité attentionnelle des stimuli (Polich, 2007). Si les voyelles étaient plus saillantes que les différences de hauteur choisies, elles auraient donc dû attirer davantage l'attention des participants et, par conséquent, évoquer une P3a plus ample et plus précoce. Ce n'était pas le cas, mais d'autres recherches employant de nouveaux contrastes phonétiques et de hauteur demeurent nécessaires pour affirmer avec certitude que les caractéristiques de nos stimuli ne jouent pas un rôle, même partiel, dans les résultats obtenus.

L'observation, dans nos expériences, d'une asymétrie systématique entre le traitement des voyelles, plus précoce que celui des hauteurs ou intervalles, est cependant intéressante puisqu'elle contredit les résultats de recherches antérieures utilisant le paradigme de classification accélérée pour examiner les interactions entre des voyelles et des hauteurs stables. En effet, Miller (1978) a observé un patron d'interférence symétrique entre les voyelles de syllabes consonne-voyelle et la hauteur sur laquelle elles étaient prononcées (voir aussi Carrell, Smith, & Pisoni, 1981). En ce qui concerne les interactions entre voyelles et tons lexicaux, une asymétrie similaire à celle que nous avons décrite a été obtenue par Repp et Lin (1990), mais elle pouvait être expliquée, dans leur cas, par une différence de discriminabilité entre dimensions. En revanche, dans notre étude, l'asymétrie ne pouvait se justifier par de tels facteurs et semble donc bien émaner des interactions entre les dimensions elles-mêmes.

La divergence entre les données de l'étude 1 de ce travail et celles de Miller (1978) pourrait s'expliquer, premièrement, par le choix d'utiliser des intervalles de deux notes au lieu de notes isolées. La tâche des participants portait donc sur le contour mélodique et non sur la hauteur absolue de la voyelle. Deuxièmement, notre matériel expérimental a été conçu de manière à favoriser un traitement des stimuli en termes de « musique vocale » : les stimuli étaient chantés de façon lyrique et les consignes insistaient sur leur aspect musical. Tomiak, Mullennix et Sawusch (1987) ont démontré que le mode de traitement appliqué aux stimuli est susceptible de modifier les interactions entre leurs dimensions. Dans leur étude, le fait de considérer des syllabes synthétiques comme de la parole conduisait à un patron d'intégralité entre la consonne et la voyelle qui les composaient, tandis que lorsque la consigne présentait ces mêmes stimuli comme non-linguistiques, un patron de séparabilité était obtenu. De manière similaire, l'intégralité asymétrique de notre étude 1 pourrait être liée au fait de traiter les stimuli comme du chant et non comme de la parole.

Nos résultats semblent donc suggérer que, au moins dans le contexte du *chant*, le traitement des voyelles précède celui du contour mélodique. Il serait, de ce fait, extrêmement intéressant de répliquer les expériences de classification accélérée avec variations de voyelles sur un matériel aux caractéristiques musicales moins marquées, par exemple en utilisant un timbre parlé et en remplaçant le contour mélodique par des variations prosodiques typiques du français (voir par exemple, Patel et al., 1998). Une telle étude permettrait de spécifier si l'intégralité entre voyelles et intervalles, d'une part, et l'asymétrie de traitement en faveur des voyelles, d'autre part, sont typiques d'un mode de traitement musical ou se généralisent au traitement de la parole.

Une comparaison directe entre des stimuli similaires, variant uniquement par leur caractère musical ou prosodique, permettrait également de mieux comprendre nos données électrophysiologiques. En effet, l'idée d'un mode de traitement musical est peu compatible avec le paradigme et les stimuli de notre étude de MMN, qui utilisait des voyelles isolées synthétisées sur des notes distantes d'un demi-ton. Par conséquent, cette troisième étude explorait les réponses cérébrales involontaires à des changements auditifs concernant l'identité de la voyelle ou sa hauteur dans un contexte fort éloigné de ce que l'on considère communément comme de la musique.

Étant donné le caractère peu musical des stimuli employés dans l'étude de MMN, l'intégration précoce entre voyelles et hauteurs pourrait reposer sur des processus non spécifiques au chant. Ces processus pourraient donc également intervenir dans le traitement de ces mêmes informations dans un contexte de parole. Nous avons, de ce fait, interprété la latence plus précoce de la MMN pour des contrastes vocaliques que pour des contrastes de hauteur par le rôle plus important des voyelles en français. Il a, par exemple, été démontré qu'en anglais, une langue où la position de l'accent lexical est variable, les distinctions vocaliques sont plus importantes que les patrons d'accentuation pour la compréhension de la parole (Fear, Cutler, & Butterfield, 1995). Par contraste, le français est une langue à accent fixe où la prosodie a pour fonction essentielle de fournir des informations émotionnelles (par exemple, Ladd, Silverman, Tolmitt, Bergmann, & Scherer, 1985) et pragmatiques ; par exemple, pour distinguer les questions des affirmations (pour une revue sur la prosodie, voir Cutler, Dahan, & van Donselaar, 1997). Il est donc d'autant plus vraisemblable que le traitement des voyelles puisse dominer celui de la hauteur ou de la prosodie en français. En revanche, dans certaines langues tonales (Maddieson, 2005), de telles différences de hauteur ont pour effet de modifier la signification des mots. Afin de mettre à l'épreuve l'hypothèse selon laquelle la précédence du traitement des voyelles sur celui des hauteurs en français est liée à la fonction linguistique plus notable des premières, il serait souhaitable de répliquer notre étude de MMN sur une population de locuteurs d'une langue tonale comme le Mandarin ou le Thaï. Alors que plusieurs études récentes ont confirmé que les locuteurs de langues tonales présentent des réponses cérébrales accrues à des différences de hauteur, y compris dans un contexte non-linguistique (p. ex., Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007), aucune à notre connaissance n'a manipulé l'information sur la voyelle et celle sur le ton au sein des mêmes stimuli. Si l'interprétation des résultats de notre étude de potentiels évoqués en termes de fonction des dimensions dans la langue est valide, les locuteurs d'une langue tonale ne devraient donc pas présenter la disparité de latence entre la MMN aux déviations de voyelle et de hauteur observée ici chez des sujets francophones.

Une dernière question en suspens est celle de l'origine neuronale de cette interaction entre le traitement des voyelles et celui des hauteurs. La sous-additivité et la similarité de distribution des réponses électrophysiologiques aux différences vocaliques et de hauteur suggèrent que ces dimensions sont traitées par le même ensemble de neurones. Cette conclusion ne peut toutefois être affirmée avec certitude sur base de la méthode choisie. Il paraît donc essentiel de reproduire notre expérience avec une méthode d'imagerie permettant une meilleure résolution spatiale, comme la magnétoencéphalographie (MEG).

Une telle étude est envisagée dans un futur proche auprès d'une population particulière : les sujets atteints d'amusie congénitale et leurs contrôles appariés. Cette étude en MEG permettra, premièrement, de spécifier les régions neuronales impliquées dans le traitement des voyelles et des hauteurs et, éventuellement, de confirmer le recouvrement des régions impliquées. De plus, l'application de notre paradigme à des participants amusiques contribuera à la compréhension de leurs déficits. Si, comme suggéré par Peretz et ses collègues (p. ex., Peretz et al., 2002), l'amusie congénitale résulte d'un déficit spécifique du traitement fin des hauteurs, déficit ayant une origine neurologique (Hyde et al., 2007; Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006), nous devrions observer chez ces individus une MMN anormale pour les différences de hauteur (P. Moreau, communication personnelle). Il est établi que les compétences linguistiques sont intactes dans l'amusie congénitale (Peretz et al., 2002). Il serait donc possible que les amusiques manifestent, en revanche, une MMN normale aux déviations de voyelles. L'observation de dissociations entre le traitement cérébral automatique de la hauteur et celui de dimensions phonologiques sur les *mêmes stimuli* constituerait une innovation importante dans la recherche sur l'amusie congénitale. En outre, les anomalies de la réponse de MMN aux déviations de hauteur chez les amusiques pourraient concerner non seulement l'amplitude et la latence de cette composante, mais également ses sources neuronales. En effet, si l'amusie congénitale est réellement manifeste dès le début de la vie de ces individus, il se peut que ces déficits aient conduit à une certaine plasticité cérébrale (pour des revues, voir par exemple Hommet, Billard, de Toffol, & Autret, 2003; Staudt, 2007). La présence d'une réorganisation cérébrale dans l'amusie congénitale est compatible avec l'observation d'une réduction de substance blanche (Hyde et al., 2006) et d'une épaisseur corticale anormale dans le cortex auditif et le gyrus frontal inférieur droit (Hyde et al., 2007) chez ces sujets. En raison de cette réorganisation potentielle, les sources neuronales du traitement des voyelles et des hauteurs pourraient donc différer chez les participants amusiques par rapport aux participants contrôles. Nous pourrions ainsi nous attendre à ce que le traitement des voyelles soit anatomiquement dissocié de celui de la hauteur chez les amusiques, par contraste avec l'intégration prédite chez les sujets contrôles.

Pour conclure sur les interactions entre voyelles et hauteurs, nous avons démontré de manière répétée, avec divers stimuli et paradigmes, que ces dimensions sont traitées de manière interactive et que cette interaction pourrait reposer sur des ressources neuronales communes. De nouvelles études restent néanmoins nécessaires pour, d'une part, établir si les processus examinés dans nos expériences sont restreints au traitement du chant ou se généralisent au traitement de la prosodie et des tons linguistiques et, d'autre part, examiner comment ces traitements sont pris en charge chez des individus présentant des déficits congénitaux du traitement musical.

3. 1. 2. La scansion des consonnes

Alors que nos trois premières études ont mis en évidence la présence d'interactions entre le traitement des voyelles et des hauteurs, les résultats des études 1 et 2 suggèrent que le traitement et la mémorisation des consonnes est séparable de celui des hauteurs. Dans l'étude 1, la classification accélérée de stimuli variant sur la consonne médiane n'a jamais conduit à un patron d'intégralité : soit nous n'avons pas observé l'interférence nécessaire à cette conclusion (consonnes occlusives), soit nous avons constaté une interférence mais sans facilitation (consonnes nasales). Les résultats d'une tâche dite de « condensation » (Pomerantz & Garner, 1973) suggèrent que cette interférence était liée à un manque d'attention sélective à chaque dimension, et non à une intégration de celles-ci (Thibaut & Gelaes, 2002). En ce qui concerne les interactions entre les traces mnésiques des phonèmes et des intervalles, nous avons observé que les stimuli dont les non-mots variaient sur une base consonantique étaient plus susceptibles de donner lieu à des conjonctions illusoires (Treisman & Schmidt, 1982) des paroles et de la mélodies que ceux variant sur une base vocalique. Ces conjonctions illusoires constituent un indice de la séparabilité des dimensions, car seule des dimensions séparables peuvent être combinées erronément (Thompson, Hall, & Pressing, 2001). En résumé, sur base de données comportementales, le traitement des consonnes apparaît relativement séparable de celui des intervalles.

Afin de spécifier si cette indépendance entre consonnes et hauteurs musicales repose sur des traitements distincts au niveau neuronal, il serait extrêmement intéressant de compléter nos résultats en utilisant une méthode telle que les potentiels évoqués ou la MEG, comme nous l'avons fait pour les voyelles. Si la séparabilité observée entre consonnes et hauteurs est due à des traitements cérébraux indépendants, nous nous attendrions à observer une additivité des réponses de MMN à des déviations de ces dimensions. Cette additivité pourrait être couplée à des sources neuronales différentes pour ces deux types de déviations.

Il reste à comprendre pourquoi les consonnes ne sont pas, comme les voyelles, traitées de manière intégrée avec la mélodie ou la hauteur. Une première hypothèse est basée sur les propriétés acoustiques distinctes des différentes catégories de phonèmes. Alors que les voyelles sont caractérisées par leurs informations spectrales relativement stables, leur sonorité et leur durée (Fry, Abramson, Eimas, & Liberman, 1962; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) qui en font des supports mélodiques inégalés (Sundberg, 1982), les consonnes présentent des variations acoustiques rapides (Delattre, Liberman, & Cooper, 1955). Elles sont aussi généralement moins sonores et plus brèves que les voyelles (Delattre, Liberman, & Cooper, 1955; Romani & Calabrese, 1998). De ce fait, les consonnes sont parfois décrites comme représentant un obstacle à la musicalité (McCrean & Morris, 2005; Scotto di Carlo, 1993). Si les propriétés acoustiques respectives des voyelles et des consonnes sont bien la cause de leur différence d'intégration avec la mélodie, des consonnes sonores, susceptibles de présenter une fréquence fondamentale facilement perceptible, devraient présenter un degré intermédiaire d'intégration avec le contour mélodique. Afin de tester cette hypothèse, nous avons comparé les patrons d'interactions dimensionnelles entre non-mots et intervalles dans une tâche de classification accélérée (Garner, 1974) selon que les variations des non-mots concernent des consonnes occlusives, supposées plus distinctes des voyelles, et des consonnes nasales, dont la sonorité et la durée les rapprochent des voyelles sur le plan acoustique. Cette manipulation a permis d'infirmer l'hypothèse acoustique : il n'y avait pas de différence d'intégration avec la mélodie selon le mode d'articulation et la sonorité des consonnes.

3. 1. 3. Voyelles et consonnes : Des statuts linguistiques différents ?

Nos résultats révélant des différences de traitement entre voyelles et consonnes indépendamment de la sonorité de ces dernières corroborent ceux d'études sur le statut de ces phonèmes dans un contexte de parole. Premièrement, le statut différent des voyelles et des consonnes est bien établi dans les langues sémitiques, où les racines consonantiques fournissent des informations sémantiques tandis que la mélodie vocalique a une fonction d'inflexion grammaticale (Boudelaa & Marslen-Wilson, 2004). La racine consonantique est d'ailleurs responsable d'effets d'amorçages importants dans ces langues (p. ex., Boudelaa & Marslen-Wilson, 2000 ; Frost, Deutsch, Gilboa, Tannenbaum, & Marlsen-Wilson, 2000), alors que ce n'est pas le cas des mélodies vocaliques (Boudelaa & Marslen-Wilson, 2004).

Ces dissociations fonctionnelles entre phonèmes consonantiques et vocaliques sont, par ailleurs, appuyées par l'étude de patients cérébro-lésés. Ainsi, Z.T., un patient aphasique billingue Arabe-Français, manifestait des erreurs altérant spécifiquement l'ordre des consonnes constituant la racine des mots (Idrissi & Kehayia, 2004 ; Prunet, Béland, & Idrissi, 2000). Les dissociations entre la production des voyelles et celle des consonnes dans l'aphasie ne sont pas limitées aux locuteurs de langues sémitiques, puisque Caramazza, Chialant, Capasso, & Miceli (2000) ont décrit une double dissociation entre la production des voyelles et celle des consonnes chez des patients aphasiques italiens. Le déficit de production des consonnes chez le patient décrit par Caramazza et al. (2000) n'était, en outre, pas modulé par leur sonorité (mais voir Romani et Calabrese, 1998), ce qui permet d'écarter l'hypothèse d'une dissociation basée sur les propriétés acoustiques des phonèmes. Les auteurs ont donc proposé que les consonnes et les voyelles seraient représentées sous forme de catégories phonétiques distinctes. Ce point de vue est renforcé par le fait qu'une stimulation électrique du gyrus temporal supérieur gauche conduit à un déficit sélectif de la discrimination des consonnes, la discrimination des voyelles étant relativement préservée (Boatman, Hall, Goldstein, Lesser, & Gordon, 1997).

Au plan développemental, Nazzi et ses collaborateurs ont observé que de jeunes enfants parviennent à discriminer deux mots variant par une seule consonne, et ce, quelle que soit la sonorité de celle-ci (Nazzi, 2005; Nazzi & New, 2007). En revanche, ces enfants se montraient incapables de réaliser la même tâche si les paires minimales se différenciaient par une seule voyelle. Ces résultats suggèrent que la fonction linguistique des consonnes est essentiellement lexicale : les consonnes serviraient à distinguer les mots entre eux. Les données issues de tâches d'apprentissage statistique présentées dans l'introduction confirment cette idée (Mehler, Peña, Nespor, & Bonatti, 2006; Newport & Aslin, 2004. Par contraste, les voyelles contribueraient plutôt aux inflexions grammaticales (Bonatti, Peña, Nespor, & Mehler, 2005), serviraient de support prosodique en raison des variations de fréquence fondamentale qu'elles permettent, et interviendraient dans l'identification des locuteurs (Owren & Cardillo, 2006).

Si l'origine de cette distinction demeure mystérieuse, les recherches sur l'évolution du langage (Fitch, 2000; Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005) et les comparaisons entre les productions vocales des humains et d'autres animaux passées en revue dans l'introduction (par exemple, Lieberman, 1968; Newport, Hauser, Spaepen, & Aslin, 2004; Owren, Seyfarth, & Cheney, 1997; Rendall, Rodman, & Emond, 1996) autorisent certaines spéculations. Ainsi, il semble que nos ancêtres préhistoriques, comme beaucoup de grands singes contemporains, n'étaient pas en mesure de produire des sons de type consonantique en raison de la forme de leur tractus vocal (Demolin, 2007) et de limitations du contrôle neuronal de ces articulateurs (Fitch, 2000). Bien que les données fossiles et les techniques actuellement disponibles ne permettent malheureusement pas de dater l'apparition du langage articulé (voir Fitch, 2000, pour une revue), l'idée d'un protolangage musical chez nos ancêtres suggérée par Mithen (2005) paraît séduisante. Combinée avec les données anatomiques (Demolin, 2007; Fitch, 2000) selon lesquelles l'apparition des voyelles aurait pu précéder celles des consonnes, d'une part, et avec les résultats suggérant que les consonnes auraient une fonction plus intrinsèquement linguistique que les voyelles (Owren & Cardillo, 2006), d'autre part, cette hypothèse peut conduire à une autre suggestion. Le chant ou protolangage des hommes préhistoriques était peut-être essentiellement basé sur les voyelles, ce qui expliquerait la fonction fondamentale de celles-ci dans le chant, y compris de nos jours. Cette évolution commune pourrait également expliquer la découverte d'une parenté entre les formants des voyelles et les intervalles musicaux appréciés dans plusieurs cultures (Ross, Choi, & Purves, 2007).

En résumé, alors qu'elles visaient initialement à mettre en évidence la spécificité du traitement des hauteurs musicales, nos recherches ont eu des retombées inattendues. Elles contribuent en effet de manière significative à la compréhension de la spécificité du *langage*, en appuyant par une approche innovante une hypothèse extrêmement prometteuse : celle de traitements et fonctions distincts des voyelles et des consonnes dans la parole. Nos résultats, en accord avec d'autres données récentes (Caramazza et al., 2000; Nazzi, 2005; Nazzi & New, 2007; Nespor, Peña, & Mehler, 2003; Owren & Cardillo, 2006), suggèrent que les voyelles et les consonnes constituent deux catégories phonétiques dissociables, le traitement des voyelles étant potentiellement pris en charge par des mécanismes moins spécifiques au langage que celui des consonnes.

3. 2. Hauteurs musicales et associations spatiales

Après avoir comparé le traitement de la hauteur musicale à celui des phonèmes, nous avons examiné les liens entre traitements musicaux et spatiaux, une idée au coeur de plusieurs études récentes (par exemple, Brochard, Dufour, & Despres, 2004; Douglas & Bilkey, 2007; Graziano, Peterson, & Shaw, 1999; Sluming, Brooks, Howard, Downes, & Roberts, 2007). Dans cet objectif, nous avons emprunté au domaine de la cognition numérique le concept désormais classique d'effet SNARC (Dehaene, Bossini, & Giraux, 1993; Fias, Brysbaert, Geypens, & d'Ydewalle, 1996). D'après Dehaene (2001), cet effet d'association spatiale de codes de réponse reflète une représentation mentale des nombres de type linéaire et orientée de gauche à droite (ou de bas en haut, Ito & Hatta, 2004).

Notre étude a non seulement confirmé l'existence d'associations spatiales similaires pour les hauteurs musicales isolées (voir aussi Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006), un effet intitulé *SPARC (Spatial Pitch Association of Response Codes)*, mais a surtout démontré que des stimuli musicaux plus complexes, comme des intervalles de deux notes, n'activaient pas cette ligne mentale supposée, sauf dans le cas particulier d'auditeurs musiciens avec un dispositif horizontal. De plus, nous avons établi que l'expertise musicale joue un rôle fondamental dans l'émergence d'associations spatiales horizontales pour les notes et les intervalles, tandis que les associations verticales pour les notes semblent indépendantes de la pratique musicale. Ces résultats ont été largement commentés dans la discussion générale de l'**étude 4**, mais la diffusion de ces découvertes a conduit à de nouveaux développements (Beecham, Wilson, & Reeve, 2006; Douglas & Bilkey, 2007; Stewart, Nasralla, Verdonschot, & Lanipekun, 2007) qui méritent d'être détaillés ici. Par ailleurs, la littérature sur les origines cognitives des effets de type SNARC n'a cessé de croître ces dernières années (par exemple, Fischer, 2006; Gevers & Lammertyn, 2005; Gevers, Ratinckx, De Baene, & Fias, 2006; Hubbard, Piazza, Pinel, & Dehaene, 2005; Müller & Schwarz, 2007a) et ces données peuvent apporter un éclairage nouveau à nos résultats.

3.2.1. Que reflètent les associations spatiales de codes de réponse ?

L'effet SNARC est l'objet d'un intérêt grandissant depuis sa découverte en 1993 par l'équipe de Stanislas Dehaene. Il a pourtant fallu plusieurs années avant que son interprétation initiale en termes de ligne mentale numérique ne soit remise en question. Ainsi, Giaquinto (2001) a très justement souligné que les associations spatiales de codes de réponse pouvaient avoir une origine culturelle, une idée suggérée par l'observation d'une inversion de l'effet SNARC chez des participants issus de cultures où on lit et écrit de droite à gauche (Dehaene et al., 1993; Zebian, 2005). Ce point de vue est encore conforté par le fait que l'association entre hauteur musicale et réponses horizontales est accrue chez des musiciens par rapport à des non-musiciens (Lidji et al., 2007 ; Rusconi et al., 2006).

En réponse à cette critique, l'un des principaux arguments pour considérer les effets de type SNARC comme issus d'une ligne mentale abstraite est leur manifestation constante, quel que soit le format de présentation des chiffres (Fias, 2001; Fias et al., 1996; Nuerk, Wood, & Willmes, 2005), la tâche (Fias, 2001; Fias et al., 1996; Fias, Lauwereyns, & Lammertyn, 2001) et la modalité de réponse (Fischer, 2004; Fischer, Warlop, Hill, & Fias, 2004; Keus & Schwarz, 2005; Schwarz & Keus, 2004). Plus spécifiquement, la représentation spatiale des nombres semble activée automatiquement, puisque l'effet SNARC, comme l'effet SPARC (Lidji et al., 2007), est observé lors de tâches indirectes, c'est-à-dire ne requérant aucun jugement explicite de magnitude ou de hauteur des notes (Fias et al., 2001; Gevers, Lammertyn, Notebaert, Verguts, & Fias, 2006).

Toutefois, alors que cette représentation a longtemps été décrite comme indépendante de l'effecteur puisqu'un effet SNARC a été observé par l'équipe de Dehaene (1993) même lorsque les participants avaient les mains croisées, de nouvelles données remettent cette assertion en question. Ainsi, nous avons observé un effet de la main utilisée lors des réponses verticales chez les musiciens (effet SPARC). De plus, dans le domaine numérique, Wood, Nuerk et Willmes (2006a) ne sont pas parvenus à répliquer l'effet SNARC rapporté par Dehaene et al. (1993) en demandant à leurs participants de croiser les mains pour répondre (main droite sur le bouton de gauche et inversement), alors que le nombre de participants testés et la puissance statistique étaient bien supérieurs à ceux de l'étude originale. D'après les auteurs, cela indique que les associations spatiales de codes de réponse ne dépendent pas uniquement d'un cadre de référence extra-corporel mais interagissent avec des informations liées à l'espace corporel, les deux référentiels étant activés simultanément. Müller et Schwarz (2007b) sont allés plus loin en montrant que, lorsque les réponses étaient verticales, l'orientation de l'effet SNARC était modifiée selon que la consigne fasse référence au bouton ou à la main utilisée pour répondre. Ces résultats semblent confirmer l'idée que les représentations spatiales activées dans ce type de tâche sont à la fois liées aux effecteurs et à l'espace extra-corporel.

L'ensemble de ces données a engendré une interprétation alternative à celle de l'activation d'une ligne mentale par les chiffres (et les autres séquences ordonnées) : les effets de type SNARC pourraient simplement refléter des facteurs stratégiques (Fischer, 2006). L'idée d'un double codage corporel et extra-corporel de la représentation des nombres (Müller & Schwarz, 2007b; Wood, Nuerk, & Willmes, 2006a) semble, en effet, fort peu économique. La vision de Fischer (2006) est plus parcimonieuse, puisque tous les effets d'associations spatiales de codes de réponse, quelle que soit la modalité ou le type de stimuli utilisés, pourraient reposer sur un choix stratégique des participants afin d'optimiser leur performance. Dans le même esprit, Giaquinto (2001) a proposé que le système cognitif ait besoin d'un référentiel pour réussir ces tâches et fasse ainsi appel à une imagerie de type visuo-spatial. Ces facteurs stratégiques expliqueraient à la fois l'extraordinaire flexibilité des effets d'association spatiale de codes de réponse, leur force et leur orientation pouvant être modulées par les consignes (par exemple, Bätchtold, Baumüller, & Brugger, 1998) ou par certains paramètres culturels (Lidji et al., 2007; Zebian, 2005), mais aussi la grande variabilité individuelle observée (Beecham et al., 2006; Wood et al., 2006a; mais voir Wood, Nuerk, & Willmes, 2006b). L'idée de traitements stratégiques, donc relativement tardifs, concorde également avec les résultats comportementaux (Keus & Schwarz, 2005; Müller & Schwarz, 2007a) et électrophysiologiques (Keus, Jenks, & Schwarz, 2005) indiquant que l'effet SNARC prend sa source à l'étape de sélection de la réponse.

Le débat sur l'origine cognitive des associations spatiales de codes de réponse est pourtant loin d'être clos. Si des processus stratégiques peuvent intervenir dans les effets observés pour les séquences non numériques (Gevers, Reynvoet, & Fias, 2003, 2004) et les hauteurs musicales (Lidji et al., 2007 ; Rusconi et al., 2006), d'autres arguments soutiennent l'hypothèse de réelles associations spatiales en ce qui concerne la cognition numérique. Premièrement, des relations entre traitements numériques et spatiaux ont été mises en évidence dans des tâches différentes de celles ayant classiquement démontré l'effet SNARC. Par exemple, dans une tâche de bissection de lignes constituées de chiffres, les lignes composées de chiffres de faible magnitude tendaient à être segmentées un peu à gauche de leur milieu, alors qu'un biais vers la droite était observé pour des lignes de chiffres de grande magnitude (Calabria & Rossetti, 2005; Fischer, 2001). La présentation visuelle de chiffres peut, de surcroît, influencer l'orientation de l'attention vers les zones de l'espace concordant avec leur magnitude (Fischer, Castel, Dodd, & Pratt, 2003). Ces données favorisent clairement l'hypothèse d'une association entre chiffres et espace au-delà de simples processus stratégiques.

L'existence d'un recouvrement anatomique entre les aires cérébrales pariétales impliquées dans les comparaisons numériques et les traitements spatiaux corrobore cette position (voir par exemple Göbel, Walsh, & Rushworth, 2001; Hubbard et al., 2005; Sandrini, Rossini, & Miniussi, 2004; Suchan et al., 2002). Quoiqu'une similarité des structures cérébrales ne soit pas systématiquement le signe d'une association des fonctions cognitives, ces constats répétés sont intrigants, surtout si on les met en relation avec des données issues de la neuropsychologie. Ainsi, des enfants atteints d'un déficit visuo-spatial ne manifestent pas d'effet SNARC dans une tâche directe de comparaison de magnitude, alors que des enfants contrôles appariés présentent l'effet (Bachot, Gevers, Fias, & Roeyers, 2005). Zorzi, Priftis et Umiltà (2002) ont, de plus, fait état d'une atteinte de la ligne mentale numérique chez des patients héminégligents, un trouble attentionnel altérant le traitement d'une partie de l'espace. Lorsqu'ils devaient indiquer la médiane de deux chiffres, ces patients présentaient un biais vers les grands chiffres, supposément situés à droite de la ligne mentale et, surtout, ce biais pouvait être corrigé par l'utilisation de prismes visuels (Rossetti et al., 2004). Le fait qu'une correction visuelle améliore la performance des patients héminégligents dans une tâche ne comportant aucun caractère visuo-spatial explicite suggère fortement que le traitement des chiffres recoure à des processus de nature spatiale.

En somme, bien que des voix se soient élevées pour affirmer que l'effet SNARC ne reflète pas une représentation mentale spatiale des chiffres mais serait plutôt dû à des facteurs stratégiques (Fischer, 2006; Giaquinto, 2001), ce qui n'est d'ailleurs pas totalement exclu, une multitude de données convergentes indiquent qu'il existe bien un lien entre traitements numériques et spatiaux. Par contre, l'origine cognitive de l'effet SPARC (Lidji et al., 2007, Rusconi et al., 2006) reste un mystère en raison du peu de données disponibles à ce jour. Des recherches équivalentes à celles effectuées sur l'effet SNARC seront nécessaires pour améliorer notre compréhension des associations spatiales pour les hauteurs musicales. Il faudrait, par exemple, varier le type de tâche utilisée et le mode de réponse afin de dissocier la contribution de composantes motrices et celles de composantes cognitives dans cet effet, explorer les aires cérébrales activées par les tâches générant l'effet SPARC et les comparer à celles impliquées dans des traitements spatiaux et, enfin, examiner l'impact de déficits du traitement de la musique (Douglas & Bilkey, 2007) et / ou du traitement spatial (Stewart & Walsh, 2007) sur cet effet. Ce mouvement a déjà commencé son essor, comme nous le verrons dans la section suivante.

3.2.2. L'avenir de l'effet SPARC

La découverte d'associations spatiales pour les hauteurs musicales a ranimé la controverse sur les relations entre compétences musicales et spatiales. Plusieurs recherches récentes ont ainsi pour enjeu la compréhension des interactions entre traitements musicaux et spatiaux. Par exemple, Beecham, Wilson et Reeve (2006) ont examiné la possibilité d'un recouvrement entre les associations spatiales pour les chiffres (effet SNARC) et celles pour les hauteurs (effet SPARC). Leurs participants effectuaient successivement une tâche sur un matériel numérique et sur un matériel musical. L'hypothèse sous-jacente à cette étude était que si les effets SNARC et SPARC émergent de processus communs, les mêmes participants devraient présenter les deux effets. Dans cette perspective, les résultats se sont avérés décevants : en plus de démontrer une variabilité individuelle extrêmement importante, les auteurs n'ont pas observé de lien entre les deux types d'effets. Ces résultats sont intéressants à deux titres. Tout d'abord, ils confirment l'idée que les effets d'association spatiale de codes de réponse sont extrêmement variables selon les individus, un point sur lequel nous reviendrons en abordant la question des liens entre amusie et compétences spatiales (Douglas & Bilkey, 2007). Ensuite, ils semblent suggérer que les associations spatiales numériques et musicales sont indépendantes. Cette conclusion paraît toutefois prématurée car la méthode choisie par Beecham et ses collègues est loin d'être optimale. En effet, les associations spatiales numériques et musicales étaient évaluées dans deux tâches différentes et successives. Un paradigme plus approprié pour tester directement l'hypothèse d'un partage de ressources spatiales pour le traitement des chiffres et des hauteurs devrait plutôt exploiter ces deux dimensions simultanément.

C'est ce qui a été réalisé par Stewart et ses collègues (Stewart et al., 2007 ; voir pour un paradigme similaire Stewart, Walsh, & Frith, 2004). Des pianistes et des participants non-musiciens devaient réaliser une séquence motrice sur un clavier en réponse à des chiffres de 1 à 5 présentés auditivement, chaque chiffre correspondant à un doigté particulier (de 1 pour le pouce à 5 pour l'auriculaire, chaque doigt étant systématiquement associé à une touche du clavier). L'aspect déterminant de cette étude pour la question qui nous intéresse réside dans la manipulation de la hauteur à laquelle les chiffres étaient prononcés. En effet, la hauteur des chiffres pouvait correspondre à celle attendue en vertu de la séquence motrice, le chiffre 1 étant prononcé sur la note associée au pouce et le chiffre 5 sur la note, plus aiguë, associée à l'auriculaire, ou, au contraire, être inversée et conduire à une situation d'incongruité entre l'information fournie par le chiffre et celle apportée par la note. Seuls les pianistes manifestaient une interférence en situation d'incongruité, un résultat interprété par les auteurs comme issu d'un couplage fonctionnel entre les systèmes auditif et moteur chez des pianistes expérimentés.

Une interprétation alternative, relayée par nos résultats et ceux de Rusconi et al. (2006), est toutefois concevable. Si les musiciens, pianistes ou non, se représentent mentalement les hauteurs musicales sur une ligne mentale orientée de gauche à droite, l'effet observé par Stewart et al. (2007) pourrait découler de l'incongruité entre les associations mentales relatives aux chiffres présentés et celles activées par la hauteur de la note, indépendamment d'un couplage auditivo-moteur. En d'autres termes, les effets SNARC et SPARC pourraient interagir, ce qui suggérerait fortement qu'ils sont bien issus de processus similaires. Les données exposées par Stewart et al. (2007) ne permettent pas de trancher entre les deux hypothèses proposées mais une étude en cours devrait répondre à cette question (L. Stewart, communication personnelle).

Dans cette étude, les notes sur lesquelles les chiffres sont prononcés ne correspondent pas aux notes successives du clavier mais sont plus distantes en termes de hauteur. Si les interactions entre chiffres et notes décrites par Stewart et al. (2007) émanent d'interactions entre les représentations spatiales liées aux notes et aux chiffres, cette situation devrait augmenter l'interférence chez les musiciens en cas d'incongruité. En effet, nous avons constaté (Lidji et al., 2007) que l'effet SPARC était plus robuste avec des notes distantes qu'avec des notes successives (Rusconi et al., 2006). Au contraire, si l'effet observé par Stewart et al. (2007) trouve bien son origine dans des interactions auditivo-motrices liées à la pratique du piano, l'interférence devrait être réduite dans la situation où les notes, bien que plus distantes en termes de hauteur, ne correspondent pas au doigté effectué.

Les non-musiciens testés par Stewart et al. (2007) ne présentaient pas d'interférence entre chiffres et hauteur, un constat qui peut aussi bien s'expliquer par l'absence d'associations spatiales horizontales automatiques dans cette population (Lidji et al. ; 2007, Rusconi et al., 2006) que par leur absence de pratique du piano. Par contre, si les notes et les chiffres activent bien des représentations spatiales communes, elles devraient interférer chez des non-musiciens, mais avec un dispositif de réponse vertical. Les résultats d'études (Franco, 2006 ; Lidji, Franco, Kolinsky, & Peretz, 2006) au cours desquelles des non-musiciens devaient juger la magnitude de chiffres prononcés à des hauteurs congruentes ou incongrues avec leur position supposée sur la « ligne mentale » vont dans ce sens. L'effet SNARC se trouvait renforcé lorsque la note et le chiffre se référaient au même côté de l'espace (par exemple, un chiffre de grande magnitude prononcé sur une note aiguë, tous deux associés à la partie haute de l'espace) par rapport à la situation d'incongruité. Ces données partielles doivent cependant être interprétées avec précaution car cette interaction n'atteignait pas le seuil de signification lorsque la tâche des sujets était de juger la hauteur de la note (effet SPARC) ni lors de tâches indirectes telles que des jugements de parité ou de timbre. En somme, si l'idée d'examiner les interactions entre les associations spatiales pour les chiffres et les notes semble prometteuse, les données disponibles à ce jour ne permettent pas d'établir avec certitude l'existence d'un lien entre ces représentations. Un enjeu important pour l'avenir sera d'approfondir l'exploration de cette question.

Une autre voie de recherche à l'avenir prometteur est celle qui relie la découverte d'associations spatiales pour les hauteurs musicales et l'amusie congénitale. Nous avons déjà mentionné dans l'introduction les données de Douglas et Bilkey (2007) selon lesquelles des participants atteints d'amusie congénitale manifestent un déficit dans des tâches classiques de rotation mentale visuo-spatiale par rapport à des non-musiciens et des musiciens appariés. Dans un autre volet de cette étude, ces auteurs ont examiné la présence d'un effet SPARC et, crucialement, d'interférences entre le traitement de la hauteur et la réalisation de la tâche de rotation mentale chez ces participants. L'hypothèse audacieuse à l'origine de ces travaux était qu'au lieu de refléter un déficit spécifique du traitement des hauteurs musicales, l'amusie puisse trouver son origine dans un facteur plus général : les compétences visuo-spatiales. L'observation de déficits de rotation mentale chez les participants amusiques et d'une corrélation entre la performance au sous test « contour » de la Batterie de Montréal d'Evaluation des Amusies (MBEA, Peretz, Champod, & Hyde, 2003) et à la tâche de rotation mentale semblent soutenir cette hypothèse. De plus, dans une autre étude (Peretz et al., sous presse), des individus amusiques déclarent éprouver plus de problèmes spatiaux que des individus non atteints.

La tâche visant à évaluer la présence d'un effet SPARC dans l'étude de Douglas et Bilkey (2007) consistait à demander aux participants de comparer la hauteur de notes (tâche directe) en indiquant si la deuxième note était plus grave ou plus aiguë que la seconde (la différence variait de 1 à 5 tons), le dispositif de réponse pouvant être compatible ou non avec la « ligne mentale verticale ». Les résultats sont surprenants : bien que les participants amusiques semblent présenter un effet SPARC, avec une performance plus altérée lorsque le dispositif de réponse était incompatible avec la représentations spatiale supposée des hauteurs, cet effet était significativement moins marqué dans ce groupe que chez les participants contrôles et particulièrement chez les musiciens. En situation de double tâche, la tâche de rotation mentale interférait significativement avec la performance à la tâche de comparaison de hauteur chez les sujets contrôles mais pas chez les amusiques. Les auteurs interprètent cet ensemble de données comme indiquant que l'amusie est fortement associée à des déficits de traitement spatial de manière générale, et que les individus amusiques présentent un lien plus faible entre le traitements musicaux et spatiaux que les individus non atteints de ce trouble.

Étant donné la publication récente de ces résultats, leur impact est difficile à évaluer pour l'instant. Il est cependant probable qu'ils seront à l'origine d'un courant de recherche prolifique. En effet, si un lien causal entre traitement spatial et musical est démontré, cette découverte ira sans doute à l'encontre de l'existence de traitements et de déficits spécifiques à la musique (Peretz & Coltheart, 2003). Une telle conclusion est toutefois prématurée pour plusieurs raisons.

Premièrement, le mode de sélection des participants amusiques et le choix de la tâche de rotation mentale utilisée par Douglas et Bilkey (2007) sont discutables. En effet, les participants étaient considérés comme amusiques sur base d'un seul sous-test de la MBEA, le test de contour qui n'est, de surcroît, pas le plus représentatif des habiletés de perception musicales. La tâche de rotation employée était une tâche classique (Shepard & Metzler, 1971) en version papier : Douglas et Bilkey (2007) ont donc calculé la performance sur l'ensemble des essais. Il serait souhaitable de répliquer ces observations avec des participants amusiques diagnostiqués avec plus de soin et avec des méthodes plus précises. Dans une étude en cours (I. Peretz et N. Gosselin, communication personnelle), des participants considérés comme amusiques sur base de la MBEA et ayant déjà fait l'objet d'investigations approfondies de leurs capacités musicales devaient effectuer une tâche de rotation mentale informatisée où la distance angulaire entre les formes à comparer était manipulée. Contrairement aux données rapportées dans Nature Neuroscience (Douglas & Bilkey, 2007), les résultats préliminaires n'indiquent pas de différence de performance entre les participants contrôles et les amusiques. Si ces données provisoires doivent être considérées avec prudence, elles sont loin de soutenir l'existence d'une corrélation systématique entre compétences spatiales et musicales.

Deuxièmement, Douglas et Bilkey (2007) ne fournissent pas d'information sur la performance individuelle des participants. Comme nous l'avons déjà souligné, les effets d'association spatiale de codes de réponse présentent une variabilité individuelle importante (Beecham et al., 2006), il est donc fort possible que certains des participants amusiques présentent bien l'effet mais que celui-ci soit masqué par la procédure de moyennage.

Troisièmement, les auteurs ont utilisé une tâche directe de comparaison de notes pour évaluer la présence d'un effet d'association spatiale de codes de réponse. Ce choix est surprenant lorsqu'on sait que les personnes souffrant d'amusie éprouvent des difficultés importantes dans ce type de tâche qui consiste, en réalité, à juger du contour mélodique d'un intervalle (Foxton, Dean, Gee, Peretz, & Griffiths, 2004). La réduction de l'effet SPARC chez les participants amusiques pourrait donc résulter d'une incapacité à réaliser la tâche, du moins pour les intervalles inférieurs à deux tons (Stewart & Walsh, 2007). Une approche plus appropriée consisterait à évaluer la présence d'associations spatiales pour les hauteurs dans une tâche indirecte, par exemple une tâche de jugement de timbre telle que celle utilisée par Rusconi et al. (2006) et notre équipe (Lidji et al., 2007). En effet, la coexistence de traitements explicites déficitaires et de capacités implicites préservées est un phénomène classique en neuropsychologie (par exemple, Lê, Raufaste, Roussel, Puel, & Démonet, 2003; Schacter & Buckner, 1998; Vandenberghe, Schmidt, Fery, & Cleeremans, 2006), y compris dans le cas de l'amusie acquise (Tillmann, Peretz, Bigand, & Gosselin, 2007). Il est donc urgent, avant de conclure à une absence d'associations spatiales verticales pour les hauteurs dans l'amusie congénitale, d'évaluer ces processus au moyen d'une tâche indirecte. Un projet de recherche en cours a pour objectif d'examiner ces questions.

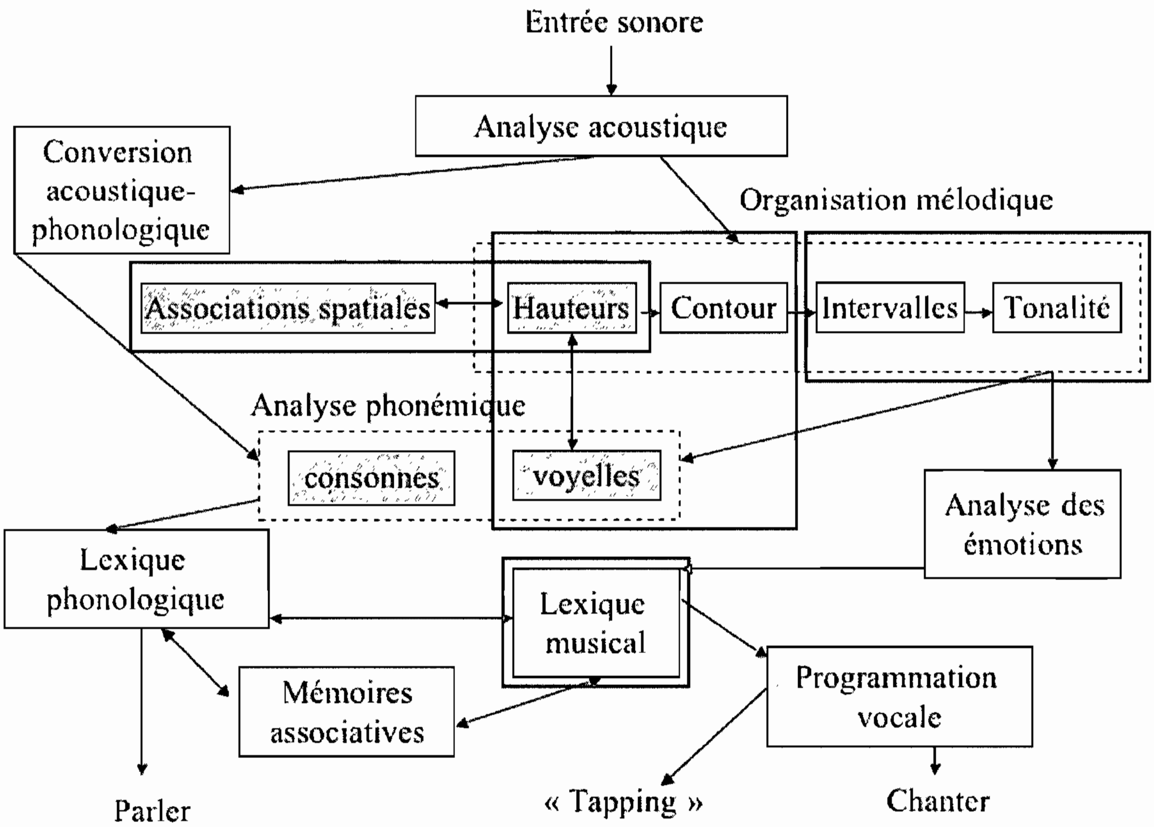
Enfin, l'existence de corrélations et d'interférences entre traitements spatiaux et musicaux rapportée par Douglas et Bilkey (2007) ne permet pas de conclure quant à la direction de l'association (Stewart et Walsh, 2007). Au lieu de démontrer que l'amusie congénitale est liée à des déficits des traitements visuo-spatiaux, une idée qui conduirait à rejeter l'hypothèse d'un déficit spécifique du traitement de la hauteur (Peretz et al., 2002), ces données peuvent aussi bien indiquer qu'un trouble des compétences musicales a des conséquences dommageables sur les habiletés visuo-spatiales. Ce dernier point de vue est compatible avec les études démontrant un effet de la pratique musicale sur ces compétences visuo-spatiales (Brochard et al., 2004; Costa-Giomi, 2004; Sluming et al., 2007).

Pour conclure, malgré les critiques développées ci-dessus, la contribution de Douglas et Bilkey (2007) est importante car elle apporte un des premiers arguments en faveur d'interactions entre processus spatiaux et traitement des hauteurs musicales chez des individus tout-venants grâce à l'observation d'une interférence de la tâche de rotation mentale sur le jugement des hauteurs. Cette idée avait déjà été avancée par Cupchik, Phillips et Hill (2001), puisque ces auteurs avaient mis en évidence une corrélation entre la performance à une tâche de rotation mentale et la capacité à effectuer une permutation de notes. Le paradigme de Douglas et Bilkey est toutefois plus élégant puisqu'il ne fait pas appel à des corrélations mais à une situation de double tâche. Ces travaux, comme ceux de Rusconi et al. (2006) et les nôtres, sont un premier pas dans la compréhension des interactions entre processus musicaux et spatiaux. Cependant, il reste encore beaucoup de chemin à parcourir pour spécifier l'étendue et la nature de ces interactions.

3. 3. Conclusion

Le traitement des hauteurs musicales est-il pris en charge par des composantes cognitives spécifiques ? Les réponses apportées dans cette thèse sont ambiguës. Le traitement des hauteurs semble bien présenter une certaine spécificité par rapport à celui de la dimension phonologique du langage, mais cette conclusion est limitée aux consonnes. Nous avons en revanche mis en évidence deux arguments en faveur de la non-spécificité de traitement de la hauteur : les voyelles et la hauteur constituent des dimensions interactives et les hauteurs musicales, comme d'autres informations faisant partie d'une séquence ordonnée, évoquent des associations de nature spatiale. Ces découvertes ont une implication cruciale sur le plan théorique : elles confirment l'idée, déjà soulignée par Peretz (2006), qu'il est fondamental de dissocier une fonction cognitive en ses différentes composantes pour examiner la spécificité de chacune d'entre elles isolément. Le modèle de traitement musical élaboré par Peretz et Coltheart (2003) peut donc désormais être complété en vertu de nos résultats (voir Figure 1b).

Figure 1b. Modèle de la perception musicale d'après Peretz et Coltheart (2003) complété à partir de nos résultats. Les composantes rythmiques ont été supprimées par souci de clarté. Les nouvelles composantes issues de ce travail sont indiquées en bleu et les nouvelles interactions entre dimensions sont encadrées en turquoise. Les composants de traitement spécifique à la musique selon Peretz et Coltheart sont entourés en rouge.



D'autres recherches demeurent nécessaires pour explorer les multiples composantes musicales qui dépassaient le cadre de nos travaux et, en particulier, les aspects rythmiques (Krumhansl, 2000; Patel, 2008). Nous avons également concentré notre attention sur la dimension perceptive, mais la production et la performance musicale constituent des objets de recherche tout aussi essentiels (pour une revue, voir Palmer, 1997). Ainsi, l'étude de la production du chant n'en est qu'à ses balbutiements (Dalla Bella, Giguère, & Peretz, 2007; Racette, Bard, & Peretz, 2006; Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006) alors qu'elle constituera certainement un outil de choix dans la compréhension des relations entre musique et langage. Si l'engouement récent pour la cognition musicale a marqué un tournant dans la recherche en psychologie, le prochain virage à amorcer sera d'explorer ces territoires encore relativement délaissés.

4. Bibliographie

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16*, 38-51.

American National Institute (1973). *American National Standard psychoacoustical terminology.* New York: American National Standard Institute.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representations of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neurosciences, 20*, 303-330.

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

Ashby, F. G., & Maddox, W. T. (1994). A response-time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38*(4), 423-466.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179.

Astesano, C., Schön, D., & Besson, M. (2004). Le langage et la musique dans le chant. *Revue de Neuropsy News, 3*, 78-83

Aulanko, R., Hari, R., Lounasmaa, O. V., Näätänen, R., & Sams, M. (1993). Phonetic invariance in the human auditory cortex. *Neuroreport, 4*, 1356-1358.

Avanzini, G., Koelsch, S., Lopez, L. , & Majno, M. (2005). *The Neurosciences and Music II: From perception to performance.* New York: Annals of the Bew York Academy of Sciences, vol. 1060.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia - A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238-251.

Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica*, 7, 80-88.

Bachot, J., Gevers, W., Fias, W., & Roeyers, H. (2005). Number sense in children with visuospatial disabilities: Orientation of the mental number line. *Psychology Science, 47*(1), 172-183.

Bangerter, A., & Heath, C. (2004). The Mozart effect: Tracking the evolution of a scientific legend. *British Journal of Social Psychology, 43*, 605-623.

Bätchtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia, 36*, 731-735.

Beecham, R., Wilson, S. J., & Reeve, R. A. (2006). *Cognitive bases of mathematical and musical competence: Shared or independent processes?* Paper presented at the 9th International Conference on Music Perception and Cognition, Bologna.

Béland, R., Caplan, D., & Nespoulous, J.-L. (1990). The role of abstract phonological representations in word production: Evidence from phonemic paraphasia. *Jounal of Neurolinguistics, 5*, 125-164.

Belin, P., Fecteau, S., & Bédard, S. (2004). Thinking the voice: Neural correclates of voice perception. *Cognitive Science, 8*(3), 129-136.

Ben Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification - Role of stimulus difference. *Perception & Psychophysics, 57*(8), 1151-1162.

Ben-Artzi, E., & Marks, L. (1999). Processing linguistic and perceptual dimensions of speech: Interactions in speeded classification. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 579-595.

Bertelson, P. (1972). Listening from left to right versus right to left. *Perception*, 1, 161-165.

Besson, M. (1998). Meaning, structure, and time in language and music. *Cahiers de Psychologie Cognitive-Current Psychology of Cognition, 17*(4-5), 921-950.

Besson, M., & Schön, D. (2001). Comparison between language and music. In I. Peretz & R. Zatorre (Eds.), *The Biological Foundations of Music* (Vol. 930, pp. 232-258). New York: Annals of the New York Academy of Sciences.

Besson, M., Faïta, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science, 9*(6), 494-498.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience, 25*(3-4), 399-410.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction-time and musical expectancy - priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403-410.

Bharucha, J. J., Curtis, M., & Paroo, K. (2006). Varieties of musical experience. *Cognition, 100*, 131-172.

Biederman, I., & Checkosky, S. F. (1970). Processing redundant information. *Journal of Experimental Psychology, 83*, 486-490.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics, 59*, 1098-1107.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D. A., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition, 81*(1), B11-B20.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice.* Cambridge, MA: MIT Press.

Blacking, J. (1995). *Music, Culture and Experience.* Chicago: University of Chicago Press.

Boatman, D., Hall, C., Golsdstein, M.H., Lesser, R., & Gordon, B. (1997). Neuroperceptual differences in consonant and vowel discrimination: As revealed by direct cortical electrical interference. *Cortex, 33*, 83-98.

Boersma, P., & Weenink, D. (2007). Praat: Doing phonetics by computer (Version 5.0). Retrieved December 10, 2007, from http:www.praat.org

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science, 16*(6), 451-459.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens and eggs. *Psychological Science, 18*(10), 924-925.

Bonnel, A. M., Faïta, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: Evidence for independent processing. *Perception & Psychophysics, 63*(7), 1201-1213.

Boudelaa, S., & Marlsen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing : Evidence from Modern Standard Arabic. *Proceedings of SWAP.* Nijmegen, pp. 23-26.

Boudelaa, S., & Marlsen-Wilson, W. D. (2004). Abstract morphemes and lexical representation : The CV-Skeleton in Arabic. *Cognition, 92*, 271-303.

Brochard, R., Dufour, A., & Despres, O. (2004). Effect of musical expertise on visuospatial abilities: Evidence from reaction times and mental imagery. *Brain and Cognition, 54*, 103-109.

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M. H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience, 18*(12), 1959-1972.

Caclin, A., Giard, M-H, Smith, B.K., & McAdams, S. (2007). Interactive processing of timbre dimensions : A Garner interference study. *Brain Research, 1138*, 159-170.

Caclin, A., McAdams, S., Smith, B.K., Giard, M.-H. (2008). Interactive processing of timbre dimensions : An exploration with even-related potentials. *Journal of Cognitive Neuroscience, 20*(1), 49-64.

Calabria, M., & Rossetti, Y. (2005). Interference between number processing and line bisection: a methodology. *Neuropsychologia, 43*, 779-783.

Calvert, S. L., & Billingsley, R. L. (1998). Young children's recitation and comprehension of information presented by songs. *Journal of Applied Developmental Psychology, 19*(1), 97-108.

Canter, G. J., Trost, J. E., & Burns, M. S. (1985). Constrasting speech patterns in apraxia of speech and phonemic paraphasia. *Brain and Language, 24*, 204-222.

Caramazza, A., Chialant, D., Capasso, R., & Miceli, G. (2000). Separable processing of consonants and vowels. *Nature, 403*, 428-430.

Carrell, T. D., Smith, L. B., & Pisoni, D. B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception & Psychophysics, 29*, 1-10.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimentional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25*(3-4), 195-210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. (2007). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport, 18*(8), 1963-1967.

Chater, N. (2003). How much can we learn from double dissociations? *Cortex, 39*(1), 167-169.

Cheek, J. M., & Smith, L. R. (1999). Music training and mathematics achievement. *Adolescence, 34*, 759-761.

Chen, S., & Melara, R.D. (2003). Effects of Garner interference between integral dimensions of auditory timbre and pitch. *Psychophysiology, 40* (S1), S33.

Cho, Y.S., & Proctor, R.W. (2003). Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychonomic Bulletin & Review*, 10, 45-73.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton & Co.

Cohen, J., Macwhinney, B., Flatt, M., & Provost, J. (1993). Psyscope - an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods Instruments & Computers, 25*(2), 257-271.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neurosciences, 22*, 319-349.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3*(3), 115-120.

Costa-Giomi, E. (2004). *Effects of starting age of music lessons on spatial performance.* Paper presented at the 8th International Conference on Music Perception and Cognition, Evanson, IL.

Crowder, R. G. (1991). Imagery and the perception of musical timbre. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure* (pp. 145-156). Washinton, DC: American Psychological Association.

Crowder, R. G., Serafine, M. L., & Repp, B. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition, 18*(5), 469-476.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cognition, 46*, 373-382.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141-201.

Dalla Bella, S., Dunlop, T., Dawe, L., Humphrey, K., & Peretz, I. (1999). The Mozart effect revisited. *Journal of Cognitive Neuroscience*, 103.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America, 121*(2), 1182-1189.

Dehaene, S. (2001). Precis of the number sense. *Mind & Language*, 16, 16-36.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122, 371-396.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 626-641.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*, 487-506.

Delattre, P., Liberman, A., & Cooper, F. (1955). Acoustic loci and transitionnal cues for consonants. *Journal of the Acoustical Society of America, 27*, 769-773.

Demolin, D. (2007). *L'évolution du langage et de la musique dans une perspective comparative.* Présentation lors du XXXème Symposium de l'APSLF: Musique, Langage et Emotion, Bruxelles.

D'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A.-M., Soressi, M., Besson, F., Maureille, B., Nowell, A., Lakarra, J., Backwell, L., & Julien, M. (2003). Archeological evidence for the emergence of language, symbolism, and music - An alternative multidisciplinary perspective. *Journal of World Prehistory, 17*, 1-70.

Di Pietro, M., Laganaro, M., Leeman, B. & Schnider, A. (2004). Receptive amusia : Temporal auditory processing deficit in a professional musician following a left temporo-parietal lesion. *Neuropsychologia, 42*(7), 868-877.

Dickson, D., & Grant, L. (2003). Physics karaoke: Why not? *Physics Education, 38*(4), 320-323.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55*, 149-179.

Donchin, E. (1981). Surprise! ... Surprise? *Psychophysiology, 18*(5), 493-513.

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience, 10*(7), 915-921.

Dowling, W. J., &. Harwood, D. L. (1986). *Music cognition*, Orlando: Academic Press.

Ehrlé, N., Samson, S., & Baulac, M. (2001). Processing of rapid auditory information in epileptic patients with left temporal lobe damage. *Neuropsychologia, 39*, 525-531.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology and Neuro-otology, 5*, 151-166.

Escoffier, N., & Tillmann, B. (2006). *Tonal function modulates speed of visual processing.* Paper presented at the 9th International Conference on Music Perception and Cognition, Bologna.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. *The Journal of Neuoscience, 15*(4), 2748-2755.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception ? *Psychological Review, 105*(3), 482-498.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America, 97*, 1893-1904.

Fias, W. (2001). Two routes for the processing of verbal numbers: evidence from the SNARC effect. *Psychological Research*, 65, 250-259.

Fias, W., & Fischer, M.H. (2005). Spatial representation of numbers. In J.I.D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 43-54). New York: Psychology Press.

Fias, W., Brysbaert, M., Geypens, F., & d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition, 2*, 95-110.

Fischer, M. H. (2001). Number processing induces spatial performance biases. *Neurology, 57*, 822-826.

Fischer, M. H. (2004). Spatial representation in number processing : Evidence from a pointing task. *Visual Cognition, 10*, 493-508.

Fischer, M. H. (2006). The future for SNARC could be STARK. *Cortex, 42*(8), 1066-1068.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature Neuroscience, 6*(6), 555-556.

Fischer, M. H., Warlop, N., Hill, R. L., & Fias, W. (2004). Oculomotor bias induced by number perception. *Experimental Psychology, 51*, 91-97.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences, 4*(7), 258-267.

Fitch, W. T. (2002). Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx. In A. Wray (Ed.), *The Transition to Language* (pp. 21-45). Oxford: Oxford University Press.

Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition, 100*(1), 173-215.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*(1477), 1669-1675.

Fitts, P. M., & Biederman, L. (1965). S-R compatibility and information reduction. *Journal of Experimental Psychology, 69*, 408-412.

Fitts, P.M., & Seeger, C.M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 46, 199-210.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, Mass. : Bradford Books, MIT Press.

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, Mass.: Bradford Books, M.I.T. Press.

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying "tone-deafness". *Brain, 127*, 801-810.

Franco, A. (2006). Les interactions entre les représentations mentales des chiffres et des notes de musique. *Mémoire de fin d'études non publié,* dirigé par R. Kolinsky, Faculté des Sciences psychologiques et de l'Education, ULB.

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. *Neuroscience and Biobehavioral Reviews, 25*, 355-373.

Frost, R., Deutsch, A., Gilboa, O., Tannenbaum, M., & Marlsen-Wilson, W. D. (2000). Morphological priming : Dissociation of phonological, semantic and morphological factors. *Memory and Cognition, 28*(8), 1277-1288.

Fry, D., Abramson, A., Eimas, P., & Liberman, A. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171-189.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377.

Gallistel, C.R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43-74.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Neuroreport, 9*, 2115-2119.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12*, 207-222.

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: Further evidence for the single-route hypothesis. *Journal of Experimental Psychology-Human Perception and Performance, 28*(4), 854-867.

Gardiner, M. F., Fox, A., Knowles, F., & Jeffrey, D. (1996). Learning improved by arts training. *Nature, 381*, 284.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, Maryland: Lawrence Erlbaum Associates.

Garner, W. R. (1977). The effect of absolute size on the separability of the dimensions of size and brightness. *Bulletin of the Psychonomic Society, 9*, 380-382.

Garner, W. R. (1978a). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8*, 98-123.

Garner, W. R. (1978b). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General, 107*, 287-308.

Garner, W. R. (1981). The analysis of unanalyzed perceptions. In M. Kubovy & J. R. Pomerantz (Eds.), *Perception and organization* (pp. 199-139). Hillsdale, N.J.: Lawrence Erlbaum.

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition and development : interactional analyses* (pp. 1-37). Hillsdale, N.J.: Erlbaum.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Gevers, W., & Lammertyn, J. (2005). The hunt for SNARC. *Psychology Science*, 47, 10-21.

Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., & Fias, W. (2006). Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychologica, 122*, 221-233.

Gevers, W., Ratinckx, E., De Baene, W., & Fias, W. (2006). Further evidence that the SNARC effect is processed along a dual-route architecture: Evidence from the lateralized readiness potential. *Exp Psychol, 53*(1), 58-68.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition,* 87, B87-B95.

Gevers, W., Reynvoet, B., & Fias, W. (2004). The mental representation of ordinal sequences is spatially organized: Evidence from days of the week. *Cortex, 40*(1), 171-172.

Giaquinto, M. (2001). What cognitive systems underlie arithmetical abilities? *Mind & Language*, 16, 56-68.

Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., & Pernier, J. (1995). Separate representation of stimulus frequency, intensity and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience, 7*, 133-143.

Göbel, S., Walsh, V., & Rushworth, M. (2001). The mental number line and the human angular gyrus. *Neuroimage*, 14, 1278-1289.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford: Blackwell.

Goldstone, R. L. (1999). Similarity. In R. A. Wilson & F. C. Keil (Eds.), *MIT Encyclopedia of the Cognitive Sciences* (pp. 763-765). Cambridge: MA: M.I.T. press.

Gottwald, R. L., & Garner, W. R. (1972). Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks. *Perception & Psychophysics, 11*, 179-182.

Gottwald, R., & Garner, W. R. (1975). Filtering and condensation tasks with integral and separable dimensions. *Perception and Psychophysics, 2*, 50-55.

Grau, J. W., & Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions - Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117*(4), 347-370.

Graziano, A. B., Peterson, M., & Shaw, G. L. (1999). Enhanced learning of proportional math through music training and spatial-temporal reasoning. *Neurological Research, 21*, 139-152.

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Gregg, J., & Scherer, R. C. (2006). Vowel intelligibility in classical singing. *Journal of Voice, 20*, 198-210.

Griffiths, T. D., Rees, A., Witton, C., Cross, P. M., Shakir, R. A., & Green, G. G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infraction: A psychophysical study. *Brain, 120*, 785-794.

Handel, S. (1989). *Listening : An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Harris, J. (2006). The phonology of being understood : Further arguments against sonority. *Lingua, 116*, 1483-1494.

Haueisen, J., & Knösche, T.R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience, 13*, 786-792.

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: a comparative perspective. *Nature Neuroscience, 6*(7), 663-668.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569-1579.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition*, B53-B64.

Hébert, S., & Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *Brain and Cognition, 46*(1-2), 169-175.

Hébert, S., Racette, A., Gagnon, L., & Peretz, I. (2003). Revisiting the dissociation between singing and speaking in expressive aphasia. *Brain, 126*, 1838-1850.

Hewson-Stoate, N., Schönwiesner, M., & Krumbholtz, K. (2006). Vowel processing evoke a large sustained response anterior to primary auditory cortex. *European Journal of Neuroscience, 24*, 2661-2671.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*, 67-99.

Hickok, G., Bellugi, U., & Klima, E. (1998). The neural organization of language: Evidence from sign language aphasia. *Trends in Cognitive Sciences, 2*, 129-136.

Hirschfeld, L. A., & Gelamn, S. A. (1994). *Mapping the Mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.

Ho, Y., Cheung, M., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology, 17*(3), 439-450.

Holender, D. (1992). Expectancy effects, congruity effects, and the interpretation of response latency measurement. In Alegria, J., Holender, D., Junça de Morais, J., & Radeau, M. (Eds), *Analytic Approaches to Human Cognition*, Elsevier.

Hommet, C., Billard, C., de Toffol, B., & Autret, A. (2003). Functional brain lateralization in children: developmental theories and implications for developmental diseases. *Revue Neurologique, 159*(11), 997-1007.

Hubbard, E.M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, 6, 435-447.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360.

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *Journal of Neuroscience, 27*(47), 13028-13032.

Hyde, K. L., Peretz, I., & Zatorre, R. (sous presse). Evidence for the role of the right auditory cortex un fine pitch resolution. *Neuropsychologia.*

Hyde, K. L., Zatorre, R. J., Griffiths, T. D., Lerch, J. P., & Peretz, I. (2006). Morphometry of the amusic brain: a two-site study. *Brain, 129*(10), 2562-2570.

Hyman, R., & Well, A. (1967). Judgments of similarity and spatial models. *Perception and Psychophysics, 2*, 233-248.

Idrissi, A., & Kehayia, E. (2004). Morphological units in the Arabic mental lexicon : Evidence from an individual with deep dyslexia. *Brain and Language, 90*, 183-197.

Ikeda, K., Hashimoto, S., Otomo, K., & Kanno, A. (2002). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters, 321*, 133-136.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory and Cognition*, 32, 662-673.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*(1), 33-72.

Jacobsen, T. (2004). Mismatch negativity to frequency changes: no evidence from human event-related brain potentials for categorical speech processing of complex tones resembling vowel formant structure *Neuroscience Letters, 362*, 204-208.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive processing of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*, 654-659.

Jamieson, R. K., Thompson, W. F., Cuddy, L. L., & Mewhort, D. J. K. (2003). Do conjunction errors in auditory recognition imply feature migration? *Canadian Journal of Experimental Psychology, 57*(2), 125-130.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex, 16*(9), 1266-1275.

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage, 15*, 733-746.

Jerger, S., Elizondo, R., Dinh, T., Sanchez, P., & Chavira, E. (1994). Linguistic influences on the auditory processing of speech by children with normal-hearing or hearing impairment. *Ear and Hearing, 15*(2), 138-160.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123*, 155-163.

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45*, 82-102.

Jones, T. C., Bartlett, J. C., & Wade, K. A. (2006). Nonverbal conjunctions errors in recognition memory: Support for familiarity but not for feature bundling. *Journal of Memory and Language, 55*, 138-155.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*(5), 770-814.

Kaganovich, N., Francis, A.L., & Melara, R.D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research, 1114,* 161-172.

Kaufmann, J. M., & Schweinberger, S. R. (2005). Speaker variations influence speechreading speed for dynamic faces. *Perception, 34*(5), 595-610.

Keele, S. W. (1970). Effects of input and output modes on decision time. *Journal of Experimental Psychology, 85*, 157-164.

Kemler-Nelson, D. G. (1993). Processing integral dimensions - the whole Vvew. *Journal of Experimental Psychology: Human Perception and Performance, 19*(5), 1105-1113.

Keus, I. M., & Schwarz, W. (2005) Searching for the functional locus of the SNARC effect: Evidence for a response-related origin. *Memory and Cognition, 33,* 681-695.

Keus, I. M., Jenk, K., & Schwarz, W. (2005). Psychophysiological evidence that the SNARC effect has its functional locus in a response selection stage. *Cognitive Brain Research, 24*, 48-56.

Kimura, D. (1964). Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology, 16*, 355-358.

Kishon-Rabin, L., Amir, O., Vexler, Y., & Zaltz, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic & Clinical Physiology & Pharmacology*, 12, 125-143.

Klein, D., Zatorre, R., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage, 13*, 646-653.

Knosche, T.R., Maess, B., Nakamura, A, & Friederici, A.D. (2005). Human communication investigated with magnetoencephalography: Speech, music and gesture. *International Review of Neurobiology, 68*, 79-120.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical "language-network" serves the processing of music. *Neuroimage, 17*(2), 956-966.

Koelsch, S., Gunter, T. C., Wittgoth, M., & Sammler, D. (2005). Interactions between syntax processing in language and music: an ERP study. *Journal of Cognitive Neuroscience, 17*, 1565-1577.

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

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

Kolinsky, R. (1989). The development of separability in visual perception. *Cognition, 33*(3), 243-284.

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (under review). Processing interactions between lyrics and tune: Vowels sing but consonants speak. *Cognition.*

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility. A model and taxonomy. *Psychological Review*, 97, 253-270.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Pschological Bulletin, 126*(2), 159-179.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 739-751.

Ladd, D. R., Silverman, K. E. A., Tolmitt, F., Bergmann, G., & Scherer, K. (1985). Evidence for the independent function of intonation contour type, voice quality and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America, 78*, 435-444.

Lê, S., Raufaste, E., Roussel, S., Puel, M., & Démonet, J. F. (2003). Implicit face perception in a patient with visual agnosia? Evidence from behavioural and eye-tracking analyses. *Neuropsychologia, 41*(6), 702-712.

Lee, L., & Nusbaum, H. C. (1993). Processing Interactions between segmental and suprasegmental information in native speakers of English and Mandarine Chinese. *Perception & Psychophysics, 53*(2), 157-165.

Levänen, S., Hari, R., McEvoy, L., & Sams, M. (1993). Responses of the human auditory cortex to changes in one vs. two stimulus features. *Experimental Brain Research, 97*, 177-183.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *Neuroimage, 20*(4), 2142-2152.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Lidji, P. (2007). Intégralité et séparabilité : Revue et application aux interactions entre paroles et mélodies dans le chant. *L'Année Psychologique, 107*(4), 659-694.

Lidji, P., Franco, A., Kolinsky, R., & Peretz, I. (2006). *Un, deux, trois, chantons! Interactions entre hauteur des notes et chiffres.* Poster présenté au 74ème Congrès de l'ACFAS, L'interdisciplinarité dans les sciences et technologies de la musique, Montréal.

Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance, 33*(5), 1189-1207.

Lidji, P., Kolinsky, R., Peretz, I., Lafontaine, H., & Morais, J. (in preparation). Integration effect and illusory conjunctions in memory for lyrics and tune: Vowels sing but consonants swing.

Lidji, P., Jolicoeur, P., Moreau, P., Connolly, J. F., Kolinsky, R. & Peretz, I. (in preparation). Early integration of vowel and pitch processing: A Mismatch Negativity study.

Lieberman, P. (1968). Primate vocalization and human linguistic abilities. *Journal of the Acoustical Society of America, 44*(6), 1574-1584.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Cambridge: Harvard University Press.

Lieberman, P., & Blumstein, S. E. (Eds.). (1988). *Speech physiology, speech perception, and acoustic phonetics.* Cambridge: Cambridge University Press.

Lieberman, P., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

Liégeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain, 121*, 1853-1867.

Lippa, Y., & Adam, J.J. (2001). Orthogonal stimulus-response compatibility resulting from spatial transformations. *Perception and Psychophysics, 63*, 156-174.

Lokhead, G. R., & Pomerantz, J. R. (1991). *The perception of structure: Essays in honnor of Wendell R. Garner*. Washinton, D.C.: American Psychological Association.

Lorch Jr., R. F., & Meyers, J. L. (1990). Regression analyses of repeated measures data in cognition research. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 16, 149-157.

Lu, C.H., & Proctor, R.W. (1995). The influence of irrelevant location information on performance: A review of the Simon and spatial Stroop effects. *Psychonomic Bulletin & Review*, 2, 174-207.

Luck, S. J. (2005). *An introduction to the event-related potentials technique*. Cambridge, Ma.: MIT Press.

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science*, 14, 296-301.

Macmillan, N. A., & Ornstein, A. S. (1998). The mean-integral representation of rectangles. *Perception & Psychophysics, 60*(2), 250-262.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioural and Brain Sciences, 21*, 499-546.

MacNeilage, P. F., & Davis, B. (2000). On the origin of internal structure of word forms. *Science, 288*, 527-531.

Maddieson, I. (2005). Tone. In M. Haspelmath, M. S. Dryer, D. Gi & B. Comrie (Eds.), *The World Atlas of Language Structures* (pp. 58-61). New York: Oxford University Press.

Maddox, W. T. (1992). Perceptual and Decisional Separability. In E. G. Ashby (Ed.), *Multidimensional Models of Perception and Cognition* (pp. 147-180): Erlbaum.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 795-817.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199-211.

Marozeau, J., de Cheveigné, A., McAdams, S., & Winsberg, S. (2003). The dependency of timbre on fundamental frequency. *Journal of the Acoustical Society of America, 114*(5), 2946-2957.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19*, 1453-1463.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography & Clinical Neurophysiology, 62*(3), 203-208.

McCarthy, J. J. (1985). *Formal problems in semitic phonology and morphology*. New York: Garland.

McCrean, C. R., & Morris, R. J. (2005). Comparisons of voice onset time for trained male singers and male nonsingers during speaking and singing. *Journal of Voice, 19*(3), 420-430.

McDermott, J., & Hauser, M. (2004). Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition, 94*(2), B11-B21.

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception, 23*(1), 29-59.

McKelvie, P., & Low, J. (2002). Listening to Mozart does not improve children's spatial ability: Final curtains for the Mozart effect. *British Journal of Developmental Psychology, 20*, 241-258.

McMullen, E., & Saffran, J. R. (2004). Music and Language: A developmental comparison. *Music Perception, 21*(3), 289-311.

Mehler, J., Peña, M., Nespor, M., & Bonatti, L. L. (2006). The "soul" of language does not use statistics: Reflections on vowels and consonants. *Cortex, 42*(6), 846-854.

Melara, R. D. (1989). Dimensional interaction between color and pitch. *Journal of Experimental Psychology: Human Perception and Performance, 15*(1), 69-79.

Melara, R. D., & Marks, L. E. (1990a). Dimensional interactions in language processing - Investigating directions and levels of crosstalk. *Journal of Experimental Psychology-Learning Memory and Cognition, 16*(4), 539-554.

Melara, R. D., & Marks, L. E. (1990b). Interaction among Auditory Dimensions - Timbre, Pitch, and Loudness. *Perception & Psychophysics, 48*(2), 169-178.

Melara, R. D., & Marks, L. E. (1990c). Perceptual primacy of dimensions - Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 398-414.

Melara, R. D., & Marks, L. E. (1990d). Processes underlying dimensional interactions - correspondences between linguistic and nonlinguistic dimensions. *Memory & Cognition, 18*(5), 477-495.

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General, 116*, 323–336.

Melara, R. D., Marks, L. E., & Potts, B. C. (1993). Early-Holistic Processing or Dimensional Similarity. *Journal of Experimental Psychology: Human Perception and Performance, 19*(5), 1114-1120.

Mendez, M. (2001). Generalized auditory agnosia with spared music recognition in a left-hander. Analysis of a case with right temporal stroke. *Cortex, 37*, 139-150.

Mikumo, M. (1998). Encoding strategies for pitch information. *Japanese Psychological Monographs*, 27. The Japanese Psychological Association.

Miller, J. (1978). Interactions in processing segmental and suprasegmental features of speech. *Perception and Psychophysics, 24*(2), 175-180.

Mithen, S. J. (2005). *The singing Neanderthals: The origins of music, language and body.* London: Weindenfeld & Nicholson.

Monaghan, P., & Shillcock, R. (2003). Connectionist modelling of the separable processing of consonants and vowels. *Brain and Language, 86*, 83-98.

Moreno, S., & Besson, M. (2006). Musical training and language-related brain electrical activity in children. *Psychophysiology, 43*(3), 287-291.

Morin, R. E., Forrin, B., & Archer, W. (1961). Information processing behavior: The role of irrelevant stimulus information. *Journal of Experimental Psychology, 61*, 89-96.

Morrongiello, B. A., & Roes, C. L. (1990). Childrens' memory for new songs - Integration or independent storage of words and tunes. *Journal of Experimental Child Psychology, 50*(1), 25-38.

Müller, D., & Schwarz, W. (2007a). Exploring the mental number line: Evidence from a dual-task paradigm. *Psychological Research, 71*, 598-613.

Müller, D., & Schwarz, W. (2007b). Is there an internal association of numbers to hands? The task set influences the nature of the SNARC effect. *Memory and Cognition, 35*(5), 1151-1161.

Näätänen, R. (1992). *Attention and Brain function.* Hillsdale, N.J.: Erlbaum.

Näätänen, R. (1995). The Mismatch Negativity - a powerful tool for cognitive neuroscience. *Ear and Hearing, 16*(1), 6-18.

Näätänen, R., & Alho, K. (1995). Generators of electrical and magnetic mismatch responses in humans. *Brain Topography, 7*(4), 315-320.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (in press). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*.

Näätänen, R., Simpson, L., & Loveless, N. E. (1982). Stimulus deviance and evoked potentials. *Biological Psychology, 14*, 53-98.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science, 10*(4), 370-373.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98*(1), 13-30.

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical decision. *Cognitive Development, 22*, 271-279.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different role of vowels and consonants in language processing and language acquisition. *Lingue e Linguaggio, 2*, 221-247.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127-162.

Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance : II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology, 49*, 58-117.

Notebaert, W., Gevers, W., Verguts, T., & Fias, W. (2006). Shared spatial representations of numbers and space : the reversal of the SNARC and the Simon effect. *Journal of Experimental Psychology : Human Perception and Performance,* 32, 1197-1207.

Nuerk, H. C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology*, 57, 835-863.

Nuerk, H. C., Wood, G., & Willmes, K. (2005). The universal SNARC effect. The association between number magnitude and space is amodal. *Experimental Psychology*, 52, 187-194.

Ohnishi, T, Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A, & Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex, 11* (8), 754-760.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America, 119*(3), 1727-1739.

Owren, M. J., Seyfarth, R. M., & Cheney, D. L. (1997). The acoustic features of vowel-like grunts calls in chacuma baboons (papio cynocephalus ursinus): Implications for production processes and functions. *Journal of the Acoustical Society of America, 101*, 2951-2963.

Paavilainen, P., Mikkonen, M., Kilpeläinen, M., Lehtinen, R., Saarela, M., & Tapola, L. (2003). Evidence for the different additivity of the temporal and frontal generators of mismatch negativity: A human auditory event-related potential study. *Neuroscience Letters, 349*, 79-82.

Paavilainen, P., Valppu, S., & Näätänen, R. (2001). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: $1+1 \approx 2$ but $1+1+1 < 3$. *Neuroscience Letters, 301*(3), 179-182.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*, 115-138.

Pantev, C., Roberts, L., Schulz, M., Engelien, A., & Ross, B. (2000). Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, 12, 1-6.

Patching, G. & Quinlan, P. (2002). Garner and Congruence Effetcts in the Speeded Classification of Bimodal Signals. *Journal of Experimental Psychology : Human Perception and Performance, 28* (4), 755-775.

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

Patel, A. D. (2007). *Language, music, and the brain: a resource-sharing framework.* Paper presented at the Conference « Language and Music as Cognitive Systems », Cambridge, UK.

Patel, A. D. (2008). *Music, Language, and the Brain.* New York: Oxford University Press.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition, 59*, 310-313.

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

Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences, 11*(9), 369-372.

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and Cognition of Music* (pp. 191-215). London: Erlbaum Psychology Press.

Patel, A. D., Peretz, I., Tramo, M., & Labrecque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language, 61*(1), 123-144.

Peretz, I. (2001). Brain specialization for music: New evidence from congenital amusia. In *The Biological Foundations of Music* (Vol. 930, pp. 153-165). New York: New York Academy of Science.

Peretz, I. (2002). Brain specialization for music. *Neuroscientist, 8*(4), 372-380.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition, 100*(1), 1-32.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*(2), 185-191.

Peretz, I., Belleville, S., & Fontaine, F. S. (1997). Dissociations entre musique et langage après atteinte crébrale: un nouveau cas d'amusie sans aphasie. *Revue Canadienne de Psychologie Expérimentale, 51*(4), 354-367.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders - The Montreal Battery of Evaluation of Amusia. *Neurosciences and Music, 999*, 58-75.

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

Peretz, I., Cummings, S., & Dubé, M. P. (2007). The genetics of congenital amusia (or tone deafness): A familiy aggregation study. *American Journal of Human Genetics, 81*, 582-588.

Peretz, I., Gagnon, L., Hébert, S., & Macoir, J. (2004). Singing in the brain: Insights from cognitive neuropsychology. *Music Perception, 21*(3), 373-390.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences, 7*(8), 362-367.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination - a neuropsychological perspective. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46*(2), 301-325.

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

Peretz, I., & Morais, J. (1989). The modularity of music. In S. McAdams & I. Deliège (Eds.), *Music and Cognitive Sciences*. London: Gordon & Breach.

Peretz, I., Radeau, M., & Arguin, M. (2004). Two-way interactions between music and language: Evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition, 32*(1), 142-152.

Peretz, I., & Zatorre, R.J. (2005). Brain organization for music processing. *Annual Review of Psychology*, 56, 89-114.

Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials. I. The nature of the response. *Electroencephalography and Clinical Neurophysiosology, 45*, 186-197.

Pinker, S. (1997). *How the mind works*. New York London: W.W. Norton & Company.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*, 201-236.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology-Human Perception and Performance, 20*(5), 976-986.

Plaut, D. C. (1995). Double dissociation without modularity - Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology, 17*(2), 291-321.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymetric sampling in time". *Speech Communication, 41*, 245-255.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*, 2128-2148.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112*, 516-540.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception & Psychophysics, 14*, 565-569.

Pomerantz, J. R., Pristach, E. A., & Carson, C. E. (1989). Attention and object perception. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 53-89). Hillsdale, N.J.: Erlbaum.

Pomerantz, J., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 422-435.

Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics, 18*, 355-361.

Posner, M. I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71*, 491-504.

Potts, B. C., Melara, R. D., & Marks, L. E. (1998). Circle size and diameter tilt: A new look at integrality and separability. *Perception & Psychophysics, 60*(1), 101-112.

Poulin-Charronnat, B., Bigand, E., Madurell, F., & Peereman, R. (2005). Musical structure modulates semantic priming in vocal music. *Cognition, 94*(3), B67-B78.

Price, C., Thierry, G., & Griffiths, T. (2006). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9*(6), 271-276.

Proctor, R.W., & Cho, Y.S. (2006). Polarity Correspondence: A general principle for performance of speeded binary classification task. *Psychological Bulletin, 132*, 416-442.

Prunet, J.-F., Béland, R., & Idrissi, A. (2000). The mental representation of Semitic words. *Linguistic Inquiry, 31*(4), 609-648.

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*(2), 242-253.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain, 129*, 2571-2584.

Rainey, D. W., & Larsen, J. D. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception, 20*(2), 173-186.

Rauscher, F. H., & Hinton, S. (2006). The Mozart effect: Music listening is not music instruction. *Educational Psychologist, 41*(4), 233-238.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365*(6447), 611.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters, 185*(1), 44-47.

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour, 57*, 583-592.

Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech-perception - a cross-linguistic study. *Journal of Phonetics, 18*(4), 481-495.

Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *Journal of the Acoustical Society of America*, 43, 1260–1266.

Romani, C., & Calabrese, A. (1998). Syllabic constraints on the phonological errors of an aphasic patient. *Brain and Language, 64*, 83-121.

Ross, D., Choi, J., & Purves, D. (2007). Musical intervals in speech. *Proceedings of the National Academy of Science, 104*(23), 9852-9857.

Rossetti, Y., Jacquin-Courtois, S., Rode, G., Otal, H., Michel, C., & Boisson, D. (2004). Does action make the link between number and space representation? Visuo-manual adaptation improves number bisection in unilateral neglect. *Psychological Science, 15*(6), 426-430.

Rumelhart, D., McClelland, J., & the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* London: MIT Press.

Rusconi, E., Kwan, B., Giordano, B.L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition,* 99, 113-129.

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics, 67*(4), 559-568.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.

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

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606-621.

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *Neuroreport, 17*(18), 1893-1897.

Samson, S. (1999). Musical function and temporal lobe structures: A review of brain lesion studies. *Journal of New Music Research, 28*(3), 217-228.

Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal-lobe lesion - Evidence for dual encoding. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*(4), 793-804.

Sandrini, M., Rossini, P., & Miniussi, C. (2004). The differential involvement of inferior parietal lobule in number comparison: a rTMS study. *Neuropsychologia,* 42, 1902-1909.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Cognitive Brain Research, 16,* 250-256.

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron, 20,* 185-195.

Schellenberg, E. G. (2001). Music and nonmusical abilities. In Zatorre R.J. & Peretz I. (Eds), *The Biological Foundations of Music* (pp. 355-371). New York : The New York Academy of Sciences.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*(8), 511-514.

Schellenberg, E. G. (2006). Long-term positive association between music lessons and IQ. *Journal of Educational Psychology, 92*(2), 457-468.

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review, 1*(2), 191-201.

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infants listeners. *Psychological Science, 7*(272-277).

Schlaug, G. (2001). The brain of musicians: A model for functional and structural adaptation. In Zatorre, R.J., & Peretz, I. (Eds). *The biological foundations of Music.* New York: Annals of the New York Academy of Sciences, 930, 281-299.

Schmithorst, V.J., & Holland, S.K. (2003). The effect of musical training on music processing: a functional magnetic resonance imaging study in humans. *Neuroscience Letters*, 348, 65-68.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106*(2), 975-983.

Schön, D., Gordon, R., & Besson, M. (2005). Musical and linguistic processing in song perception. In G. Avanzini, S. Koelsch, L. Lopez & M. Majno (Eds.), *The Neurosciences and Music II: From Perception to Performance* (Vol. 1060, pp. 71-81). New York: Annals of the New York Academy of Sciences.

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

Schröger, E. (1995). Processing of auditory deviants with changes in one-stimulus versus 2-stimulus dimensions. *Psychophysiology, 32*(1), 55-65.

Schröger, E. (1996a). A neural mechanism for involuntary attention shifts to changes in auditory stimulation. *Journal of Cognitive Neuroscience, 8*, 527-539.

Schröger, E. (1996b). Interaural time and level differences: Integrated of separated processing? *Hearing Research, 96*, 191-198.

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behavior Research Methods, Instruments, & Computers, 30*(1), 131-145.

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line: comparing SNARC effects with saccadic and manual responses. *Perception and Psychophysics, 66*(4), 651-664.

Scotto di Carlo, N. (1993). Diction et musicalité. *Médecine des Arts, 5*, 4-11.

Scotto di Carlo, N. (2007a). Effect of multifactorial constraints on opera-singing intelligibility (I). *Journal of Singing, 63*, 1-13.

Scotto di Carlo, N. (2007b). Effect of multifactorial constraints on opera-singing intelligibility (II). *Journal of Singing, 63*, 559-567.

Scotto di Carlo, N., & Germain, A. (1985). A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica, 42*(4), 188-197.

Selkirk, E. O. (1982). The syllable. In H. Van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 337-384). Dordrecht: Foris Publications.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.

Semenza, C., Bencini, G. M., Bertella, L., Mori, I., Pignatti, R., Ceriani, F., Cherrick, D., & Caldognetto, E. M. (2007). A dedicated neural mechanism for vowel selection: A case of relative vowel deficit sparing the number lexicon. *Neuropsychologia, 45*(2), 425-430.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*(6), 695-703.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition, 16*(3), 285-303.

Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody text integration in memory for songs. *Journal of Memory and Language, 25*(2), 123-135.

Seron, X., & Pesenti, M. (2001). The number sense theory needs more empirical evidence. *Mind & Language*, 16, 76-88.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1*, 54-87.

Shepard, R. N. (1982a). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305-333.

Shepard, R. N. (1982b). Structural Representations of Musical Pitch. In Deutsch, D. (Ed.), *The Psychology of Music* (pp. 343-390). New York: Academic Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-730.

Shepp, B. E., Barrett, S. E., & Kolbet, L. L. (1987). The development of selective attention: Holistic perception versus resource allocation. *Journal of Experimental Child Psychology, 43*(2), 159-180.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM - Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145-166.

Simon, J.R., Sly, P.E., & Vilapakkam, S. (1981). Effect of compatibility of S-R mapping on reactions toward the stimulus source.*Acta Psychologica*, 47, 63-81.

Sinnott, J. M., & Gilmore, C. S. (2004). Perception of place-of-articulation information in natural speech by monkeys versus humans. *Perception & Psychophysics, 66*(8), 1341-1350.

Sinnott, J. M., & Williamson, T. L. (1999). Cab macaques perceive place of articulation from formant transition information? *Journal of the Acoustical Society of America, 106*(2), 929-937.

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

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *Journal of Neuroscience, 27*(14), 3799-3806.

Sommers, M. S., Moody, D. B., Prosen, C. A., & Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (Macaca fuscata). *Journal of the Acoustical Society of America, 91*(6), 3499-3510.

Sperber, D., & Hirschfeld, L. A. (2004). The cognitive foundations of cultural stability and diversity. *Trends in Cognitive Sciences, 8*(1), 40-47.

Squires, N. K., Squires, K. C., & Hillyards, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiosology, 38*, 387-410.

Staudt, M. (2007). (Re-)organization of the developing human brain following periventricular white matter lesions. *Neuroscience and Biobehavioral Reviews, 31*(8), 1150-1156.

Steele, K. M., Bass, K. E., & Crook, M. D. (1999). The mystery of the Mozart effect : Failure to replicate. *Psychological Science, 10*(4), 366-369.

Steele, K. M., Dalla Bella, S., Peretz, I., Dunlop, T., Dawe, L. A., Humphrey, G. K., Shannon, R. A., Kirby, J. L., & Olmstead, C. G. (1999). Prelude or requiem for the 'Mozart effect'? *Nature, 400*(6747), 827-827.

Stefurak, D. L., & Boynton, R. M. (1986). Independence of memory for categorically different colours and shapes. *Perception & Psychophysics, 39*, 164-174.

Steinke, W., Cuddy, L. L., & Holden, R. (1997). Dissociation of musical tonality and pitch memory from nonmusical cognitive abilities. *Canadian Journal of Experimental Psychology, 51*, 316-335.

Steinke, W., Cuddy, L. L., & Jakobson, L. (2001). Dissociations among functional subsystems governing melody recognition after rigth-hemisphere damage. *Cognitive Neuropsychology, 18*(5), 411-437.

Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *Neuroimage*, 20, 71-83.

Stewart, L., Nasralla, P., Verdonschot, R., & Lanipekun, J. (2007). *Automatic perception-action coupling in pianists.* Paper presented at the Society for Music Perception and Cognition 2007, Montréal.

Stewart, L., von Kriegstein, K., Warren, J. D., & Griffiths, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain, 129*(10), 2533-2553.

Stewart, L., & Walsh, V. (2007). Music perception: Sounds lost in space. *Current Biology, 17*(20), R892-R893.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception and Psychophysics*, 66, 183-195.

Stöcker, C., Sebald, A., & Hoffmann, J. (2003). The influence of response-effect compatibility in a serial reaction time task. *The Quarterly Journal of Experimental Psychology*, 56A, 685-703.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.

Suchan, B., Yaquez, L., Wunderlih, G., Canavan, A. G., Herzog, H., Tellmann, L., Hömberg, V., & Seitz, R. (2002). Neural correlates of visuospatial imagery. *Behavioural Brain Research, 131*, 163-168.

Sundberg, J. (1982). Perception of singing. In D. Deutsch (Ed.), *The Psychology of Music,* (pp. 59-98). New York: Academic Press.

Suthers, R. A. (1994). Variable asymetry and resonance in the avian vocal tract: A structural basis for individually distinct vocalizations. *Journal of Comparative Physiology, 157*, 457-466.

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language Processing in Chinese* (pp. 151-172). Amsterdam: Elsevier.

Tagliabue, M., Zorzi, M., & Umiltà, C. (2002). Cross-modal re-mapping influences the Simon effect. *Memory & Cognition*, 30, 18-23.

Takegata, R., Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *Neuroreport, 12*(3), 525-529.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters, 266*(2), 109-112.

Takegata, R., Syssoeva, O., Winkler, I., Paavilainen, P., & Näätänen, R. (2001). Common neural mechanism for processing onset-to-onset intervals and silent gaps in sound sequences. *Neuroreport, 12*(8), 1783-1787.

Takeuchi , A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, 113, 345-361.

Tervaniemi, M. (1999). Preattentive and attentive brain mechanisms of musical sound perception. *Psychophysiology, 36*, S17-S18.

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews, 43*(3), 231-246.

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., & Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *Neuroimage, 9*(3), 330-336.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., & Näätänen, R. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Human Brain Mapping, 10*(2), 74-79.

Thibaut, J. P., & Gelaes, S. (2002). The development of analytic and holistic processing. *L'Année Psychologique, 102*(3), 485-522.

Thomas, R. D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology, 40*(4), 318-341.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 128-140.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Mood, arousal and the Mozart effect. *Psychological Science, 12*(3), 248-251.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*(1), 46-64.

Tijsseling, A. G., & Gluck, M. A. (2002). A connectionist approach to processing dimensional interaction. *Connection Science, 14*(1), 1-48.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107*, 885-913.

Tillmann, B., Peretz, I., Bigand, E., & Gosselin, N. (2007). Harmonic priming in an amusic patient: The power of implicit tasks. *Cognitive Neuropsychology, 24*(6), 603-622.

Tomasello, M. (1998). *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, N.J.: Erlbaum.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes - Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81*(3), 755-764.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Toro, J. M., & Trobalòn, J. B. (2005). Satistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*(5), 867-875.

Tramo, M. J., Shah, G. D., & Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of Neurophysiology, 87*, 122-139.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechnanisms? *Cognition, 100*(1), 73-99.

Trehub, S. E., & Trainor, L. J. (1998). Singing to infants: Lullabies and play songs. *Advances in Infancy Research, 12*, 43-77.

Trehub, S. E., Thorpe, L. A., & Morrongiello, B. A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development, 58*, 741-749.

Trehub, S. E., Schellenberg, E. G., & Kamenetsky, S. B. (1999). Infants' and adults' perception of scale structure. *Journal of Experimental Psychology: Human Perception and Performance, 25*(4), 965-975.

Treisman, A. M., & Gelade, G. (1980). Feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107-141.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI* (pp. 333-361). Hillsdale, NJ: Erlbaum.

Ueda, K. & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using multidimensional scaling technique. *Journal of the Acoustical Society of America*, 82, 1193-1200.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (2001). What do double dissociations prove? *Cognitive Science, 25*(1), 111-172.

Vandenberghe, M., Schmidt, N., Fery, P., & Cleeremans, A. (2006). Can amnesic patients learn without awareness? New evidence comparing deterministic and probabilistic sequence learning. *Neuropsychologia, 44*, 1629-1641.

Vu, K.P.L., Proctor, R.W., & Pick, D.F. (2000). Vertical versus horizontal spatial compatibility: Right-left prevalence with bimanual responses. *Psychological Research*, 64, 25-40.

Vu, K.P.L., Proctor, R.W., & Urcuioli, P. (2003). Transfer effects of incompatible location-relevant mappings on a subsequent visual or auditory Simon task. *Memory & Cognition*, 31, 1146-1152.

Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, 13, 75–81.

Wallace, W. T. (1994). Memory for music - Effect of melody on recall of text. *Journal of Experimental Psychology: Learning Memory and Cognition, 20*(6), 1471-1485.

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

Warren, J. D., Warren, J. E., Fox, N. C., & Warrington, E. K. (2003). Nothing to say, something to sing: Primary progressive dynamic aphasia. *Neurocase, 9*, 140-155.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics, 64*(2), 198-207.

Weeks, D.J., Proctor, R.W.; & Beyak, B. (1995). Stimulus-response compatibility for vertically oriented spatial stimuli and horizontally oriented responses: Evidence for spatial coding. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*, 367-383.

Winkler, I., Lehtokoski, A., Alku, P., Vaino, M., Czigler, I., Csépe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory representations. *Cognitive Brain Research, 7*, 357-369.

Wolff, C., & Schröger, E. (2001). Human pre-attentive change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters, 311*, 37-40.

Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin, 59*(2), 83-95.

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

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics, 15*, 501-508.

Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception & Performance, 104*(1), 3-20.

Wood, G., Nuerk, H. C., & Willmes, K. (2006a). Crossed hands and the SNARC effect: a failure to replicate Dehaene, Bossini and Giraux (1993). *Cortex, 42*(8), 1069-1079; discussion 1119-1023.

Wood; G., Nuerk, H. C., & Willmes, K. (2006b). Variability of the SNARC effect: Systematic indvidual differences or just random error? *Cortex, 42*, 1119-1123.

Xu, Y., Gandour, J. T., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., Li, X., & Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping, 27*, 173-183.

Ylinen, S., Huotilainen, M., & Näätänen, R. (2005). Phoneme quality and quantity are processed independently in the human brain. *Neuroreport, 16*(16), 1857-1860.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*(441-517).

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

Zatorre, R., Belin, P., & Penhune, V. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*(1), 37-46.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14*, 1908-1919.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Zatorre, R. J., & Gandour, J. T. (in press). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society.*

Zatorre, R. J., & Samson, S. (1991). The role of the right temporal-lobe in short-term retention of pitch information. *Journal of Clinical and Experimental Neuropsychology, 13*(1), 67-68.

Zatorre, R.J., & Peretz, I. (2001). *The biological foundations of Music.* New York: Annals of the New York Academy of Sciences, 930.

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: the SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture, 5*(1-2), 165-190.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature, 417*, 138-139.

ANNEXE

Contribution des auteurs

1. Perceptual interactions between lyrics and tunes: Vowels sing but consonants speak.

Régine Kolinsky : Proposition de la problématique et de la méthode, recherche et analyse de la littérature, supervision lors la conceptualisation du matériel expérimental et lors des analyses, interprétation des résultats, rédaction.

Pascale Lidji : Recherche et analyse de la littérature, choix et préparation du matériel expérimental, collecte et analyse des données, interprétation des résultats, rédaction.

Isabelle Peretz : Supervision lors de la création du matériel synthétique, participation aux interprétations, corrections lors de la rédaction.

Mireille Besson : Contribution au développement de la problématique, participation aux interprétations, corrections lors de la rédaction.

José Morais : Supervision lors de la conceptualisation du matériel expérimental et de l'interprétation des résultats, participation à la revue de littérature et à la rédaction.

2. Integration and illusory conjunctions in memory for lyrics and tune: Vowels sing but consonants swing.

Pascale Lidji : Recherche et analyse de la littérature, développement de la méthode et du matériel expérimental, collecte des données, analyse des résultats, interprétation, rédaction.

Régine Kolinsky : Supervision du développement de la problématique, de la sélection du matériel expérimental et des analyses, interprétation, rédaction.

Isabelle Peretz : Supervision de la rédaction.

Hélène Lafontaine : Contribution à la sélection et à la préparation du matériel, collecte des données, contribution aux interprétations.

José Morais : Supervision de la sélection du matériel expérimental, corrections lors de la rédaction.

3. Integrated pre-attentive processing of vowel and pitch: A Mismatch negativity study.

Pascale Lidji : Recherche et analyse de la littérature, développement du paradigme expérimental, collecte des données, analyse et interprétation des résultats, rédaction.

Pierre Jolicoeur : Participation au développement du paradigme, supervision de l'analyse des résultats, contribution à l'interprétation et à la rédaction.

Régine Kolinsky : Supervision lors des études préliminaires, corrections lors de la rédaction.

Patricia Moreau : Participation à l'acquisition, à l'analyse et à l'interprétation des données, contribution à la création des figures, corrections lors de la rédaction.

John Connolly : Participation au développement du paradigme et suggestions pour l'analyse et l'interprétation des résultats.

Isabelle Peretz : Proposition de la méthode, supervision lors du développement du paradigme, de l'acquisition des données, de l'analyse des résultats, de leur interprétation et de la rédaction, corrections et suggestions lors de la rédaction.

4. Spatial Associations for Musical Stimuli: A Piano in the Head

Pascale Lidji : Recherche et analyse de la littérature, développement de la méthode, sélection et préparation du matériel, collecte des données, analyse des résultats, interprétation, rédaction.

Régine Kolinsky : Supervision lors du choix du matériel, de la collecte et de l'analyse des données, participation aux interprétations et à la rédaction.

Aliette Lochy : Suggestions lors du développement de la méthode, contribution aux interprétations, corrections lors de la rédaction.

José Morais : Proposition de la problématique, supervision du choix du matériel et de l'analyse des données, participation aux interprétations et à la rédaction.