Université de Montréal

Le chant des amusiques : prédictions d'une dissociation entre les habiletés perceptives et vocales

par

Catherine Roquet

Département de psychologie

Faculté des arts et sciences

Mémoire présenté à la Faculté des études supérieures et postdoctorales

en vue de l'obtention du grade de Maître ès Sciences (M.Sc.)

en psychologie

Janvier 2014

Université de Montréal

Faculté des études supérieures et postdoctorales

Ce mémoire intitulé :

Le chant des amusiques : prédictions d'une dissociation entre les habiletés perceptives et vocales

Présenté par :

Catherine Roquet

a été évalué par un jury composé des personnes suivantes :

Pierre Jolicoeur, président-rapporteur

Isabelle Peretz, directrice de recherche

Sylvie Hébert, membre du jury

Université de Montréal

Résumé

L'objectif de cette étude était d'évaluer l'influence des habiletés perceptives sur les capacités de production vocale dans l'amusie congénitale. Treize amusiques et douze contrôles appariés ont réalisé quatre tâches : deux tâches de discrimination perceptive et deux tâches de production vocale. Les stimuli utilisés pour les tâches étaient des enregistrements vocaux provenant des participants, rendant les tâches plus écologiques et enlevant le besoin pour les participants de modifier le timbre des stimuli lorsqu'ils chantent. Les résultats ont démontré que, malgré le fait que les contrôles aient surpassé la performance des amusiques dans toutes les tâches, il y avait beaucoup plus de variabilité dans les performances des amusiques que prévu. La moitié des amusiques avaient des performances égales à celles des contrôles sur les deux tâches perceptives. D'autres amusiques montraient des performances égales ou semblables à celles des contrôles sur au moins une des tâches d'imitation vocale. Ces résultats mènent à croire qu'il serait possible que ces deux types d'habiletés musicales soient dissociables.

Mots-clés : amusie congénitale, perception de hauteur, chant, production, perception.

Abstract

Our goal was to examine to what extent vocal and perceptual pitch-matching abilities were related in congenital amusia. To do this, we asked 13 amusics and 12 matched controls to perform four tasks, including two pitch perception tasks and two vocal imitation tasks. We controlled for any timbral translation by recording the participants singing and using it as stimuli across most tasks. Results showed great variability in both perceptual and vocal imitation tasks in amusics, while controls had good performances on all tasks. We illustrated how some amusics could retain good perceptual pitch-matching abilities while being unable to perform well in the vocal tasks, and how some amusics could perform well in the vocal imitation tasks. These results help illustrate the potential for these abilities to be independent. However, further studies are required to fully understand their relation.

Keywords : congenital amusia, pitch perception, singing, production, perception.

Table des matières

Liste des tableaux

Liste des figures

Remerciements

J'aimerais prendre le temps de remercier ma superviseure, Isabelle Peretz, qui a toujours été disponible pour moi pendant mes études. Ce fût un réel privilège que de travailler au sein de votre laboratoire. Vous avez toujours été très ouverte, encourageante et compréhensive et je vous en remercie grandement.

J'aimerais également souligner l'engagement et le soutien de mon ancien collègue, Sean Hutchins. Pour m'avoir enseigné tout ce dont j'avais besoin de connaître pour exceller dans le domaine de la recherche, ainsi que pour avoir été un excellent mentor pendant mes premières années au BRAMS : un chaleureux merci. Pour votre soutient au fil des années et votre aide pendant la finalisation de ce mémoire, je tiens à mentionner que cela me fût indispensable et je m'en vois touchée et privilégiée.

Je tiens aussi à remercier mes amis et collègues de laboratoire, qui m'ont offert une belle atmosphère de travail et de belles soirées pendant mes études. Un merci particulier à Nathalie Gosselin, qui m'a permis de rentrer dans le merveilleux univers du BRAMS et qui a toujours cru en moi, ainsi qu'à Nicolas Robitaille, pour son aide pendant mon expérimentation.

Finalement, une mention spéciale pour mon amie Marie-Élaine Lagrois. Marie, tu m'as offert ton amitié dès la première journée où nous sommes toutes les deux rentrées au BRAMS et ton support pendant toutes ces années. Merci.

Introduction

Toutes les cultures à travers le monde créent et honorent instinctivement la musique, le plus souvent par le biais du chant, et encore plus en situation de groupe (e.g. chant tribal, chorales, messes) (Merriam, 1964; Mithen, 2006).Cependant, il existe une partie de la population qui est insensible à la musique. Ces individus souffrent d'un trouble neurologique développemental au niveau de leur cortex auditif, qui affecte leur capacité de discrimination de la hauteur des notes («pitch») (Peretz, 2001). Ce trouble, connu sous le nom d'amusie congénitale, semble aussi être associé à une capacité de production vocale déficitaire. Très peu d'études ont été menées sur les capacités de chant des amusiques. La relation entre leurs capacités perceptives et leur capacité d'imitation de l'intonation vocale est encore moins connue. Elle n'est pour le moment pas prise en compte dans les classements des différents profils d'amusiques. L'objectif principal de ce projet de recherche sera d'essayer d'éclairer ces relations et les différents profils d'amusiques avec l'aide de tâches perceptives et vocales.

1.1. Le développement d'un mode de communication musical

Le chant possède une caractéristique profondément sociale et permet de créer de meilleurs liens sociaux, ce qui pourrait expliquer pourquoi la musique pourrait avoir une valeur adaptative (Huron, 2001; Mithen, 2006; Wallin, Merker, & Brown, 2001). Les capacités musicales se développent habituellement spontanément dès la naissance (Tillmann, Bharucha, & Bigand, 2001; Trehub & Trainor, 1998). En grandissant, les individus acquièrent des connaissances implicites sur la structure des échelles musicales (Tillmann et al., 2001). Les bébés peuvent commencer à produire des chansons dès l'âge de 18 mois (Ostwald, 1973). Ces chansons contiennent certaines des composantes les plus importantes pour la justesse du chant: un rythme répétitif et un contour mélodique (Ostwald, 1973). D'autres composantes nécessaires pour un chant juste comme la stabilité de la tonalité ainsi qu'un tempo régulier sont ensuite acquises vers l'âge de cinq ans (Dowling & Harwood, 1986). Plusieurs études suggèrent que la précision du chant augmente avec l'âge, soulignant ainsi le caractère développemental du chant (Geringer, 1983; Goetze, 1985; Gould, 1969; Green, 1990; Klemish, 1974; Yarbrough, 1991).

1.1.2 Les caractéristiques du chant dans la population normale

La majorité des gens chantent bien (Dalla Bella & Berkowska, 2009; Dalla Bella, Giguère, & Peretz, 2007; Pfordresher & Brown, 2007). Pourtant, beaucoup de gens se trouvent gênés face à la situation de chanter devant d'autres, sans le pouvoir encourageant et rassurant de faire partie d'une foule. Même si nous écoutons presque tous de la musique tous les jours et chantons lorsque nous sommes seuls, pour le plaisir, nous avons tendance à juger nos propres habiletés de chant comme étant moins bonnes que la moyenne. La capacité de chant trouvée chez la plupart des gens serait en fait bien meilleure que ce qu'ils en jugent (Dalla Bella et al., 2007; Pfordresher & Brown, 2007). Dalla Bella et collaborateurs (2007) ont démontré que certains chanteurs amateurs sont même capables d'atteindre une performance semblable à celle de chanteurs professionnels et la proportion de gens qui chantent réellement faux serait limitée à dix à vingt pourcent de la population (Pfordresher & Brown, 2009; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010). Paradoxalement, le chant dans notre société est vu comme une capacité spéciale, où seuls quelques individus (soit des professionnels ou des gagnants de concours), sont perçus comme de bons chanteurs. Pourtant, les chanteurs qui passent à la radio sont aussi des gens normaux, des gens qui ont peut-être un talent musical plus élevé que la moyenne mais qui ont souvent accès à des outils de correction de la voix : des logiciels conçus pour modifier les erreurs dans le chant automatiquement, ce qui peut donner une perception déformée de ce que le chant devrait vraiment être en élevant le seuil d'acceptabilité des erreurs de justesse dans le chant.

La voix, contrairement à la plupart des instruments, possède une modularité caractéristique qui est très difficile à corriger (Hutchins & Peretz, 2011, 2012). Les sons que l'on produit avec notre appareil vocal sont complexes, ils incluent plusieurs harmoniques, certaines fréquences non-harmoniques et beaucoup de fluctuations dans l'intonation (qu'elles proviennent du vibrato ou d'un faible contrôle de la hauteur de note) (Hutchins & Peretz, 2011). Il existe dans la voix une plus grande propension à l'erreur que dans les sons produits par des instruments et nous avons tendance à avoir une capacité limitée de discrimination pour ces erreurs, ou plus précisément, il existe un effet de générosité vocale : où une note va être jugée plus juste si elle est chantée que si elle est jouée, même si elle détient le même degré d'erreur (Hutchins, Roquet, & Peretz, 2012). Pourtant, les logiciels de corrections d'intonation sont moins utilisés pour les pistes instrumentales que pour corriger la voix. Que motive cette volonté de perfection?

1.2. Les caractéristiques du chant 'faux'

Il existe différentes façons de caractériser les erreurs de justesse dans le chant, ou ce qu'on appelle le chant 'faux'. Certains auteurs utilisent des critères fixes : ils comparent la hauteur d'une note produite à sa cible et la caractérisent comme 'fausse' si elle dévie de plus ou moins un demi-ton ou un quart de ton (Hutchins & Peretz, 2012; Hutchins, Zarate, Zatorre, & Peretz, 2010; Pfordresher & Brown, 2007). D'autres auteurs comparent les performances chantées à un groupe contrôle présélectionné pour ensuite déterminer si les performances sont dans la « norme » ou non (Satoh, Takeda, & Kuzuhara, 2007; Schön, Lorber, Spacal, & Semenza, 2004). Finalement, certains auteurs préfèrent comparer les performances à des modèles musicaux présélectionnés où l'on compare la performance des individus à une partition musicale (Dalla Bella et al., 2007; Tremblay-Champoux, Dalla Bella, Phillips-Silver, Lebrun, & Peretz, 2010). Pour le contexte la présente recherche, nous utiliserons le même critère que les études de Hutchins et Peretz, 2012. Nous considèrerons donc une note comme étant fausse lorqu'elle dévie de la cible de plus de +/- 50 cents (+/- un quart de ton). Ce critère a été utilisé dans cette étude avec succès, et rend l'analyse plus sensible. Par précaution, ainsi que pour s'assurer que ce critère ne défavorise pas le groupe amusique, nous analyserons aussi les données avec un critère de +/- 100 cents.

1.2.1. Théories et mécanismes du chant 'faux'

Joyner (1969) propose l'existence de trois opérations nécessaires pour pouvoir bien chanter. Premièrement, il faut être capable de discriminer les différentes hauteurs de notes l'une de l'autre. Deuxièmement, il faut pouvoir retrouver l'organisation mélodique d'une chanson. Enfin, il faut détenir un organe vocal qui puisse produire fidèlement les notes souhaitées (Joyner, 1969). À ces trois opérations (perception, mémoire et production), Pfordresher et Brown (2007) en ajoutent une quatrième, la capacité de pouvoir percevoir la hauteur d'une note et la tranformer en une réponse motrice vocale appropriée : une transformation sensorimotrice. Ainsi, Pfordresher et Brown (2007) estiment que la cause du chant faux est principalement reliée à la capacité d'imitation vocale d'une note. Cet acte semble simple mais regroupe en fait plusieurs démarches complexes. Selon Hutchins et Peretz (2011), il faut d'abord percevoir la note voulue, qu'elle provienne de notre mémoire ou de l'extérieur. Ensuite, il faut élaborer le plan moteur qui produit le son désiré pour finalement exécuter ce plan moteur. Un problème à l'un de ces trois niveaux pourrait suffire à fausser le chant. La plupart de ces processus ne proviennent pas de gestes conscients mais plutôt de gestes instinctifs. Trouver la cause du chant faux est donc un problème particulièrement complexe.

Pour Hutchins et Peretz (2011), les causes les plus vraisemblables des problèmes de production vocale sont trouvées à ces trois étapes. Cependant, aucun consensus n'est encore atteint sur la relation entre les problèmes de perception et de production vocale. Certaines études trouvent une corrélation entre les habiletés de discrimination de hauteur de notes et les habiletés de production vocale (Estis, Dean-Claytor, Moore, & Rowell, 2011; Moore, Keaton, & Watts, 2007; Watts, Moore, & McCaghren, 2005) tandis que d'autres n'ont pas retrouvé ces résultats (Amir, Amir, & Kishon-Rabin, 2003; Bradshaw & McHenry, 2005; Dalla Bella et al., 2007; Moore, Estis, Gordon-Hickey, & Watts, 2008; Pfordresher & Brown, 2007). Plusieurs de ces études ont aussi rapporté des individus ayant des capacités d'imitation vocale beaucoup plus élevées que leurs capacité de discrimination de hauteur de notes (Amir et al., 2003; Dalla Bella, Giguère, & Peretz, 2009; Watts et al., 2005). Hutchins et Peretz (2011) suggèrent que le manque d'homogénéité de ces résultats provient du fait que les instruments et les tests utilisés pour qualifier les capacités de perception de hauteur de notes sont peu comparables et mesurent différents aspects de la même habileté. De plus, la présence (ou l'absence) d'une corrélation ne signifie pas que la relation est causale et ne spécifie pas la direction de la causalité, si elle existe. Hutchins et Peretz (2011) concluent que la présence de déficits perceptifs ne semble pas être la cause majeure du chant faux. À l'inverse, ces déficits ont plutôt tendance à être présents lorsque la performance vocale est amoindrie (Hutchins & Peretz, 2011).

À l'étape de planification motrice lors de l'imitation d'une note jouée par un instrument, le problème pourrait aussi provenir d'une difficulté à «traduire» le timbre de la cible. En effet, il a été démontré que le timbre d'une note affecte notre perception de sa hauteur (Krumhansl & Iverson, 1992; Melara & Marks, 1990; Pitt, 1994; Warrier & Zatorre, 2002). Même si on perçoit correctement la hauteur de note de la cible à imiter, le timbre de la cible peut être un facteur déterminant dans le processus de traduction de cette hauteur pour ensuite déterminer la planification motrice requise pour l'imiter (Hutchins & Peretz, 2011). Nous sommes d'ailleurs meilleurs pour imiter des notes qui proviennent d'enregistrements de notre propre voix (Moore et al., 2008; Watts & Hall, 2008).

Finalement, le problème pourrait se trouver au niveau du contrôle moteur du chanteur. Certains individus ne possèderaient pas la coordination requise ni l'étendue vocale nécessaire pour exécuter le plan moteur approprié (Hutchins & Peretz, 2011). Coordonner son organe vocal pour pouvoir produire une hauteur de note précise est un geste complexe (Berkowska & Dalla Bella, 2009; Hutchins & Peretz, 2011; Pfordresher & Brown, 2007). Cependant, certains chercheurs ont démontré que des mauvais chanteurs pouvaient maintenir une aussi bonne stabilité de note que de bons chanteurs (Pfordresher & Mantell, 2009).

Ainsi, il est vraisemblable qu'il n'y ait pas une seule cause de problèmes de production vocale. Hutchins & Peretz (2012) proposent par exemple que les causes les plus fréquentes soient des problèmes de contrôle moteur ainsi que des problèmes liés à l'abstraction du timbre. La relation entre ces causes et leurs conséquences spécifiques n'est pas encore claire et demande plus d'investigations.

Bien que les déficits perceptifs ne semblent pas être la cause majeure des problèmes de production vocale, Hutchins & Peretz (2011) mentionnent qu'il existe une exception dans le cas des amusiques congénitaux puisque dans cette population clinique, leurs problèmes perceptifs caractéristiques sont la plupart du temps accompagnés d'un déficit de la justesse chantée (Hutchins & Peretz, 2011). Toutefois, certains amusiques conservent une bonne production vocale (Dalla Bella et al., 2009; Loui, Guenther, Mathys, & Schlaug, 2008; Tremblay-Champoux et al., 2010). Les liens entre les capacités de perception et de production musicale des amusiques ne sont pas encore bien établis. L'éclaircissement de ceux-ci pourrait en fait être un élément-clé dans la découverte des causes des déficits de production vocale dans la population générale.

1.3. L'amusie

Les amusiques sont des individus qui n'ont jamais développé d'habiletés musicales normales, malgré leur niveau normal d'éducation et d'exposition à la musique (Ayotte, Peretz, & Hyde, 2002). L'amusie congénitale se caractérise principalement par un déficit en perception de hauteur de notes (Peretz et al., 2002) et affecte approximativement 4% de la population (Kalmus & Fry, 1980). Les amusiques sont aussi caractérisés comme ayant généralement de pauvres habiletés en production de chant (Dalla Bella et al., 2009; Hutchins et al., 2010), la cause étant probablement reliée à leurs déficits de perception (Hyde & Peretz, 2004) les rendant alors moins habiles pour reproduire des notes cibles (Ayotte et al., 2002; Dalla Bella et al., 2009; Hutchins et al., 2010; Tremblay-Champoux et al., 2010). Cependant, certaines études suggèrent qu'une sous-population d'amusiques aurait tendance à mieux chanter par rapport à ce que prédirait leur déficit perceptif (Dalla Bella et al., 2009; Hutchins et al., 2010; Liu, Patel, Fourcin, & Stewart, 2010; Loui et al., 2008) et se trouveraient même dans l'étendue normale des habiletés d'imitation de hauteur de note (Dalla Bella et al., 2009; Hutchins et al., 2010). Hutchins et Peretz (2012) ont postulé que les déficits perceptifs ne seraient pas nécessairement la première cause des problèmes de production en chant. Ils suggèrent que le chant faux («poor singing») serait plutôt lié aux problèmes moteurs ou sensorimoteurs qu'aux déficits perceptifs de hauteur (Hutchins & Peretz, 2012). Ainsi, ces auteurs avancent la dissociation des processus de perception et de production vocale : les problèmes de production de chant pourraient exister sans déficit de perception, et un déficit de perception notable ne serait pas forcément associé à un problème de production vocale.

De plus, des mesures d'électroencéphalographie ont montré que des différences d'intonation même faibles (jusqu'à un quart de ton) provoquent une réaction électrophysiologique normale et inconsciente chez les amusiques (Moreau, Jolicœur, & Peretz, 2013; Peretz, Brattico, Järvenpää, & Tervaniemi, 2009). Ces résultats démontrent l'information sensorielle sur l'intonation est présente dans le cortex auditif des amusiques, même si elle n'est pas traitée consciemment. Il se pourrait donc qu'un certain niveau de traitement perceptif de l'intonation reste intact chez certains individus amusiques, suffisamment pour ne pas compromettre la production vocale au niveau perceptif.

1.3.1. Le chant des amusiques

La première recherche à explorer plus profondément le chant chez les amusiques fût celle de Dalla Bella et collègues (2009). Dans cette étude, le chant d'une chanson familière produit par onze amusiques diagnostiqués à l'aide de la Batterie Montréalaise d'Évaluation de l'Amusie (BMEA) (Peretz, Champod, & Hyde, 2003) a été analysé avec un critère fixe pour révéler les erreurs de hauteur de note. Les résultats ont révélé que neuf amusiques sur onze possédaient un déficit de production vocale : les performances comportaient plusieurs erreurs de hauteur ainsi que plusieurs erreurs d'intervalle et de contour (Dalla Bella et al., 2009). De plus, ces déficits de production vocale étaient corrélés avec leurs déficits de discrimination de hauteur de note (Dalla Bella et al., 2009). Deux des amusiques possédaient toutefois une production vocale équivalente à celle des contrôles et leurs performances ne semblaient pas être reliées à leur déficit perceptif (Dalla Bella et al., 2009).

Loui et al. (2008) ont démontré que des amusiques pouvaient imiter vocalement la direction des intervalles qui leur étaient présentés, même si les intervalles produits étaient déficitaires. Par contre, lorsque les auteurs ont demandé à leurs participants de faire la même tâche en devant nommer la direction de l'intervalle, ils étaient beaucoup moins habiles (Loui et al., 2008). Hutchins et al. (2010) ont trouvé un grand éventail de capacités de production vocale chez les amusiques. Ces auteurs ont aussi trouvé que les amusiques pouvaient produire des réponses vocales qui étaient corrélées de manière cohérente avec la hauteur de chaque note cible, même si la réponse comme telle était inexacte.

La première question à laquelle ce projet tentera de répondre est de savoir si le chant des amusiques s'améliore lorsque le timbre des cibles est plus naturel. Nous pourrions également avoir une meilleure idée des capacités d'imitation perceptive d'intonation («perceptual pitch matching») des amusiques en utilisant un instrument qui contourne le système moteur. Enfin, nous pourrions démontrer que certains amusiques conservent en effet leur capacité d'imiter des notes vocalement, tout en ayant un déficit perceptif.

Accord des coauteurs

Article :

Roquet, C., Hutchins, S., & Peretz, I. Vocal and instrumental pitch imitation abilities of amusics using a self-matching paradigm

Déclaration des coauteurs :

À titre de coauteur de l'article identifié ci-dessus, je suis d'accord pour que **Catherine Roquet** inclus cet article dans son mémoire de maîtrise qui a pour titre : Le chant des amusiques : prédictions d'une dissociation entre les habiletés perceptives et vocales

Sean Hutchins Signature Date

Isabelle Peretz Signature Date

Article

Vocal and instrumental pitch imitation abilities of amusics using a self-matching paradigm

Catherine Roquet (1), Sean Hutchins (1) and Isabelle Peretz (1).

(1) BRAMS, Department of Psychology, Université de Montréal, Canada

Running title: Self-Matching abilities in congenital amusia

Keywords: congenital amusia, singing, pitch perception, vocal imitation

Abstract

Poor singing characterizes congenital amusia, as do poor perceptual deficits. Here, we seek to illustrate how these relate and if they are dependent or independent abilities. We tested thirteen amusics and twelve matched controls on four pitch matching tasks, two of which involved imitation and two that required perceptual discrimination. In order to minimize the effect of timbre on pitch-matching, we used the participant's own voice recordings as targets. The imitation tasks asked participants to imitate previously-recorded instances of their own voice at different pitch heights. One discrimination task involved pitch-matching with a slider, which allows participants to match target pitches (a synthesized voice) via finger presses, in order to rule out any impairment in vocal-motor abilities on the data. The other discrimination task involved making a same/different judgment between two of their sung tones (which could be shifted in pitch by a quarter-tone or semitone). Our results showed that both amusics and controls were more accurate at pitch matching with the slider than with their own voice. For the perceptual tasks, controls were more accurate at matching pitches than amusics. Controls were also more accurate at all vocal pitch matching tasks than amusics. Interestingly, vocal pitch matching results for amusics showed greater variability in imitation abilities than expected: four amusics had performances closer or equal to the controls' performance in at least one vocal self-matching task. This could demonstrate a dissociation of perception and production abilities, and highlights the importance of timbral similarity on pitch matching measurements.

Singing is one of the most universal forms of musical expression shared throughout cultures and can be argued to be a fundamental part of what it is to be human (Mithen, 2006). The study of developmental processes related to singing has been of great interest to researchers, although most of these studies concern normal development (Dowling, 1999; Hannon & Trehub, 2005; Tillmann et al., 2001; Trehub, 2001). The ability to sing develops naturally and automatically, it is subserved by multiple cognitive processes and accordingly, singing abilities are normally distributed in the general population (Dalla Bella et al., 2007). Many studies have stressed the great complexity of vocal production; to sing a specific note, one has to determine the exact dispositions of the different parts of the vocal apparatus and carry out a motor command (Berkowska & Dalla Bella, 2009; Hutchins & Peretz, 2011; Pfordresher & Brown, 2007). Any variation in this process could lead to the production of an inaccurate pitch. Difficulties with pitch production, including pitch accuracy, pitch stability and pitch contour are the common metrics which characterize inaccurate singing (Dalla Bella et al., 2007, 2009; Tremblay-Champoux et al., 2010). While poor singing is natural in the general population it can also be indicative of congenital amusia.

Congenital amusia is a rare developmental disorder that is characterized by deficits in pitch perception. These deficits result in an impaired ability to discriminate small differences of pitch (Peretz et al., 2002). In terms of vocal pitch production, congenital amusics have poor singing abilities compared to the general population (Ayotte et al., 2002; Dalla Bella et al., 2007, 2009; Tremblay-Champoux et al., 2010). The cause of poor vocal pitch production abilities in congenital amusia is thought to be linked to the perceptual problems with pitch discrimination (Ayotte et al., 2002; Hyde & Peretz, 2004) but could also be due to poor vocal-motor control (Dalla Bella et al., 2007; Hutchins & Peretz, 2012; Hutchins et al., 2010).

In the general population, we find poor singers who have normal abilities to make fine-grained pitch discrimination (Dalla Bella et al., 2007). This suggests that normal pitch perception does not lead to accurate pitch production. Moreover, this suggests that it may be possible to find good singers in a population marked by poor pitch perception, namely congenital amusics. Importantly, Hutchins et al. (2010) demonstrated that some amusics showed a significant correlation between their produced pitches and the target pitch. These amusics' vocal imitation responses showed a clear linear trend with the height of the different target pitches presented to them, even though these responses had pitch errors that were, on average, over a semitone off from the target pitch. Studies by Loui et al. (2008) and Dalla Bella et al. (2009) have also shown that amusics have vocal production abilities that could not be predicted by their pitch perception deficits. Since some unconscious processing of pitch differences are evident in amusia (Moreau et al., 2013; Peretz et al., 2009), it might be possible that this unconscious representation is enough to guide their vocal responses, even if it they cannot consciously discriminate between pitches smaller than a semitone (Hyde & Peretz, 2004; Moreau et al., 2013; Peretz et al., 2009). Moreover, these findings suggest that conscious auditory perception may dissociate from motor-action processes.

Hutchins & Peretz (2012) conducted a study that compared perceptual and vocal pitch matching abilities. For the pitch matching task, participants used a 'slider'. The slider allowed participants to 'slide' their finger across a touch-sensitive surface to modify the frequency of a sound linearly, in order to match a reference sound. This enabled participants to bypass the vocal-motor processes in the task and thus measured pitch-matching ability independently from their ability to control their own voice. For the vocal production tasks, the participants heard synthetized tones or previously recorded self-vocalizations and were asked to vocally imitate them. They found that all participants were better at matching pitches using the slider compared to vocally matching synthetized tones or their own voice. They were however, better at vocally matching their own voice compared to vocally matching the synthesized tones. They concluded that this indicated a problem in the vocal-motor process. Moreover, they found that performance on the perceptual judgment task did not differ between natural voice stimuli and synthesized tones. This supports the idea that timbral translation is an important factor that might hinder the vocal imitation performance of participants. Finally, Hutchins & Peretz (2012) found that performance on the vocal imitation tasks were not related to performance on the matching task using the slider.

The extent of the relationship between pitch perception abilities and vocal production abilities is still unclear. Neuroimaging studies have suggested a possible dissociation (Loui, Alsop, & Schlaug, 2009), as found in other areas of perception and production such as language (Hickok & Poeppel, 2007). We propose to clarify this link using both perceptual pitch-matching and vocal pitch-matching tasks within the congenital amusic population, using a self-matching paradigm. In this paradigm, target pitches will be recordings of the participants own voices at different pitch heights. These will be used for most tasks, providing a timbral control. We will not be using the participants own voices recordings for the slider task, since it has been shown that natural vocal stimuli does not affect the accuracy of pitch matching when using a slider (Hutchins & Peretz, 2012).

Given that we know very little about the relationship between perception and production in amusics, the goal of the present study was to compare the perceptual pitch-matching and vocal imitation abilities of amusics and paired controls. To achieve this, we asked amusics to perform four tasks: a pitch-matching task using a slider to manually match the pitch, a vocal imitation task using the participants' self-recorded target pitches, a perceptual pitch-matching task using a same-different paradigm with the same self-recorded targets and a vocal imitation task using the same self-recorded targets. Obviously, controls are expected to perform better on all tasks compared to amusics. We predicted that all amusics will be unable to pitch match within one semitone of the targets on all perceptual tasks (Foxton, Dean, Gee, Peretz, & Griffiths, 2004; Hyde & Peretz, 2004) and all vocal production tasks (as found previously in Ayotte et al., 2002; Dalla Bella et al., 2009; Tremblay-Champoux et al., 2010). However, among amusics, we should be able to find cases who can vocally match pitches they cannot discriminate.

Methods

Participants

Participants were 13 amusics (9 women, 4 men) and 12 matched controls (8 women, 4 men). Controls were matched on the basis of age, gender, education level and music education level. One participant served as a control for two twin amusics, hence the unequal numbers of participants in each category.

All amusics had 3 year(s) or less of formal musical training (*M* = 1 year), defined as regular structured lessons on any instrument or the voice in a private or small setting over the course of at least 6 months. Amusics ranged in age from 23 to 72 (*M* = 58.2 years). Amusics reported a mean of 0.31 years of informal singing experience and did not report any formal singing training.

Controls reported more formal music training (maximum 8 years, *M* = 2.58) but this did not differ significantly from amusics' experience ($t = -2.00$, $p = .064$). Controls ranged in age from 23 to 74 (*M* = 58.1 years). Controls reported a mean of 2.18 years of informal singing experience.

All participants had a standard audiometric test to ensure their hearing was normal (at around 30 dB HL at 1000 Hz, the standard for their age). The Montreal Battery of Evaluation of Amusia (Peretz et al., 2003; Peretz et al., 2008) was used to determine and confirm the status of amusia, where participants falling two standard deviations or more below the mean global score of the control group are considered amusics. None of our participants had any diagnosed auditory deficits or neurological disorders. All participants were francophone except for *ATF*, who was anglophone. See Table 1 for complete participant characteristics.

(Table 1 about here)

Stimuli and equipment

Two separate sessions were conducted for stimuli acquisition and the four tasks. We will refer to these as session 1 and 2. Session 1 was conducted in the lab and included the recording of the targets as well as the Slider and Self-Matching tasks, whereas session 2 was conducted at the *UNF*'s Imaging Center as a part of the neuroimagery component of this study and includes the Scanner Pitch Discrimination and Scanner Vocal-Imitation tasks.

Session 1

Slider pitch matching

The slider is an instrument-based pressure sensor first used by Hutchins and Peretz (2012) which presents complex wave tones designed to imitate the timbre of the human voice on the syllable /a/. FM vibrato with a frequency of 5Hz and an amplitude of +/- 7 cents is added to the tones. Tones are presented to the participants in Max/MSP (Cycling ’74, San Francisco, CA) through DT 990 Pro headphones (Beyerdynamic, Heilbronn, Germany). This instrument reacts to finger presses across its pressure sensor (Infusion Systems, Montreal, Canada) and then presents sounds based on the position of the finger across the sensor. The Slider can register up to 1024 individual positions across its sensor, playing a slightly different pitch at each position. We chose five target tones, equally spaced along the sensor (low, medium-low, middle, medium-high, high). The five target tones were B3 (246.94 Hz), C#4 (277.18 Hz), D#4 (311.13 Hz), F4 (349.23 Hz) and G4 (392.00 Hz) for women, an octave lower for men. They were fixed and presented at random in 100 trials. Participants were given practice trials as well as a chance to familiarize themselves with the Slider at the start of the task.

Participants pressed the spacebar to begin each trial; they were then presented with a continuous target sound through their headphones. They had to find the same tone on the slider, and were told they could respond as many times as they wanted until the best answer was found. Once they pressed the Slider, the original target pitch would silence and the participant would hear the Slider tone at the position his finger had pressed. Once he released the Slider, the original target tone came back in the headphones. Participants were encouraged to try and match the pitch they heard as close as they could, until they could not hear any difference between the tone they pressed on the slider and the tone they heard in their headphones. We counted the final tone they played as being their final answer.

Due to time constraints, not all 25 participants were able to finish the whole 100 trials. However, all participants performed a minimum of 40 trials. The average number of trials completed was 74.5 for amusics and 88 for controls.

All data from this task was recorded and saved in individual .txt files by Max/MSP, including target tones and the frequencies pressed on the Slider at each millisecond.

Recording targets

The next step was to ask participants to record five samples of their voice at different pitch levels on the syllable /ba/ for 2 or 3 seconds each. Since most of the participants have musical difficulties or are unfamiliar with how they can modulate pitch height with their vocal cords, we would guide them through a simple structure, making them ascend from a low tone, to a medium-low tone, to a middle (or "comfortable") tone, to a medium-high and then a high tone. We also offered vocal warm-ups and the experimenter would provide examples by going through the procedure and singing in front of the participant. This was to ensure that participants felt at ease with the task. Tones were self-selected and participants were instructed to make sure they never sang outside of their comfortable vocal range. Each pitch level was recorded three times, through Max/MSP (Cycling ’74, San Francisco, CA). All vocal stimuli and subsequent vocal responses were recorded with a Neumann TLM 103 microphone (Georg Neumann GmbH, Berlin, Germany). After the recordings, the experimenter chose the best sample for each tone (on the basis of pitch stability, clarity of voice and height differentiation from other tones), which became the targets. These targets were then normalized for amplitude and trimmed to get rid of silences at the beginning and end of the recording. These targets were used in the Self-Matching task. These targets, and their manipulated versions, were used in session 2 as well. Figure 1 shows the range of original targets produced by all participants. All participants recorded five original targets except for the control *CRS* who, at the time of the scanner session, had not completed the Self-Matching task and had only recorded three targets in a previous experiment. This participant came back after the scanner session and completed the Self-Matching task with her three self-recorded targets.

(Figure 1 about here)

Self-Matching

The previously recorded self-chosen targets were presented to the participants through headphones in a random order with Max/MSP for 100 trials, where each target was presented 20 times. Participants were instructed to press on the space bar, after which the targets would play for their duration (approximately 2-3 seconds). Participants were then asked to try and imitate the note as best as they could. Vocal responses were recorded through a Neumann TLM 103 microphone (Georg Neumann GmbH, Berlin, Germany). As in the Slider task, they were told they could try as many times as they liked until they felt they had given the best answer. Participants could press the enter key to listen to the target again. To close a trial, participants would press the space bar, which then ended the voice recording and save the .aiff file. Data were also stored in individual .txt files for each trial.

Most participants were able to finish the whole task, with at least 87 trials for amusics and 79 for controls. The average of trials completed was 97.1 for amusics and 96.2 for controls.

This task was always done last to make sure participants would not use singing strategies during the Slider pitch-matching task. Participants were reimbursed $60 for their time.

Analyses

All vocal recordings (including targets and responses) were analysed using a Matlab (The MathWorks Inc., Natick, Massachussetts, USA) implementation of YIN (De Cheveigné & Kawahara, 2002). Analyses included data about frequency which was converted into cents (relative to D4=0). Only the participants' final response was analysed, but the number of responses made was compiled.

To analyse pitch matching accuracy, we looked at both signed and absolute measures of the sung tones in cents and compared them to the pitch of targets using a criterion of +/- 50 cents. This criterion has been used in recent studies as a valid threshold where participants can discriminate when a note is out of tune (Hutchins & Peretz, 2012; Hutchins et al., 2012). However, since this criterion has not been used yet with amusics, we decided to take the opportunity to confirm that the use of this criterion with this population is appropriate by comparing it with the usual +/- 100 cents criterion (Dalla Bella et al., 2007; Pfordresher & Brown, 2007).

Session 2

Stimuli

We manipulated the recorded targets to create the stimuli for session 2. This manipulation was done with the help of Melodyne (Celemony Software GmbH, Munich, Germany). Four deviations were created from each original target by transposing them in steps of 50 cents (where 100 cents = one semitone), creating four variations (+50 cents, +100 cents, -50 cents, -

100 cents). This yielded a total of 25 stimuli for each participant, except for *CRS*, whose targets yielded a total of 15 stimuli.

Scanner Pitch Discrimination and production

Stimuli were arranged in pseudo-random order in four preset lists. Participants were allocated a list in alphabetical order for the Scanner Pitch Discrimination task and the reverse for the Scanner Vocal-Imitation task. Only *CRS* had specific lists, compensating for the fact that she had only 15 stimuli. Both tasks were presented with Eprime (Psychology Software Tools, Inc., Sharpsburg, Pennsylvania, USA).

Before entering the scanner room, participants were presented with practice trials for both tasks to familiarize them with the scanner sound as well as the pace of stimuli presentation and the response modalities.

For each trial of the Scanner Pitch Discrimination task, participants were presented with an original target (no manipulation) and then presented with either the same tone or a manipulated version of the target. Participants were asked (visually) to answer the question "Is the second tone same or different?" and had to respond using a button press. Each stimulus was separated by interfering scanner noise for approximately 2.5 seconds (Figure 2). The peak noise levels during the scanning protocols were 138 dB. The task consisted of 30 trials.

(Figure 2 about here)

In the Scanner Vocal Imitation task, participants were presented with an original target (no manipulation) and then given a visual cue to sing back what they heard. Participants were asked to try to imitate as best as they could. This task was run after the pitch discrimination task for all participants. The task consisted of 30 trials. The sound was recorded from a fiber optics microphone, attached to the head cage of the MRI table. The microphone was set at about an inch of the participant's mouth. The sound was then relayed to a laptop in the control room and recorded with Audacity (Audacity Development Team, Creative Commons Attribution License, version 3.0).

Analyses

Participants produced a minimum of 28 responses (M = 29.8). For one amusic participant (*AJG),* we do not have any recording as the fiber optics microphone was broken.

Results

In the lab

Controls performed very well on both tasks, reaching an average of 91.80% accuracy on the Slider task and an average of 91.77% accuracy on the Self-Matching task. They could match pitch within an average of 39.20 cents for the slider task and within an average of 28.83 cents on the self-matching task (see Table 2). Amusics did much more poorly on both tasks, although their accuracy results on the Slider task was surprising. Amusics could find the corresponding pitch within half a semi-tone on an average of 71.99% of trials for the Slider task and on an average of 57.15% of trials for the Self-Matching task. Their final response deviated from the target pitch at an average of 111.88 cents for the Slider task and at an average of 113.46 cents on the Self-Matching task (see Table 2).

For the Self-Matching task, participants chose lower target pitches (where the average target height for all participants was -540.6 cents relative to the middle D#4) than the experimenter-selected Slider target pitches (*M* = 0 cents, or D#4), $t = 10.15$, $p < .001$) (see Figure 1). The target distribution was significantly skewed (skewness = .37, *SE* = 519.03), with a longer tail on the right side of the distribution. Target tones chosen by men and woman differed significantly on average ($t = 7.04$, $p < .001$), but not their ranges and skewness. Target tones chosen by amusics and controls did not differ significantly on average ($t(121)= -1.32$, p =0.191), range and skewness. Participants' mean target skewness was 0.23, significantly greater than 0, $t(24) = -2.12$, $p = .022$.

Since performance in both tasks can be analysed in many ways, we looked at many different variables. These included the average of the final tone pitch error (the deviation from the target pitch, in cents) in absolute values (to avoid any sharp and flat errors cancelling each other out), the proportion of accurate final responses within +/- 50 cents, the average trial duration, the average absolute value of the initial to final change in pitch error (whether participants improved their pitch error from the onset of their final response to its offset) and the average signed final tone pitch error (to see if participants tended to be consistently flat or sharp in their answers). We also analysed the average number of incorrect pitch changes produced (when participants changed their responses further from the target, not closer) but only within the Slider, as participants rarely made more than one response in the Self-Matching task. Both amusics and controls showed a number of differences between the Slider and Self-Matching tasks. Both groups performed somewhat better with the Slider than the Self-Matching task, and controls performed better in both tasks.

We performed eight separate 2 x 2 mixed-design analyses of variances (ANOVAs) on the factors of group (between-subjects; amusics versus controls) and response modality (within-subjects; Slider versus Self-Matching), using all of these dependent variables (all values are reported with Greenhouse-Geisser corrections, which adjusts the degrees of freedom). For interactions, we used t-tests to pinpoint the effect, and adjusted the results with a Bonferonni correction of .0125. Since our main goal in this study was to underline inherent differences between our amusic group and our control group, group differences will have a greater importance in our report. Five ANOVAs revealed significant main effects of response modality, four ANOVAs revealed significant main effects of group and one ANOVA revealed significant group by response modality interaction. We also ran these ANOVAs adding the factor of target pitch (whether the target was a low, medium-low, middle, medium-high or high tone), turning them into 2 x (2 x 5) mixed-design ANOVAs. No significant main effects of target pitch were found. This replicates the findings of Hutchins and Peretz (2011). Thus, we decided to ignore this factor for all subsequent analyses.

(Figure 3 about here)

Final response error

We found no main effect, indicating that participants' final responses were not significantly different between the Self-Matching task and the Slider task. We found a significant group difference, where controls' final responses were closer to the target than those of amusics, $F(1, 23) = 8.02$, $p = .009$, $\eta^2 = .26$ (see Figure 3). Due to the predicted variability of responses in the amusic group, we illustrate individual participant's answers in Figure 4.

(Figure 4 about here)

Proportion of correct responses

No main effect was found but there was a significant group difference. Controls made more correct responses (i.e. final responses within +/- 50 cents of the target pitch) than amusics, $F(1, 23) = 14.07$, $p = .001$, $\eta^2 = .38$ (see Figure 3). Figure 4 shows the scatterplot of the individual participant's answers.

Total trial duration

We found a significant main effect for response modality, where participants took significantly longer to complete trials in the Slider task than in the Self-Matching task, $F(1, 23) = 42.23$, $p < .001$, $\eta^2 = .65$ (see Figure 3). This can be explained by the fact that participants made many attempts at perfecting their responses in the Slider task: they responded on average 1.1 times in the Self-Matching task, compared to 14.16 times on average in the Slider task. There were no significant group differences.

Initial to final error change

A significant main effect of response modality shows that participants were making significantly greater changes in pitch from the initial response to the final response in the Slider task than in the Self-Matching task, $F(1, 23) = 68.68$, $p < .001$, $\eta^2 = .75$. This could be explained by the fact that most participants would start at a random position on the Slider and then proceed to correct their response. There were no significant group differences.

Signed final response error

No significant main effects or interactions were found for either tasks, and there were no significant group differences. This means that participants had no tendency to be consistently sharp or flat in their responses.

Incorrect pitch change responses

Analysing this variable for the Slider task only, we found a significant group difference. Controls' responses were significantly more likely to move towards the target than amusics, $t(13) = 3.37$, $p = .005$. This suggests that amusics could not perceive correctly if they were moving closer to or farther from the target.

Correlations

To further ascertain the consistency of performance across tasks, we correlated participants' performances on each variable for each task (see Table 3). We found a significant correlation for final response error between the Slider and Self-Matching task, $r(25) = .655$, $p < .001$. However, this effect only held for amusics, $r(13) = .590$, $p = 0.034$. We also found a correlation for accuracy between the Slider and Self-Matching task, $r(25) = .525$, $p = .007$. These results show that participants who had a better performance on one of the tasks tended to be more accurate on the other. Interestingly, this effect only held for controls, $r(12) = .742$, $p = 0.006$, and was not found for amusics, $r(13) = .303$, $p = 0.315$.

(Table 3 about here)

Other analyses showed no significant differences in participants' accuracy according to gender.

We also plotted the average error in cents of the Slider task against the average error in cents of the Self-Matching task, for all participants, to examine whether there was a relationship between these two variables. For amusics a relationship between performance on the Slider task and the Self-Matching task was correlated (see Figure 5). For controls, there was little relationship in performance between these two tasks.

(Figure 5 about here)

In the scanner

We first looked at accuracy measures (the proportion of correct responses) for both the Pitch Discrimination and Vocal-Imitation tasks. There was a performance dip for both amusics and controls. Controls performed well on both tasks, reaching an average of 71.66% accuracy on the Discrimination task and an average of 84.07% accuracy on the Imitation task, where they could match pitch within an average of 42.18 cents (see Table 2). Amusics performed poorly on both tasks and could only hear whether or not there was a difference between two notes correctly on an average 56.41% of trials for the Discrimination task and could match pitches vocally within +/- 50 cents on an average of 42.73% of trials for the Imitation task, where they matched pitch at an average of 111.02 cents from the target (see Table 2). Since this analysis was combined, we took out *AJG* since his data from the Vocal-Imitation task were missing. We performed a single 2 x 2 mixed-design ANOVA on the factors of group (amusics versus controls) and response modality (Discrimination vs. Imitation). For interactions, we used t-tests to pinpoint the effect, and adjusted the results with a Bonferonni correction of .0125.

A significant group difference shows that controls performed significantly better, $F(1, 22) = 17.01$, $p < .001$, $\eta^2 = .44$. We also found a significant interaction between response modality and group, $F(1, 22) = 10.95$, $p = .003$, $\eta^2 = .33$, which shows that this effect held only for the Vocal-Imitation task ($t(12) = -4.98$, $p < .001$) (See Figure 6). This is probably due to the great variability of participants' performances. Figure 7 shows the scatterplot of the individual data.

(Figure 6 about here)

(Figure 7 about here)

We performed correlations between these two data sets to see how consistent participants' performances were across these two tasks. We found a significant correlation, showing that participants who had a better performance on one of the tasks tended to be more accurate on the other as well, $r(24) = .621$, $p = .001$. However, when divided in groups, the significant correlation only held for the control group, $r(12) = .610$, $p = .035$. The correlation for amusics was not significant ($r(12) = .454$, $p = .138$).

It is interesting to note that, when we analysed the accuracy variable in the Vocal-Imitation task with both the +/- 50 cents and +/- 100 cents criterion, both measurements yielded significant group differences ($t(22) = -4.98$, $p < .001$; $t(14) = -2.91$, $p = .011$): a result that shows that group differences are detectable no matter what criterion is chosen.

Vocal pitch errors

We then looked at how far vocal pitch was from target (in cents) in the Vocal-Imitation performed in the scanner and compared it to the vocal responses recorded in the Self-Matching task in the lab. This was to ensure that performance in both these vocal production tasks were consistent. Here as well, we took out *AJG*. We performed two 2 x 2 mixed-design ANOVAs on the factors of group (amusics versus controls) and location (Scanner Vocal-Imitation versus Self-Matching in the lab).

Proportion of correct responses

We found a main effect of location, where participants were more vocally accurate in the lab (better at matching the target pitch to within +/- 50 cents), $F(1, 22) = 13.90$, $p = .001$, $\eta^2 = .39$ (See Figure 8). This shows that being in the scanner did impair performance. A significant group difference showed that controls' performances were more accurate than those of amusics, $F(1, 22) = 24.34$, $p < .001$, $\eta^2 = .53$.

(Figure 8 about here)

Indeed, we see evidence of performance dips on both the scanner tasks. Amusics' performance dropped 14.42% from the Self-Matching to the Scanner Vocal-Imitation task. Controls' performances dropped 7.7% from the Self-Matching to the Scanner Vocal-Imitation task.

Absolute final response error

We found no significant main effect of location, showing that even if participants were less accurate in the scanner, they tended to produce the same kind of pitch error on average. For the Self-Matching task in the lab, controls had an average error of 28.83 cents compared to an average error of 42.18 cents for the Imitation task in the scanner. Amusics had an average error of 113.46 cents in the lab compared to an average error of 111.02 cents in the scanner. We found a significant group difference showing that controls were significantly closer to target than amusics, $F(1, 22) = 9.16$, $p = .006$, $\eta^2 = .29$ (See Figure 8).

We performed correlations for both accuracy and pitch error variables, to see how consistent participants' performances were across testing locations. We found a significant correlation for both accuracy, $r(24) = .863$, $p < .001$, and pitch error, $r(24) = .753$, $p < .001$. These results show that participants who had better performances on one task tended to be more accurate on the other. The correlation found for accuracy held for both amusics, $r(12) = .783$, $p = .003$, and controls, $r(12) = .818$, $p = .001$. However, when divided in groups, the significant correlation found for average pitch error only held for the amusics, $r(12) = .696$, $p = .012$.

Scanner Pitch-Discrimination

Since group differences in the Discrimination task were previously not detected in the scanner probably due to the high variability of responses, we examined the data separately. Participants' responses ("Same" versus "Different") were used to calculate the hits minus false alarms rate. This was then used to calculate the d' for each participants' performance. The d-prime values were entered in a 2 x 2 mixed-design ANOVA on the factors of group (amusics versus controls) and difference (whether the pitch change was 50 or 100 cents). Amusics performed below chance for the trials with 50 cents deviation from target, with a d-prime score of 0.52 and a hits minus false alarm rate of 0.163. They also performed poorly for the trials with 100 cents deviation from target, with a d-prime score of 1.05 and a hits minus false alarm rate of 0.352.

We found a significant main effect for difference, where participants were significantly more accurate in trials where the pitch change was 100 cents, $F(1, 23) = 23.41$, $p < .001$, $\eta^2 = .50$. We also found a significant interaction between difference and group, $F(1, 23) = 4.36$, $p = .048$, $\eta^2 = .16$, which shows that this effect held for both amusics ($t(12) = -3.38$, $p = .005$) and controls ($t(11) = -3.67$, $p = .004$) (see Figure 9). There was a significant group difference, where controls were more accurate than amusics, $F(1, 23) = 12.29$, $p = .02$, $\eta^2 = .22$.

(Figure 9 about here)

Slider versus Scanner Pitch-Discrimination

Lastly, we compared the amusics and controls' performances on the Slider and Scanner Pitch-Discrimination task. We examined the proportion of correct responses in each task (whether they could match pitches to within +/- 50 cents in the case of the Slider task, and whether they could hear pitch deviations from the target of +/- 50 or 100 cents in the case of the Discrimination task), since this is the only measure that could be defined across both tasks. We performed a single 2 x 2 mixed-design ANOVA on the factors of group (amusics versus controls) and response modality (Slider versus Discrimination).

A significant main effect of response modality reveals that participants were more accurate in the Slider task (in the lab), $F(1, 23) = 22.32$, $p < .001$, $\eta^2 = .49$ (See Figure 10). This confirms previous results and shows that the scanner impairs performance. The significant group difference shows that controls' performances were more accurate than those of amusics on both tasks, $F(1, 23) = 5.33$, $p = .030$, $\eta^2 = .19$. The interaction between group and response modality was not found to be significant ($F(1, 23) = 0.364$, $p = .552$, $\eta^2 = .016$).

(Figure 10 about here)

As a follow-up to this, we performed correlations comparing both accuracy measures, to see how consistent participants' performances were across tasks. We found a significant correlation showing that participants who had better performances on one task tended to be more accurate on the other, $r(25) = .686$, $p < .001$. However, when divided in groups, the significant correlation only held for amusics, $r(13) = .682$, $p = .010$.

Amusic subgroups

Due to the surprisingly good average accuracy scores for amusics on the Slider task, we examined amusics' performances more closely. Results showed that amusics were divided into two clear subgroups for this task. Seven amusics (see Table 1) on thirteen were clearly above the group's average (71.99%) and could match pitches within +/- 50 cents on more than 82% of trials, reaching an average of 93.28% accuracy, whereas the other six amusic participants (see Table 1) had accuracy scores below 64% and could only successfully complete 47.15% of trials on average, thus creating a severely pitch-deaf subgroup within the amusics (see Table 2 for results on all tasks). Interestingly, when we compared the MBEA score of these two subgroups, we found no significant differences. However, when we examined their performance on the Pitch Change Detection (*PCD*) task, we found that two members of the above-average subgroup could discriminate pitches to within a quarter semitone (25 cents) at 94.4% hits (both had the same score). As a subgroup, these seven above-average amusics could discriminate pitches accurately to within half a semitone (50 cents) with an average of 89.9% hits. On the other hand, the severely pitch deaf amusics could only match pitches accurately above chance from differences of one semitone or more, which is in line with all main findings on congenital amusia (Foxton et al., 2004; Hyde & Peretz, 2004).

Good amusic singers

We also found a subset of four amusics who could match pitches with their voice with an average accuracy score of 84.16% on the Self-Matching task and had performances falling within the control group's range of performances (from 70% to 100%). They could match pitches on this task to within an average of 45.89 cents. Consistent with the trend seen in all groups, these participants scored lower on the Scanner Vocal-Imitation with an average accuracy score of 64.08%, where they could match pitches to within an average of 45.71 cents. However, it is interesting to note that all four remained within the controls performances range (from 9.3 to 99.73 cents of error, and from 36.67% to 100% accuracy).

Discussion

We found that amusics, on average, performed poorly compared to controls on all tasks, except for the slider task. Controls were able to match pitch accurately on all tasks. Excluding the surprising findings for the Slider task, these results confirm our main hypotheses.

In general, participants performed better on the slider task than the other tasks, except for the severely pitch-deaf amusics, who performed better on the Self-Matching task. Control performances on the Self-Matching task were very similar to their performances on the Slider task. Our controls performed similarly to students in a previous study by Hutchins and Peretz (2012).

As hypotheses predicted, we found group differences for all main variables when analyzing the Slider and Self-Matching tasks (which were the final response error and accuracy). These are the variables that define performance most precisely.

When we plotted the average errors in cents of both the Slider and Self-Matching task (see Figure 5), we found that there was a relationship between both for all amusics (regardless of which subgroup they belong to) but this was not seen for controls. This could imply that the level of pitch perception deficit does determine the amusics' performances on vocal pitch-matching. However, the absence of a relationship between these two variables for the controls suggests that this is not the case for them. This is similar to what was found among musicians and non-musicians (Hutchins & Peretz, 2012).

There were differences between the Slider and Self-Matching tasks for total trial duration as well as response count. On the Slider task, amusics spent an average of 24.77 seconds on each trial and making an average of 17.53 different responses on each trial, whereas controls spent an average on 18.39 seconds on trials, and made an average of 10.50 different responses for each trial. On the Self-Matching task, these numbers were lower. Amusics spent an average of 6.77 seconds on each trial, and made an average of 1.12 attempts at a response, controls spent 6.71 seconds on average on each trial and made an average of 1.1 attempts. These differences may be due to a general lack of attempts in the Self-Matching task for both groups, even though participants were told and encouraged to make as many attempts as they wanted they rarely made more than one attempt. A similar finding was reported by Hutchins & Peretz (2012). This could be due to a general sense of unease towards singing.

For the amusics, performances on the Slider task were highly variable and divided the amusics into two clear groups: amusics and severely pitch-deaf amusics. Results show that both controls and the amusic subgroup can accurately match pitch using the slider, whereas the severely pitch-deaf amusics cannot. These results cannot be explained by demographics, or working memory abilities as these were similar between the groups. These results could also be linked to the fact that hearing the target tone consistently and after each attempt at a response significantly helped both amusics and controls. Also, this effect cannot be explained from how differently the working memory is involved in the Slider versus the Pitch Discrimination task, since this sub-group was also performing significantly better at the Pitch Discrimination task than were the severely pitch deaf amusics, and the Pitch Discrimination task does involve more working memory. We did find that their above performance on these perception tasks could be predicted by their Pitch Change Detection (*PCD*) task scores, which all participants complete before studies in our lab.

Despite the heterogeneity of the amusics, there is still evidence that amusia is related to a fine grained pitch problem. This is shown in the severely pitch-deaf amusics' performances on the Slider task as well as their results on the Self-Matching and Scanner Vocal-Imitation tasks, where their average error would exceed a semitone (see Table 2). Results of the Scanner Pitch Discrimination task show that severely pitch-deaf amusics had very low sensitivity when detecting a pitch difference of one semitone. These finding provide additional support for the hypothesis that amusia is linked to deficits in pitch discrimination, when the tones differ by less than one semitone (Hyde & Peretz, 2004; Moreau et al., 2013; Peretz, Brattico, & Tervaniemi, 2005).

The finding of high performers within amusics on the Slider task was unexpected. This above-average amusic subgroup shows slightly better performance on both vocal-imitation tasks than the severely pitch-deaf amusic subgroup, yet their scores fall much below those of the controls and their average error shows that they cannot match pitch as closely as controls (see Table 2). However, their scores on both the Slider and Discrimination tasks fall within the range of the controls' performances. These results, along with the fact that they had very high *PCD* scores, show that these amusics have spared pitch perception abilities in simple tonal contexts. It would be interesting to know if they have spared pitch perception abilities in complex musical contexts. These amusics show evidence of having spared pitch discrimination while having impaired vocal imitation abilities.

Performing auditory tasks in a noisy environment is distracting and dampened auditory perception abilities for both groups. Although showing reduced performances on these scanner tasks, both controls and amusics remained consistent with their performances in the lab, as shown in the correlations between the performances of the Slider and Scanner Pitch Discrimination tasks, and the performances on the Self-Matching and Scanner Vocal-Imitation tasks.

When we examined the relationship between the performances of both tasks done in the lab, we found a significant correlation for controls, but this was not found for amusics. Similarly, when we looked at the relationship between both scanner tasks, we found that the correlation was significant for controls, but not for amusics. This means that in both locations, the performance on the auditory pitch-matching and pitch discrimination tasks were not related to the performance on the vocal-imitation tasks, but only in the case of amusics. This is probably due to the higher variability in performance, seen in the existence of the severely pitch-deaf amusic subgroup and the good amusic singers. However, it does present some evidence in the case that these abilities are distinct and dissociated within congenital amusia.

Good amusic singers

We identified four ‘good singers’ within our amusic sample. These amusics showed spared vocal imitation abilities and could perform as well as some controls on the vocal tasks. It is interesting to note that even though their accuracy scores dipped when performing in the scanner, their average error stayed stable, showing that they are in fact quite able to match pitch vocally to within +/- 50 cents. Three of them were also part of the above-average amusic subgroup found for the Slider task. This would suggest that their low MBEA scores might reflect a music-specific problem, probably more closely related to melodies and higher-level auditory processing than a pitch-perception deficit. These results would also suggest that they behave similarly to controls and could be labeled as normal for the singing related tasks. More tests are needed to examine why they performed badly on the MBEA and where their deficit truly lies. *AML* is the only 'good singer' who shows impaired performance on both the Slider and Discrimination tasks. Her results could lead to evidence of a dissociation between impaired pitch discrimination and spared vocal imitation. Combining data from this study with the forthcoming neuroimaging study will help us examine if this behavioural dissociation can be seen in relation to the neuroimaging data.

Individual cases

One amusic, *ADM*, consistently performed as well as controls (see Table 2 for her performances on all tasks) with scores that were much higher than her group's averages and that easily fell within or surpassed the controls' average performances. Despite being on the cusp of meeting diagnostic criteria for amusia according to the MBEA (with a global score of 73.89%), *ADM* can consistently match pitches accurately on all tasks, which means she has normal perceptual and vocal production abilities. *ADM* was part of the amusic subgroup that was above-average and, using the *PCD* task, was able to discriminate pitches as small as a quarter semitone (57.8% hits minus false alarms). This would suggest that, if she is truly amusic, her disorder is music-specific. More tests are needed as a follow-up to determine if she is truly amusic and why she scored low on some MBEA scales.

A second interesting case, *AMB*, performed well on all tasks except the Scanner Vocal-Imitation task, even though her score on this task was above the average for amusics and fell within the controls performance range (yet she performed much below the controls' average on this task, see Table 2). This amusic is part of the amusic subgroup that is not severely pitch-deaf found for the Slider and Scanner Pitch Discrimination tasks. Her scores at the *PCD* task showed that she had a 94.4% hits minus false alarms score in discriminating pitches with differences to within a quarter semitone. This is surprising as she shows very low scores on all scales of the MBEA, with a global score of 56.67%.

Next we have the case of *AML*, who can be categorized as a good singer even though she is amusic (see Table 2 for her results). The lower score on the Scanner Vocal-Imitation task reflects what was found in the main analyses: that most participants found the Scanner Vocal-Imitation task harder than the Self-Matching task, yet it is still well above her group's average accuracy score. However, she performs very poorly on the perceptual tasks. This is consistent with her MBEA results, with a global score of 60%. She also had very low scores on the *PCD*, only reaching a score above chance at the one semitone difference level.

As another interesting participant, *CAH* consistently under-performed in all tasks even though he qualifies as a control (see Table 2 for his results). He shows very poor perceptual skills for both the Slider task and the Pitch Discrimination task, as well as clear deficits in vocal production skills in the Self-Matching and Vocal-Imitation tasks. This is surprising since he obtained a perfect or very near to perfect score on five of the MBEA scales and had a global score of 93.33%. His scores never went below 80% on any of the scales. Is it also interesting to note that his *PCD* task results showed that he had a 94.4% hits minus false alarms score in discriminating pitches with differences to within a quarter semitone (25 cents). It is clear that this participant needs further examination before qualifying as a control in a future study.

One of the major advantages of the current study over previous studies was that we provided a more ecologically valid setting because we used the participant's own voice across many tasks. A trend towards more natural stimuli has been used effectively in previous studies (Hutchins et al., 2010; Tremblay-Champoux et al., 2010).

This study was the first part of another study which is currently in progress. The goals of the future study are to find neural correlates to these behavioural findings and to be able to pinpoint the amusics' behaviour on each task to a specific neuronal activity. These neuroimaging data will also help us conclude if there is a possibility of a dissociation between perceptual pitch perception and vocal production abilities. If there is evidence of amusics having good performances on vocal production tasks, with the presence of their characteristic pitch perception deficits, this might give more credence to the poor vocal-motor control theory. The study will also investigate tractography in order to pinpoint where and if amusics have irregularities in their neural pathways and how that might affect and cause their musical deficits. We hope to find clear patterns to help understand the cause and development of congenital amusia.

In conclusion, these data confirm that most amusics have impaired perceptual pitch-matching abilities compared to controls, across different perceptual tasks using different self-chosen target heights and response modalities. Amusics were also impaired in their vocal production pitch-matching abilities, compared to controls, across different self-chosen target heights. A few amusics showed production abilities close to controls' performances, which also demonstrates the inherent variability in amusics' abilities. This study showed evidence of amusic individuals who performed as well as controls on the perceptual tasks, and were as impaired as other amusics on the vocal production tasks. This study also showed evidence of amusic individuals performing as well as controls on vocal imitation tasks, although their performances on the perceptual tasks were not necessarily impaired. These results, coupled with previous work (see Hutchins et al. 2010 for similar findings) underline more variability than was previously thought to exist in the amusic population and demands the attention of a comprehensive study.

Table 1

Table 1

Participant demographic information, musical background and global scores obtained on the MBEA (in percentages) with standard deviations relative to controls' mean score in parentheses.

	Age (yrs)	Gender	Education (yrs)	Musical Education (yrs)	Music Experience (yrs)	MBEA (%)
Amusics						
A-ML*	72	F	15	0	0	60 (-6.83)
A-AS	71	F	14	2	0	62.78 (-6.23)
A-NB*	67	F	13	3	2	66.67 (-5.39)
A-DM	64	F	18	1	0	73.89 (-3.84)
A-EL	62	F	19	3	0	60.55 (-6.71)
A-GsA*	56	F	15	0	0	61.67 (-6.47)
A-GnA*	56	F	25	0	0	70 (-4.68)
A-MB	30	F	20	0	0	56.67 (-7.54)
A-SK	23	F	16	0	0	75.56 (-3.48)
A-EB*	71	M	16	0	0	77.22 (-3.12)
A-BL*	68	M	14	1	0	60 (-6.83)
A-JG	63	M	19	1	0	63.33 (-6.11)
A-TF	54	M	19	2	2	70 (-4.68)
M	58.23	–	17.15	1	0.31	66.03 (-5.53)
Controls						
C-AJ	72	F	14	2	7	90.56 (-0.26)
C-RS	69	F	18	3	0	93.33 (0.34)
C-GB	63	F	12	0	0	82.78 (-1.93)
C-CL	63	F	16	5	10	90 (-0.38)
C-RD	62	F	12	0	0	86.67 (-1.09)
C-MB	53	F	19	8	0	98.89 (1.53)
C-JD	27	F	22	5	2	94.44 (0.58)
C-LD	23	F	19	2	0	96.67 (1.05)
C-JC	74	M	13	0	0	91.11 (-0.14)
C-AH	69	M	15	3	3	93.33 (0.34)
C-YB	63	M	19	3	2	87.22 (-0.98)
C-GD	59	M	15	0	0	96.11 (0.94)
M	58.08	–	16.17	2.58	2.18	91.76 (4.65)
t-tests	n.s.	–	n.s.	n.s.	n.s.	$p < .001$

n.s. refers to a non-significant difference ($p > .05$)

*Amusics part of the severely pitch-deaf subgroup

Table 2

Table 2. Overview of performances on all tasks including accuracy and (when available) average error in cents for all groups, subgroups and special cases

	Slider		**Self-Matching**		**Discrimination**	**Imitation**	
	Error (in cents)	Accuracy	Error (in cents)	Accuracy	Accuracy	Error (in cents)	Accuracy
Controls	39.20	91.80%	28.83	91.77%	71.66%	42.18	84.07%
Amusics	111.88	71.99%	113.46	57.15%	56.41%	111.02	42.73%
Severely pitch-deaf amusics	179.18	47.15%	158.15	52.18%	41.70%	133.22	40.52%
Above-average amusics	54.19	93.28%	75.16	61.40%	69.03%	88.82	44.94%
Good singers	45.89	87.26%	28.39	84.16%	71.24%	45.71	64.08%
ADM*	23.68	97.5%	22.43	89%	83.33%	21.86	89.66%
AMB*	40.33	97%	31.56	82.83%	83.33%	53.97	53.33%
AML*	72.62	60.61%	23.12	90.82%	48.28%	37.47	63.33%
ASK	46.94	93.94%	36.43	74%	70%	69.53	50%
Special cases							
CAH	108.87	36%	40.62	70%	46.43%	91.8	36.67%

**Amusics belonging to both the 'good singers' and the 'special cases'*

Table 3.

Table 3

Correlations between main variables for the Slider (Y axis) and Self-Matching (X axis) tasks, across all participants.

Variable	1	2	3	4	5	6
1. Final response error	0.655**	-0.571**	-0.317	-0.100	0.545**	-0.089
2. Proportion of correct responses	-0.536**	0.525**	0.319	0.163	-0.384	0.111
3. Total trial duration	0.259	-0.325	0.362	0.291	0.110	0.245
4. Initial to final error change	0.092	-0.297	0.019	-0.106	-0.154	-0.064
5. Signed final response error	-0.714**	0.606**	0.151	0.006	-0.602**	-0.014
6. Incorrect pitch change responses	0.353	-0.503*	0.166	0.219	0.017	0.257

*$p < .05$, **$p < .01$

Figure 1. Histogram of target tones chosen by amusics and controls serving as stimuli for the Self-Matching, Scanner Pitch Discrimination and Scanner Vocal-Imitation tasks, sorted by frequency in bins of 100 cents. D#4 serves as 0, as it represents the middle note of the slider. Men's target pitches are controlled for octave.

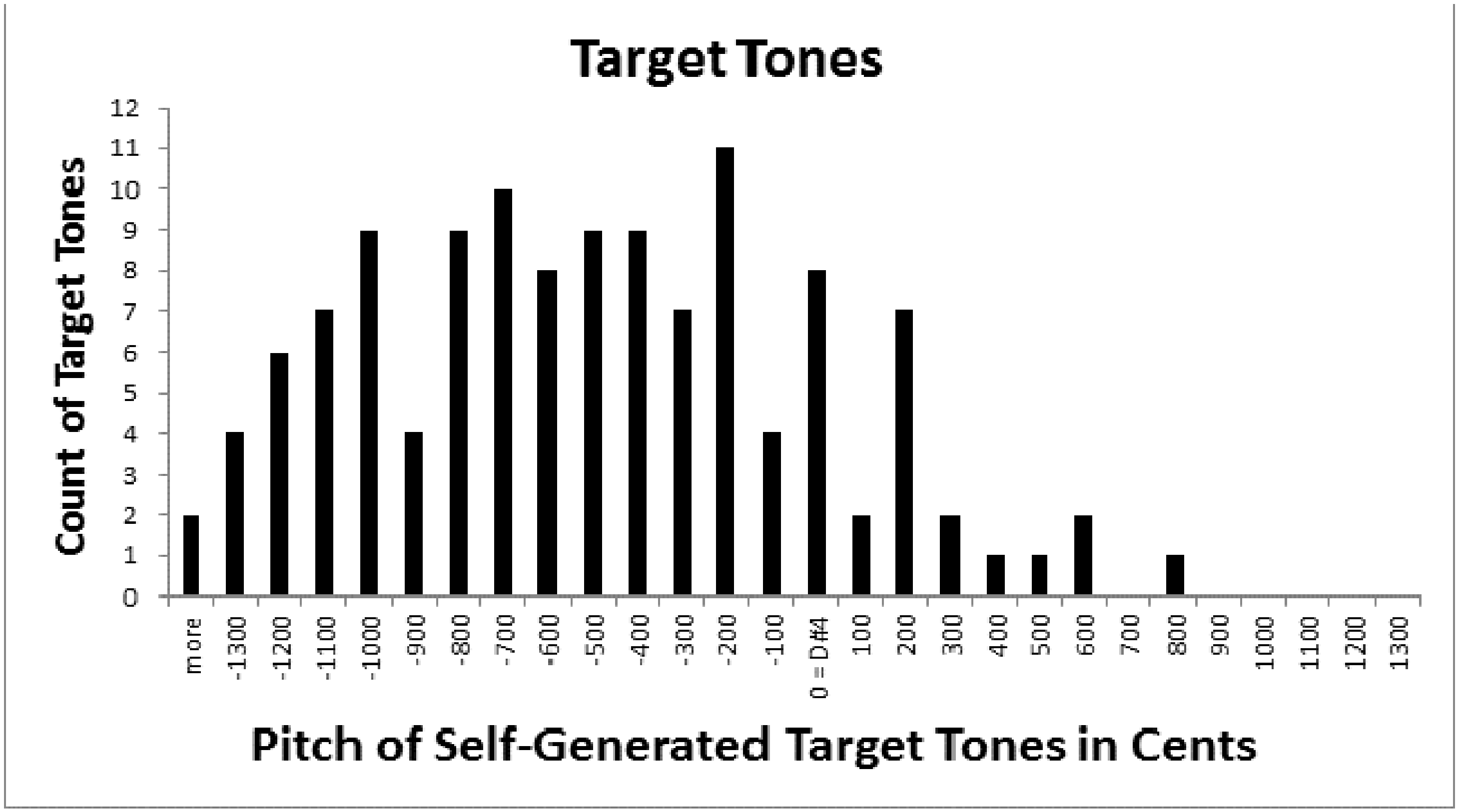


Figure 2. Graphic representation of the Scanner Pitch Discrimination and Scanner Vocal-Imitation fMRI protocols.

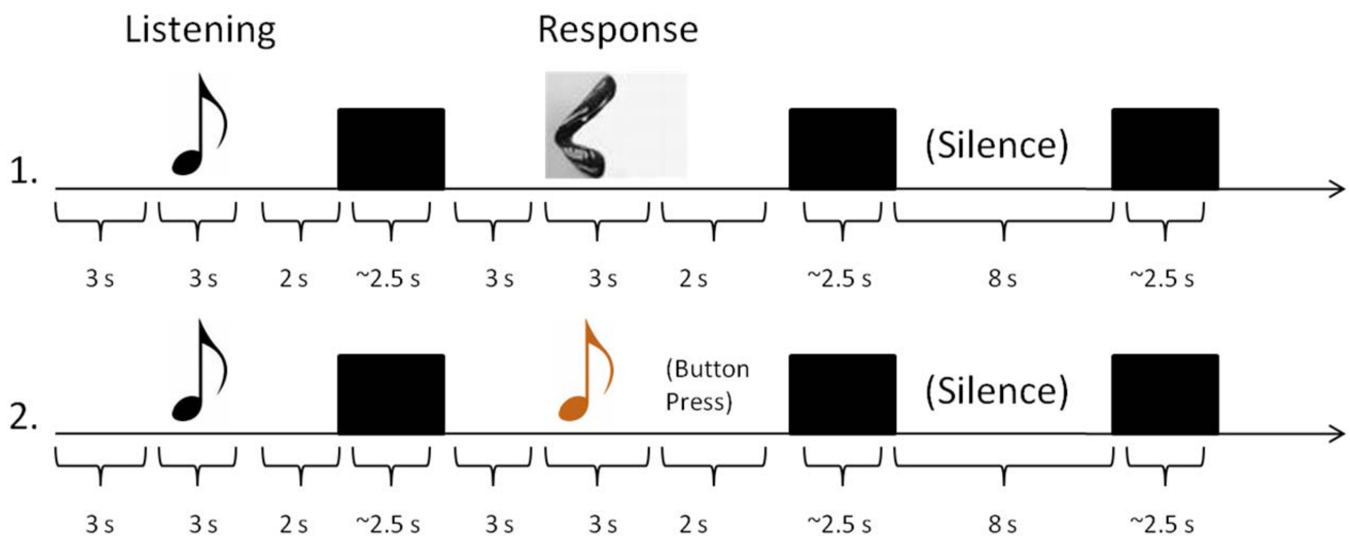
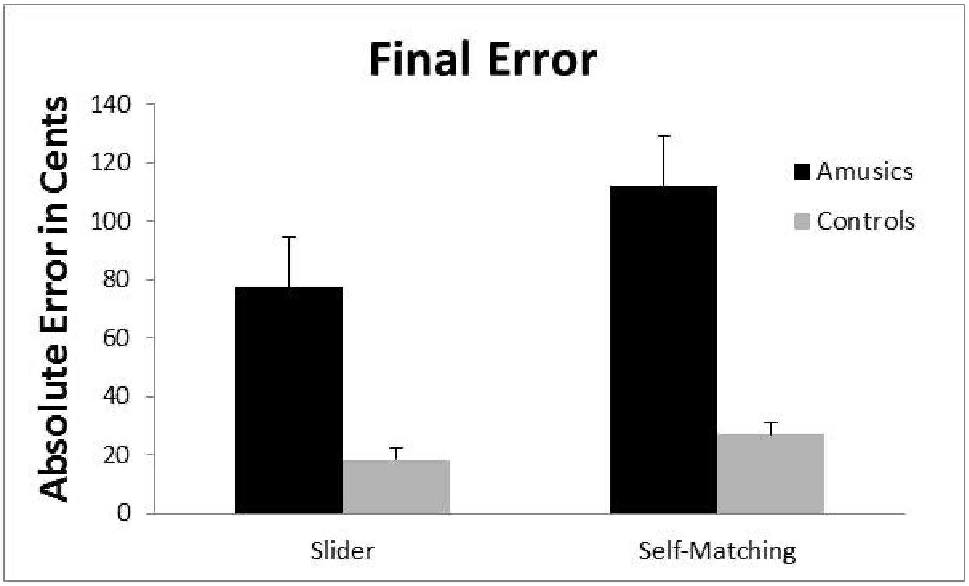
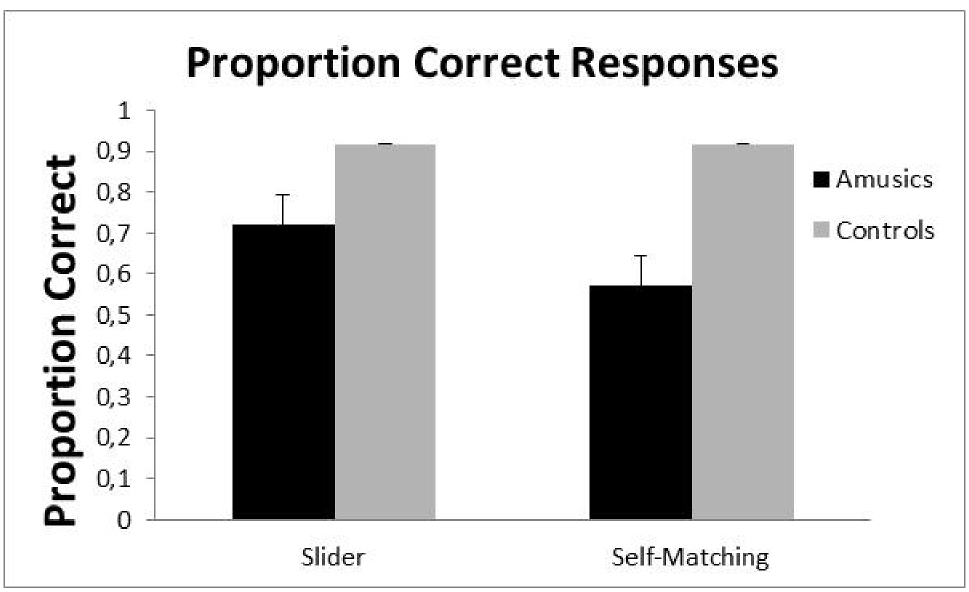


Figure 3. The absolute average value of the pitch error during the final tone produced, the proportion of accurate final responses and the average total trial duration shown for amusics and controls in both the Slider and Self-Matching tasks from session 1, with standard error bars.





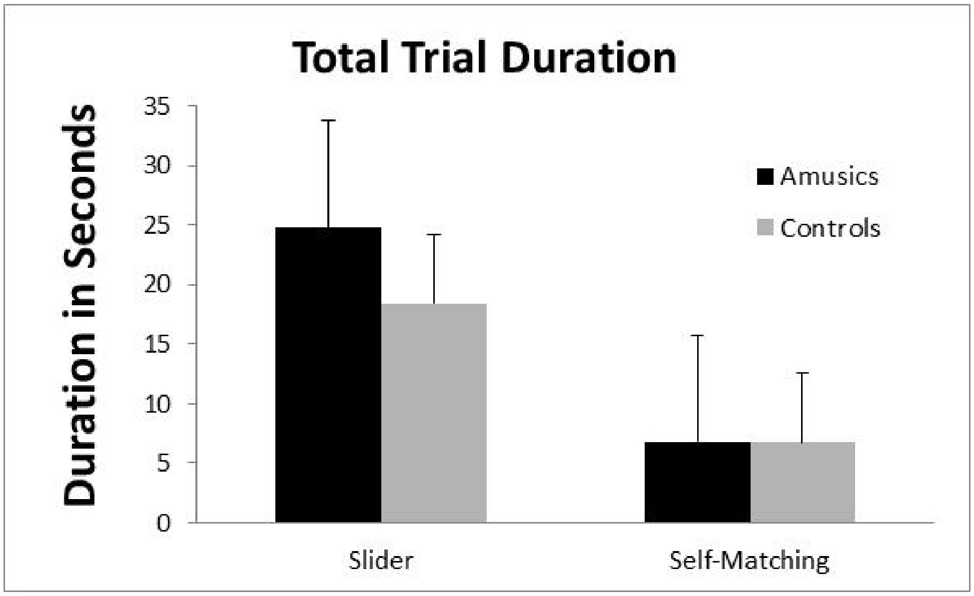

Total Trial Duration
Duration in Seconds
35
30
25
20
15
10
5
0
Amusics
Controls
Slider
Self-Matching

Figure 4. The scatterplot of all individual scores from the Slider and Self-Matching tasks of session 1 for the a) average final error and b) proportion of accurate final responses.

A)

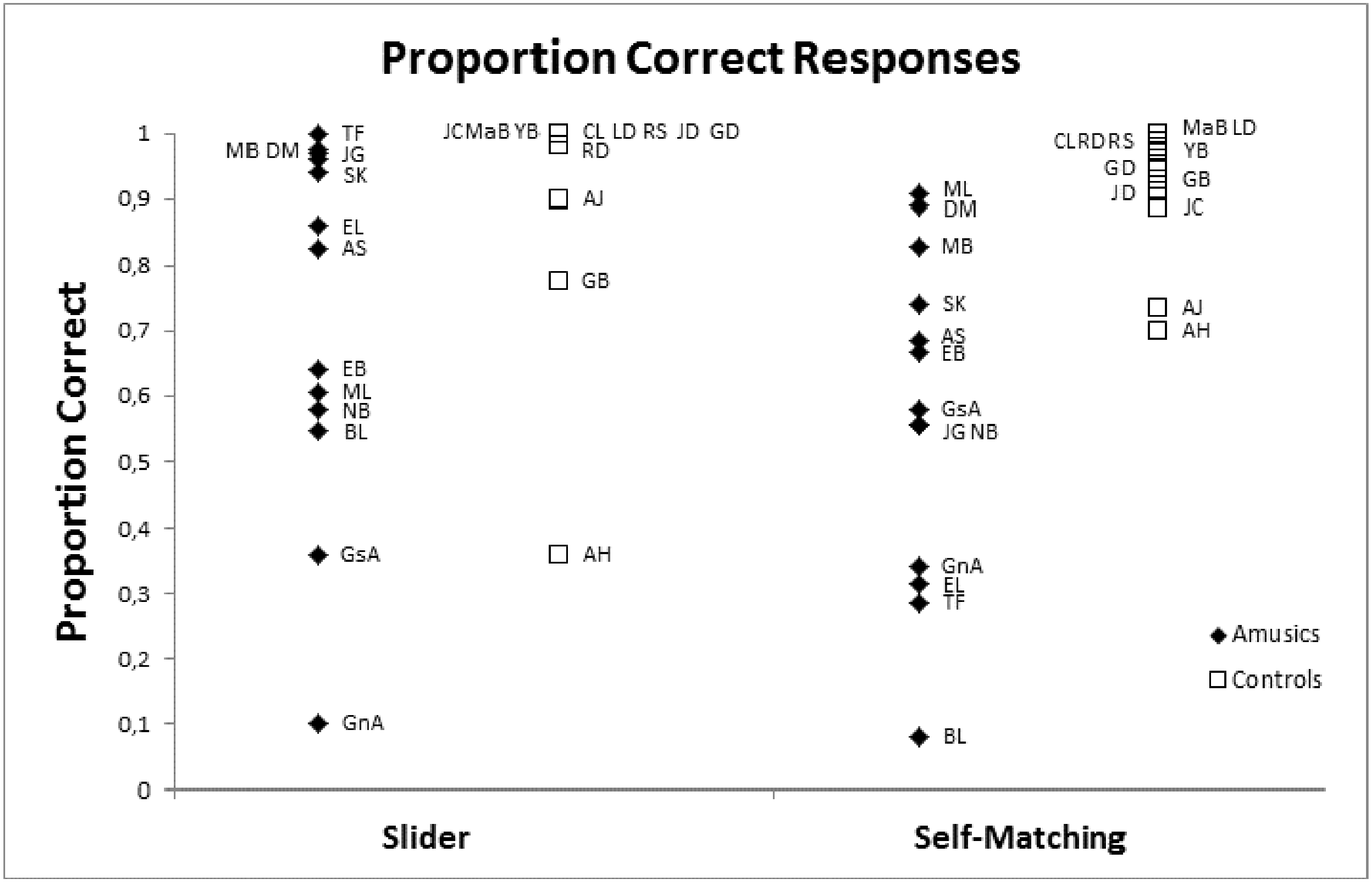
B) 
Proportion Correct Responses
Proportion Correct
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
Slider
Self-Matching
TF
MB DM
JG
SK
EL
AS
EB
ML
NB
BL
GsA
GnA
JCMaB YB
CL LD RS JD GD
RD
AJ
GB
AH
ML
DM
MB
SK
AS
EB
GsA
JG NB
GnA
EL
TF
BL
MaB LD
CLRDRS
YB
GD
GB
JD
JC
AJ
AH
Amusics
Controls

Figure 5. Average absolute Slider error plotted against the average absolute Self-Matching error for both amusics (black diamonds) and controls (white squares).

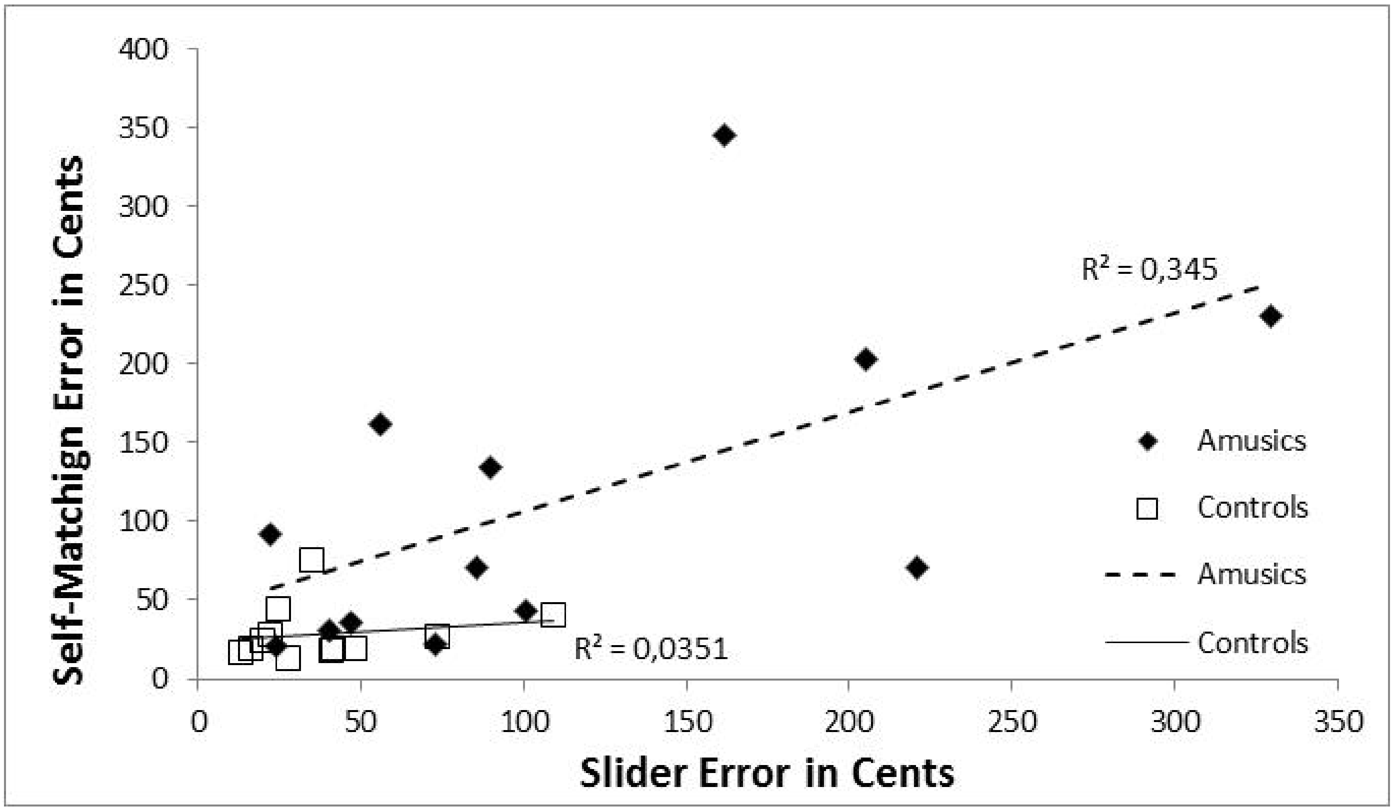


Figure 6. The proportion of accurate final responses for amusics and controls in both the Scanner Pitch Discrimination and Vocal-Imitation tasks from session 2, with standard error bars.

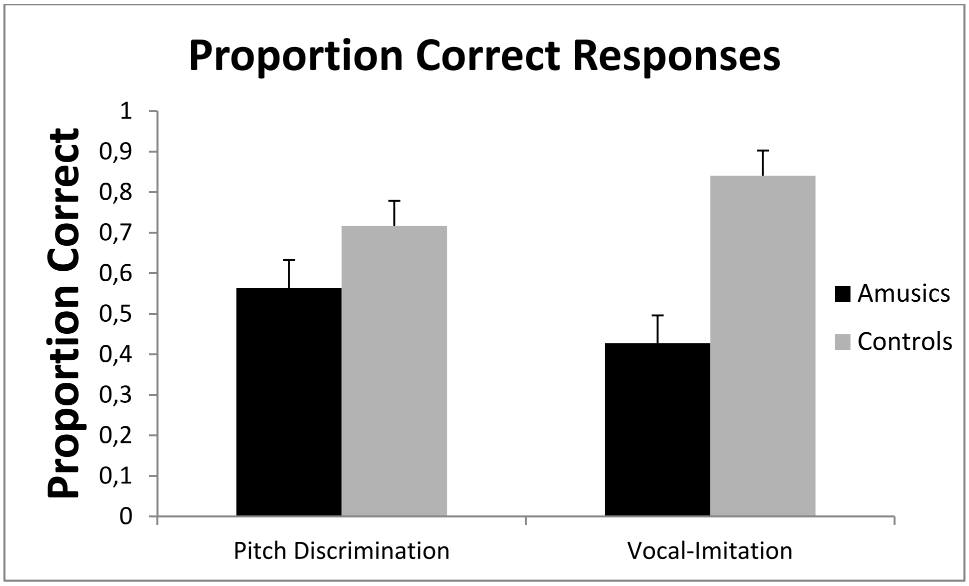


Figure 7. The scatterplot of all individual scores from the Scanner Pitch Discrimination and Vocal-Imitation tasks of session 2 for the proportion of accurate final responses.

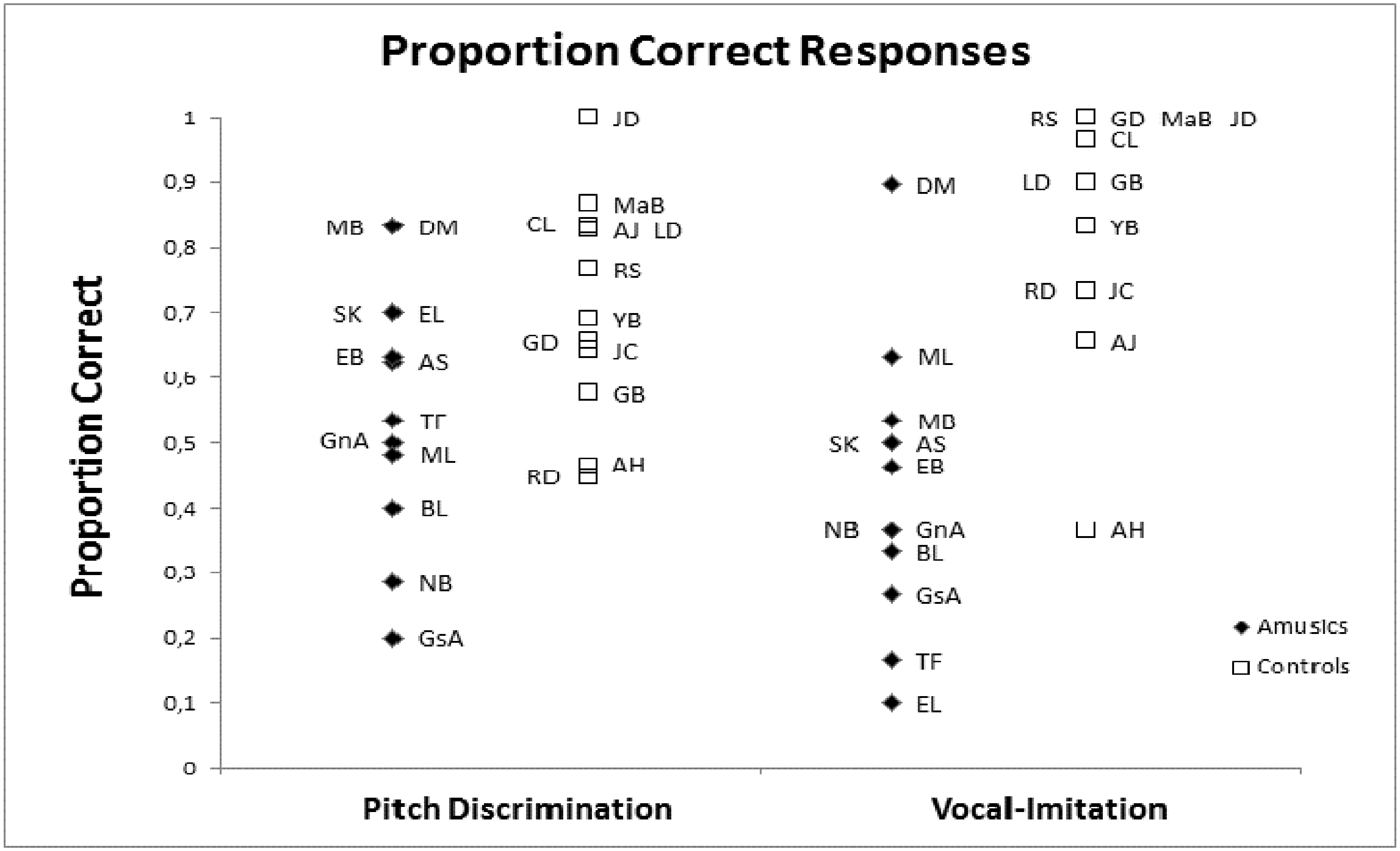
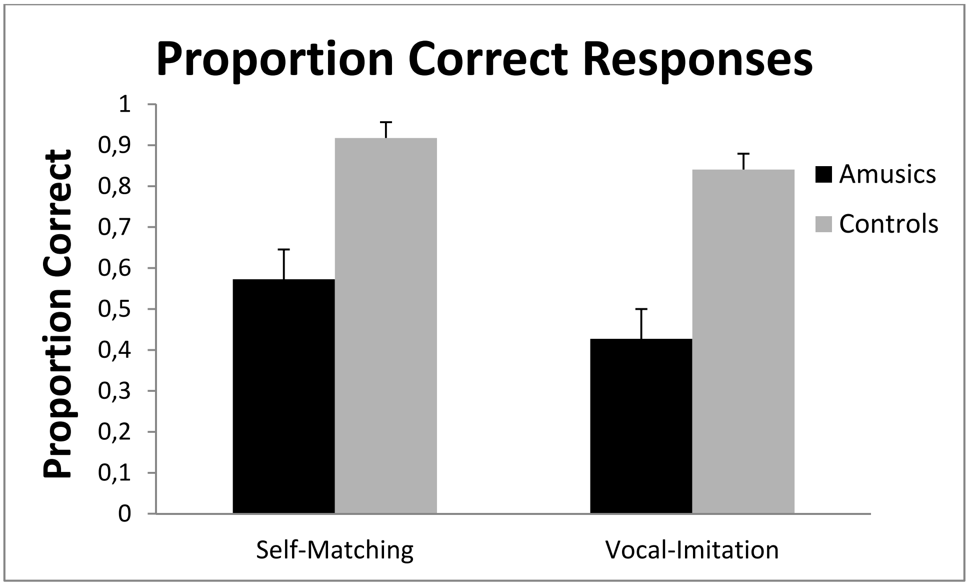


Figure 8. The proportion of accurate final responses and the absolute average value of the pitch error produced for amusics and controls in both the Self-Matching and Scanner Vocal-Imitation tasks from session 1 and 2, with standard error bars.



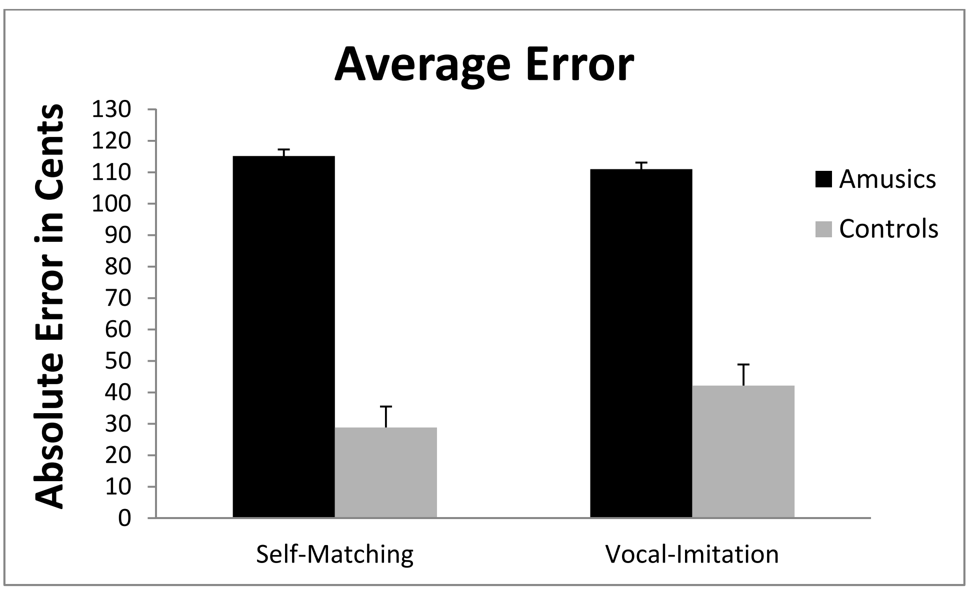


Figure 9. The average d-prime score of amusics and controls on the Scanner Pitch Discrimination task for each pitch difference (50 cents or 100 cents), with standard error bars.

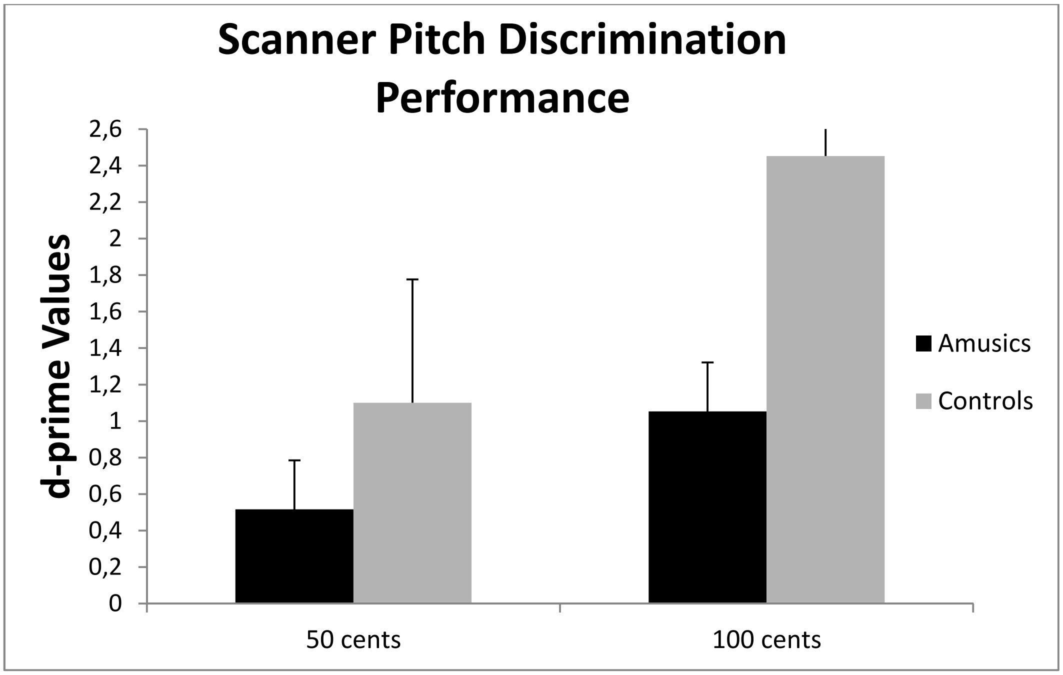
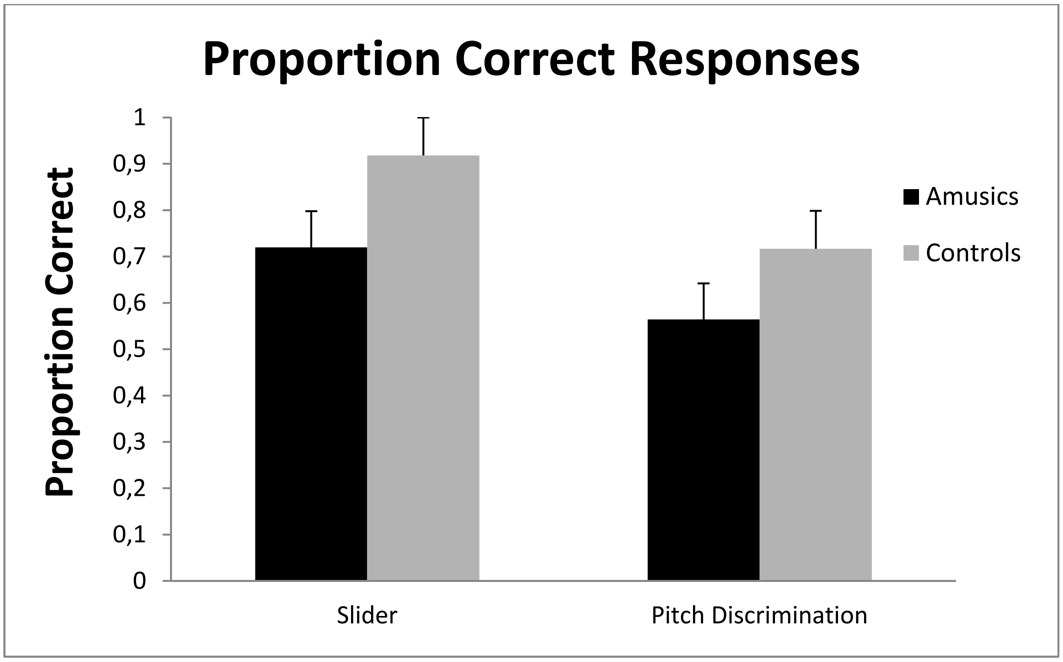


Figure 10. The proportion of accurate final responses for amusics and controls comparing the Slider and Scanner Pitch Discrimination perceptual tasks from session 1 and 2, with standard error bars.



Discussion générale

L'objectif général de ce mémoire était d'étudier la relation entre les capacités de perception et de production vocale des individus souffrant d'amusie congénitale. Plus précisément, les tâches présentées dans l'article nous ont permis de documenter ces habiletés à l'aide de différents protocoles. Les résultats ont démontré que les deux groupes (amusiques et contrôles) étaient meilleurs pour imiter des cibles avec le *slider* qu'avec leurs voix, même si cet instrument leur était non-familier. Les contrôles étaient plus précis que les amusiques sur toutes les tâches, même si quelques amusiques pouvaient atteindre des performances semblables à celles des contrôles sur certaines tâches. Il y avait une grande diminution de la performance pour les deux groupes lorsqu'ils devaient imiter vocalement leur propre voix à l'intérieur du scanneur, mais les performances sont restées constantes. Le même effet s'est produit au sein des tâches perceptives, où la tâche dans le scanneur était plus difficile sans affecter la constance des performances. Les deux groupes abordaient la tâche du *slider* très différemment des tâches vocales, étant beaucoup plus à l'aise à faire des tentatives sur le *slider* qu'avec leur voix, expliquant ainsi les différences marquées de duration des essais. Les résultats démontrent aussi un éventail très variable des habiletés d'imitation perceptive et d'imitation vocale et impliquent différentes causes sous-jacentes pour ces déficits. Celles-ci seront étudiées de manière plus complète dans une étude future.

Dans une perspective future, ces résultats seront approfondis au niveau structurel, fonctionnel, et au niveau de la connectivité neuronale. Aucune étude n'a encore pu combiner toutes ces informations provenant de techniques de neuroimagerie, incluant l'IRM de diffusion. Les données existantes proposent que l'amusie puisse être décrite en tant que syndrome de déconnexion neuronal, où les voies de matière blanche qui mènent l'information musicale (principalement l'information de hauteur de note) des aires auditives du cerveau aux aires frontales sont réduites (Hyde, Zatorre, Griffiths, Lerch, & Peretz, 2006). Cela suggère que les habiletés de production et de perception de hauteur de note seraient des habiletés dissociables (Hutchins et al., 2010; Loui et al., 2009; Loui et al., 2008), activant différentes aires spécifiques des cortex auditifs primaire et secondaire, plus précisément dans les régions neuronales englobant le faisceau arqué (Hyde, Peretz, & Zatorre, 2008; Johnsrude, Penhune, & Zatorre, 2000; Kleber, Veit, Birbaumer, Gruzelier, & Lotze, 2010; Zarate & Zatorre, 2008). Nos études futures pourront démontrer les causes sous-jacentes des habiletés musicales appauvries en termes de leurs fondements neuronaux (volume neuronal, anatomie structurelle et connectivité des voies neuronales d'intérêt) et possiblement illustrer une dissociation entre les processus de perception et de production dans le chant.

En conclusion, cette étude a permis de clarifier les habiletés comportementales de perception et de production vocale chez les amusiques. Ceci nous aide dans une perspective d'élaboration d'une théorie formulant les causes des déficits de production vocale, chez les amusiques et la population normale. Ces résultats nous aident aussi à établir une meilleure définition de la relation entre ces habiletés et leur influence l'un sur l'autre.

Références

Amir, O., Amir, N., & Kishon-Rabin, L. (2003). The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America, 113*, 1102.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia A group study of adults afflicted with a music-specific disorder. *Brain, 125*(2), 238-251.

Berkowska, M., & Dalla Bella, S. (2009). Acquired and congenital disorders of sung performance: A review. *Advances in Cognitive Psychology, 5*, 69.

Bradshaw, E., & McHenry, M. A. (2005). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice, 19*(3), 431-439.

Dalla Bella, S., & Berkowska, M. (2009). Singing proficiency in the majority. *Annals of the New York Academy of Sciences, 1169*(1), 99-107.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2007). Singing proficiency in the general population. *The journal of the Acoustical Society of America, 121*, 1182.

Dalla Bella, S., Giguère, J.-F., & Peretz, I. (2009). Singing in congenital amusia. *The Journal of the Acoustical Society of America, 126*, 414.

De Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America, 111*, 1917.

Dowling, W. J. (1999). *The development of music perception and cognition*: Academic Press.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition* (Vol. 19986): Academic Press New York.

Estis, J. M., Dean-Claytor, A., Moore, R. E., & Rowell, T. L. (2011). Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. *Journal of Voice, 25*(2), 173-180.

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*(4), 801-810.

Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education, 31*(2), 93-99.

Goetze, M. (1985). *Factors affecting accuracy in children's singing*: University of Colorado at Boulder.

Gould, A. O. (1969). Developing specialized programs for singing in the elementary school. *Bulletin of the Council for Research in Music Education*, 9-22.

Green, G. A. (1990). The effect of vocal modeling on pitch-matching accuracy of elementary schoolchildren. *Journal of Research in Music Education, 38*(3), 225-231.

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences of the United States of America, 102*(35), 12639-12643.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402.

Huron, D. (2001). Is music an evolutionary adaptation? *Annals of the New York Academy of Sciences, 930*(1), 43-61.

Hutchins, S., & Peretz, I. (2011). Perception and action in singing. *Progress in brain research, 191*, 103.

Hutchins, S., & Peretz, I. (2012). A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology-General, 141*(1), 76.

Hutchins, S., Roquet, C., & Peretz, I. (2012). The Vocal Generosity Effect: How Bad Can Your Singing Be? *Music Perception, 30*(2), 147-159.

Hutchins, S., Zarate, J. M., Zatorre, R. J., & Peretz, I. (2010). An acoustical study of vocal pitch matching in congenital amusia. *The Journal of the Acoustical Society of America, 127*, 504.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science, 15*(5), 356-360.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia, 46*(2), 632-639.

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.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123*(1), 155-163.

Joyner, D. R. (1969). The monotone problem. *Journal of Research in Music Education, 17*(1), 115-124.

Kalmus, H., & Fry, D. (1980). On tune deafness (dysmelodia): frequency, development, genetics and musical background. *Annals of human genetics, 43*(4), 369-382.

Kleber, B., Veit, R., Birbaumer, N., Gruzelier, J., & Lotze, M. (2010). The brain of opera singers: experience-dependent changes in functional activation. *Cerebral Cortex, 20*(5), 1144-1152.

Klemish, J. (1974). Treating the uncertain singer through the use of the tape recorder. *Bulletin of the Council for Research in Music Education*(37), 36-45.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 739.

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain, 133*(6), 1682-1693.

Loui, P., Alsop, D., & Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *The Journal of Neuroscience, 29*(33), 10215-10220.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action–perception mismatch in tone-deafness. *Current Biology, 18*(8), R331-R332.

Melara, R. D., & Marks, L. E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics, 48*(2), 169-178.

Merriam, A. (1964). *The anthropology of music*: Northwestern Univ Press.

Mithen, S. (2006). The 'Singing Neanderthals': the origins of music, language, mind and body. *Cambridge Archaeological Journal, 16*(1), 97-112. doi: Doi 10.1017/S0959774306000060

Moore, R. E., Estis, J., Gordon-Hickey, S., & Watts, C. (2008). Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice, 22*(4), 399-407.

Moore, R. E., Keaton, C., & Watts, C. (2007). The role of pitch memory in pitch discrimination and pitch matching. *Journal of Voice, 21*(5), 560-567.

Moreau, P., Jolicœur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain and cognition, 81*(3), 337-344.

Ostwald, P. F. (1973). Musical behavior in early childhood. *Developmental Medicine & Child Neurology, 15*(3), 367-375.

Peretz, I. (2001). Brain specialization for music. *Annals of the New York Academy of Sciences, 930*(1), 153-165.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. t. (2002). Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron, 33*(2), 185-191.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain, 132*(5), 1277-1286.

Peretz, I., Brattico, E., & Tervaniemi, M. (2005). Abnormal electrical brain responses to pitch in congenital amusia. *Annals of neurology, 58*(3), 478-482.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences, 999*(1), 58-75.

Peretz, I., Gosselin, N., Tillmann, B., Cuddy, L. L., Gagnon, B., Trimmer, C. G., . . . Bouchard, B. (2008). On-line identification of congenital amusia. *Music Perception: An Interdisciplinary Journal, 25*(4), 331-343.

Pfordresher, P. Q., & Brown, S. (2007). Poor-Pitch Singing in the Absence of" Tone Deafness". *Music Perception, 25*(2), 95-115.

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics, 71*(6), 1385-1398.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America, 128*, 2182.

Pfordresher, P. Q., & Mantell, J. (2009). Singing as a form of vocal imitation-Mechanisms and deficits.

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.

Satoh, M., Takeda, K., & Kuzuhara, S. (2007). A case of auditory agnosia with impairment of perception and expression of music: cognitive processing of tonality. *European Neurology, 58*(2), 70-77.

Schön, D., Lorber, B., Spacal, M., & Semenza, C. (2004). A selective deficit in the production of exact musical intervals following right-hemisphere damage. *Cognitive neuropsychology, 21*(7), 773-784.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2001). Implicit learning of regularities in Western tonal music by self-organization *Connectionist Models of Learning, Development and Evolution* (pp. 175-184): Springer.

Trehub, S. E. (2001). Musical predispositions in infancy. *Ann N Y Acad Sci, 930*, 1-16.

Trehub, S. E., & Trainor, L. (1998). Singing to infants: Lullabies and play songs. *Advances in infancy research, 12*, 43-78.

Tremblay-Champoux, A., Dalla Bella, S., Phillips-Silver, J., Lebrun, M.-A., & Peretz, I. (2010). Singing proficiency in congenital amusia: Imitation helps. *Cognitive neuropsychology, 27*(6), 463-476.

Wallin, L., Merker, B., & Brown, S. (2001). *The origins of music*: The MIT Press.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & psychophysics, 64*(2), 198-207.

Watts, C., & Hall, M. (2008). Timbral influences on vocal pitch-matching accuracy. *Logopedics Phonatrics Vocology, 33*(2), 74-82.

Watts, C., Moore, R., & McCaghren, K. (2005). The relationship between vocal pitch-matching skills and pitch discrimination skills in untrained accurate and inaccurate singers. *Journal of Voice, 19*(4), 534-543.

Yarbrough, C. (1991). Inaccurate Singers: An Exploratory Study of Variables Affecting Pitch-Matching. *Bulletin of the Council for Research in Music Education, 107*, 23-34.

Zarate, J. M., & Zatorre, R. J. (2008). Experience-dependent neural substrates involved in vocal pitch regulation during singing. *Neuroimage, 40*(4), 1871-1887.