AN EAR FOR PITCH

On the Effects of Experience and Aptitude in Processing Pitch in Language and Music

Salomi S. Asaridou

ISBN 978-90-76203-64-5

The cover of this thesis was designed by Marina Kyriazi.

The research reported in this dissertation was supported by an International Max Planck Research School fellowship funded by the Max Planck Society, Munich, Germany.

An Ear for Pitch

On the Effects of Experience and Aptitude in Processing Pitch in Language and Music

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op dinsdag 28 april 2015
om 14.30 uur precies

door
Salomi Savvatia Asaridou
geboren op 6 augustus 1983
te Panorama, Griekenland

Promotoren:

Prof. dr. James M. McQueen
Prof. dr. Peter Hagoort

Manuscriptcommissie:

Prof. dr. Carlos H.M. Gussenhoven
Prof. dr. Mireille Besson (Aix Marseille Université, Frankrijk)
Dr. Daniel J. Acheson (MPI)

An Ear for Pitch

On the Effects of Experience and Aptitude in Processing Pitch in Language and Music

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. Th.L.M. Engelen,
according to the decision of the Council of Deans
to be defended in public on Tuesday, April 28, 2015
at 14.30 hours

by
Salomi Savvatia Asaridou
born August 6 1983
in Panorama, Greece

Supervisors:

Prof. dr. James M. McQueen
Prof. dr. Peter Hagoort

Doctoral Thesis Committee:

Prof. dr. Carlos H.M. Gussenhoven
Prof. dr. Mireille Besson (Aix Marseille Université, France)
Dr. Daniel J. Acheson (MPI)

Acknowledgments

As you set out for Ithaca
hope the voyage is a long one,
full of adventure, full of discovery…

C.P. Cavafy, Ithaca

Throughout the journey to my Ithaca, my PhD, I would bring this beloved poem to mind, to empower and inspire me. At moments when the data was uninterpretable, participants not showing up, when scripts had to be debugged and long lab meetings to be endured, these verses would remind me to appreciate every moment of this journey. Having happily reached the island, I am grateful for all the people that helped me get there a little wiser and more experienced than when I set off.

These were, first and foremost, my supervisors, James and Peter.

James, I consider myself incredibly lucky to have had you as my primary supervisor. Not only are you a great scientist and thinker, an eloquent writer and speaker, you are also an inspiring mentor. Your door was always open, metaphorically and literally, encouraging me to seek your advice on any matter. Thank you for trusting my research ideas and for being so supportive.

Peter, thank you for your guidance and mentorship. Your scientific and leadership achievements in the field are impressive and I feel privileged to have worked with you. I am especially grateful that you so generously supported my risky and rather expensive research project when my fellowship funding was running low.

I would like to thank my reading committee members, Carlos Gussenhoven, Mireille Besson and Dan Acheson for taking the time to read and evaluate this dissertation. It means a lot to me to have Mireille Besson in my committee since it

was her work that kindled my interest in language and music and inspired my initial doctoral thesis proposal.

Next, I would like to thank my co-authors. Atsuko Takashima, “thank you” seems hardly enough to express my gratitude. I am deeply indebted to you for teaching me everything I know about fMRI analysis, for training me to operate the MRI scanner, for helping me out with testing, for providing new ideas for data analyses and interpretation, and for doing all this in the most patient, respectful and considerate way.

Dan Dediu, I am thankful for your valuable thoughts and input on my Dutchinese experiment. Your unending enthusiasm about tones, genes, pitch perception and sound learning is really contagious. It made me always look forward to our meetings.

Hubert Fonteijn thank you for patiently teaching me how to perform resting-state fMRI and DTI analysis, for helping me troubleshoot when something went wrong, and for brainstorming about my complicated results.

Many thanks to my collaborators: Michele Gubian, without your magic algorithm I wouldn’t have had such controlled and precise stimuli; Pieter van Groenestijn, thank you for helping me to program the online Dutchinese rating questionnaire; Wencui and Haiteng for helping me translate the questionnaire in Mandarin; Lin Wang for recruiting students in Beijing for the rating study; Danchao, Jiyoun and Makiko, thank you for spending your valuable time piloting my experiments; Hubert Voogd and Pascal de Water thank you for your immense help with programming my experiments. I would also like to thank my Mandarin speakers: Wencui, Lin, Zeshu, Mingyuan, Guyan, Xumin, Zilin and my Dutch speakers, Patrick, Arina, Alma, Saskia, Joost, Huib, Matthias and David for lending their voices to the Dutchinese words that haunted my participants; Carmen, Merel and Alex for helping out with Chapter 3 stimulus recordings. Yiya Chen, thank you for hosting my experiment in Leiden and helping me recruit the ever elusive Dutch-Cantonese bilinguals. I am also grateful to Regine Kolinsky and Pascale Lidji for

making their Vowel-Interval Speeded Classification stimuli available to me. Paul Gaalman, thank you for teaching me how to operate the scanner at the DCCN and for making the MRI lab such a fun and cosy place to work in.

I owe a big thank you to all my participants, these brave young individuals who endured multiple long and challenging sessions in the testing cubicle and the MRI scanner for my experiments. Thank you for your diligence and enthusiasm. This thesis wouldn't have been possible without you.

As an IMPRS student, I had the privilege of being affiliated with multiple research groups across the MPI and the Donders Institutes. I would like to thank all the members of the Neurobiology of Language department at the MPI, the members of the Sound Learning group, and the members of the Language Division Meeting at the DCC, for countless stimulating meetings and discussions.

I would like to thank the International Max Planck Research School (IMPRS) for Language Sciences for funding my PhD project. I am very grateful to Dirkje, Rachel and Els for all their support and guidance throughout my PhD. Special thanks to all my IMPRS colleagues, my co- workers, administrative staff, secretaries, librarians, technicians and lab managers at the Spinozagebouw, the MPI, the DCC and DCCN.

Alina and Haiteng, I'm so excited to have you as my paranymphs! Thank you for being such great friends throughout my PhD journey. You were always there, whether I needed a scanning buddy on a Friday evening, a last-minute translation, a matlab script, or a place to stay. It means a lot to me to have you by my side on my defense.

A big thank you to my dear friends in the Netherlands: Rosa and Sho- chan, thank you for all the wonderful moments we shared, our conversations, our trips, our parties, our dinners, our highs and lows; Sarita, I always looked forward to meet you for hot cocoa, a walk, or even when we had to do our advanced math homework together; Katja, it was so great being your paranymph and spending time with you in the US; also thank you and Remon for helping me translate the summary of my thesis in Dutch; Wencui, thank you for being so generous, warm and open-hearted; Carmen, thank you for always being such a sweet and considerate friend; Jiyoun for all the cookie-baking/ dumpling-making memories and your sweet smile; Ewelinko thank you for cheering up my mornings and evenings and, together with Ely and Giovanni, making our apartment feel like home. Huadong, Lin, Kevin, Christina, Iske, Inge, Helen, Maarten, Thordis, Zeshu, Mingyuan, Cheng Hua, William, Claudia, Virgilia I had so much fun hanging out with you; Binyam and Amit, I already miss our Indian dinners; Maraki it was wonderful going to concerts with you and spending time with your family in Utrecht; Manolio and Tulay, thank you for being a "piece" of family in the Netherlands.

I am also deeply thankful to my friends scattered around the world who supported me by visiting me in the Netherlands, writing me nice emails, skyping with me for hours on end, by sending me beautiful postcards or packages filled with goodies. Especially big hugs to Athina, Marina, Thomas, Katerina G., Katerina P., Despina, Nikiforos, Michalis, Jessamyn, Karen, Trevor, Annoula, Yi-Chun, Thanasis, Maria, Christina, Alexandra and Antonis.

Finally, I would like to thank my family in Greece and Switzerland for their unconditional love and encouragement. Peti, Claudia, Reni, und Nane viele Dank für Ihri Liebi und Unterstützig. Anna and Sophie, I can't thank you enough for everything you've done for me throughout the years. Είστε οι καλύτερες θείες του κόσμου! Mom, dad, Chloe and Dimitri, your love has been the most essential travel equipment on the pursuit of my Ithacas. Thank you. Αυτή η διατριβή είναι αφιερωμένη σε σας.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Individual variation in perceptual, motor and cognitive skills is omnipresent among humans (Thornton & Lukas, 2012). Indeed, we are all acquainted with this variation in our daily lives, as we come across individuals who excel in drawing while others can barely draw a straight line; or individuals that can taste subtle differences between different types of wine when others cannot; or individuals who are very fast at memorizing new information whilst others have to study for hours and still do not accomplish the same result.

Speech and music are domains in which such individual variation is extremely common. It is not a coincidence that there is an idiomatic expression in English which captures this phenomenon: “to have an ear for something”. According to the Cambridge Idioms Dictionary the idiom refers to the following: “if someone has an ear for music, poetry, or languages, they are good at hearing, repeating, or understanding these sounds” (*Cambridge Idioms Dictionary*, 2006) while the Cambridge Dictionary of American Idioms defines it as: “to be especially good at hearing and repeating sounds […] Usage notes: said especially of speech or music” (*Cambridge Dictionary of American Idioms*, 2003). So, why is it that some individuals have an ear for sounds, be it linguistic or musical sounds, while others don’t? This is the core question of the current dissertation. Two obvious reasons why people may vary in “having an ear” are experience, in the form of training with or exposure to sounds, and aptitude, referring to the innate predisposition or talent to process sounds. In a series of experiments, using behavioral and neuroimaging approaches, the differential contributions of experience and aptitude are evaluated with the ultimate goal of gaining better understanding of what underlies individual variation in sound processing.

What is the role of experience in sound processing?

Experience, whether in the form of simple exposure to or explicit training with linguistic or musical sounds, is expected to explain a large proportion of the individual variation in sound processing. Undoubtedly, an individual who has received musical training will outperform naïve individuals in identifying musical intervals while an individual that speaks Mandarin will be better at identifying Mandarin tones compared to those who have never heard Mandarin before. A more interesting question is then: will musical training also offer such an advantage in identifying non-native Mandarin tones? And will experience with Mandarin be beneficial for identifying musical intervals? These are the questions addressed in Chapters 2 and 3.

A central assumption throughout this thesis is that speech and music rely to some extent on shared, domain-general, sound-processing mechanisms. That is, both domains use the same properties in the acoustic signal (primarily pitch and timbre) as their building blocks and hence they employ the same subcortical and cortical auditory processing networks to process them. According to this assumption, individuals who have musical experience will be better not only in processing musical sounds but also in processing linguistic sounds. An increasing number of studies lend support to this assumption by showing transfer effects from music to speech, such that individuals who have received musical training are better at perceiving (identifying and discriminating), learning and producing native and non- native speech sounds (Besson, Chobert, & Marie, 2011b; Kraus & Chandrasekaran, 2010). However, for the assumption to hold one would also have to have evidence for speech – to – music transfer effects. That is, if the benefits of musical training on speech are caused by sharpening shared domain-general processing resources then one would expect that linguistic experience should offer similar benefits.

Chapter 2 offers a comprehensive review of the evidence for bidirectional speech-music influences as well as of the theoretical frameworks that attempt to explain them. The review focuses on pitch-processing studies, that is, studies which used pitch, the perceptual attribute of sound frequency, in either linguistic or musical contexts to investigate transfer effects. According to *ANSI (1994)* “Pitch [is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high”, which is a neutral definition with respect to domain. Pitch carries important information in both language and music (Plack & Oxenham, 2005) while, in contrast to timbre, it can also be studied in a domain-general way with sine-wave tones. These properties make it ideal for investigating transfer effects from one domain to the other.

The reviewed literature suggests that the two domains make use of shared pitch processing mechanisms. More specifically, there is behavioral and electrophysiological evidence that experience with pitch in the musical domain influences the processing of pitch in speech, for example in the processing of lexical tone (Bidelman, Gandour, & Krishnan, 2011b; Magne, Schön, & Besson, 2006; Marie, Kujala, & Besson, 2012; Parbery-Clark, Skoe, & Kraus, 2009). Interestingly, the effect is bidirectional: experience with lexical tone in the speech domain in tone language speakers can in turn influence the processing of pitch in music (Bidelman, Gandour, & Krishnan, 2011a; Bidelman, Hutka, & Moreno, 2013; Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011a; Wong et al., 2012). Thus, there is empirical and theoretical support that both speech and music shape the way individuals process sounds via shared, domain-general sound processing resources

Despite the fact that there is evidence for the effect of language experience on sound processing, it is less robust than the equivalent effect of music experience. As discussed in detail in Chapter 2, there are several reasons for this asymmetry in the literature, one of which is the difficulty to define what constitutes language experience. In studies investigating pitch processing, language experience is defined simply as speaking a tone language. Tone language speakers' performance in processing pitch in a non- linguistic context is in this case compared to non-tone language speakers or musicians. The reported transfer effects in these studies are weak compared to studies on musical training effects. Musical training is much more effortful and demanding compared to speaking ones native language; it forces the individual to think about sounds and their manipulation in an explicit way that tone language experience does not. The fact that tone language speakers may not have explicit auditory expertise comparable to that of musicians could be responsible for the lack of robust speech-to-music transfer effects.

One of the conclusions of Chapter 2 is that plain linguistic experience with a tone language may not suffice for an individual to show advanced pitch processing. Based on that, the study in Chapter 3 sets out to test speech- to-music transfer effects in an alternative group of "auditory experts", namely early bilinguals who have learned a tone language and a non-tone language. These individuals have gained sound expertise by learning from a very young age onwards to perceive and produce two very diverse linguistic inventories (Cantonese and Dutch) that make different use of pitch (at the lexical level in Cantonese vs. the intonational level in Dutch). This bilingual experience requires more explicit manipulation of linguistic sounds, since individuals need to learn two diverse phonemic inventories, to use the appropriate sounds for each one and inhibit the inappropriate ones depending on the language at hand (Campbell & Sais, 1995). This in turn makes the experience more comparable to musical training, compared to "normal" tone language experience. Importantly, early bilinguals (or simultaneous bilinguals as they are some times referred to) acquire both languages from infancy, a fact that has long-term plasticity effects in their brain structure and function as well as in their linguistic and cognitive development (see Costa & Sebastián-Gallés, 2014 for a review).

A group of early bilinguals in Cantonese (a tone language) and Dutch (a non-tone language) was therefore compared to a group of monolingual Dutch speakers. The hypothesis tested in Chapter 3 was that due to their experience with two languages with different use of linguistic pitch, early Cantonese-Dutch bilinguals should outperform their Dutch peers in tasks that require non-linguistic pitch processing. Three different tasks that use non-linguistic pitch and tap different levels of speech and music interactions were administered: 1) a speeded classification task which assesses simultaneous processing of melodic and phonological information in sung speech, 2) a music interval identification training task which assesses the ability to learn new (music) sound categories, and 3) a series of pitch perception tasks which assess the ability to discriminate changes in pitch direction and intervals. It was expected that the variation in pitch processing performance would be explained by the participants' linguistic experience.

What is the role of aptitude? Neural correlates of successful vs. less successful sound learners

Chapters 2 and 3 emphasize the role of experience in the form of exposure and training which contributes to individual variation in sound processing. However, it is often the case that performance varies greatly even among individuals that have the same experience. Thus, another important factor that has to be taken into consideration is aptitude or predisposition for learning new sounds. Accumulating evidence from neuroimaging studies, comparing successful to less successful sound learners, has led to the suggestion that differences in aptitude arise as a result of differences in brain structure and function (Zatorre, 2013). In other words, variation in brain morphology may lead to variation in behavior, such that individuals who happen to have a larger auditory cortex or more myelinated fiber connections between auditory and frontal cortices, for example, may have an advantage in processing and learning linguistic and musical sounds.

Chapters 4 and 5 focus on the neural correlates of aptitude for non- native sound learning. Chapter 4 investigates which cortical and subcortical areas along the sound processing pathway are underlying individual variation in sound learning performance. Chapter 5 in turn explores how resting-state functional connectivity between brain areas involved in sound learning contributes to this variation. In the studies in both these chapters, Dutch native speakers without any prior experience with Mandarin tones were trained to learn twenty-four new words over the course of five different sessions. Participants learned to match auditory-presented words that differed minimally in pitch contour to pictures of ordinary objects. The training was thus lexical, that is it required form-meaning mappings instead of simple phonological categorization. In order to successfully match the words to their corresponding pictures, participants had to learn to discriminate and identify the non-native tones (pitch contours modeled after Mandarin tones) in the words. Their learning performance, measured as the percentage correct word-picture mappings, was assessed in each session.

Using pitch contours served a double purpose: not only it did it allow keeping the acoustic property studied throughout the thesis constant, but this non-native contrast also happens to be very difficult for Dutch native speakers. Based on that as well as previous literature using the same paradigm (Chandrasekaran, Sampath, & Wong, 2010; Wong, Perrachione, & Parrish, 2007), it was predicted that this training would give rise to large individual variation in learning performance. The training paradigm was designed to allow the identification of the whole spectrum of variation in sound learning performance from very high to medium and very low in a continuous rather than a bimodal way.

In Chapter 4 individual variation in learning performance is juxtaposed with participants' variation in neuronal responses to the non- native tones as measured with functional magnetic resonance imaging (fMRI). Participants' hemodynamic responses to the non-native sounds were recorded before and after receiving the sound training with an fMRI adaptation paradigm. FMRI adaptation or repetition suppression measures the reduction in the hemodynamic response to a stimulus when all or part of its features is presented repeatedly (Grill-Spector, Henson, & Martin, 2006). It has been suggested that the repetition suppression effect is indicative of less effortful or more efficient neuronal processing of the stimulus characteristics (Grill-Spector et al., 2006).

The aim of the study in Chapter 4 was to measure repetition suppression to non-native tones and test whether larger repetition suppression, and thus more efficient processing, is associated with tone learning performance. Although previous studies have shown that brain activation to non-native sounds differs in successful compared to less successful learners (Golestani & Zatorre, 2004; Wong, Perrachione, et al., 2007), this is the first study to investigate the role of processing efficiency by looking at fMRI adaptation in the whole brain. By measuring repetition suppression at two time points it was possible to evaluate both the contribution of pre-existing differences in processing (in the pre-training fMRI session) as well as the contribution of training (in the post-training fMRI session).

Although the role of specific brain areas to sound processing is important for understanding variation in learning performance, it is of equal importance to take into account the role of connectivity between these areas. Previous studies on the role of structural connectivity have revealed that variability in language learning performance is associated with variability in white matter connectivity in fiber tracts such as the left arcuate fasciculus and the left extreme capsule/inferior fronto-occipital fasciculus (Lopez-Barroso et al., 2011; López-Barroso et al., 2013; Wong, Chandrasekaran, Garibaldi, & Wong, 2011). Connectivity between brain areas, however, is not restricted to their direct anatomical connections.

The dynamic aspect of functional brain connectivity can be investigated with resting-state connectivity. Resting-state connectivity as measured with fMRI refers to the temporal correlations in the spontaneous fluctuations in the blood oxygenation level-dependent signal (BOLD) between brain areas (Biswal, Yetkin, Haughton, & Hyde, 1995). This connectivity is measured when the brain is at rest, that is, when participants are not engaged in any structured activity in the MRI scanner. Resting-state connectivity is not random but it reflects consistent connectivity patterns between brain areas that form functional networks (Smith et al., 2009). As such, resting-state fMRI can serve as a useful tool to study how connectivity contributes to learning and plasticity as well as individual variation in cognitive performance (Guerra-Carrillo, Mackey, & Bunge, 2014).

Chapter 5 considers the role of resting-state connectivity in non-native sound learning performance. Resting-state data was collected from the same set of participants recruited in the fMRI adaptation study (Chapter 4). Using the same experimental set up, resting-state fMRI was recorded before and after participants completed the five-session behavioral training in non-native tones. A seed-based correlation approach was taken with bilateral auditory cortex areas and left ventrolateral prefrontal cortex (VLPFC) areas as seeds in the analysis. These areas were selected based on previous literature on tone learning which has shown their increased involvement in tone processing after training (Wong et al., 2007). Two hypotheses were put to test: the first one was that the strength of resting-state connectivity between areas involved in spectrotemporal processing (both speech and pitch processing) and higher order language areas (VLPFC) would change as a result of training. The second hypothesis was that connectivity changes would vary as a function of individual variation in non-native word learning performance. Resting-state connectivity has been previously studied in relation to the individual's ability to learn to discriminate and identify a non-native phonetic contrast (Ventura- Campos et al., 2013). By looking at non-native sound contrasts embedded in words and connected to meaning, the study in Chapter 5 is the first to test the role of resting-state connectivity in mapping sounds to meaning performance. That is, it does not simply focus on non-native phonetic sound perception but on the role of neuronal communication at the lexical crossroads between phonetic, phonological and semantic processing.

Outline

The studies reported in this thesis explore the differential contribution of linguistic experience (Chapters 2 and 3) and neuronal predispositions (Chapters 4 and 5) in “having an ear for” pitch. Chapter 2 reviews the literature on how linguistic experience with pitch can shape domain-general pitch processing, in a way comparable to musical experience. Putting the assumptions of the theoretical frameworks reviewed in Chapter 2 into test, Chapter 3 investigates how experience with two diverse languages in early bilinguals can influence different levels of non-linguistic pitch processing. Chapter 4 focuses on the role of individual differences in neuronal processing efficiency, measured with fMRI adaptation, in learning non-native tones. Finally, Chapter 5 explores how differences in resting-state connectivity patterns contribute to individual variation in non-native sound-to-meaning learning performance. Chapter 6 provides a general discussion of the results presented in this thesis along with conclusions and suggestions for future directions.

CHAPTER 2

SPEECH AND MUSIC SHAPE THE LISTENING BRAIN: EVIDENCE FOR SHARED DOMAIN-GENERAL MECHANISMS

After: Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*(321), 1-14. doi:10.3389/fpsyg.2013.00321

Abstract

Are there bi-directional influences between speech perception and music perception? An answer to this question is essential for understanding the extent to which the speech and music that we hear are processed by domain-general auditory processes and/or by distinct neural auditory mechanisms. This review summarizes a large body of behavioral and neuroscientific findings which suggest that the musical experience of trained musicians does modulate speech processing, and a sparser set of data, largely on pitch processing, which suggest in addition that linguistic experience, in particular learning a tone language, modulates music processing. Although research has focused mostly on music on speech effects, we argue that both directions of influence need to be studied, and conclude that the picture which thus emerges is one of mutual interaction across domains. In particular, it is not simply that experience with spoken language has some effects on music perception, and vice versa, but that because of shared domain-general subcortical and cortical networks, experiences in both domains influence behavior in both domains.

There are two ways to approach the comparison of language and music: either by providing a long list of their differences, or a surprisingly long list of their commonalities. In recent years, the latter way has been far more popular than the former. This is not an attempt to underrate the uniqueness of each domain in the human cognitive repertoire. Language and music are undoubtedly systems with distinct representations, structure, and utility. Nevertheless, commonalities do emerge when one considers that they share the same basic building blocks. For both perception of speech and perception of music, the starting point is the temporally organized acoustic signal (Besson, Faita, Czternasty, & Kutas, 1997; McMullen & Saffran, 2004; Patel, 2008). Despite the fact that speech primarily makes use of timbral while music makes use of pitch contrasts, pitch information is also relevant to speech, and timbral contrasts are also used in music, whilst both organize the acoustic signal in distinct sound categories (Patel, 2008).

One might nonetheless wonder: why is a comparative approach to language and music interesting? After all, no matter what the similarities are, a jazz improvisation piece will always be easily distinguishable from a homily. Apart from purely theoretical reasons for taking a comparative approach (see Besson & Schön, 2011) a great incentive for emphasizing the shared properties of language and music stems from accumulating evidence showing that experience with one of the two induces plastic changes to the brain's structure and function. It has been long argued, for example, that the musician's brain provides a model for plasticity (Münte, Altenmüller, & Jäncke, 2002). What has drawn even more attention to this topic is the fact that extensive music training enhances auditory processing not only within but also beyond this domain, to general auditory and speech processing. This finding is of great value to our understanding of auditory perception mechanisms and their plastic properties. In particular, it indicates that at least some auditory mechanisms are domain-general in nature, and thus are not special to either music or speech processing.

The spotlight of attention so far has been mostly on the effects of musical training and experience on linguistic processing. However, there are two terms in the music and language equation, and although focusing on the consequences of music experience on speech is justifiable, it is of equal importance to investigate what happens when the terms are reversed. Indeed, given the aforementioned similarities between the two domains and the idea that influences of music on speech arise because of shared, domain-general auditory mechanisms, it is likely that linguistic experience will have an effect on music processing. Asking whether the influences are bidirectional thus offers an important test of the claim about domain-general processes and should help to define where in the processing stream those general mechanisms end and where domain-specific mechanisms begin.

In the present paper we will review the evidence for bi-directional influences between speech and music. While language and music influence each other at multiple levels from sounds and melodies to semantics and syntax, in this review we will focus on the level of sound processing. We begin by summarizing the extensive evidence on the effects of musical experience on linguistic sound processing and then discuss existing theoretical frameworks that seek to explain these data. That discussion leads to the theories' predictions concerning the effects of linguistic experience on musical behavior, and then a review of the smaller body of findings about such effects. We will discuss behavioral data and describe the brain structures which appear to be involved in music and speech processing, making the case that there are resources shared across domains. We also cover the evidence on mutual interactions between speech and music, as well as on structure- function associations in the brain. We then discuss the challenges that will need to be faced by future research in this area. We conclude that there is convincing evidence that speech and music interact in shaping the auditory brain and in jointly determining aspects of perceptual behavior in both domains.

Effect of music on speech

There is a wide range of research focusing on the effect of music on speech processing. At the behavioral level, there is evidence that musical aptitude correlates uniquely with L2 phonological production and perception abilities in adults (Slevc & Miyake, 2006) as well as in children (Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008). This correlation between the ability to perceive, discriminate and process music sounds, on the one hand, and the ability to perceive and pronounce non-native speech sounds in musically naïve individuals on the other, suggests that common processing mechanisms mediate both. Musical aptitude can also predict performance in linguistic tone discrimination task in non-tone-language speakers (Delogu, Lampis, & Belardinelli, 2010). Non-tone-language speakers that score highly on melodic perception tasks also score higher in tonal discrimination tasks (Delogu et al., 2010). Furthermore, individuals with music training are better in discriminating and imitating Mandarin tones than nonmusicians, even when the task requires categorical rather than pure auditory perception (Gottfried, Staby, & Ziemer, 2001). This performance superiority cannot be attributed to absolute pitch abilities (Lee & Hung, 2008).

Domain-general sound processing abilities have been found to be predictors of lexical tone learning performance (Wong et al., 2008; Wong & Perrachione, 2007). Participants' performance in a non-lexical, pitch contour identification task was predictive of their ability to use pitch in a Mandarin- like word learning paradigm (Wong & Perrachione, 2007). Thus, the ability to perceive and represent pitch movement which is important in music facilitated learning lexical tone, or linguistic pitch patterns. These findings support the view that speech processing depends, at least in part, on domain-general processes shared with music.

Subcortical auditory processing is also shaped by music training. Information about the malleability of subcortical auditory processing mechanisms primarily comes from electrophysiological studies using the FFR (Frequency Following Response) component. FFR is a brain-wave that is elicited preattentively and originates in the inferior colliculus in the rostral brainstem. It encodes the waveform of the $f0$ of an auditory stimulus in a phase-locked manner (Worden & Marsh, 1968). Wong, Skoe, Russo, Dees, and Kraus, (2007) compared FFR responses elicited by musicians and nonmusicians while listening to linguistic pitch patterns. They found that musicians' FFR responses followed pitch contours with greater fidelity than nonmusicians'. In addition, musicians' auditory brainstem responses encode spectral characteristics of the speech signal (vowel formants) with greater precision compared to nonmusicians when participants are listening to degraded speech (Bidelman & Krishnan, 2010; Parbery-Clark et al., 2009). Such enhancement of subcortical encoding of formant characteristics in speech can take place as early as three years of age, according to a recent study (Strait, O'Connell, Parbery-Clark, & Kraus, 2013). Long-term domain- specific training can therefore augment subcortical sound processing mechanisms (Wong, Skoe, et al., 2007). This superior subcortical neural representation of speech and music stimuli correlates positively with the amount of music training received, suggesting that it is primarily shaped by experience rather than innate abilities (Musacchia, Sams, Skoe, & Kraus, 2007).

Extensive training and experience with music leads not only to subcortical changes but also to plastic changes in the activation of the cerebral cortex, possibly by sharpening cortical preattentive and attentive pitch processing networks. When presented with speech, musicians showed higher bilateral middle temporal gyrus activation compared to non-musicians (Oechslin, Meyer, & Jäncke, 2010). The higher the training in music, the lower the activation found in primary auditory cortex areas, an indication of more efficient processing of acoustic information (Oechslin et al., 2010). Schön, Magne, and Besson (2004) used an EEG paradigm in which the $f0$ of the final syllable of a sentence was manipulated to create prosodically incongruous stimuli. Musicians showed advanced pitch contour processing of the sentences, as reflected by task performance accuracy as well as EEG recordings (Schön et al., 2004). Similar results are obtained when comparing children who have received musical training to their musically naïve peers (Magne et al., 2006).

Facilitated cortical pitch processing, as revealed by EEG data, in both music and language found in these "early musicians" constitutes positive transfer from one domain to the other. A longitudinal study by Moreno and colleagues (2009) controlled for existing predispositions in "early musicians", or effects of cognitive motivation and/or maturation that might have affected the results mentioned above. Children randomly assigned to receive music training outperformed their matched peers who had received an equally intense and interesting painting training, both in terms of accuracy but also in their electrophysiological responses to speech stimuli (Moreno et al., 2009). Moreover, this enhancement is not limited to native language processing but extends to foreign languages as well. French musicians were faster and more accurate than nonmusicians in detecting prosodic pitch violations in Portuguese, a language not spoken by either group (Marques, Moreno, Castro, & Besson, 2007).

It has been shown that musical training not only facilitates lexical tone processing but also segmental processing, such as for example the processing of consonants (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011). Interestingly, these facilitation effects cannot be merely due to attention (Marie et al., 2011). Musicians outperform non musicians in phonetic categorization and their superior performance is associated with higher left Planum Temporale (PT) activation (Elmer, Meyer, & Jäncke, 2012). In addition, musicians' electrophysiological responses to phonetic cues such as Voice Onset Time (the time between the release of articulatory closure and initiation of voicing) differ from nonmusicians', although no differences are detected in behavioral performance (Ott, Langer, Oechslin, Meyer, & Jäncke, 2011). This advantage was further observed in children who, after being randomly assigned to a musical training group, improved in Voice Onset Time (VOT) and syllable duration processing with one year of training (Chobert, François, Velay, & Besson, 2012). After two years of music training, children also improved their speech segmentation skills (François, Chobert, Besson, & Schön, 2013). Because the children were randomly assigned to the music training group and because of the longitudinal design, it can be concluded that the beneficial effects are due to the training and not pre- existing differences between groups.

This plethora of evidence showing that music training or aptitude can influence linguistic behavior casts doubt on whether music and speech are fully modular, encapsulated systems. Data from double dissociations in neuropsychological patients (i.e. patients with preserved speech production or comprehension but impaired tonal pitch abilities, and patients with spared tonal or singing but impaired speech abilities) previously led to the conclusion that music is subserved by components that are domain-specific and neuroanatomically distinct (Peretz, 2006, 2009). Peretz and Coltheart (2003) have proposed such a model in which a domain-general "acoustic analysis" module is the first to receive and process the acoustic input. Depending on the nature of the input, this module feeds it forward to a music-specific module ("contour analysis"), to a language-specific module ("acoustic-to-phonological conversion"), or to a module which has not yet been confirmed to be either musical or linguistic ("rhythm and meter") (Peretz & Coltheart, 2003). Although this model assumes that there is a common acoustic processing module, its role is not well defined and only forward flow of information from that module to further processing nodes is allowed. The literature reviewed above, however, suggests that there are either feedback connections from music processing levels to basic acoustic processing levels or direct connections between the domain-specific modules. With compelling evidence against strict modularity increasing, a number of theoretical frameworks that can account for language-music relationships have emerged.

Theoretical frameworks

We have reviewed behavioral, cortical and subcortical data showing that music training influences linguistic processing. How can these effects be explained? Several frameworks have been proposed, either referring to shared mechanisms between music and language, or even going beyond that to explain how transfer phenomena occur.

Sharpening of shared auditory skills

One of the most parsimonious accounts for transfer effects is one where music and language share the same auditory processing infrastructure. The argument made is that as this infrastructure becomes more efficient as a result of music experience, this leads to more efficient speech processing. The basic assumption is that the auditory system is malleable and changes with experience. This is supported by a variety of evidence ranging from animal studies to sensory deprivation and perceptual learning effects in humans (for a review see Kraus & Banai, 2007). The fact that music training retunes sound encoding even at its most basic subcortical level reinforces the view that domain-specific experience sharpens domain-general auditory mechanisms (Kraus & Banai, 2007; Kraus & Chandrasekaran, 2010; Krizman, Marian, Shook, Skoe, & Kraus, 2012; Skoe & Kraus, 2011). It is proposed that music training enhances these skills primarily through top-down feedback connections from cortical to subcortical sound encoding structures (Kraus & Chandrasekaran, 2010). Musicians learn to guide their attention to meaningful information in the acoustic signal, which in turn leads to improved sensory encoding of this information. Considering the overlap between the acoustic and cognitive demands for music and language, it has been suggested that similar listening skills are required for processing both of them, and hence to the observed transfer effects (Kraus & Chandrasekaran, 2010).

The shared sound category learning mechanism hypothesis (SSCLMH)

According to Patel (2008), music and language make use of domain- specific categories which exploit different attributes of sound. However, it is hypothesized that the mechanism for sound category learning is common across the two domains. The influence of music training on language can therefore be attributed to the sharpening of an underlying domain-general sound learning mechanism. Patel proposes that statistical learning could be such a mechanism, serving both domains and being indifferent to the nature of the final product, that is, to the characteristics of the acoustic signal being exploited. Such a domain-general learning mechanism for language and music has also been put forward by McMullen and Saffran (2004). While reviewing data on the ontogeny of language and music in human infants, they conclude that both domains rely on the same learning mechanisms, namely extraction of an abstract set of rules through statistical learning, in order to form "native" sound categories (McMullen & Saffran, 2004).

Beyond shared mechanisms

Besson, Chobert, and Marie (2011a; 2011b) agree that there is a common mechanism processing the same acoustic parameters in speech and music. If long-term experience with music only sharpened shared acoustic processing abilities in language, then this would indicate that a domain- general processing mechanism account would suffice. However, in order for a theoretical account to be complete, transfer effects should be taken into consideration. If long-term experience in one domain not only sharpens common characteristics but also domain-specific characteristics, this would indicate that experience can transfer from one domain to the other. Evidence in favor of this account should demonstrate that experience in music should facilitate not only domain-general but also domain-specific processing in language. The fact that musicians are better in segmental processing of a nonnative language (Marie et al., 2011) is an example of transfer as defined in this framework.

Lastly, Patel's OPERA hypothesis builds up on Kraus and Chandrasekaran's (2010) account, in order to specifically explain how music training facilitates subcortical speech processing (Patel, 2011). Although this hypothesis is mainly concerned with the effect of music on brainstem plasticity, it can serve as a framework for other levels of plasticity pertaining to music and speech. "OPERA" is an acronym composed from the initial letters of five conditions necessary for transfer to occur. These, according to Patel, are the following: 1) Overlap, the fact that training has to tap into a common neural circuit for music and speech, 2) Precision, the demands for processing precision should be high in order to trigger top-down tuning, 3) Emotion, refers to the importance of the emotional rewards that music offers, 4) Repetition, the simple learning principle which is a *sine qua non* for plasticity to occur, and 5) Attention, refers to the importance of engaging focused attention while training. According to the OPERA hypothesis, whenever those prerequisites are fulfilled, music training induces plastic changes that can in turn impact speech processing (Patel, 2011).

Can language experience have an effect on music?
Predictions deriving from the theoretical frameworks

None of the above frameworks assumes that the influence of music on language should be unidirectional. On the contrary, bidirectional influences are inherent in shared auditory skills accounts, since they attribute the effects of music on speech to the sharpening of skills mediating both domains. If this mechanism (a common auditory processing or learning mechanism) is shared between music and language, language experience should influence music perception. However, each account makes different predictions with respect to how these influences can occur.

According to the shared auditory skills accounts, language experience can and does induce plastic changes to auditory processing and through that to music processing (Kraus & Banai, 2007; Krishnan, Gandour, & Bidelman, 2012). Nonetheless, it could be argued that these changes would mostly result from bottom-up statistical learning instead of the top-down nature of learning in music, and might also be more dependent on sensitive periods. The same holds for the SSCLMH (Patel, 2008). Patel (2008) states that there is, as yet, no evidence against the possibility that the mechanism for sound category learning is common across the two domains. Any experience or training that would increase the efficiency of the sound category learning mechanism should be beneficial for both music and language. It should be noted that contrary to the shared auditory skills accounts, the SSCLMH predicts that individuals with either music or linguistic experience should be better in learning new sound categories. It is therefore not automatically assumed that a domain-general sound processing device improves and manifests itself in music and language but rather that the learning device is more resourceful, and this can only be manifested when new learning is required.

Things get more complicated with frameworks that go beyond shared resources and attempt to include transfer effects in their interpretation of music-language interactions. Although bidirectional influences are not ruled out, and although in theory transfer effects from language to music should be possible, the thresholds for these effects to be detected become higher. That is, the demands on language experience or training are higher. Let us consider the OPERA hypothesis, for example. As summarized above, there are five conditions that have to be met in order for language to affect the neural encoding of music, at least in a subcortical processing level (Patel, 2011). The Overlap and Repetition conditions are assumed to be met in an individual who speaks a tone language. However, the Precision, Emotion and Attention conditions might not be met, at least not in the same way as they would be met in music training. Although precision is required for using pitch in a tone language, the demands are not comparable to those for music. There is experimental evidence that pitch is neither necessary nor sufficient for speech perception: Mandarin is intelligible even in the absence of pitch variation (Patel, Xu, & Wang, 2010) while plenty of contextual and grammatical cues are available in the signal aiding speech comprehension (Liu et al., 2012; Y Xu, 1994). This difference in precision demand is very important for plasticity- induced fine tuning of the auditory system to take place (Patel, 2012). If the precision demands on auditory encoding placed by music are much higher than those placed by speech perception, one should expect no or very weak effects of language experience on music processing (Patel, 2012). With respect to the rest of the OPERA conditions, it is difficult to define how emotionally rewarding speaking a tone language can be. Although language is a vehicle for communication of emotions, that alone does not automatically mean that the emotion criterion is satisfied. Lastly, the demand for focused attention is one that cannot be met when language experience is defined as tone language experience. Although focused attention is imperative for music training, if not with respect to sounds, then certainly with respect to motor coordination, language acquisition is something that happens effortlessly and naturally (Kuhl, 2004). Under these assumptions, one would have to define language experience differently, in order to observe transfer phenomena. Some alternatives would be to look at trained phoneticians, multilingual individuals, or simultaneous interpreters (see Elmer, Hänggi, Meyer, & Jäncke, 2011) where precision, focused attention and executive control are important in a manner more comparable to music.

Despite the fact that defining language experience and finding its effects might be more complicated in comparison to music, there are no theoretical reasons to exclude this possibility. In the following section we provide an overview of studies that have examined effects of language experience on music and sound processing. Evidence is presented according to different processing levels, ranging from behavioral to brain structure studies.

Evidence of bidirectional influence

Behavioral evidence

There is clear behavioral evidence of bidirectional influences between speech and music. In an earlier section, we discussed musicians' superior processing of segmental and subsegmental (VOT) speech cues. What was not mentioned, however, is the fact that perception of acoustic features is not enhanced equally but instead interacts with linguistic experience. In a cross-linguistic experiment in Japanese and Dutch speakers, Sadakata and Sekiyama (2011) showed that although discrimination and identification of nonnative temporal and spectral speech contrasts (Japanese consonants and Dutch vowels respectively) was better in musicians, there were stimuli for which musicianship had no advantageous effect. This, according to the authors, is a constraint posed by linguistic experience, namely the effect of a change in the weighting of perceptual cues as individuals develop their native language categories (Sadakata & Sekiyama, 2011). Linguistic influences are thus already present in the studies on musicians. Linguistic experience interacts with music experience, shaping and restricting the perception of the acoustic signal.

Whether domain-specific experience with language has domain-general consequences has been partially addressed by studying tone language speakers' ability to process pitch in a non-linguistic, musical context. It appears that tone language speakers' fine-grained pitch processing ability can transfer to music. When tested in music perception, speakers of Mandarin outperform English speakers in detecting contour and interval changes in simple melodies (Bradley, 2012) while speakers of Cantonese are better than English speaking nonmusicians in melody discrimination and tonal memory (Bidelman et al., 2013). Tone-language speakers perform better than non-tone- language speakers in musical interval production and perception tasks (Pfordresher & Brown, 2009) as well as in pitch discrimination tasks (Bidelman et al., 2013; Giuliano, Pfordresher, Stanley, Narayana, & Wicha, 2011b). This superiority is more pronounced in small pitch excursions (Giuliano et al., 2011b) but not when these are much smaller than the excursions occurring naturally in the respective tone language (Bidelman et al., 2013).

Experience with a tone language seems to provide a perceptual attunement to pitch contours (Stevens, Keller, & Tyler, 2011). Thai speakers outperformed native English speakers in discriminating contours in speech and filtered speech, in both Thai and English. They were also faster than their control group in detecting contour characteristics in music stimuli (Stevens et al., 2011). Another study, however, found that tone-language speakers were significantly worse than non tone-language speakers in detecting downward pitch differences in simple melodies (Peretz, Nguyen, & Cummings, 2011). Since this disadvantage occurred only when the direction of the interval was descending, the authors claim that it is signaling interference from language experience (falling tones in Mandarin are larger in pitch excursion than rising ones). Those biases were present at the most difficult excursions (near threshold) leading to the conclusion that speech strategies are employed when the non-speech context is highly demanding (Peretz et al., 2011). Response biases for falling and rising pitch contours have been found before in Mandarin speakers and were interpreted as above in the framework of statistical learning (Bent, Bradlow, & Wright, 2006). The evidence might thus seem conflicting, since tone-language experience sometimes enhances pitch perception while at other times it poses limitations or biases. Nonetheless, these findings are consistent with the fact that linguistic experience shapes sound processing either by enhancing or by restricting it depending on the specific sound attribute and the level of processing studied.

It is also of interest to examine the consequences of sound perception deficits. Individuals with tone deafness have difficulties in fine-grained pitch discrimination, particularly detecting pitch changes smaller than one semitone. This deficiency cannot be attributed to lack of musical training, brain lesions (which differentiate Congenital Amusia [CA] from acquired amusia), low IQ or level of education, hearing impairment, or another identifiable neurological or psychiatric disorder (Stewart, 2008). Are these pitch deficits specific to music or are they domain-general?

New findings suggest that the deficit is not as domain-specific as it was originally thought to be, since individuals with tone deafness show impaired linguistic pitch perception. Their ability to discriminate pitch variation in an unfamiliar language, namely Mandarin, is significantly worse than that of controls (Nguyen, Tillmann, Gosselin, & Peretz, 2009). This finding suggests that lexical tone discrimination is mediated by the same (in this case impaired) pitch system as music (Nguyen et al., 2009). Impaired pitch processing has been found at a suprasegmental level as well. Tone deaf individuals fail to differentiate statements from questions when intonation is the only source of information they can rely upon (Liu, Patel, Fourcin, & Stewart, 2010). Furthermore, they appear to have phonological and phonemic awareness deficits, deficits that lie outside the narrow domain of music (Jones, Lucker, Zalewski, & Brewer, 2010).

It was not until recently that the incidence of tone deafness in tone language speakers was examined systematically. One of the main findings is that tone deafness does occur in tone language speakers, despite the fact that in principal they should be more “trained” with processing fine-grained pitch information (Jiang, Hamm, Lim, Kirk, & Yang, 2010; Nan, Sun, & Peretz, 2010). What is striking is that some tone deaf Mandarin speakers also have difficulties discriminating Mandarin tones (Jiang et al., 2010; Nan et al., 2010). These individuals confuse lexical tones in words and also fail to discriminate between statements and questions, thus exhibiting both segmental and suprasegmental pitch processing deficits (Jiang et al., 2010). Although these deficits arise mostly in laboratory conditions (Liu et al., 2012), lexical tone and intonation difficulties in Mandarin speakers suggest that the disorder has domain-general consequences. Tone deafness is thus a domain-general rather than purely musical disorder, a fact that offers support for theoretical frameworks which propose common auditory processing mechanisms for music and language.

Subcortical and cortical evidence

At the subcortical level, results show domain-general pitch processing benefits arising from domain-specific experience with language. In one such experiment, tone language speakers' FFR responses to pitch changes were compared to non-tone language speakers, musicians and nonmusicians (Bidelman et al., 2011a). Results showed that experience with linguistic pitch enhanced FFR encoding of musical pitch patterns. Despite the fact that there was an influence of domain on the features extracted from pitch patterns in the study, there was nonetheless transfer between domains suggesting that brainstem neurons are amenable to plastic changes and that this has domain- general consequences.

Interestingly, neuroplasticity in pitch processing at this subcortical level of sound encoding is not restricted to the domain in which pitch contours are relevant (Krishnan, Gandour, & Bidelman, 2010; Krishnan, Gandour, Smalt, & Bidelman, 2010). Strong effects of context which arise in other studies (see Nan, Friederici, Shu, & Luo, 2009; Tervaniemi et al., 2009) do not seem to influence brainstem responses. This finding led Krishnan et al. (2010) to conclude that language and music are "epiphenomenal" with respect to subcortical pitch encoding and that the encoding mechanism has evolved to capture information in the acoustic signal that is of relevance in each domain, in order to facilitate higher-order cortical processing of pitch across domains.

The question that arises however is whether enhanced subcortical encoding of pitch has any consequences for musical pitch perception at a behavioral level. In order to provide an answer, Bidelman, Gandour, and Krishnan (2011b) compared Mandarin speakers, musicians and nonmusicians' FFR responses and perceptual discrimination performance using musical pitch stimuli. They found that tone language experience enhances subcortical pitch processing in a manner similar to musical experience. However, this was not evident at a behavioral level. Although Mandarin speakers performed better than nonmusicians, the FFR response accuracy was a successful predictor of behavioral performance only for the musician group. Thus, while subcortical pitch encoding is sharpened in tone language speakers, this is a necessary but not sufficient condition for perceptual advantages to occur in behavior (Bidelman et al., 2011b).

Evidence concerning cortical processing suggests that language experience can have the same advantageous effects as music in processing pitch in domain-specific or domain-general contexts. Chandrasekaran, Krishnan, and Gandour (2007) tested Mandarin and English speakers using an oddball paradigm with Mandarin tones, and found that the MisMatch Negativity (MMN) elicited by the Mandarin speakers was significantly larger in amplitude. This result suggests that long-term experience with linguistic pitch patterns will enhance processing of similar pitch patterns at a cortical preattentive level. This holds even when non-speech homologues are used, as long as they preserve the language relative pitch pattern (Chandrasekaran, Krishnan, & Gandour, 2007a). What is also of great interest is the fact that experience with linguistically relevant acoustic information such as phoneme duration, which is important in some languages, can generalize to perception of sound duration in a non-linguistic context (Marie et al., 2012; Tervaniemi et al., 2006).

In an investigation of the electrophysiological responses to pure tones presented in a discrimination task and a pitch interval discrimination task, it was shown that tone language experience influenced the timing of the neuronal response to pitch differences (earlier in tone language speakers), and the distribution of processing (more focal in tone-language speakers and more widely distributed in non-tone-language speakers) (Giuliano et al., 2011b). Finally, a study, using a refined design, directly compared the effect of tone language and music experience in the preattentive processing of pitch contours resembling those of tone languages (Chandrasekaran, Krishnan, & Gandour, 2009). Mandarin native speakers were compared to English speaking musicians and English speaking nonmusicians using Iterated Rippled Noise (IRN) stimuli (iterations of adding a delayed copy of white noise sample to itself which produces a pitch sensation) to create dynamic pitch trajectories that were analogues of lexical tones but lacked the formant structure of real speech (Chandrasekaran et al., 2009). The stimuli included between- and within-tone category conditions to control for categorical perception vs. auditory perception effects. Mandarin speakers had significantly larger MMN responses than musicians and nonmusicians in both conditions, while musicians had significantly larger MMN responses than nonmusicians. No categorical perception effects were evident at the preattentive level in Mandarin speakers. These results demonstrate that there is experience dependent auditory cortical plasticity that generalizes from specific experiences to domain-general abilities, but also that this plasticity remains more sensitive to the specific context in which it was acquired.

The neural correlates of tone deafness can also help to elucidate the cortical processing of speech and music. Tone deaf individuals' electrophysiological responses to inappropriate intonation during speech intonation differ significantly from those of normal individuals (Jiang et al., 2012). Whereas appropriate vs. inappropriate intonation elicits N100 and P600 ERP effects in control participants, such effects are absent in tone deaf participants (Jiang et al., 2012). The absence of a P600 effect in detecting incongruence between linguistic syntax and intonation is reminiscent of the absence of the same effect when incongruence between a note and its tonal context (musical key) fails to be detected in the same group (Peretz, Brattico, Järvenpää, & Tervaniemi, 2009). These electrophysiological findings are in accordance with behavioral data (see section 5.1) and strongly suggest that there is an overlap in neuronal resources used for speech and music.

Although an fMRI study on speech processing and tone deafness has yet to be conducted, evidence from the music domain show abnormal activations to pitch changes in fronto-temporal areas (Hyde, Zatorre, & Peretz, 2011). In order to find which node in this fronto-temporal network is underlying the pitch perception-production deficits observed in tone deafness, transcranial direct current stimulation (tDCS) was used to selectively “block” activation in specific brain areas (Loui, Hohmann, & Schlaug, 2010). Inferior frontal and superior temporal areas were interrupted with tDCS in normal participants during a pitch perception and production task. The results revealed that the left posterior inferior frontal gyrus (IFG) and the right posterior superior temporal gyrus (STG) stimulation affected performance most strongly. When these areas are interrupted, the pitch performance profile of normal individuals resembles that of tone deaf individuals (Loui et al., 2010). Interestingly, these areas seem to be part of a shared network for processing pitch in language and music in Mandarin-speaking musicians. Nan and Friederici (2012) found that in these individuals, who have extensive experience with pitch in both domains, processing pitch incongruities engages the right STG and the left IFG (BA 45). While the right STG is thought to be involved in perceptual pitch processing, the left IFG is responsible for processing pitch at a higher cognitive level irrespective of domain.

To summarize, neural evidence seems to support the view that resources between language and music are shared. Key stages of auditory processing, ranging from subcortical pitch encoding in the inferior colliculus to higher order pitch pattern representation in the STG, are modulated by linguistic experience in a way comparable to music experience. This is in agreement with common processing mechanism accounts. Moreover, the fact that the strongest evidence comes from subcortical sources indicates that bidirectional effects are more prominent in early auditory stages where the auditory signal is processed independent of its linguistic or musical function.

On-line speech and music processing interactions

As we have seen in the previous sections, speech and music processing are inter-dependent, at least over time (musical experience shapes later linguistic processing, and language experience shapes later musical processing). These inter-dependencies are open to two interpretations, however. One possibility is that speech and music compete for the same resources but remain independent processes. The other possibility is that they rely on the same resources but are actually processed concurrently, in an integrated, holistic way. In order to investigate these two alternatives, one has to look at instances where music and speech are processed simultaneously, as in sung speech.

To investigate simultaneous processing of speech and music, Kolinsky, Lidji, Peretz, Besson, and Morais (2009) conducted a speeded classification experiment where participants heard two nonwords, differing in their last vowel, sung on an ascending or descending interval. Participants were asked to classify the stimuli according to a specified dimension: melodic (ascending or descending interval), or phonological (according to vowel identity). They were much faster in their classifications when the two dimensions varied in a redundant way (when pitch interval and phoneme identity varied consistently together), and much slower when the variation was orthogonal (when both dimensions varied inconsistently), compared to baseline (when only the task relevant dimension varied). This is evidence that the two dimensions interact; participants could not filter out irrelevant variations in one dimension when processing the other, while, importantly, they gained in performance when this variation was redundant, indicating that the two are processed integrally (Kolinsky et al., 2009). Note, however, that although integrality was observed for vowels and pitch intervals, it was not found when the vowels were replaced by consonants.

Recent MEG and EEG data support the shared pitch-vowel processing evidence, by showing that the source of increased neuronal response to vowels compared to non-vowels coincides with the source of increased activation to pitch compared to non-pitch stimuli (Gutschalk & Uppenkamp, 2011). This common source was identified as the antero-lateral HG in the Superior Temporal Plane. The same region showed a selective adaptation effect to vowel identity, placing at least part of vowel perception as early as in the Primary Auditory Cortex (Gutschalk & Uppenkamp, 2011).

This language-music interference effect was also found in a task with real words sung on simple melodies. It took participants significantly longer to judge whether two words or two melodies are the same, when the irrelevant dimension would vary within pairs (Gordon, Schön, Magne, Astésano, & Besson, 2010). As in the Kolinsky et al. (2009) study, asymmetric interference was found, with more interference from word processing on melodic judgments than the other way around (Gordon et al., 2010).

Following up on these results, Lidji, Jolicoeur, Moreau, Kolinsky, and Peretz, (2009) examined whether the vowel-interval interaction occurs preattentively. If pitch and vowels are processed independently, then a MMN ERP response to a simultaneous deviation in both attributes should have amplitude equal to the sum of the MMN ERPs elicited to each one respectively. What they found was that the MMN amplitude to the simultaneous (double deviant) manipulation of vowel and pitch was not additive, providing evidence for the interaction and not the independence account (Lidji et al., 2009). The same interaction was found for consonant- pitch double deviants' elicited MMNs, suggesting that, at a preattentive level, consonants are also processed by the same resources as pitch (Gao et al., 2012). Furthermore, Gordon et al. (2010) report that the amplitude of the electrophysiological responses to double deviant pairs of sung real words are not additive, as the independence account would have predicted. Moreover, the different melody condition elicited a negativity component (300-500ms), very similar to the N400 in the different word condition. It was suggested that this might denote violations of “semantic” expectations induced by change in music comparable to semantic violations in language.

The interaction account is also supported by fMRI data. When participants are asked to pay attention to music (simple melodies) and language (real words) simultaneously in sung stimuli, the interaction employs a bilateral network including the middle and superior temporal gyri, the insula, the anterior and posterior cingulates, and the inferior frontal gyri (Schön et al., 2010). Interestingly, there is a quantitative rather than a qualitative difference between the cerebral networks involved in speech and song processing (Schön et al., 2010; Tierney, Dick, Deutsch, & Sereno, 2013). In an fMRI adaptation study, the left mid-STS showed greater adaptation when lyrics and music were repeated compared to conditions where at least one of them differed (Sammler et al., 2010). Activation to song seems to be following a continuous processing course, with more integrated sound processing occurring in the mid section, and more domain-specific processing of lyrics in the anterior section of the STS (Sammler et al., 2010).

Song has been described by Peretz (2009) as a “natural alliance” between language and music. It has been also suggested that singing might have played an intermediate role in the evolution of language in humans (Masataka, 2007). We have just reviewed results from studies looking at this music-language alliance in order to shed more light on the underlying processes involved when speech and music sounds are processed simultaneously. The evidence is in favor of interaction, at least up until the level of phonetic perception of speech. Indeed, experiments focusing on the interaction at the level of melodic and semantic processing failed to find evidence for interactions (Besson, Faita, Peretz, Bonnel, & Requin, 1998; Bonnel, Faita, Peretz, & Besson, 2001). Processing of sung speech results in behavioral and neural effects that are not equal to the sum of the effects of lyrics and melody separately.

Although this section is devoted to interactions observed during on-line processing of music and language, it is worth mentioning that there is also evidence in favor of interaction from offline, long-term experience effects. We have already mentioned results showing an interaction between music training and native language representations in non-native speech perception (Sadakata & Sekiyama, 2011). Another study has examined the interactive effects of musical and linguistic experience by looking at how these different experiences affect learning an unfamiliar tone language. Cooper and Wang (2012) tested tone identification and sound-to-meaning learning performance in English-speaking musicians and non-musicians as well as in Thai-speaking musicians and non-musicians. If the effects of musical and linguistic experience were independent and linearly additive, Thai musicians should perform best given that they have both types of experience. On the contrary, the Thai-speaking musicians not only were outperformed by the English-speaking musicians in both tasks but were also outperformed by the Thai non-musicians in the sound-to-meaning learning task. These findings demonstrate that, in isolation, musical and linguistic experience have beneficial effects on tone identification and sound-to-meaning mapping. However, in individuals who have acquired both types of experience, such as Thai musicians, music and language interact: the beneficial effect of music is restrained by interference from the native language on the non-native tones and the beneficial effect of language is in turn restrained by music interference. While English speakers simply relied on low level sound processing, which was enhanced in those who were musicians, Thai speakers could not prevent interference from higher level processing calling on tone categories from their native language. The study confirms that there is dynamic interplay of linguistic and non-linguistic pitch experience in tone perception.

Overlapping functional and structural correlates of speech and music

Another way of gaining insight to shared resources between speech and music is by investigating shared brain areas and how they are shaped by experience in these two domains. If they employ common neural mechanisms, then we should expect an overlap in the structural consequences of this extensive experience. In this section, we will review findings on two cortical areas important for sound processing in both language and music, the IFG and the Auditory Cortex. The reader should bear in mind that the studies cited have not been conducted so as to directly compare language and music and also that they did not use designs that can fully dissociate functional from structural changes in neuronal populations within a brain region. As Price and Friston (2005) have noted: “there is a many-to-many mapping between cognitive functions and anatomical regions”. While we acknowledge that there are many issues with respect to spatial precision and function-to- anatomy mapping in neuroimaging studies, we still would argue that it is worth examining the function-structure relationship resulting from linguistic and musical experience.

The left IFG shaped by language and music

Accumulating neuroimaging evidence suggest that the left IFG serves as a hub for processing structured sequences across language, music, and action (Fadiga, Craighero, & D’Ausilio, 2009). This area is well known to be involved in language, with BA44 and BA6 activated during phonological processing, BA44 and BA45 during syntactic processing, and BA45 and BA47 during semantic processing (Hagoort, 2005). As far as action is concerned, BA44 is part of the mirror neuron network for observation and motor imitation of action (Molnar-Szakacs, Iacoboni, Koski, & Mazziotta, 2005). As mentioned in Section 5.2, the left IFG is found to be part of a shared language- music pitch network in Mandarin speaking musicians, one that is engaged in cognitive pitch representation processing in both domains (Nan & Friederici, 2012).

Sluming, Matthew, and Cezayirli (2002) found that experienced symphony orchestra musicians had increased Grey Matter (GM) density in Broca's area. In a subsequent study, a significant difference between the musicians and controls was observed in the GM of the left Pars Opercularis (POP, BA44) (Abdul-Kareem, Stancak, Parkes, & Sluming, 2011). Significant positive correlations were found between GM in the left POP and years of music training and performance in the musician group (Abdul-Kareem et al., 2011; Gaser & Schlaug, 2003). These findings can be attributed to extensive action-related sound processing in musicians, involving components of the mirror neuron system (Abdul-Kareem et al., 2011). Conversely, individuals with impaired pitch processing have significantly less grey GM concentration in the left Pars Orbitalis in the IFG (area BA 47) (Mandell, Schulze, & Schlaug, 2007) as well as increased cortical thickness in the right homologue of the same area (Hyde et al., 2007). These morphological measures correlated with individuals' performance in musical tasks (Hyde et al., 2007; Mandell et al., 2007).

Golestani, Price, and Scott (2011) studied the brains of another group of individuals who have extensive experience with sound processing: phoneticians. They found, among other things, that GM volume in the left POP was larger in phoneticians and that the number of years of experience in phonetic transcription could predict successfully the left POP's surface area with a similar trend for the volume measure (Golestani et al., 2011). On the other hand, poor phonetic perceivers of a non-native vowel contrast have more white matter density in their right POP (Sebastián-Gallés et al., 2012), which could be part of a compensatory mechanism (Wong, Perrachione, et al., 2007).

In sum, the left IFG has greater volume in individuals whose profession requires sdetailed monitoring, production, and manipulation of music or language sounds, while in individuals with poor sound skills a decrease or an increase in its right homologue is observed. Importantly, volume and surface measures in the IFG correlate with the amount of experience with sound processing as well as the degree to which this is poor or impaired.

The role of the Auditory Cortex in language and music

Naturally, when discussing sound processing in either language or music, the main area of interest is the auditory cortex including the Primary Auditory Cortex (PAC) and belt areas in the supratemporal plane. The PAC lies roughly at Heschl's gyrus (HG) and its adjacent sulci although there is big inter- and intra-individual variability (Da Costa et al., 2011). The auditory cortex, specifically the left lateral HG and PT, is engaged in the acoustic analysis of linguistic sounds (Obleser, Zimmermann, Van Meter, & Rauschecker, 2007) as well in the production of melodies and sentences (Brown, Martinez, & Parsons, 2006) while the same regions bilaterally are important for pitch processing (Barker, Plack, & Hall, 2012). One would therefore expect that experience with linguistic or music sounds would have an effect on the morphology of these auditory regions.

Consistent with this assumption, several studies report greater GM density in Heschl's Gyri of musicians (Bermudez, Lerch, Evans, & Zatorre, 2009; Gaser & Schlaug, 2003; Schneider et al., 2002) found that GM volume in the anteromedial HG bilaterally was larger in both professional and amateur musicians compared to non-musicians, with the total volume of the right HG being larger in professional musicians only. The anatomical differences in the amHG were positively correlated with participants' neurophysiological responses to pure tones as well as musical aptitude measures (Schneider et al., 2002).

By performing a whole-brain volumetric analysis in male keyboard players, Gaser and Schlaug (2003) found that GM volume in the left HG differed according to musician status (naïve, amateur, professional), while both gyri showed significant differences in a more liberal threshold in agreement with Schneider et al. (2002). In a less homogeneous group of musicians, Bermudez et al. (2009) found differences in

GM in the right posterolateral HG. GM density in the right PAC also correlates with relative pitch judgment performance in a music transformation task in individuals with variable musical training (Foster & Zatorre, 2010). Increased volume in the right Heschl's gyrus after receiving instrumental training has further been reported in children using a longitudinal design with random assignment of children to training conditions (Hyde et al., 2009). This increase correlated with behavioral measures of melodic and rhythmic abilities (Hyde et al., 2009).

Bermudez et al. (2009) also performed a cortical thickness analysis that revealed greater cortical thickness in the PT (BA 42, posterior to PAC) bilaterally in musicians. A previous study measuring GM volume had found that the right PT and Planum Polare (PP) (BA 52, anterior to PAC) had significantly greater GM density in musicians (Bermudez & Zatorre, 2005). Interestingly, tone deaf individuals have less GM in the left STS (adjacent to PT) although there is no correlation between this morphological measure and pitch performance (Mandell et al., 2007). However, cortical thickness in the right STG (close to BA 22) does correlate negatively with music pitch performance with tone deaf individuals having significantly greater thickness in that region(Hyde et al., 2007).

In the search for neuroanatomical markers of experience with a tone language, Crinion and colleagues (2009) compared Chinese speakers (both native and second-learners of Chinese to control for ethnicity) to multilingual non-Chinese speakers. Regions in the auditory cortex, specifically the right PP in the anterior superior temporal lobe showed significantly more GM in Chinese speakers (Crinion et al., 2009). Greater white matter (WM) density was found in the right HG and just posterior to the left HG in phoneticians (Golestani et al., 2011). Heschl's gyri were reportedly larger in phoneticians, while gyrification was greater in the left but not the right hemisphere compared to controls. Neither volume nor gyrification correlated with phonetic transcription experience, leading to the conclusion that the morphology of this structure is innately defined (Golestani et al., 2011). However, a recent study contradicts this conclusion. By looking at early Spanish-Catalan bilinguals who learn to master two different phonological systems from birth, Ressel et al. (2012) found that bilinguals had greater GM and WM density in both Heschl's gyri. Since, contrary to phoneticians, bilinguals cannot be self-selected, it is assumed that there is a causal link between language experience and HG differences (Ressel et al., 2012).

Heschl's gyrus structure also correlates with learning new linguistic sounds. Performance in a "Mandarin-like" word learning task correlated positively with grey and white matter density in the left HG (Wong et al., 2008). Successful learners had larger left HG volume and learning speed correlated with GM in the left HG as well (i.e. the faster the learning, the greater GM) (Wong et al., 2008). Apart from linguistic pitch, when learning a nonnative phonetic contrast, fast learners have increased volume and white matter density in the left HG (Golestani, Molko, Dehaene, LeBihan, & Pallier, 2007).

To conclude, despite the differences between the samples recruited, the measures used and the analysis methods between these studies, their results suggest that morphological differences in auditory areas constitute structural correlates of language and music aptitude and experience or lack thereof.

Summary

Music and language expertise appear to correlate with differences in brain anatomy, especially in regions that play an important role in sound processing. As with most neuroanatomical studies, there are two caveats in interpreting the results. The first one is related to causal links between brain structure and experience. Given the fact that there is great inter-individual variability in the regions discussed, and that it is very difficult to control for those prior to training initiation in expert individuals, self-selection cannot be ruled out. That is, individuals with greater HG surface might have a propensity to be better sound learners and become musicians or phoneticians. Although there is evidence against self-selection (see Hyde et al., 2009 and Ressel et al., 2012) it remains an open question whether the structural differences observed in IFG and PAC are the cause or the effect of musical and/or linguistic experience. An experimental way to surpass this obstacle is by conducting longitudinal studies where participants are randomly assigned to music training.

The second caveat lies in the sort of arguments presented by Price and Friston (2005). Grey or white matter density, volume, and cortical thickness constitute quite crude measures of brain plasticity. They cannot dissociate quantitative (same neuronal populations but different degree/number that light up) from qualitative (dissociable neuronal populations) differences as the mechanisms underlying plasticity changes. We therefore ought to be cautious when claiming that the same regions are being shaped by music and speech. Even if the exact same anatomical regions show changes with both types of training without knowing the underlying mechanism we might be looking at independent phenomena (different neuronal populations that are shaped by music and speech but lie within the same anatomical region). Neuroanatomical evidence need to be combined with more sensitive measures looking at functional activation differences, for example using multivariate pattern recognition methods in fMRI data (see Staeren, Renvall, De Martino, Goebel, & Formisano, 2009).

Challenges in looking at the equation from the language perspective

Having presented evidence in favor of bidirectional influences between language and music, let us consider the main challenges or limitations when looking at the language-music equation from the perspective of effects of linguistic experience.

First, what constitutes "language experience"? This is one of the major methodological challenges in this research area. What kind of experience with linguistic sounds can qualify as being comparable to music training? A plethora of studies have focused on tone language speakers, mostly due to the fact that tone languages primarily make use of pitch in order to convey lexical information. Since pitch is a sound property that is shared between language and music, tone language speakers have been regarded as comparable to musicians. Speakers of quantity languages, in which vowel duration information plays an important role, have been studied as well with respect to their sensitivity to sound duration in non-linguistic contexts (Marie et al., 2012; Tervaniemi et al., 2006). Early bilinguals have also been considered to have special linguistic experience based on the fact that they have learned to manipulate different phonetic inventories from an early age on (Krizman et al., 2012; Ressel et al., 2012). Other candidate populations include professional phoneticians, simultaneous interpreters, and multilingual individuals, with different advantages and disadvantages for each group.

This methodological difficulty is in fact two-fold, as the lack of a strict definition for language experience leads to great heterogeneity in the populations recruited. Contrary to musicians, where heterogeneity, though of course also present in the wider population, can at least be controlled within an experiment (for example one can recruit pianists from a specific conservatory, following the same curriculum and training, having achieved the same level of performance etc.), all the aforementioned linguistic groups differ fundamentally in their expertise, making experimental control very difficult. Acquisition of expertise is in some cases achieved implicitly, by exposure to speech input (in the case of tone or durational language speakers, and in bilinguals), while in other cases it is achieved explicitly, by formal training (in the case of phoneticians and simultaneous interpreters). As a result, the level of linguistic expertise cannot be defined as systematically as in musicians. Lastly, in each group a set of distinct sound properties are "trained" more than others and this increases the difficulty of making appropriate comparisons or predictions.

Experimental designs are affected substantially by this heterogeneity. This means that finding tasks and measures that are “fair” or sensitive enough to capture any advantages of language experience on sound processing is not an easy endeavor. For example, there are studies reporting enhanced sound processing in tone language speakers in electrophysiological measures in the absence of behavioral advantages (Bidelman et al., 2011b; Guiliano et al., 2011). Finding measures that are sufficiently sensitive depends not only on the heterogeneity of the groups under investigation but also on the fact that the effects that are being investigated are likely to be quite small.

The selection of stimuli is also crucial, especially when comparing “language experts” with musicians. Let us take, for instance, studies that focus on tone language speakers’ pitch perception abilities. It has been consistently shown that the context in which pitch stimuli are embedded influences their processing (Bidelman et al., 2011b; Nan et al., 2009). Pitch information can serve multiple functions in language (lexical, syntactic, prosodic and/or pragmatic information) compared to music, and the context can bias its perception and neural processing accordingly. Finding “context- free” pitch stimuli is difficult but imperative in order to achieve an objective assessment of the effect of language experience on pitch processing. Such attempts have been made with respect to pitch (see e.g. the IRN in Chandrasekaran et al., 2007b) but not to other sound properties. Of course, language is more than tones, as music is more than pitch intervals. Both domains are multi-faceted and thus hard to parse or fit into neat categories without sacrificing their richness and ecological validity.

Another major difficulty when looking at linguistic experience and how it might affect sound perception is the extent to which this experience taps into or “trains” top-down processing mechanisms. According to the Reverse Hierarchy Theory (Ahissar, Nahum, Nelken, & Hochstein, 2009) perception is by default guided by higher-order mechanisms, leading to divergence or convergence of low-level information into higher-order categories. Perceptual attunement depends on the engagement of higher- order cortical structures that search backwards for the most informative low- level population with respect to the task in hand (Ahissar et al., 2009). Perceptual learning is therefore taking place when the signal to noise ratio from lower level input increases as a function of attention and training. Music experience triggers top-down mechanisms, since attention and purposeful repetition are essential elements of music training (Patel, 2011). In contrast, when acquiring one's native language, little explicit focus is placed on phonology and other sound properties of the speech signal (contrary to what's happening when learning a second language). This explicit training to pay attention to sounds offers a great advantage to musicians over tone language speakers, for instance.

Perceptual attunement is not the only benefit music training offers. Other higher-order cognitive functions such as auditory working memory, IQ, and executive functions are also enhanced in musicians and contribute to their behavioral performance superiority (Degé & Schwarzer, 2011; Moreno et al., 2012; Schellenberg, 2004, 2006; Strait, Kraus, Parbery-Clark, & Ashley, 2010). Although it is an empirical question whether this is also true for tone language speakers, there are few theoretical reasons to assume that this is the case (though see Bidelman et al., 2013).

A solution to the problems associated with explicit training would be to focus on individuals with linguistic experience that has been acquired involving top-down mechanisms. Early bilinguals or multilinguals could be an example of such individuals. It is top-down processing in bilinguals (Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006) that makes a difference in their sound processing abilities compared to monolingual tone language speakers. Recent findings have shown that bilinguals are less susceptible to the distorting effects of background noise when listening to speech (Krizman et al., 2012), something that has been consistently shown in musicians (Parbery-Clark et al., 2009) and children receiving music training (Strait, Parbery-Clark, Hittner, & Kraus, 2012). Similarly to musicians (Strait et al., 2010), these beneficial effects of bilingualism could be mediated by enhanced top-down mechanisms such as auditory cognitive abilities and executive functions (Krizman et al., 2012). When one has to reflect on language sounds and to learn to dissociate, manipulate, and inhibit different sound systems from a sensitive period on, more top-down processing involvement would be expected. The same would also hold for phoneticians or interpreters, who have extensive linguistic experience more comparable to purposeful music training.

Despite all the aforementioned challenges, we believe that this line of research should continue. One cannot have a complete account of the effect of music on language unless the inverse effect is also systematically studied to inform existing theoretical frameworks.

Conclusion: Speech and music in interaction

We have reviewed the literature on music and speech, by taking a less common stance and focusing primarily on the effect of language experience on music, or, more correctly, on sound processing. We have presented behavioral, electrophysiological, and neuroimaging data revealing the effects of language experience on music and sound processing, and evidence of on- line interactions across domains, and we have presented findings on associations between experience in the two domains and differences in brain structure. Consistent with a shared auditory skills account, language experience shapes sound perception, by augmenting it or in some cases restricting it. Building up on the shared auditory skills framework, we reviewed the literature on tone deafness and saw that this impairment affects both musical and linguistic pitch processing. Data on song processing added to the picture of what is actually shared when linguistic and music sounds are processed simultaneously, while neuroanatomical data was presented on the infrastructure involved in both domains. Furthermore, we have seen that experience with pitch in a linguistic context can enhance music pitch processing. In other words, there can be positive transfer from the speech domain to music, as defined by Besson et al. (2011a, 2011b). Additionally, evidence for enhanced subcortical pitch encoding in tone language speakers suggests that language experience can, under certain circumstances, meet the OPERA hypothesis requirements (Patel, 2011).

Evidence of language on music effects is sparser than of the reverse. There is need for more research to broaden our understanding of bidirectional language-music effects. For example, the "Shared sound category learning mechanism hypothesis" (Patel, 2008) has not yet been addressed from the language perspective, to the best of our knowledge. Future research aiming to test this hypothesis will need to look into whether learning music categories might be modulated by linguistic experience or expertise. The existing frameworks should also try to accommodate observed phenomena. For instance, in some cases, we have seen that although neuronal sound mechanisms show a clear language experience advantage in performance, no such advantage exists in behavior (Bidelman et al., 2011b). The same pattern has been observed in musician studies (Ott et al., 2011). The theoretical accounts do not yet make predictions about these differences.

There are many other missing pieces in this puzzle. What we wanted to demonstrate, however, is that some of the pieces can only be revealed by looking at the effect of language experience on sound processing. We hope that this review will motivate future research that considers the effects of both linguistic and musical experience, as well as their mutual interactions.

The existing data, however, already offer strong support for a shared auditory skills account of speech, music, and sound processing (Besson et al., 2011a, 2011b; Kraus & Chandrasekaran, 2010; Patel, 2008, 2011). In particular, the evidence points to a synergistic account: music and linguistic experience influence sound processing beyond their narrow domains, and while doing so they mutually interact. As Zatorre and Gandour (2008) have suggested, the synergy probably lies in the interplay between the sensory encoding of sound and the abstract representation of sound, that is, between domain-general, low-level acoustic processes and domain-specific, higher-level cognitive processes. Synergy at this stage of processing would result in the four bidirectional phenomena that have been reviewed: Interactions over time, where prior music experience influences current linguistic behavior and prior language experience influences current musical behavior; interactions across domains in on-line processing; shared underlying brain structures; and sub- cortical and cortical changes shaped by speech and music experience, acting in concert.

Acknowledgements

We would like to thank Mireille Besson, Richard Kunert, and two anonymous reviewers for providing us with constructive comments and suggestions.

CHAPTER 3

EFFECTS OF EARLY BILINGUAL EXPERIENCE WITH A TONE AND A NON-TONE LANGUAGE ON SPEECH – MUSIC INTEGRATION

After: Asaridou, S. S., Hagoort, P., & McQueen, J. M. (submitted). Effects of early bilingual experience with a tone and a non-tone language on speech- music integration.

Abstract

We investigated music and language processing in a group of early bilinguals who spoke a tone language and a non-tone language (Cantonese and Dutch). We assessed online speech-music processing interactions with a speeded classification task in which participants judged sung pseudowords either musically (the direction of the musical interval) or phonologically (the identity of the sung vowel). We also assessed longer-term effects of linguistic experience on musical ability both with a task in which participants had to learn to identify musical intervals and with pitch-perception tasks. Our hypothesis was that due to their experience in two different languages using lexical and intonational tone, the early Cantonese-Dutch bilinguals would differ from Dutch control participants. In online processing, the early bilinguals processed speech and music more holistically than controls. This effect seems to be driven by experience with a tone language, in which integration of segmental and pitch information is fundamental. Regarding longer-term effects of linguistic experience, we found no evidence for a bilingual advantage in either the music-interval learning task or the pitch- perception tasks. Together, these results suggest that being an early bilingual in a tone and non-tone language does not have any measurable longer-term effects on pitch and music processing, but does have consequences for how speech and music are processed jointly.

Speech and music share fundamental building blocks: sounds. Despite the fact that sounds are organized in distinct representational systems for speech and music using different aspects of the acoustic signal (Patel, 2008), the cortical and subcortical mechanisms processing them seem to be overlapping (Bidelman, Gandour, & Krishnan, 2011b; Marie, Kujala, & Besson, 2012; Parbery-Clark, Skoe, & Kraus, 2009; Schön et al., 2010). It therefore comes as no surprise that transfer effects between speech and music arise, such that auditory expertise in music or in speech modulates processing in the other domain (for a review, see Asaridou & McQueen, 2013). There is an asymmetry, however, in the number of studies on music-to-language versus language-to-music effects, with fewer of the latter type. The present study examines language-to-music effects in a novel way. Previous studies of this type have focused largely on tone-language speakers. The assumption is that such speakers have been "trained" in pitch, a sound property fundamental for music, albeit in a linguistic context. The question usually addressed is the following: Can tone-language speakers' ability to process pitch in speech transfer to pitch processing in a non-linguistic or musical context? Here, we ask this question with respect to Cantonese-Dutch bilinguals, who may have particular auditory expertise in pitch processing because of their experience with pitch as a cue to multiple lexical and intonational distinctions.

A number of studies have found evidence that tone-language speakers are better on musical pitch tasks than non-tone-language speakers. Mandarin speakers outperform English speakers in detecting contour and interval changes in simple melodies (Bradley, 2013); Thai speakers are faster than English speakers in detecting contour characteristics in music stimuli (Stevens et al., 2011); and Cantonese speakers are better than nonmusician English speakers in skills such as melody discrimination and tonal memory (Bidelman et al., 2013; Wong et al., 2012). Tone-language speakers are also superior in pitch discrimination tasks (Bidelman et al., 2013; Giuliano et al., 2011b) and their advantage is not only evident in perception but also in the production of music intervals (Pfordresher & Brown, 2009).

But tone-language experience can also induce limitations to pitch perception. Mandarin speakers, for example, show biased responses when asked to identify non-linguistic pitch contours with small pitch excursions, identifying them more often as rising than as falling (Bent et al., 2006). In accordance with that, Mandarin speakers are significantly worse than non- tone-language speakers in detecting downward pitch changes in simple melodies (Peretz et al., 2011). These limitations are caused by interference from language experience, given that rising tones in Mandarin consist of small frequency excursions while falling tones have large excursions and are therefore less finely tuned in Mandarin speakers (Bent et al., 2006; Peretz et al., 2011).

There is cortical and subcortical electrophysiological data showing sharpened pitch encoding in tone-language speakers (Bidelman, Gandour, & Krishnan, 2011a; Chandrasekaran, Krishnan, & Gandour, 2007; Giuliano et al., 2011). However, although precision in neuronal pitch encoding is necessary for perceptual advantages to occur, it is not sufficient. Bidelman, Gandour, and Krishnan (2011a) found that tone-language experience enhances subcortical pitch processing in a manner similar to musical experience. While English speaking musicians also showed better pitch perception alongside brainstem responses, tone-language speakers did not show such behavioral advantage (Bidelman et al., 2011a). Cooper and Wang (2012) also failed to find a tone-language advantage in tone identification whereas they did find an advantage of musical training.

One potential source of this discrepancy between the effects of tone- language and music experience is the explicit knowledge musicians acquire through purposeful training. Explicit knowledge, such as a sophisticated vocabulary for different sound phenomena and concepts, allows musicians to develop a metacognitive understanding of sound. This is of importance for transfer effects to occur, as attentional mechanisms are involved in perceptual attunement to stimulus properties (Ahissar et al., 2009; Patel, 2011). Music experience triggers attunement to auditory features through attentional exercises and purposeful repetition. These are both essential elements in music training (Patel, 2011) but are rather unnecessary in natural first language acquisition.

With this background in mind, we set out to investigate language-to- music transfer effects in individuals in whom metacognitive sound processing has been enhanced not through music training, but through linguistic experience: early bilinguals (cf. Rodriguez-Fornells, De Diego Balaguer, & Münte, 2006). By 'early bilinguals' we refer to individuals who have been exposed to two languages from birth. Early bilinguals can be considered "auditory experts" since they have increased metalinguistic phonological skills (see Campbell & Sais, 1995); that is, they have learned to reflect on language sounds and learn to dissociate, manipulate, and inhibit different phonetic systems from early infancy. Bilinguals are additionally less susceptible to background noise when listening to speech (Krizman et al., 2012), an effect also found in musicians (Parbery-Clark et al., 2009; Strait et al., 2012). These effects are apparently mediated by enhanced attentional mechanisms such as auditory cognitive abilities and executive functions (Krizman et al., 2012; Strait et al., 2010).

Early experience with two languages induces plastic changes in the brain's function (Kim, Relkin, Lee, & Hirsch, 1997) and structure (Mechelli et al., 2004). One of the most interesting findings is that early bilinguals have greater grey and white matter density in the primary auditory cortex compared to late bilinguals (Ressel et al., 2012). This has been attributed to early exposure to different phonetic inventories. Early second language onset is thus an important cause of plastic changes in the human cortex.

What then are the consequences of such plastic changes in sound processing performance? Our hypothesis was that linguistic experience, that is, speaking two languages from an early age on, will influence sound processing outside the domain of language. We chose a bilingual population with experience in a tone language (Cantonese) and a non-tone language (Dutch), assuming that this is a circumstance under which pitch processing is put under the most pressure, due to the diversity in the use of pitch information in this case (see Singh & Foong, 2012). Dutch uses pitch to signal intonational distinctions at the sentence level and to signal lexical stress distinctions (’t Hart, 1998). Cantonese also uses pitch for sentence-level intonation (Bauer & Benedict, 1997), but uses it especially to convey lexical meaning (e.g. to distinguish meanings in words differing only in tone). Moreover, Cantonese has a more complex tone repertoire than the more frequently studied Mandarin language. The six tones in Cantonese include both contour and level tones, and hence require fine-grained F0 processing abilities (Khouw & Ciocca, 2007). Learning to master control over such different phonetic inventories could potentially train Cantonese-Dutch bilinguals in a manner comparable to music training.

Whereas the majority of prior studies have focused on transfer effects, that is, the beneficial effect of past experience in the linguistic domain on current processing in the music domain, in this study we set out to test two other speech-music effects as well. The first is the online interaction that occurs when speech and music are processed simultaneously in sung speech. The goal was to investigate how experience with two different linguistic pitch systems influences this interaction. The second is whether there are long-term consequences of learning Cantonese and Dutch in early bilingualism on the ability to learn new sounds. In particular, we wanted to test the Shared Sound Category Learning Mechanism (SSCLM) hypothesis put forward by Patel (2008). According to this hypothesis, music and speech may rely on the same learning mechanism, one which extracts regularities and creates the respective sound categories. We wanted to test whether the efficacy of the SSCLM is increased in early bilinguals, given the fact that it is put under higher pressure to perform during a sensitive period of language development.

We assessed online interactions using a speeded classification task introduced by Kolinsky, Lidji, Peretz, Besson, & Morais (2009) and based on Garner interference (W. Garner & Felfoldy, 1970). In this task, participants listen to sung words and classify them according to a pre-specified dimension: musical or phonological. In the musical dimension participants judge the music interval as either ascending or descending, and in the phonological dimension the judge the identity of the vowel in the sung word. Kolinsky et al. (2009) found that music and vowels are processed integrally and thus participants cannot filter out irrelevant variation in one dimension when performing a judgment on the other. That is, they are faster when both dimensions vary consistently (e.g. ascending interval paired with vowel x) and slower when they vary inconsistently (e.g. all possible combinations between intervals and vowels). The ability to ignore irrelevant variation in the acoustic signal seems to play an important role in the acquisition of nonnative phonemes where perceptual interference from irrelevant dimensions from one's native language can arise (Iverson, Kuhl, Akahane-yamada, & Diesch, 2003).

We had two alternative hypotheses with regards to this task. On the one hand, we expected that Cantonese-Dutch bilinguals will have mastered the ability, from early in development, to switch between two language systems that make different uses of pitch information. In order to accomplish this, they would have to learn to ignore, for each of their languages, pitch variation that was irrelevant in that language (but relevant in the other language). According to this hypothesis, Cantonese-Dutch bilinguals should exhibit less interference than Dutch control participants when asked to perform such a task, because the task requires simultaneous evaluation of phonemic and pitch variation. On the other hand, it has been found that tone-language speakers show more interference when processing pitch and phonemes simultaneously (Lee & Nusbaum, 1993; Repp & Lin, 1990). Specifically, Lee and Nusbaum (1993) found that, while performing a speeded classification task, Mandarin speakers showed interference from constant pitch information whereas English speakers did not. Due to their linguistic experience, tone-language speakers may have more highly- developed processing strategies, such that they integrate segmental and suprasegmental information more than non tone-language speakers do (Lee & Nusbaum, 1993). An alternative hypothesis would therefore be that, as speakers of a tone language, it may be harder for Cantonese-Dutch bilinguals to ignore pitch information compared to participants who do not speak Cantonese, even when the pitch information is non-linguistic or irrelevant.

In order to test the SSCLM hypothesis (Patel, 2008), we used a training task in which participants had to learn new sound categories in the music domain. Participants were trained to associate ascending music intervals with colors, that is, to extract regularities from the intervals and form basic, abstract color-coded categories (this task was adapted from Hove, Sutherland, & Krumhansl, 2010). It was hypothesized that Cantonese-Dutch bilinguals would perform better than their controls, since their SSCLM is more "trained" in learning sound categories in different languages. We also administered a rhythmic pattern category-learning task, as a control for general learning capacities (see Hove et al., 2010).

Lastly, we wanted to test whether Cantonese-Dutch bilinguals will show language-to-music transfer effects in pitch processing, as previously shown in tone-language speakers (Bidelman et al., 2013; Wong et al., 2012). We therefore included a set of pitch perception tasks testing different levels of pitch representation. The purpose of these tasks was to detect whether bilinguals' experience with a tone language and a non-tone language transfers to non-linguistic pitch processing and, if so, determine the level of pitch representation at which this transfer is demonstrated.

Methods

Participants

Bilingual participants. The bilingual group comprised 21 Cantonese- Dutch bilingual speakers. From those, two participants were excluded, one on the basis of music experience and the other for failing a hearing screening test. The data from nineteen participants were therefore analysed, including 10 males and 9 females, aged between 17 and 35 years (mean=24.16, SD= 5.55).

The majority of participants were tested at Radboud University Nijmegen (five were tested at Leiden University). Participants were recruited through advertisements posted on university webpages and through colleagues. The bilinguals were all raised in Cantonese-speaking environments, as in all cases both their parents were Cantonese speakers. Participants were rewarded with 30 euro gift cards. All participants signed an informed consent form prior to participation and had no self-reported neurological or psychiatric disorders.

Control participants. Forty-three university students were recruited for the control group from Radboud University's research participation system database. Participants were all native speakers of Dutch. It should be noted that the control participants were not monolingual speakers, having learned on average more than two languages in addition to Dutch. Exclusion criteria were 1) failing the hearing screening, 2) being an early bilingual, 3) having more than 3 years of music experience, 4) speaking or learning a tone language or a Dutch tone dialect, and 5) being older than 35 years. A total of eleven participants were excluded for not meeting the criteria: four participants for failing the hearing screening, three for being raised bilingual, one for having music experience, and two for learning and speaking Mandarin and a Limburg tone dialect respectively. The remaining sample consisted of 32 participants, 6 males and 26 females, aged between 18 and 35 years (mean= 22.12, SD=3.37). One participant did not show up for the second part of the experiment but data from her first part were used. Participants were rewarded with course credit or 30 euro gift cards.

Due to a measurement error, data from 30 participants in the Vowel- Interval speeded classification task (see below) were rendered unusable. Therefore, an additional control group of native Dutch speakers was recruited for this task. This final group consisted of 22 participants (after excluding one participant for music experience, one for learning a tone language, one for exceeding the age limit, and two for corrupted data). Participants included one male and 21 females, aged between 18 and 24 years (mean= 19.27, SD=1.98). Participants were rewarded with course credit.

Procedure

Testing took place in a sound-proof booth where participants sat comfortably in front of a computer screen. Stimuli were presented and responses were recorded on a Philips computer running an in-house software program. Auditory stimuli were presented at a comfortable intensity level over a pair of Monacor MD-4300 stereo headphones. Response recordings were performed using an IMG Stage line DM-5000LN Dynamic Microphone. Task order was randomized, with the exception of the Vowel-Interval speeded classification task which for the bilingual sample was always administered at the beginning of the session so as to match the additional control sample's conditions. The entire testing procedure took around 200 minutes, split in two sessions of 100 minutes each. Participants had a 10- minute break whenever they desired during the session. Due to time constraints, not all participants completed all the tasks (see Table 1).

Table 1.

Total Number of Participants that Completed Each Task.

	Bilingual	Control
Task	N	N
Vowel-Insterval Speeded Classification	17	24
Learning to Identify Music Intervals	19	26
Learning to Identify Rhythmic Patterns	19	27
Pitch Change Detection	17	27
Pitch Direction Discrimination	16	28
Simple	16	32
Transposed	18	20

Vowel-Interval Speeded Classification Task

In the vowel-interval speeded classification task, participants heard two disyllabic nonwords (/dalɔ̃/ and /dalø/) differing in their last vowel and sung by a professional French baritone on two intervals, one ascending from the first syllable to the second (F3-F3#), and one descending (F3-A2). They were asked to classify as fast and as accurately as possible the stimuli according to a specified dimension; in the melodic task they were instructed to ignore the words, focus on the melody and classify the intervals as going up or down, and in the phonological task they were instructed to ignore the melody, focus on the words and classify them as "dalo" or "dale".

There were three conditions in which the two dimensions in the stimuli varied differently: 1) the Redundant condition, when the music interval and phoneme identity varied consistently, which should result in performance gain, 2) the Orthogonal condition, when the variation in the irrelevant dimension was inconsistent, which should cause interference, and 3) the Baseline condition, when the irrelevant dimension was kept constant and only the relevant dimension varied. We measured interference by comparing performance in the orthogonal condition to baseline and gain by comparing performance in the redundant condition to baseline. Response buttons were used and participants' accuracy and reaction times were measured. There were three blocks in each task corresponding to three conditions in which the two dimensions in the stimuli varied differently. Kolinsky et al. (2009) kindly gave us their materials. For a detailed description of stimulus generation procedures see Kolinsky et al. (2009) (auditory examples are available online at: http://www.brams.umontreal.ca/plab/research/Stimuli/kolinsky_et_al/index.html).

Participants had a short practice session followed by 72 trials for each block and the whole task took around 30 minutes to complete. Presentation order of tasks and conditions was counterbalanced, and order of trials within conditions was pseudo-randomized. Unlike in Kolinsky et al. (2009), there was no time out beep 2500ms after stimulus presentation. However, we excluded from the analyses reaction times larger than 3000ms.

Learning to Identify Music Intervals

In this task, participants were trained to associate three ascending music intervals, major 2nd, perfect 4th, and perfect 5th, starting from two different reference tones, C4 and F4#, with three arbitrary color labels. The intervals were composed using the Logic Express 9 program (http://www.apple.com/support/logicexpress/) on a MacBook, using piano timbre. The task consisted of a training and a testing phase. Participants were given detailed instructions in which they were introduced to the concept of a pitch interval with emphasis on its relative nature.

In the training phase, participants were presented with a color on a computer screen and were instructed to generate the interval corresponding to the color by trial and error, pressing the space bar for the reference tone and one of three marked keys on the keyboard. Each key corresponded to a specific interval. Feedback was provided, in the form of "correct" or "incorrect" appearing on the screen, so that participants could correctly associate the color label with its matching interval. After the training phase was completed, an identification test followed to assess learning. Participants heard the trained intervals while the three color options appeared on the computer screen. They had to match each interval to its correct color label by clicking on the respective key on a button box corresponding to the color alignment on the screen. The training and testing phases each consisted of 96 trials in random order.

Participants also performed a control task where they had to learn to associate three rhythmic patterns with three color labels. Similarly to Hove et al.'s (2010) study, all the patterns were in duple meter, consisted of a total of seven tones repeated twice, with one having 1:1 ratio using eighths, the other a 2:1 ratio using triplets, and one with 3:1 ratio using 16ths as distinguishing features. Sequences were presented either in a moderate or in a slow tempo. The patterns were presented in marimba timbre. The concept of a rhythmic pattern was introduced and its relative nature was explained and emphasized during instruction. In the training phase participants were presented with a color on a computer screen and were instructed to generate the rhythmic pattern corresponding to the color by trial and error, by pressing one of three marked keys on a keyboard. Each key corresponded to a specific rhythmic pattern. Feedback was provided, in the form of "correct" or "incorrect" appearing on the screen, in order for participants to correctly identify which color label corresponds to which interval. After training, participants performed an identification test which followed the same structure as the one in the interval condition. Also as in the interval condition, the training and testing phases each consisted of 96 trials and trial presentation was randomized. The entire task took around 40 minutes to complete.

Pitch Perception Tasks

Four two-alternative forced choice (2AFC) tasks were administered to assess pitch perception. The first two tapped into the initial levels of pitch representation and the remaining two into higher pitch pattern representations (Foxton, Dean, Gee, Peretz, & Griffiths, 2004). Sine wave tones created with Praat (Boersma & Weenink, 2009) were used for all the pitch perception tasks, following the procedures described by Foxton et al. (2004).

We used response accuracy as the dependent measure. Sine wave sounds were used in all the tasks and included 20ms onset/offset frequency ramps. The inter-stimulus interval was always 1100ms and the inter-trial interval 2000ms long. All tasks were preceded by a short practice session to familiarize participants with the stimuli and procedure.

Pitch change detection. Participants were presented with two pairs of sine wave sounds. One pair always consisted of sounds that had the same frequency

(500Hz) while the other pair consisted of sounds that differed in frequency (ranging from 490Hz to 510Hz). Each pure tone was 250ms long with 100ms gap between a pair. The largest difference in frequency was 20Hz and the smallest 2Hz which translates to about 70 – 17 cents difference in the "different" pair (1 semitone =100 cents). Participants had to detect which pair was the "different" one by pressing a key on a button box. Half of the different pairs were presented first and half were presented second while the trial order was completely randomized for all participants. A total of 80 trials took around seven minutes to complete.

Pitch direction discriminations. Participants were presented with two sine wave sounds, one with an upward and the other with a downward glide. Each glide consisted of the initial and final tone, each 250ms long, connected with a linear ramp of 100ms duration. The frequency interval between the initial and final glide frequencies ranged from 4Hz – 50Hz (20 cents – 170 cents, 1 semitone =100 cents). Participants were asked to indicate which of the two sounds was the one with the upward glide by pressing a button box. The order of upward glide presentation was counterbalanced while the trial order was completely randomized for all participants. A total of 80 trials took around seven minutes to complete.

Steady contour pitch sequence task. Participants were presented with pairs of four sine wave tone sequences that were either exactly the same or differed in one tone. The difference was such that one tone (either the second or third) in the second sequence was different in frequency from the respective tone (the tone in the same position) in the first sequence. Importantly, the different tone did not violate the melodic contour in the sequence. That is, the direction of the intervals in the two sequences was kept the same (ascending or descending) so that the different tone in the second sequence was different from the first only in terms of absolute frequency (Figure 1). Each tone in the sequence was 250ms long. Participants were asked to report whether the sequences were same or different by pressing a key on a button box. Half of the pairs were same and half different and their presentation order was counterbalanced, while the trial order was completely randomized for all participants. A total of 112 trials took around ten minutes to complete.

Steady contour transposed pitch sequence task. This task was identical to the steady contour sequence task with one change: this time the second sequence was always transposed half an octave up or down in frequency. Thus, the second sequence could either have identical frequency intervals as the first or it could contain a different interval that did not violate contour. Participants were asked to ignore the absolute pitch change in the second, transposed, sequence and report whether the sequences were same or different by pressing a key on a button box.

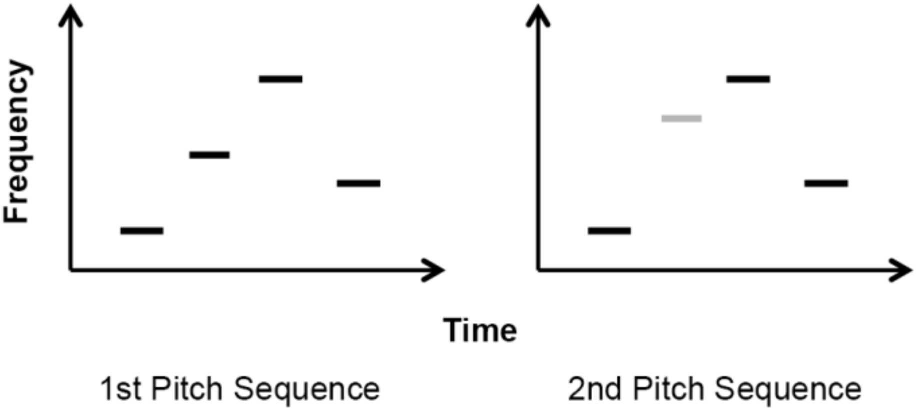


Figure 1. Steady contour pitch sequence task depiction: the 2nd sequence differs from the first in the 2nd tone, which has a different frequency but does not violate the contour.

Control Tasks

Hearing screening was performed with an Oscilla USB-330 audiometer using the random automatic hearing test at 20 dB in 11 frequencies ranging from 125 Hz to 8 KHz in both ears. Since the frequencies used in the experiments were never below 250 Hz or above 4 KHz, participants' performance in very low and very high frequencies was disregarded. Participants failed the hearing screening when they could not identify frequencies ranging from 250 Hz to 4 KHz at an intensity larger than 30 dB in either of the two ears.

The Raven's Standard Progressive Matrices test was used to assess general, non-verbal intelligence (1960 Edition) . Participants' handedness was assessed using a shortened version of the Edinburgh Handedness Inventory (Oldfield, 1971) and their short-term memory using the forward digit span adapted from the Dutch version of the Wechsler Adult Intelligence Scale (WAIS).

Background information was collected about the participants' language and music experience. Participants filled in an online version of the Language history questionnaire for bilingual research [http://www.personal.psu.edu/pul8/questionnaire/L2_questionnaire.html (Li, Sepanski, & Zhao, 2006)]. They also filled in an in-house questionnaire about their music experience and preferences.

Results

Control Measures

The two groups did not differ significantly in age, intelligence (as measured by the Raven's test of progressive matrices), or music experience (see Table 2). However, they differed significantly in their digit span scores (t=6.14, p<.001), with the control group having higher scores than bilinguals, and in number of languages spoken (t=-3.21, p=.002), with the bilingual group having learned significantly more languages than the control group. Despite the discrepancy between the two groups in their digit span scores, it is assumed that this did not affect the present results since Digit span did not correlate with any of the measures of interest, apart from accuracy in the control Rhythm Training Identification task for bilinguals (r=.491, p=.033).

Table 2.

Independent T-tests on Demographic measures between groups.

	Bilingual		Control				
Demographics	M	SD	M	SD	95% CI for Mean Difference	p	df
Age	24.15	5.55	22.12	3.37	-4.53, .472	.109	49
Raven's	55.88	3.80	56.19	3.12	-1.71, 2.32	.763	47
Digit Span	7.00	1.52	9.90	1.69	1.95, 3.85	.000**	49
Music Experience*	3.31	7.93	5.81	8.88	-2.47, 7.47	.318	49
Languages	3.61	.697	2.72	1.03	-.886, .276	.002**	45

Note. *in months, **p<.01

Vowel- Interval Speeded Classification Task

Phoneme identity discriminability analysis. Since the vowel stimuli were French, we first wanted to test how discriminable they were for our participants. Whereas /ø/ is part of both the Dutch and the Cantonese phoneme inventory, the nasalized /ɔ̃/ is not. Therefore, we conducted two mixed-model ANOVAs, one on accuracy scores and one on RTs of the phonological baseline condition, with group as a between-participant factor. The results of the analysis on accuracy did not reveal any effect of phoneme identity [F(1, 39)=2.19, p>.05] or main effect of group [F(1, 39)=.832, p>.05]. However, analysis on RTs revealed a significant main effect of phoneme identity [F(1, 39)=15.42, p<.001, η_p^2 =.283], as participants needed more time to respond to /ɔ̃/ (M=761ms) compared to /ø/ (M=728ms, pairwise comparison p<.001, Bonferroni corrected). No other main effect of group or interaction reached significance. Since the two groups did not respond differently to the two phonemes in either speed or accuracy, we assume that vowel discriminability was matched across groups.

Order effects. Although task and condition order was counterbalanced between participants, we performed one-way ANOVAs on percentage correct and RTs with order as a between-participant factor. Results did not reveal any effect of task and condition order.

Accuracy analysis. Kolinsky et al. (2009) report excluding participants with an error rate larger than two standard deviations above the mean. Although the overall error rate in our participants was fairly low (2.07% for controls and 4.56% for bilinguals), we performed our analyses on accuracy by applying the Kolinsky et al. exclusion criterion, separately for each group. Seven control and four bilingual participants were hence excluded. The accuracy data from the remaining participants is summarized in Figure 2.

Percentage error was analyzed using a 2 (Group: Control, Bilingual) x 2 (Task: Melodic and Phonological) x 3 (Conditions: Baseline, Redundant, Orthogonal) mixed-model ANOVA. Analysis revealed a significant main effect of group [F(1, 28)=21.56, p>.001, η_p^2=.435], with control participants having significantly lower error rates (M=1.36%) than bilingual participants (M=3.52%, pairwise comparison: p<.001, Bonferroni corrected). Contrary to Kolinsky et al.'s (2009) results, we did not find a main effect of task [F(1,28)=2.21, p=.148], suggesting that for our participants the two dimensions x task interaction [F(1, 28)=9.85, p=.004, η_p^2 =.260], driven, as revealed in pairwise post hoc comparisons, by the bilingual participants' making more errors in the Melodic task (p=.005, Bonferroni corrected).

There was also a main effect of condition [F(2, 56)= 21.35, p<.001, η_p^2=.433], since, as expected, participants made overall more errors in the orthogonal condition compared to the other two conditions (p<.001, Bonferroni corrected). The task x condition interaction was also significant [F(2, 56)=4.53, p=.015, η_p^2 =.139]: Participants in the Melodic task made overall more errors in the orthogonal condition compared to baseline (p<.001) and redundant condition (p=.002, Bonferroni corrected). In the Phonological task, participants made significantly fewer errors in the redundant condition compared to the baseline (p=.008, Bonferroni corrected) and orthogonal conditions (p=.008, Bonferroni corrected). Furthermore, the group x condition interaction was significant [F(2, 56)=4.75, p=.012, η_p^2=.145] with control participants making fewer errors in the redundant conditions compared to the orthogonal conditions (p=.013, Bonferroni corrected) and with bilinguals making more errors in the orthogonal conditions compared to the baseline (p<.001, Bonferroni corrected) and the redundant conditions (p<.001, Bonferroni corrected). Lastly, the 3-way interaction between group, task, and condition, was also significant [F(2, 56)=4.53, p=.015, η_p^2 =.139]. Pairwisecomparisons revealed that the interaction was driven by the bilingual speakers having significantly larger error rates than the control speakers in the orthogonal condition of the Melodic task (p<.001, Bonferroni corrected) and in the redundant Melodic condition (p=.008, Bonferroni corrected). Marginal differences between groups are also found in the baseline Melodic condition (p=.065, Bonferroni corrected), with bilinguals making more errors Bonferroni corrected), with controls making more errors than bilinguals.

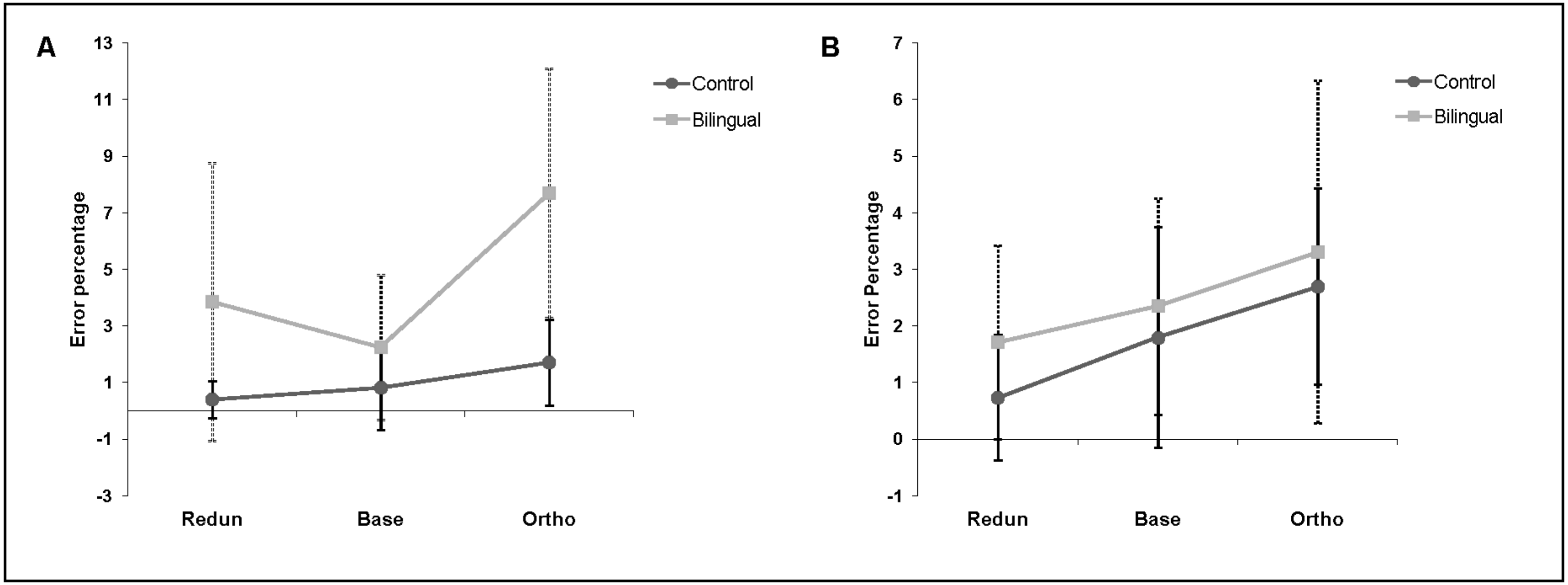


Figure 2. A. Accuracy scores for the Melodic Interval Speeded Classification task and B. Accuracy scores for the Phonological Speeded Classification task in the Redundant (Redun), Baseline (Base) and Orthogonal (Ortho) Conditions. Error bars represent standard error of the mean.

RT analysis. In all analyses, the RTs were estimated by subtracting 750 ms from the RTs measured from the beginning of the stimuli, as the crucial transition between the two notes and vowels was centered at 750 ms after stimulus onset. Only correct responses were analyzed. Any responses given 2500 ms after stimulus presentation were excluded from the analysis.

We performed the RT analysis twice: once including all participants and once excluding participants with more errors than two standard deviations above the mean. Since the results remained the same, we report only those based on the whole dataset, as summarized in Figure 3.

Discriminability of the melodic and phonological dimensions was tested with a paired-sample t-test on the baseline condition RTs for both tasks. The comparison did not reveal any significant difference between baseline RTs for the two tasks for either the control [t(23)=-1.435, p=.165] or the bilingual participants [t(16)=-.859, p=.403].

RTs were analyzed using a 2 (Group) x 2 (Task: Melodic and Phonological) x 3 (Conditions: Baseline, Orthogonal, Redundant) mixed- model ANOVA. The analysis revealed a significant main effect of condition [F(2, 78)=60.47, p<.001, η_p^2 =.608], as participants were faster in the Redundant condition (M=645 ms) and slower in the Orthogonal condition (M=790 ms), compared to Baseline (M= 713 ms, all pairwise comparisons p<.001, Bonferroni corrected) (see Figures 2a and 2b). No main effect of group or task was found (F<1), however, there was a significant task x condition interaction [F(2, 78)=6.49, p=.002, η_p^2 =.143] with slowest reaction times for the Orthogonal Melodic condition (M=806 ms) and fastest for the Redundant Phonological condition (M=638 ms).

We conducted a separate analysis on gain RTs, calculated by subtracting Redundant RTs from Baseline, and interference RTs, calculated by subtracting

Baseline from Orthogonal RTs. The 2 (Group) x 2 (Task) x 2 (Effect: Gain, Interference) mixed ANOVA on RTs revealed a significant Task x Effect interaction [F(39,1)=16.880, p<.001] with gain in performance being larger in the Phonological task (M=106ms) compared to the Melodic task (M=29ms, pairwise comparison p=.005, Bonferroni corrected) and interference larger in the Melodic task (M=124ms) compared to the Phonological task (M=30ms, pairwise comparisons p=.001, Bonferroni corrected). There was also a marginal Group effect [F(39, 1)=2.911, p=.096] which was significant in the "trimmed" RT dataset [F(28,1)=4.255, p=.049] and which was driven by overall larger interference and gain in bilinguals.

In order to make our analyses more comparable to those in Lee & Nusbaum (1993), we also performed separate RT analyses for each group. We conducted two separate 2 (Tasks) x 3 (Conditions) repeated measures ANOVAs on RTs. There were different patterns for the control and bilingual participants. Both groups showed a main effect of condition [Dutch: F(2, 46)= 26.53, p<.001, η_p^2 =.536; bilinguals: F(2, 32)= 33.513, p<.001, η_p^2=677]. However, the task x condition interaction was significant only for bilinguals [F(2, 32)= 5.57, p=.008, η_p^2 =258]. Bilinguals did not benefit from redundancy in the Melodic condition but showed robust interference effects (p=.002, Bonferroni corrected). In contrast, in the Phonological condition the bilinguals did not show interference (p=.183, Bonferroni corrected) but did show highly significant redundancy gains (p<.001, Bonferroni corrected).

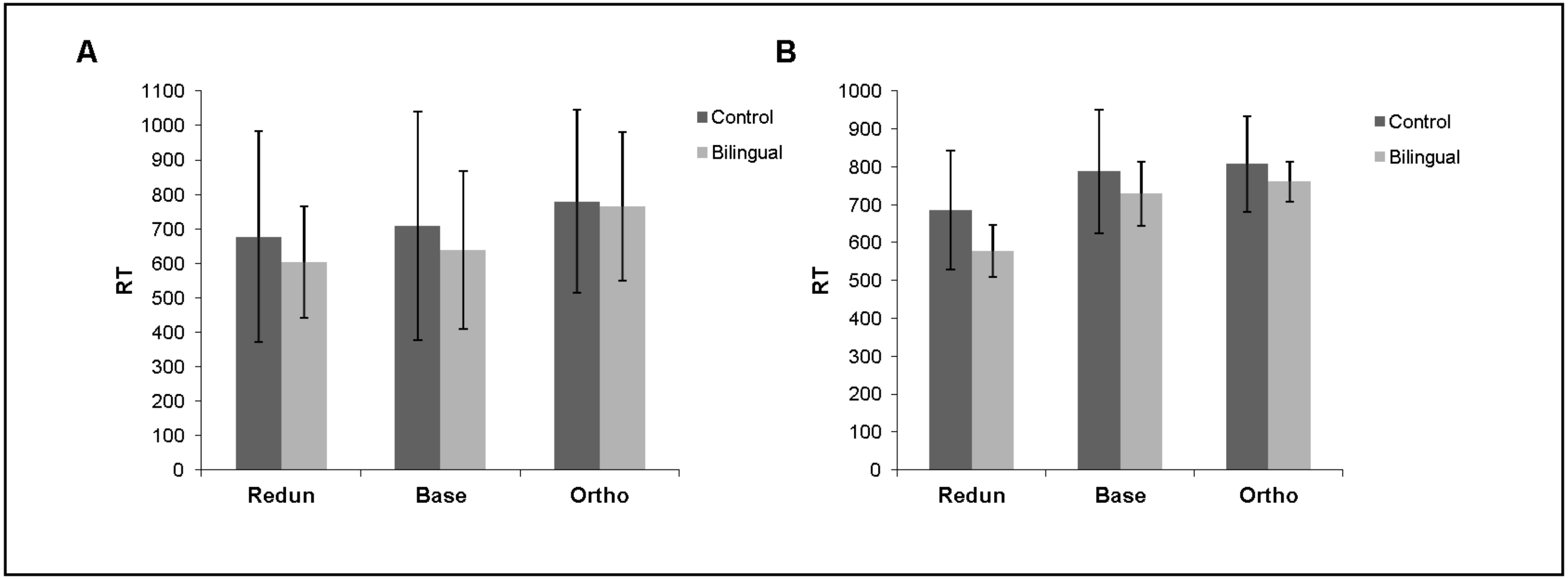


Figure 3. A. RTs for the Melodic Interval Speeded Classification task and B. RTs for the Phonological Speeded Classification task in the Redundant (Redun), Baseline (Base) and Orthogonal (Ortho) Conditions. Error bars represent standard error of the mean.

Learning to Identify Music Intervals

Overall percentage correct for each task was analyzed using a 2 (group: controls, bilinguals) x 2 (task: Pitch, Rhythm) mixed-model ANOVA. Although there was a main effect of task [F(1, 43)=42.06, p<.001, η_p^2 =.494], no task x group interaction was found. Participants overall found the Rhythm task (M=65.77%) much easier than the Pitch interval task (M=52.05%). A significant main effect of group [F(1, 43)= 5.96, p=.019, η_p^2 122] was found, with control (M=63.05%) participants overall outperforming the bilinguals (M=54.77%) (see Figure 4).

Data from the pitch interval identification test were analyzed using a 2 x 2 x 3 mixed-model ANOVA, with group (controls, bilinguals) as a between- participants factor, reference tone (C4 or F4#), and interval type (M2, P4, P5) as within-participants factors, and the percentage of correct responses as the dependent variable. There was no main effect of interval [F(2, 86)= 2.49, p=.089] and no main effect of reference note [F(1, 43)= 1.63, p=.208]. However, the reference note x interval interaction was highly significant [F(2, 86)=59.42, p<.001, η_p^2 =.580]. Similar to Hove et al. (2010), participants recognized M2 better from C4 compared to F4#, that is, produced with a lower pitch, while they recognized P5 better from F4#compared to C4 (i.e. produced with a higher pitch), indicating some sort of absolute pitch strategy. There was no main effect of group [F(1, 43)=2.23, p>.05, η_p^2 =.049]. There was however a group x interval interaction [F(2, 86)=4.52, p=.014, η_p^2 =.095]: The control participants were better at identifying M2 compared to the other two interval types, while the bilinguals were slightly better at identifying P5ths. No other interaction reached significance.

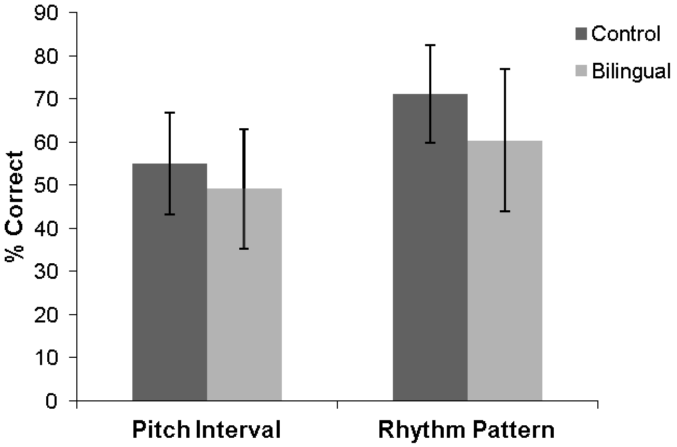


Figure 4. Mean group performance for Pitch Interval and Rhythm Pattern Identification. Error bars represent standard error of the mean.

The same analysis was performed for the rhythm pattern learning task, with a 2 (Group) x 2 (tempo: Slow, Moderate) x 3 (pattern: 1:1, 1:2, 1:3) mixed- model ANOVA on percentage correct responses. Analysis revealed a significant main effect of group [F(1, 43)=6.66, p=.013, η_p^2 =.134] with the control group (M=71.12%) outperforming the bilinguals (M=60.43%, pairwise comparisons p=.013, Bonferroni corrected). We found a main effect of tempo [F(1, 43)=9.89, p=.003, η_p^2 =.187] as participants found the moderate tempo (M=68.09%) easier than the slow tempo (M=63.45%, pairwise comparison p=.003, Bonferroni corrected). There was also a significant main effect of pattern [F(2, 86)=6.96, p=.002, η_p^2 =.139]. Surprisingly, participants found the 1:3 pattern (M=71.20%) the easiest to recognize compared to the simple 1:1 pattern (M=59.74%, pairwise comparison p<.001, Bonferroni corrected). Furthermore, a significant tempo x pattern interaction [F(2, 86)=4.21, p=.018, η_p^2 =.089] indicated that pattern 1:1 was easier to recognize in the slow tempo while patterns

1:2 and 1:3 were easier in moderate tempo.

Since the two groups differed significantly in their Digit Span performance and since this measure correlated with the bilingual's group performance in the Rhythm pattern identification task, we ran another 2 (Group) x 2 (tempo: Slow, Moderate) x 3 (pattern: 1:1, 1:2, 1:3) mixed-model ANOVA with Digit Span scores as a covariate. The effect of the covariate was significant [F(1, 40)=7.96, p=.007, η_p^2=.166)] but with short-term memory factored out, now the main effect of group was not significant [F(1, 40)=.452), p>.05, η_p^2=.011].

Moreover, performance in the Rhythm task is an index of how good participants are in such a learning paradigm, controlling for cognitive factors that could drive performance in the Pitch interval learning task. Therefore, we also repeated the 2 x 2 x 3 mixed-model ANOVA on the Pitch learning data, including overall performance in the Rhythm task as a covariate. The main effect of group did not change after the addition of this covariate, that is, it remained not significant.

Pitch Perception Tasks

Group differences in pitch perception were tested using Independent-samples t-tests with accuracy as the depended measure. Overall, there was no difference between the control and the bilingual groups in any of the pitch perception tasks (see Figures 5 and 6). No group difference was found in pitch-change detection accuracy [controls (M=23.81, SD=11.85) and bilinguals (M=24, SD=11.74), t(42)=-.051, p > .05]. An item analysis showed that the larger the pitch excursion in the different pair, the fewer the errors participants produced [r = -.897, p <.001], indicating that the task was indeed measuring pitch-change perception sensitivity. We performed a similar item analysis on the pitch-direction discrimination data which showed that the larger the pitch excursion in either pair (upward or downward), the fewer the errors committed by participants [r = -.568, p

<.001], as expected. However, no significant group difference was found [controls (M=12.64, SD=11.34), bilinguals (M=10.12, SD=10.07), t(42)=.736, p > .05] (see Figure 5).

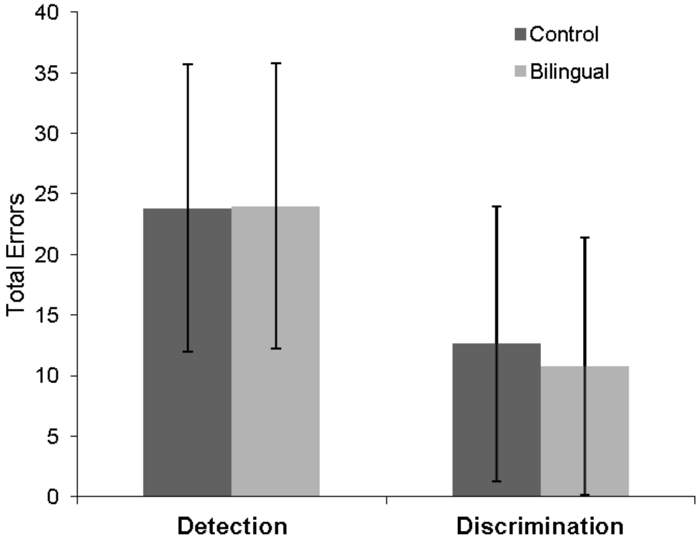


Figure 5. Pitch-change detection and pitch-direction discrimination mean group accuracies. Error bars represent standard error of the mean.

Comparisons between control (M=34.77, SD=11.41) and bilingual participants (M=33.89, SD=12.35) in the steady contour pitch sequence task were also not significant [t(48)=.256, p > .05]. Likewise, for the Transposed version of this task, no differences were found [controls (M=52.6, SD=6.58) and bilinguals (M=48.11, SD=12.87), t(36)=1.374, p > .05] (see Figure 6).

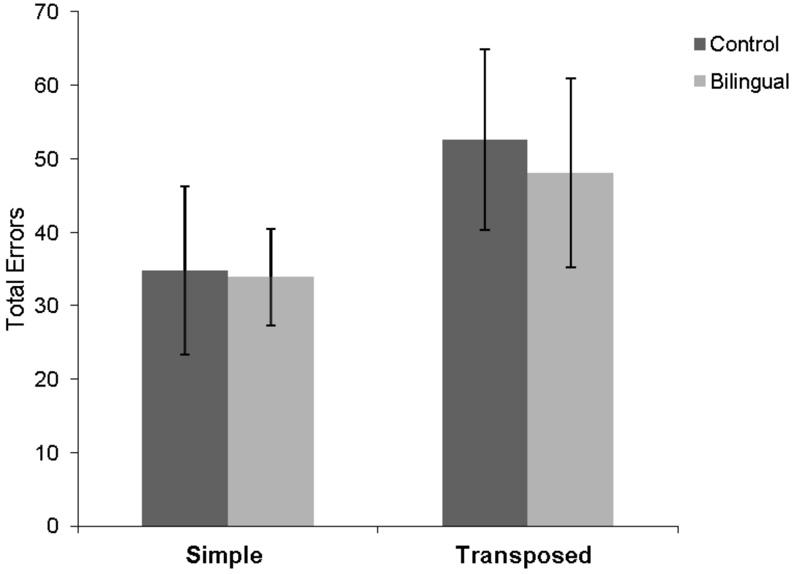


Figure 6. Mean group performance for the steady-contour pitch sequence tasks (simple and transposed). Error bars represent standard error of the mean.

The pitch sequence tasks required a Same/Different judgment and, although they have a 2AFC format like the other pitch tasks, they are more susceptible to response biases as they also have characteristics of a reminder task (the second stimulus has to be compared to the first) (MacMillan & Creelman, 2005). We therefore computed and compared Hit rates, False Alarm (FA) rates, and d prime. Results from these measures can be found in Table 3 and are not presented in detail here since participants did not differ significantly in any of them.

Table 3.

Independent T-tests on Simple Tone Sequence and Transposed Tone Sequence d prime measure

.

	Bilingual			Control			95% CI for Mean Difference		
d prime	M	SD	n	M	SD	n		t	df
Simple	1.152	.880	16	1.285	.772	32	-.366, .631	.536	46
Transposed	.267	.309	18	.176	.349	20	-.090, -.090	-.843	36

Discussion

In this study we tested the effect of early bilingualism in a tone and a non-tone language on musical pitch perception. Our hypothesis was that early exposure to diverse tonal systems (lexical tone in Cantonese and intonational tones in both Cantonese and Dutch) would sharpen attentional mechanisms involved in tone perception, and hence facilitate the processing of tone in the musical domain. With respect to online speech-music interactions, our results demonstrate that Cantonese-Dutch bilinguals process musical pitch and phonology in a more holistic way than the Dutch control participants. That is, Cantonese-Dutch bilinguals showed more interference (in speed and accuracy) with varying speech information while performing musical judgments and larger gains (significant only in RTs) when the melodic information varied consistently while performing phonological judgments.

The two groups did not differ in their ability to learn new sound categories. We also failed to observe a bilingual tone-language advantage in the pitch-perception tasks.

Online Processing Interactions between Speech and Music

Previous studies have shown that when speech and music dimensions are processed simultaneously they interact, such that irrelevant variation in one dimension cannot be filtered out and interferes with processing the other dimension, while consistent variation offers performance gains in the form of faster processing (Gordon et al., 2010; Kolinsky et al., 2009). Speech and music information are therefore intimately coupled during online processing. Not only do they share resources but they are also perceived integrally.

When asked to perform a task requiring processing of multidimensional stimuli carrying both, melodic and phonological information, our Cantonese-Dutch bilingual participants showed more interference from irrelevant variation in phonology while performing melodic judgments as indicated by their error rates and reaction times, compared to controls. Using similar multidimensional stimuli, Lee and Nusbaum (1993) found that irrelevant variation in segmental information slowed down processing of (non lexical) constant pitch information in Chinese speakers but not in English speakers. Although our Dutch control participants showed interference from phonological variation, the effect was larger for the Cantonese-Dutch participants. Contrary to Lee and Nusbaum (1993), who also found interference from orthogonal steady pitch variations in phonetic performance in Chinese but not English speakers, we found the opposite pattern: the control participants showed interference from music intervals while performing phonological judgments but the Cantonese-Dutch bilinguals did not. This disparity could be due to differences in the pitch stimuli (steady pitch instead of sung music intervals) and/or the segmental stimuli used. In particular, Lee and Nusbaum (1993) used syllables that formed actual words in Mandarin, thus carrying semantic in addition to phonological information for Chinese listeners. Caution is therefore warranted when comparing findings across these studies.

In the more comparable Kolinsky et al. (2009) study, participants showed asymmetric interference similar to our participants (i.e., more interference in the melodic task). Kolinsky et al. interpret the asymmetry in terms of processing levels, arguing that interval processing (or musical mode of processing) occurs at a later stage compared to phonetic processing and may be more demanding. This could also be the case for the overall asymmetric interference we find in the melodic task.

Interference alone, however, is not enough to argue that two dimensions are processed integrally. Lee and Nusbaum (1993) did not include a redundant condition to measure redundancy gains, which according to Garner & Felfoldy (1970) are necessary to argue that two dimensions are perceived integrally. Participants in our study showed redundancy gains in both tasks, indicating that the two stimulus dimensions, music and speech, were processed integrally. However, the gains were larger for the phonological compared to the melodic task. This asymmetry in gains was more pronounced in the Cantonese-Dutch bilinguals in the RT analysis, and they also had lower error rates in the redundant phonological condition (although the group difference was only marginally significant). This is in contrast with the symmetric gains found in Kolinsky et al. (2009). The disparity could be due to the fact that the phonemes used were non-native for our listeners and native for the French listeners in Kolinsky et al. (2009). This might have lead to discrepancies in baseline performance in the two studies and subsequently to the asymmetric gains in the present study.

Asymmetric redundancy gains are difficult to interpret solely in the context of stages of processing. If music processing occurs later than phonological processing, its redundant variation should not offer any performance facilitation. According to Garner (1983), in the absence of discriminability differences between two dimensions, asymmetric gains or interference can be attributed to selective attention differences acting on top of different levels of perceptual processing. The asymmetric redundancy gain in our data could therefore mean that it is easier for participants to attend selectively to phonological information when fewer processing resources are taken up by redundant melodic variation. The gain is asymmetric because selective attention is less effective when processing melodic information, at least if it is always processed after phonological information.

The asymmetry in gain during the phonological task versus interference in the melodic task was larger for our Cantonese-Dutch bilingual group compared to the control group. We hypothesize that the bilinguals could not avoid attending to irrelevant phonetic information while processing pitch information because of their linguistic experience. They therefore processed the two dimensions in sung words in a more integrated way than control participants. When processing a tone language, attending simultaneously to pitch and phonetic information is essential. Although such simultaneous processing of pitch and speech information should be familiar to Dutch speakers through intonation, pitch variation in that case is usually more dynamic compared to intervals consisting of two music tones. Cantonese, in contrast, includes level tones, which are discriminated in terms of F0 (Khouw & Ciocca, 2007). It is thus plausible that the observed group differences are due to the differences in the participants' linguistic background.

We did not find evidence for the alternative hypothesis that, being early bilinguals, our participants should have better inhibition and switching abilities, and therefore would show less interference than controls. One could argue that their performance in the phonological task, where they showed no interference but significant redundancy gains, was due to superior attentional control. However, if that were the case, we should have observed the same pattern in the melodic task. Since that was not the case, we favor the interpretation that our results are attributable to the Cantonese-Dutch bilinguals' tone-language experience, which has shaped the way they integrate phonological and pitch information.

Learning New Sound Categories

We also tested whether early bilingualism with a tone and a non-tone language has an influence on the ability to learn new sound categories. In particular, we wanted to test the SSCLM hypothesis (Patel, 2008) which suggests that speech and music share the same sound category learning mechanism. Our prediction was that the bilingual group would be better in learning music interval categories due to a more "trained", and thus efficient, SSCLM. Contrary to our expectation, we did not observe a bilingual advantage in learning performance. If anything, the control group performed better, at least when differences in associative learning and short-term memory were not taken into account.

We can only speculate about why we could not observe the predicted bilingual advantage. One of the possible reasons is the short duration of the training. Participants had to learn to associate three music intervals with three color labels in only 96 trials. This might have been too little exposure to yield any group differences. In Cooper and Wang's (2012) tone training study, for example, no significant differences were found in the first training session but there was an advantage for musicians and tone language speakers over non musicians in the final training session. A second reason for the absence of group effects here could be that participants did not understand the concept of a music interval category. Although special care was taken to explain the music interval as a relative difference between tones, the fact that participants relied on absolute pitch strategies (i.e. identifying M2 in the lowest key and P5 in the highest key) indicates they failed to grasp the categorical nature of the interval. Again, having multiple learning sessions could have helped participants to understand music intervals and to develop more efficient learning strategies. Furthermore, we cannot exclude the possibility that the bilinguals' expertise in intonational and lexical tone might not have been relevant enough to increase the efficiency of the SSCLM in the music domain. Finally, the comparison between the experimental and control groups in the present study was not as large as it could have been. The control participants spoke multiple languages. Thus, while the controls did not speak a tone language, and they were not early bilinguals, any difference in the effect of linguistic experience between them and the Cantonese-Dutch bilinguals was smaller than it would have been if the controls were monolingual speakers.

Future studies (e.g., with more training sessions and with a strictly monolingual control group) are needed in order to draw any firm conclusions about the SSCLM hypothesis.

Pitch Perception

Finally, we also tested bilinguals' pitch-perception abilities by administering four pitch-perception tasks. Each task aimed at different levels of pitch representation (Foxton et al., 2004) ranging from simple pitch-change detection and pitch-direction discrimination to pitch sequence processing.

Despite tone language benefits documented previously in the literature, in this study we failed to find an advantage of pitch perception in bilingual speakers of a tone and non-tone language. Bilinguals and controls did not differ in their response accuracy in any of the levels of pitch representation.

Again, caution should be taken in interpreting these results. Participants performed tasks on pure sine wave tones with which they probably did not have any prior experience, in comparison to acoustically richer musical or lexical tones. Furthermore, the majority of pitch excursions in the pitch change detection and pitch direction discrimination were smaller than a semitone. The smallest difference in Cantonese level tones is in the order of a semitone (Khouw & Ciocca, 2007). Given that the language- experience advantage decreases as the pitch excursions tested become smaller than those occurring naturally in the respective tone language spoken (Bidelman et al., 2013; Giuliano et al., 2011b), we could speculate that this was the reason we did not find group differences.

In addition, the tone sequences in the sequence task did not have any melodic structure and thus lacked contextual information that could have aided participants (e.g. dominant key information). As a previous study comparing Cantonese to English speakers has shown, music pitch perception benefits of tone language speakers are limited to the integration of musical tones over a larger melodic context (Wong et al., 2012). The absence of such a context may have posed significant challenges, especially in the transposed version of the pitch sequence task, as revealed from accuracy scores which were at chance level for both groups. The fact that changes were never contour violations added to the task difficulty, perhaps obscuring group differences that could have arisen under other conditions.

Another reason why there was no difference between groups in the pitch tasks could be, as noted with respect to the learning task, the contrast between the experimental and control groups was smaller than it could have been. Furthermore, our Cantonese-Dutch bilingual participants were living in the Netherlands and thus were immersed in a Dutch-speaking environment instead of a Cantonese one. In their study, Bidelman and colleagues (2013) found a positive correlation between the amount of exposure to Cantonese tones and perceptual pitch advantages in their Cantonese participants. This could also explain why our Cantonese-Dutch participants did not outperform their Dutch peers in pitch perception tasks.

The effect sizes for speech-music transfer effects reported in tone- language speakers are very small even in studies with large sample sizes (see Wong et al., 2012). It thus seems reasonable to conclude that, if they exist, the effects of speaking a tone and a non-tone language on musical pitch processing are not large enough to be detected in our tasks. This in turn suggests that there may be limitations to the role that experience can play in shaping pitch processing abilities. Furthermore, individual differences in pitch processing aptitude might influence performance by interacting with experience, and/or those differences may tend to be larger that those due to experience. Lastly, the fact that the transfer effects are difficult to detect indicates that although the two domains cannot be completely modular, some aspects of music pitch processing may be relatively impenetrable to the effects of linguistic experience. Further research with more power, a better control group, and more sensitive behavioral measures could nevertheless reveal effects on musical pitch processing stemming from bilingual experience with pitch.

Conclusions

To conclude, we tested whether early experience with a tone language and a non-tone language (in a group of early Cantonese-Dutch bilinguals) has effects on music processing abilities. We found no bilingual advantage in learning music interval categories or in pitch perception tasks. But we did find evidence for more holistic processing of sung stimuli in the Cantonese- Dutch bilinguals compared to Dutch controls. This is, to the best of our knowledge, the first study to show that being bilingual in a tone language and a non-tone language influences online processing interactions between music and speech. Not only did the bilinguals find it more difficult to filter our irrelevant phonological information while judging melodic intervals, but they also showed greater performance gains when melodic information varied consistently with phonological information. Although greater pitch interference in tone-language speakers has been demonstrated before (Lee & Nusbaum, 1993), this is the first time that redundancy gains, crucial for the integrality argument (W. Garner & Felfoldy, 1970), have also been shown in tone-language speakers. We interpret these results as arising from the bilinguals' experience with their tone language, where pitch and segmental information are integral properties of speech segments and spoken words.

This is fundamentally different from the case of non-tone languages, where pitch is primarily an independent feature added on top of segmental information. The necessity of processing segmental and suprasegmental information integrally in a tone language thus appears to transfer to situations where speech and music are processed jointly.

Acknowledgments

We are grateful to Regine Kolinsky and Pascale Lidji for making their Vowel-Interval Speeded Classification stimuli available to us; to Hubert F. J. M. Voogd for his assistance in programming the experiments; to Yiya Chen for her help in recruiting participants and hosting the experiment in Leiden; to Jiyoun Choi and Makiko Sadakata for their input regarding the tasks; to Alex Brandmeyer for technical assistance; to Carmen Kung and Merel van Goch for their help with stimulus recordings; to all the Cantonese-Dutch bilinguals who participated in the study.

CHAPTER 4

REPETITION SUPPRESSION IN THE LEFT IFG PREDICTS TONE LEARNING PERFORMANCE

After: Asaridou, S. S., Takashima, A., Dediu, D., Hagoort, P., & McQueen, J. M. (under revision). Repetition suppression in the left IFG predicts tone learning performance.

Abstract

Do individuals differ in how efficiently they process non-native sounds? To what extent do these differences relate to individual variability in sound learning aptitude? We addressed these questions by assessing the sound- learning abilities of Dutch native speakers as they were trained on non-native tone contrasts. We used fMRI repetition suppression to the non-native tones to measure participants' neuronal processing efficiency before and after training. Although all participants improved in tone identification with training, there was large individual variability in learning performance. A repetition suppression effect to tone was found in the bilateral inferior frontal gyri (IFG) before training. No whole-brain effect was found after training; an ROI analysis, however, showed that, after training, repetition suppression to tone in the left IFG correlated positively with learning. That is, individuals who were better in learning the non-native tones showed larger repetition suppression in this area. Crucially, this was true even before training. These findings add to existing evidence that the left IFG plays an important role in sound learning and indicate that individual differences in learning aptitude stem from differences in the neuronal efficiency with which non-native sounds are processed.

Learning a second language can be a demanding enterprise, especially when it comes to learning a non-native phonology. Individuals vary greatly in their ability to learn to perceive and produce non-native speech sounds (Chandrasekaran et al., 2010; Golestani & Zatorre, 2009; Hanulíková, Dediu, Fang, Bašnaková, & Huettig, 2012). Although several experience-related factors such as age at which the non-native phonology is acquired (Flege, Yeni-Komshian, & Liu, 1999), the amount of exposure to the non-native language (Flege, Bohn, & Jang, 1997), the overlap between native and non- native phonology (Best, McRoberts, & Goodell, 2001), or the amount of music education received (Wong & Perrachione, 2007) might all be contributing to this variability, they cannot fully account for it. What is then driving these individual differences?

It has been proposed that individual differences in language learning aptitude arise, to some extent, as a consequence of individual differences in the functional properties of underlying brain mechanisms (Zatorre, 2013). These neuronal predispositions interact with language experience, making some individuals more successful learners than others. A number of training studies have shown that successful learners of non-native speech contrasts process sounds differently compared to less successful learners (Golestani & Zatorre, 2004; Ventura-Campos et al., 2013; Wang, Sereno, Jongman, & Hirsch, 2003; Wong, Perrachione, et al., 2007). These processing differences can even sometimes be observed before the commencement of training (Wong, Perrachione, et al., 2007)

The fact that learning attainment correlates with the post-training neuronal activation associated with the non-native sounds is interpreted as showing that training increases processing efficiency in successful learners (Golestani & Zatorre, 2004). The more specific question that then arises is: Are successful learners processing these sounds more efficiently?

fMRI adaptation is a good measure of neuronal processing efficiency. fMRI adaptation or repetition suppression refers to the reduction observed in the BOLD response when a stimulus or stimulus properties are repeatedly presented (Grill-Spector et al., 2006). Although the neurophysiological mechanisms underlying adaptation phenomena are still not fully understood (Segaert, Weber, de Lange, Petersson, & Hagoort, 2013), repetition suppression can be interpreted as a neuronal marker of increased processing efficiency (Grill-Spector et al., 2006), such that the more efficient the processing of a stimulus, the greater the BOLD suppression.

A training study by Chandrasekaran, Kraus, and Wong, (2012) provided evidence that repetition suppression to non-native sounds reflects individual differences in the efficiency with which individuals process non-native sound information. fMRI adaptation was measured in the inferior colliculus, a region in the brainstem which encodes sound frequency (Yan, Zhang, & Ehret, 2005), before participants received training in non-native Mandarin tones. Individuals who showed repetition suppression to tonal contours in the inferior colliculus prior to training initiation were subsequently better learners of tones (Chandrasekaran et al., 2012). Although the implications of these findings are very interesting, the study focused exclusively on the inferior colliculus. Pitch processing, however, involves a number of cortical and subcortical areas along the auditory pathway, including the thalamus, the primary and secondary auditory cortices (Javad et al., 2014), as well as frontal areas (Nan & Friederici, 2012). Moreover, auditory learning and tuning of subcortical areas relies heavily on their feedback connections to cortical resources (Bajo, Nodal, Moore, & King, 2010).

Processing efficiency might therefore be reflected in the activity of a specific node in the pathway, or in the orchestration of multiple nodes, that is, efficiency might be instantiated in a stronger connection between the nodes along the pitch processing pathway.

In the current study, therefore, we investigated adaptation effects across the entire pitch processing pathway and asked how they relate to individual variation in tone learning. Using a learning paradigm, we trained Dutch native speakers in non-native pitch contours, modeled after Mandarin tones, over the course of five separate sessions. Participants' repetition suppression to the non-native tones was measured at two different time points, before and after training. Standard Dutch (the official language taught in school and used in public discourse) does not use tones at the lexical level. Given this, and the results of previous studies using tone training in English speakers (Chandrasekaran et al., 2012; Wong, Perrachione, et al., 2007), we anticipated large individual variation in learning performance.

The purpose of the study was three-fold. First, we were interested in which area(s) along the pitch processing pathway show repetition suppression to tone, when other acoustic properties (voice and phonemes) vary randomly. Given the hierarchical nature of pitch processing (Javad et al., 2014), we expected that the regions involved in abstracting tonal pitch contours over and above other varying acoustic information would include the bilateral superior temporal gyri/sulci and the inferior frontal gyri (Nan & Friederici, 2012; Wang et al., 2003; Wong, Perrachione, et al., 2007). These areas act in concert, with superior temporal areas being involved in the sensory processing of varying pitch (Javad et al., 2014) while the inferior frontal areas, especially in the left hemisphere, being involved in higher- order, decision-making aspects of pitch processing (Nan & Friederici, 2012).

The second purpose of the study was to test whether repetition suppression to tone is associated with differences in tone learning success. In other words, we wanted to assess the hypothesis that successful learners should process tones more efficiently and therefore show larger repetition suppression when a tone is repeated compared to less successful learners, especially after training. Previous language learning studies have demonstrated that activation in the left inferior frontal gyrus after training is associated with successful tone (Wong, Perrachione, et al., 2007) and phonetic learning (Golestani & Zatorre, 2004; Ventura-Campos et al., 2013). Although there is, to the best of our knowledge, no prior fMRI adaptation study correlating repetition suppression to tone with tone learning performance, a study looking at non-native phonetic category learning found a positive correlation between repetition suppression to non-native phonemes and performance in the left

IFG (Myers & Swan, 2012). Given the existing literature and the evidence for the involvement of the left IFG in tone perception, we expected that repetition suppression in this area (at least after training) would be associated with tone learning performance.

The third purpose of the study was to assess how repetition suppression effects are influenced by changes introduced by learning in the connectivity patterns between pitch processing areas. That is, we were interested in the dynamic changes in feed-back and feed-forward connections along the pitch processing pathway that could mediate perceptual learning (Ahissar et al., 2009; Bajo et al., 2010). For that reason, we performed functional connectivity analyses looking at cortical and subcortical areas (i.e., the inferior colliculus and the auditory thalamus). These areas are involved in pitch processing through afferent and efferent connections to the cortex (Javad et al., 2014).

Methods

Participants

Forty young adults (15 males, mean age=22.62, SD=3.16) participated in the study. They were native speakers of Dutch, recruited from the Radboud University and Max Planck Institute for Psycholinguistics databases in Nijmegen, the Netherlands. Left-handed as well as participants with neurological, speech or language disorders were excluded from the sample. Participants were all screened for hearing with an Oscilla USB-330 audiometer (Inmedico©, Denmark) using the random automatic hearing test at 20 dB in 11 frequencies ranging from 125 Hz to 8 KHz in both ears. All were able to detect frequencies ranging from 250 Hz to 4 KHz at intensity higher than 30 dB in both of the two ears. None of the participants had had experience with a tone language and/or with a tonal dialect spoken in the Dutch province of Limburg. All participants gave written informed consent prior to the experiment (local ethics committee CMO region Arnhem–Nijmegen, The Netherlands) and were compensated with 60 euro or 6 course points.

Stimuli

In the training study there were 24 Dutch-Chinese hybrid monosyllabic nonsense words. These hybrids (hereafter "Dutchinese") were Dutch in the sense that they were pseudowords with phonemes which followed Dutch phonotactic rules, and Chinese in that Mandarin tone contours were superimposed on the syllables. By using hybrid stimuli we made sure participants did not have to learn anything about Mandarin segmental phonology while at the same time we could create minimal quadruplets differing only in pitch contour with all the other variables (e.g. word duration, intensity, vowel length, production rate etc.) kept constant. The idea was to make the pitch contour the only acoustic information available in the stimuli for participants to dissociate words within a quadruplet.

Seventeen pseudowords with a Consonant-Vowel-Consonant (CVC) structure were created, six of which were used for the training paradigm (see Table1). The remaining 11 words were used in the Tone Discrimination and Tone Identification tasks. We recorded eight Dutch native speakers (four men and four women) reading aloud the list of pseudowords at a pace and pitch of their preference. Similarly we recorded eight native speakers of Chinese (four men and four women) uttering the word "mi" on four citation-style Mandarin tones: high level Tone 1 (T1), low rising Tone 2 (T2), low dipping Tone 3 (T3), high falling Tone 4 (T4). Recordings were made in a soundproof booth using Adobe Audition software at a 44100 Hz sampling rate. The hybrid stimuli were then created automatically by superimposing the Mandarin pitch contours on the Dutch utterances using the Functional Data Analysis (FDA) method for speech analysis and re-synthesis [http://lands.let.ru.nl/FDA/index.htm (Gubian, 2011.)].

Table 1

IPA Transcriptions of the Hybrid Words Used in the Experiment

Task	Dutch CVC	IPA transcription
Dutchinese Training	baaf*	[baːf]
	din	[dɪn]
	jor	[jɔr]
	moep*	[mup]
	nuuk	[nyk]
	wum*	[ʋʏm]
Tone Discrimination	dul	[dʏl]
	goel	[χul]
	luug	[lyχ]
	rof	[rɔf]
	tar	[tɑr*]
	ziem	[zim]
Tone Identification	beem	[beːm]
	nal	[nɑl]
	seek	[seːk]
	wot	[ʋɔt]

Note. * Words used in the fMRI adaptation task.

Stimulus Ratings

We conducted a rating study in order to identify the Dutchinese hybrid tokens in which native Mandarin speakers could most correctly and reliably identify the intended Mandarin tone. Twenty-nine Mandarin Chinese speakers were asked to recognize the tone in the hybrid word and rate its naturalness. We then selected the hybrid words spoken by four different hybrid Dutch- Mandarin pairs of speakers (hereafter four "Dutchinese" speakers) that were most accurately identified and had received the highest naturalness rating.

Dutchinese Training

The training was designed based on Chandrasekaran et al. (2010), adapted to five sessions of training. Neurobehavioral Systems Presentation software (www.neurobs.com) was used for stimulus presentation and response recording. The participants' task was to learn twenty-four word-picture associations over the course of the five training sessions. Each session would start with the training part followed by the testing part. During training, participants were presented with one of the colored pictures of everyday items [from the Snodgrass & Vanderwart set, (Rossion & Pourtois, 2004)] on a computer screen and heard their Dutchinese names from a pair of headphones. In order to facilitate learning, the presentation was blocked per CVC (6 CVC = 6 blocks) and sub-blocked per Dutchinese speaker. All the items were presented twice for each speaker sub-block with a total of 32 stimuli-pairs per block (1 CVC x 4 tones x 2 repeats x 4 speakers) and a total of 192 training trials.

Participants were thus trained in each minimal quadruplet for each block. In order to boost their memory with an emphasis on the tonal differences as the discriminating factor between phonemically identical words, after each block they received a mini-quiz consisting of 16 trials (1 CVC x 4 tones x 4 speakers) in which they were presented with the four pictures on the screen, heard one word at a time and had to click with the mouse on the picture that corresponded to the word. Upon clicking a picture they would hear the word again and get visual feedback on their response (either the printed word "correct" if they were right or the correct picture if they were wrong) (see Figure 1). The training data were not analyzed.

During the testing part, participants were presented with one word at a time and had to click on the corresponding picture from the whole set of twenty-four presented on the screen. The total number of trials was 96 (6 CVC x 4 tones x 4 speakers) and no feedback was provided during this part. In the final session (Session 6), participants performed a generalization test, which was identical to the regular testing part with the exception that the Dutchinese speakers uttering the words were new (i.e., the other four hybrid speakers).

Participants' response accuracy was recorded (percentage of correct picture-word matches). As in Chandrasekaran et al. (2010), we took accuracy in the final generalization test as participants' final learning score. Each training-testing session lasted around 30 minutes in total.

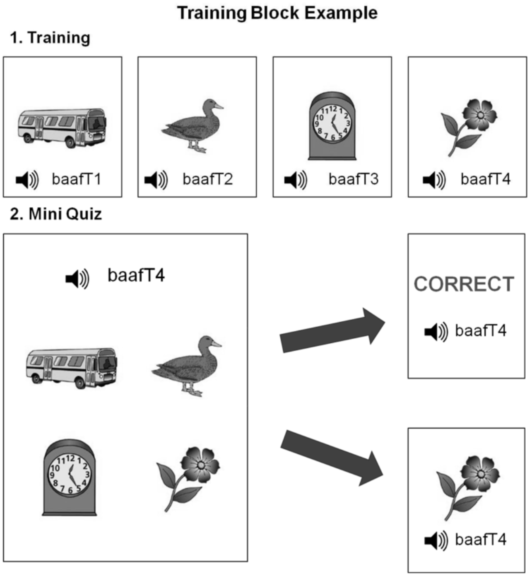


Figure 1. Example of a Dutchinese training block in which the participant is asked to learn the association between words in the minimal quadruplet baafT1, baafT2, baafT3 and baafT4 and their matching pictures.

Tone Discrimination and Identification tasks

Participants completed two tone perception tasks prior to training initiation and after training completion (designed after Chandrasekaran et al., 2010). The purpose of these tasks was to ensure that the lexical training indeed trained participants in the non-native tone contrasts instead of just tapping into simple associative learning abilities. In the tone discrimination task, participants listened to minimal pairs of Dutchinese words and had to report whether or not the words differed in tone. The pairs were CVC words chosen from six minimal tone quadruplets and were different from the ones participants were trained on (see Table 1 for the IPA transcription). All the words were uttered by the same female Dutchinese hybrid speaker so that the only acoustic difference between a pair was the pitch contour. The words were presented using in-house software through headphones with 500 ms ISI and participants were instructed to press one of two buttons on a button box as soon as they had made their same-different decision. The task included eight practice trials with feedback in the beginning and 144 test trials including all possible combinations of tones. Button and trial order were counterbalanced across participants. Response accuracy was recorded.

In the tone identification task, participants listened to single Dutchinese words and had to indicate the direction of the pitch contour in the word. There were three possible directions: upwards (indicated by an upward pointing arrow), downwards (indicated by a downward pointing arrow) and flat (indicated by a horizontal flat arrow). The words used in this task were different from the ones used in the discrimination and training tasks (see Table 1), and consisted of five CVC words uttered by a female and two male speakers. After a fixation cross, the word was presented through headphones together with the three arrows were presented on the screen. Participants were instructed to listen carefully and click the button corresponding to the correct arrow. The task included 18 practice trials with feedback in the beginning of the test and 135 test trials. Response accuracy was recorded.

Control tasks

Since learning abilities are influenced by general intelligence and memory abilities, we administered two control tasks to assess these abilities in our sample. We used Raven's Advanced Progressive Matrices Test (1998 Edition, set II) to assess non-verbal general intelligence, and the Backward Digit Span subtest adapted from the Dutch version of the Wechsler Adult Intelligence Scale (WAIS) to assess working memory. Participants were also asked to fill out a post-study questionnaire about their language and music background as well as their motivation and the learning strategies they used during the training.

fMRI Adaptation task

During the fMRI adaptation task, participants were lying in the scanner and were presented with Dutchinese words through in-ear MR compatible earbuds (Sensimetrics S14 system). The presented words were a subset of the Dutchinese words they were trained on (“baaf”, “moep”, “wum”) uttered by two female speakers. To reduce any influence of expectation, prediction and attention on our fMRI adaptation effects (Segaert et al., 2013), we used a slow event-related instead of a block design while participants were asked to perform a task that was orthogonal to our measure of interest. As in Chandrasekaran et al (2012), they performed an intensity judgment in each trial, that is, they reported whether the intensity of the presented word had changed or remained the same compared to the previous one. The task ensured participants were attending to the words during the experiment.

Each trial began with a white fixation cross presented for a jittered interval of 3-7 s after which the fixation cross turned blue for 1 s followed by the word presentation. After another jittered interval of 3-7 s, participants were presented with the two response options on the screen ("same-different") and had to press the corresponding button with their right index or middle finger (see Figure 2B). The intensity changed by 65 ±10 dB in 7% of catch trials. At the same time however, the tone in the presented words was repeated in 50% of the trials while the other acoustic dimensions varied pseudorandomly. The stimulus list was created using the MIX algorithm (www.mrc-cbu.cam.ac.uk/personal/maarten.van-casteren/Mix.htm). The total number of trials was 364 (including 20 null event trials in which no stimulus was presented) and the task lasted around 35 minutes. The sound volume was adjusted to the participants' comfort level of volume over scanner noise prior to task initiation. The fMRI Adaptation task took place twice, once before the Dutchinese training, on Session 1 (pre-training), and again after completing the Dutchinese training, on Session 7 (post-training). A post-scanning questionnaire was administered after the last fMRI session to identify participants who could have become aware of the tone manipulation.

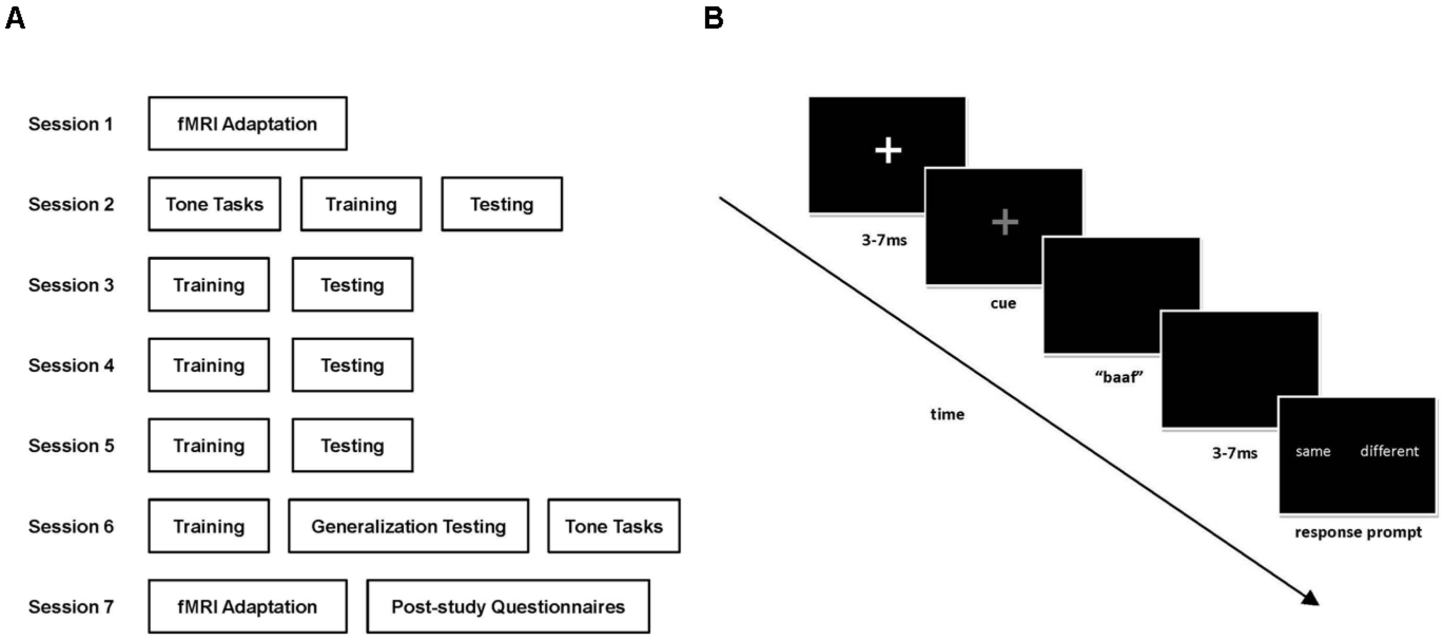


Figure 2. A. Outline of the experimental procedure. B. Example of a trial in the fMRI adaptation experiment.

Image Acquisition

MRI data were acquired on a Siemens 3T MAGNETOM Trio Tim MR system (Siemens Healthcare, Erlangen, Germany) using a 32-channel head-coil. We used multi-echo planar imaging (EPI) for the functional T2*-weighted images where a single excitation was followed by multiple acquisition times. We opted for this type of sequence since it reduces artifacts caused by signal dropout, which usually affect the inferior frontal and temporal areas we were interested in (Poser, Versluis, Hoogduin, & Norris, 2006). We used a repetition time (TR) of 2.25 s with four acquisition times (TEs) at 17.0 ms (TE1), 26 ms (TE2), 35 ms (TE3), 45 ms (TE4), with 90° flip angle, accelerated with GRAPPA parallel imaging (accelerating factor 4). We acquired 35 axial slices per volume in ascending order, with 3mm slice thickness, 224mm field of view (FOV), 0.51 mm slice gap, matrix size 64 x 64. This allowed us to acquire almost the whole brain, with the exception that the cerebellum was not scanned in most participants. We also acquired a high resolution T1-weighted anatomical image using a magnetization prepared rapid gradient echo sequence with the following parameters: TR: 2.3 s, TE 3.03 ms, 8° flip angle, 192 slices, 1.0 x 1.0 x1.0 mm^3 voxel size, 256 mm FOV, and matrix size 256 x 256, accelerated with GRAPPA parallel imaging (accelerating factor 2).

Procedure

The experiment consisted of seven separate sessions that lasted a total of seven hours (see Figure 2A). On Session 1, participants performed the pre- training fMRI-adaptation task in the scanner. Resting-state fMRI and DTI scans were also collected during that session, but will not be discussed here. On Session 2, participants came to the behavioral lab and performed the tone perception tasks (discrimination and identification) as well as the first Dutchinese training-testing task. On Sessions 3, 4 and 5, participants performed the Dutchinese training-testing task only. On Session 6, they performed the last Dutchinese training and generalization testing, followed by the tone perception and the general control tasks (Raven and Backward Digit Span). The training sessions took place on separate days with no more than three days between sessions. On Session 7, participants came to the MRI lab for the post-training fMRI-adaptation task. Resting-state fMRI and an anatomical scan were also recorded. The time between Session 6 and Session 7 was not more than three days. Participants were asked to fill out the post-study questionnaire upon completion of the study.

Behavioral Analyses

The behavioral analyses were carried out using the IBM SPSS 19 statistical package. For the Dutchinese training task, participants' response accuracy in matching the Dutchinese words to their corresponding pictures was analyzed using repeated-measures ANOVA, with Session (x 5) as factor and percentage correct as the dependent measure. All post-hoc pairwise comparisons were Bonferroni corrected. The tone discrimination and identification tasks were analyzed using paired-sample T tests to compare mean response accuracy (percentage correct) before and after training. We also performed pairwise correlations between the final learning score and the tone perception tasks as well as the general control tasks, music training duration, and motivation.

fMRI Analyses

Preprocessing

One participant was excluded from the imaging analyses because a brain anomaly was found (as assessed by a radiologist). Seven participants were further excluded from the fMRI analyses (three did not fulfill the inclusion criteria, being either left-handed or had neurological/speech/language disorders, and four due to technical problems).

Since we used a multi-echo sequence (i.e. acquired four echoes per TR) we combined the echoes before applying any preprocessing by following the echo-weighting procedure described in Poser et al. (2006). Firstly, all the first echo volumes acquired were realigned to the first volume of the first echo. All the volumes of all the remaining echoes were subsequently realigned to the first echo and resliced. Next, the first 30 acquired volumes were smoothed with a 3mm Gaussian kernel and used to calculate the optimal echo-weighting parameters (optimal contrast to noise ratio) for combining the echoes. The weighting parameters were subsequently applied to combine the echoes in all the remaining volumes. A mean functional image and a text file with the realignment parameters were created as well.

The next preprocessing steps were performed using SPM8 (www.fil.ion.ucl.ac.uk). The first five functional volumes for each participant were discarded from further analysis to remove nonequilibrium effects of magnetization. The mean functional image was co-registered to the participant’s T1-weighted anatomical image using normalized mutual information, and the registration parameters were subsequently applied to all the functional images. . The anatomical image was segmented into grey, white matter and cerebrospinal fluid and the normalization parameters from the segmentation procedure implemented in SPM8 were used for normalizing and transforming the structural and functional images to standard Montreal Neurological Institute (MNI) space (2x2x2 voxel size). Lastly, all functional images were convolved with a Gaussian smoothing kernel of full-width 8mm at half maximum.

fMRI Adaptation statistics

The statistical analysis was performed using a standard general linear model (GLM) approach in SPM8. The model included four experimental factors: tone, voice, CVC and session in a 2 (tone repeat, tone change) x 2 (voice repeat, voice change) x 2 (CVC repeat, CVC change) x 2 (pre-training session, post-training session) factorial design, which resulted in eight different conditions per session (see Table 2). Each trial was defined by the trial preceding it; that is, a trial was classified as belonging, for example, to the Tone repeat & Voice repeat & CVC repeat (TreVreCre) condition if it shared the same tone, voice and CVC with the previous trial and the Tone change & Voice Change & CVC change (TcVcCc) condition if all three features changed. The first trial, null event trials, and volume change trials were modeled in separate regressors. Events were modeled after a stick-function (0 s duration), time- locked to word onset, and convolved with the canonical hemodynamic response function. The six realignment parameters, their derivatives and the squared derivatives (in total 24) were also included in the models as regressors of no interest. Data were high-pass filtered at 128Hz cut-off and the GLM was estimated using the Restricted Maximum Likelihood (ReML) algorithm in SPM8. T-contrast images for the 16 experimental conditions vs. implicit baseline were estimated for each participant and were subsequently entered in a second level random effects analysis with random subject effects for population inferences. Since we were interested in adaptation to tone, over and above voice and consonantal information, we estimated the repetition suppression effect to tone with the following contrast: (TreVreCre + TreVcCre + TreVreCc + TreVcCc)– (TcVreCre + TcVcCre + TcVreCc + TcVcCc) masked exclusively by the repetition suppression effect to voice [(TreVreCre + TreVreCc + TcVreCre + TcVreCc) – (TreVcCre + TcVcCre + TreVcCc + TcVcCc)] and CVC [(TreVreCre + TreVcCre + TcVreCre + TcVcCre) – (TreVreCc + TreVcCc + TcVreCc + TcVcCc)] (mask uncorrected at p=.05). Results were initially voxelwise thresholded at p=.001 (uncorrected) and subsequently suprathreshold cluster extent was tested using random field methods (Hayasaka & Nichols, 2003) , corrected for multiple comparison at FWE p=.05 .

Table 2.

fMRI Adaptation Experimental Conditions

Factors			
Conditions	Tone	Voice	CVC
TreVreCre	Repeat	Repeat	Repeat
TreVcCre	Repeat	Change	Repeat
TreVreCc	Repeat	Repeat	Change
TreVcCc	Repeat	Change	Change
TcVreCre	Change	Repeat	Repeat
TcVcCre	Change	Change	Repeat
TcVreCc	Change	Repeat	Change
TcVcCc	Change	Change	Change
Null Events			
TNI			

Note. T= Tone, V= Voice, C= CVC, re= repeat, c= change, Null Events= Trials with 20s of silence and black screen. Trials of No Interest (TNI)= include the first trial and the trials with volume change.

Region of interest (ROI) analysis

We performed a region of interest analysis (ROI) on anatomically predefined regions along the auditory processing pathway. The ROI analysis aimed to increase sensitivity in detecting repetition suppression effects in brain areas that have been reported to process acoustic changes. The ROIs included Heschl's gyri (HG), superior temporal gyri (STG) and inferior frontal gyri (IFG) bilaterally (Schönwiesner et al., 2007). We also chose to include the left inferior colliculus (IC) based on the findings by Chandrasekaran et al. (2012), and the medial geniculate thalamic nuclei (MGB) since they relay acoustic information from the IC to cortical auditory areas (Javad et al., 2014). The cortical ROIs were defined using the AAL template (Tzourio-Mazoyer et al., 2002) provided by WFU PickAtlas toolbox (Maldjian, Laurienti, Kraft, & Burdette, 2003) and transformed into MNI space in MarsBaR (http://marsbar.sourceforge.net/). The subcortical ROIs were defined as spheres using the MNI coordinates reported by Mühlau et al. (2006) (5mm radius sphere around -6, -33, -11 for the left IC and 8mm radius sphere around ±17, -24, -2 for the thalamus) constructed in MarsBaR. The mean beta estimates from the single subject GLM analysis for each of these ROIs were extracted with MarsBaR and further processed in SPSS. Repetition suppression to tone was estimated as described for the whole brain analysis and analyzed in a 2 x 2 repeated measures ANOVA with tone (repeat, change) and session (pre-training, post-training) as factors. Pairwise Pearson's correlations between repetition suppression to tone in the different ROIs and the final learning score (Generalization test) were estimated in order to investigate whether individual variability in learning correlated with the size of the repetition suppression to tone effect.

Psychophysiological interaction (PPI) analyses

In order to investigate changes in functional connectivity induced by learning, we performed psychophysiological interaction (PPI) analyses in SPM8 for a number of seed regions. We selected the seed regions (volume of interest; VOIs) that, according to the literature, are involved at different stages of pitch processing: the IC, MGB, and HG (Javad et al., 2014). Since we were also interested in top-down connectivity, we also included the left IFG as a VOI. These were anatomically defined as described in the ROI section. We first estimated the physiological factor by extracting the average BOLD signal time- course from the VOIs. The psychological factor was then defined as the repetition suppression to tone effect (tone change conditions > tone repeat conditions) and was used to estimate the interaction term (seed region x effect of tone repetition). Lastly, a new GLM analysis was performed for each participant and VOI, with the 16 experimental conditions, the physiological, the psychological and the psychophysiological interaction terms as regressors, and the 24 realignment parameters as regressors of no interest. The individual contrast images for the interaction terms were then entered in one-sample t- tests at the second level for group inferences to test for the functional connectivity difference between the two experimental conditions (tone change vs. tone repeat).

Results

Behavioral Results

The behavioral analysis of participants' learning scores (percentage correct) yielded a significant effect of Session [F(1.605, 49.750)=97.187, p<.001(Greenhouse-Geisser corrected), η_p^2 =.758]. All the post-hoc comparisons were highly significant (Table 3) indicating that participants improved over the course of training.

Participants also improved in Pitch Discrimination accuracy [t(30)=-4.219, p<.001] and Pitch Identification accuracy [t(30)=-4.244, p<.001] after training compared to before (Table 4). Although all participants improved, as expected, their performance varied considerably as indicated by their learning trajectories (see Figure 3).

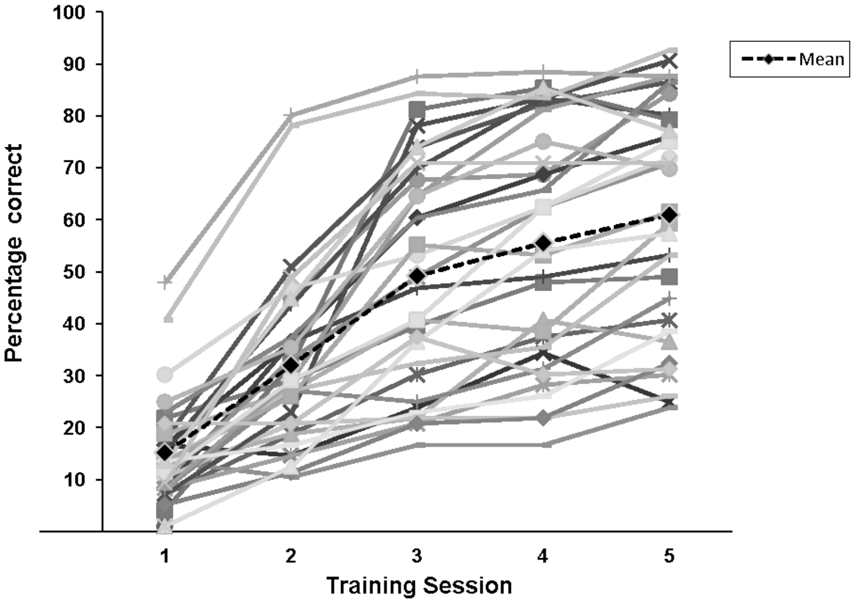


Figure 3. Individual and mean learning scores (word-picture matching accuracy) over the five training-testing sessions.

Table 3.

Post-hoc Comparisons for the Effect of Session

	Comparisons				
Final Learning Score	Session 1	Session 2	Session 3	Session 4	Session 5
Mean difference					
Session 1	-	-16.829**	-33.887**	-40.234**	-45.638**
Session 2	16.829**	-	-17.057**	-23.405**	-28.809**
Session 3	33.887**	17.057**	-	-6.348**	-11.751**
Session 4	40.234**	23.405**	6.348**	-	-5.404*
Session 5	45.638**	28.809**	11.751**	5.404*	-

Note. ** p<.001, * p=.005, *P* Bonferroni corrected for multiple comparisons.

Table 4.

Paired T-tests on Tone Discrimination and Identification Accuracy

	Pre-Training		Post-Training						
Accuracy	M	SD	M	SD	n	95% Confidence Interval	r	t	df
Discrimination	93.99	6.34	96.34	5.02	31	-3.49, -1.21	.87**	-4.21*	30
Identification	64.06	16.73	72.04	22.77	31	-11.82, -4.14	.90**	-4.24*	30

Note. ** p<.001

The correlation between the final Dutchinese learning score (Generalization) and participants' pre-training Pitch Discrimination and Identification accuracy was highly significant (r=.603, p<.001 and r=.770, p<.001 respectively) (Table 5). No correlation was found between the final Dutchinese learning score and participants' Backward Digit Span score, Raven's score, music education duration, music education onset or self-reported motivation (Table 6). We can therefore conclude that learning attainment was specific to sharpening participants' tone processing abilities rather than the result of general cognitive or musical abilities.

Table 5.

Correlations between Participants' Final Learning Score and Tone Perception Measures

Measure	Pre Tone Discrimination	Pre Tone Identification	Post Tone Discrimination	Post Tone Identification
Learning Score	.603**	.770**	.603**	.805**
Pre Tone Discr		.546**	.876**	.586**
Pre Tone Ident			.599**	.904**
Post Tone Discr				.613**

Note. **p<.001

Table 6.

Correlations between Participants' Final Learning Score and Control Measures

Measure	Backward DS	Raven	Length Music Education	Onset Music Education	Motivation
Learning Score	.189	.150	.250	.218	.280
Backward DS		.217	-.036	-.035	.303
Raven			-.042	-.181	-.091
Music Edu				.647*	.179
Onset Music					-.080

Note. *p<.001

Imaging Results

Whole brain

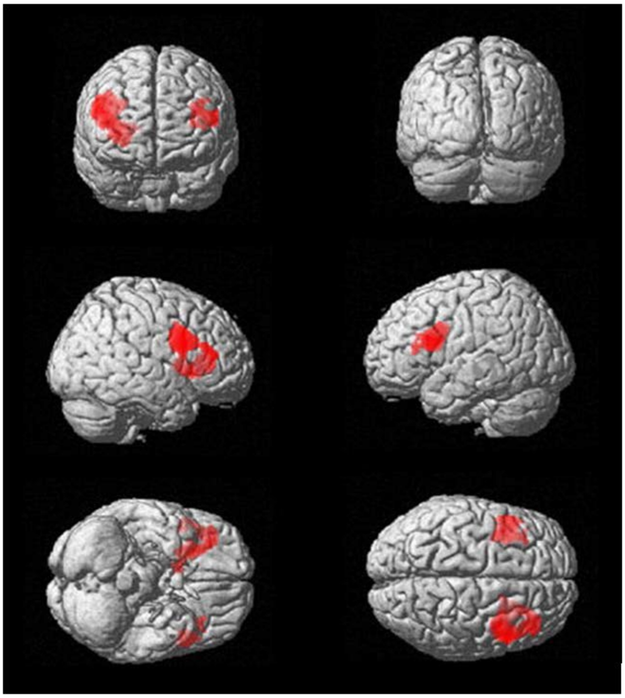


Figure 4. Repetition suppression to tone in the pre-training session. Significantly less activation with tone repetition was found in the left Pars Opercularis (POp) and Pars Triangularis (PTr), and in the right POp and Precentral Gyrus (uncor. p<.001, FWE cluster corrected).

Whole brain comparison results are summarized in Table 7. For the pre-training session, whole brain comparisons yielded significant repetition suppression effects to tone in the bilateral IFG (Figure 4). More specifically, the pars Opercularis (POp) and pars Triangularis (PTr) in the left IFG, and the POp and precentral gyrus in the right hemisphere were significantly less activated in trials where the tone was repeated compared to trials where the tone hadchanged. Overall, we did not observe repetition suppression to other acoustic stimulus dimensions (voice and CVC) and no repetition enhancement effects.

Table 7.

Whole Brain Analysis Results.

Contrast	Region	H	No of voxels	MNI Coordinates			T	Z	cluster p(FWE)
Pre-training Session				x	y	z			
				44	8	20	4.83	4.77	<.001 (2.347e-07)
Tone Change> Tone Repeat	Precentral Gyrus	R	1447	40	6	32	4.63	4.57	
	Pars Opercularis			48	14	30	4.22	4.18	
	Pars Opercularis			-48	10	22	4.65	4.6	
	Pars Triangularis	L	537	-36	28	14	4.57	4.52	.001
	Pars Opercularis			-40	6	28	3.97	3.94	
Voice Change> Tone Repeat	n.s.								
CVC Change> Tone Repeat	n.s.								
Post-training Session									
Tone Change> Tone Repeat	n.s.								
				-8	-60	26	4.40	4.36	
Voice Change> Tone Repeat	Precuneus	L	430	-10	-54	16	4.18	4.14	.002
				-4	-60	16	3.96	3.92	
CVC Change> Tone Repeat	n.s.								
Post-training > Pre-training (conjunction)									
Repetition	Pars Opercularis	L	1016	-48	12	14	4.85	4.79	<.001

	Precentral Gyrus			-46	8	30	4.76	4.7	(7.4126e-06)
				-40	4	36	4.63	4.58	
	Supplementary Motor Area	L	709	-2	24	38	4.5	4.45	<.001
				-4	16	42	4.48	4.44	(1.1821e-04)
	Mid Cingulum			-10	12	46	4.4	4.36	
	Pars Triangularis Pars Opercularis	R	334	44	22	26	4.14	4.1	.007
				46	30	26	3.92	3.88	
				48	14	32	3.59	3.57	
	~Inferior Colliculus ~Thalamus	L	265	-4	-30	0	3.96	3.92	.016
				-12	-42	4	3.72	3.69	
				2	-26	10	3.3	3.28	
Change	Mid Cingulum Anterior Cingulum	R	422	0	22	40	4.25	4.21	.002
		L		-6	26	36	4.19	4.15	
		L		-4	30	26	3.94	3.9	
	Thalamus	R L	233	2	-10	8	4.1	4.06	.025
				0	-18	10	3.78	3.75	
				-14	-14	12	3.27	3.25	

For the post-training session, we did not find any significant effect for repetition suppression or enhancement to tone. The only whole brain effect that was significant in the post-training session was a repetition suppression effect to voice in the precuneus. A comparison across sessions indicated increase in activation to tone repetitions in the post-training session compared to the pre-training session, but this was not specific to tone; a conjunction analysis showed that the same areas, including the bilateral POp, the left supplementary motor area (SMA), the left thalamus and the IC were also more active in the post- training session for repetition of voice and CVC (see Table 7, Figure 5A). Thus, the absence of post-training repetition suppression to tone seemed to be driven by an overall activation increase in response to any repeated acoustic information (tone, voice, or CVC). A similar conjunction analysis was performed on post > pre training activation to tone, voice and CVC change. It revealed more activation for post- compared to pre-training in the anterior cingulate cortex (ACC), mid cingulum and thalamus (Figure 5B).

None of the participants understood the tone repetition manipulation, as evident from their responses to the post-scanning questionnaire. Instead, they were all convinced that they were performing a task about intensity changes and had difficulties retrieving the words or the number of speakers they heard while in the scanner.

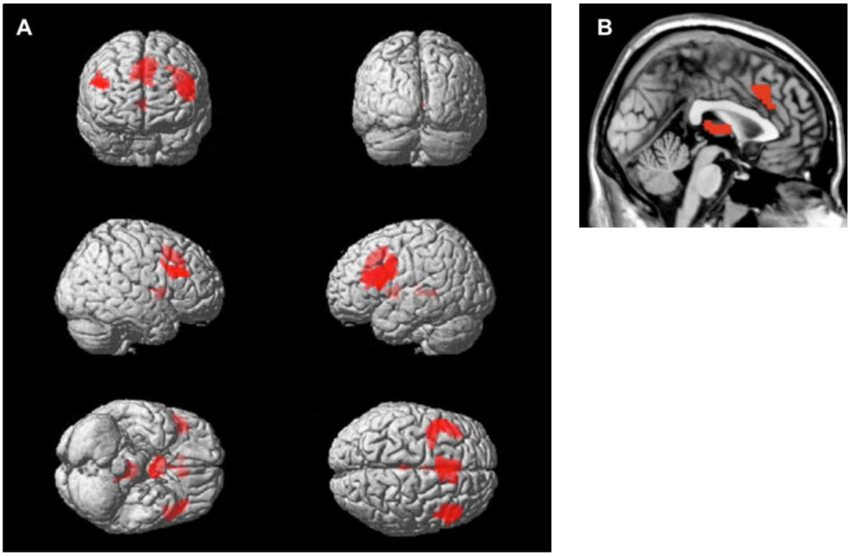


Figure 5. Conjunction analysis results (uncor. p<.001, FWE cluster corrected). A. Conjunction analysis of Post-training vs. Pre-training Tone Repetition, Voice Repetition, CVC Repetition. The bilateral POp, the left supplementary motor area (SMA), left Thalamus and IC were more active for any acoustic repetition in the post-training session. B. Sagittal view of conjunction Analysis for Post-training vs. Pre-training Tone Change, Voice Change, CVC Change. Increased activation in the anterior cingulate cortex (ACC), Mid Cingulum and Thalamus to changing acoustic information after training.

Repetition suppression effect along the auditory pathways

The ROI analysis aimed to increase sensitivity in detecting repetition suppression effects in brain areas that have been reported to process acoustic changes. The repeated measures ANOVA on the extracted beta estimates revealed a significant effect of Session with overall more activation to the stimuli on the post-training compared to the pre-training scanning session in the left and right IFG (POp, PTr, POrb), the right superior temporal gyrus (STG) and the thalami. A significant effect of Condition (Tone Change vs. Tone Repeat) was found in the bilateral IFG and thalami, with more activation for Tone Change compared to Tone Repeat condition (see Figure 6A, 6B). A significant Session x Condition interaction was found in the right Heschl's gyrus and right POp. The interaction was driven by a large repetition suppression effect in the pre-training session and a much weaker effect in the post-training session

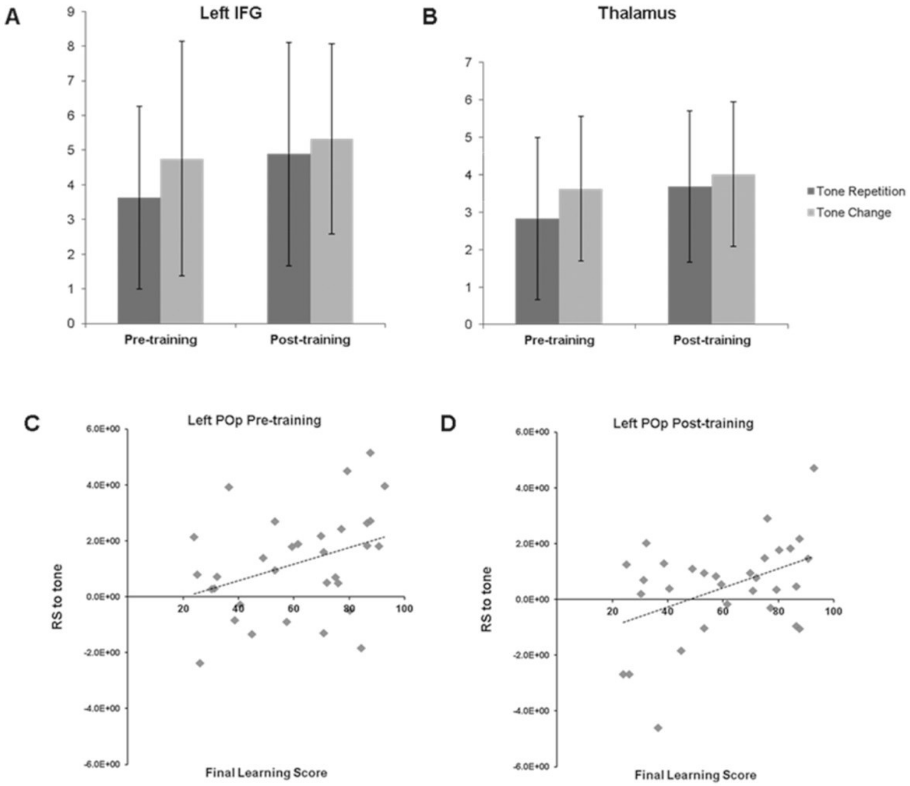


Figure 6. Mean activations (arbitrary scale) to tone repetition and tone change A. in the left IFG and B. in the thalamus, pre and post-training. Error bars denote one standard error around the mean. . C. Scatter plots of repetition suppression (RS) to tone and final learning score in the left POp for the pre-training and D. post-training session.

There was a significant correlation between participants' final learning score and repetition suppression to tone in the left IFG (r=.432, p=.014 for POp and r=.424, p=.016 for PTr) after training (see Figure 6D). Interestingly, participant's repetition suppression to tone in the left POp correlated with their final learning score even in the pre-training session (r=.361, p=.042) (see Figure 6C). This correlation seems to be driven by the fact that good learners' left POp deactivated more when a tone was repeated, compared to less good learners in the pre-training session (correlation between learning score and activation to tone repetition: r=-.384, p=.03). Apart from a marginal positive correlation between learning score and repetition suppression in the right HG (r=.324, p=.071) after training, no other correlations reached significance.

Functional connectivity along the auditory pathway

The purpose of the PPI analyses was to explore connectivity changes among auditory language areas as a result of tone learning. We therefore focused on areas that are part of the pitch processing pathway ranging from subcortical (IC) to higher order cortical brain regions (IFG). With the contrast of tone change vs. tone repeat as the psychological factor, no cluster survived the whole brain comparison in the pre-training session. However, in the post- training session, we found a significant increase in connectivity between the right HG and left POp with tone repetition (Figure 7, peak local maximum [-36, 18, 20], p=.021, FWE corrected). That is, after training had taken place, the strength of the association between activity in right HG and left POp was greater on tone repetition trials than on tone change trials. This post-training connectivity between right HG and left POp, however, did not correlate with participants' learning attainment. No other area showed significant connectivity changes in the post-training session.

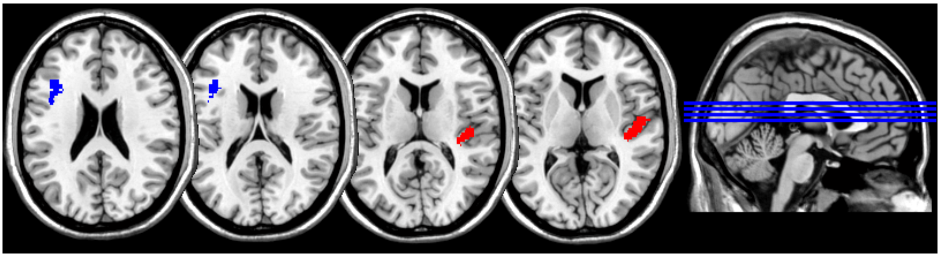


Figure 7. Multislice view of the cluster in left POp (blue) that showed significant increase in connectivity with the right HG (red) seed region in the PPI analysis (uncor. p<.001, FWE cluster corrected).

Discussion

We investigated individual variation in non-native tone learning performance by measuring fMRI adaptation to tones before and after administering a multi-session tone training procedure. Our behavioral results demonstrate that Dutch native listeners were able to learn to associate words that differed minimally in pitch contour with meaning, since their performance improved significantly with training. Based on participants' post-training improvement in tone discrimination and identification tasks, we can be confident that these results do not reflect simple associative learning but are specific to learning the non-native contrast. At the same time, we observed large individual variability in the participants' learning trajectories, replicating previous studies that used a similar paradigm (Chandrasekaran et al., 2012; Wong et al., 2007). The fact that the participants' final learning scores correlated positively with their ability to accurately discriminate and identify tone patterns before training supports the notion of pre-existing differences in learning aptitude, such that the learners who processed tone contours more efficiently benefited more from the tone training.

Overall, our Dutch native listeners showed repetition suppression to non-native tones in the bilateral IFG, including the right precentral gyrus and bilateral POp and PTr, prior to training. This was in accordance with our expectations, since bilateral IFG deactivation has been consistently reported in studies of fMRI adaptation to repeated auditory information. With respect to spoken language, IFG deactivation has been found in spoken sentence repetition (Hasson, Nusbaum, & Small, 2006), phonological feature repetition (Vaden, Muftuler, & Hickok, 2010), in repetition of non-native consonants (Myers & Swan, 2012) and with repetition of phonemes of the same phonetic category (Myers, Blumstein, Walsh, & Eliassen, 2009). A linear decrease (repetition suppression) in these areas is also observed when musical notes are repeated in short melodies (Brown et al., 2013) or when the perceived voice gender is repeated (Charest, Pernet, Latinus, Crabbe, & Belin, 2013). It thus seems that the IFG is sensitive to percepts of acoustic information, especially in cases where explicit judgments on this information are required (Hasson et al., 2006).

It is possible that our participants used their knowledge of intonation and prosody while processing the non-native pitch contours. Although Standard Dutch does not use pitch at the lexical level, it does use rising and falling pitch contours at the suprasegmental prosodic level (t'Hart, 1998). A recent meta-analysis has shown that the bilateral PTr is activated when processing affective prosody and the bilateral POp for linguistic prosody, while the right precentral gyrus is involved in both (Belyk & Brown, 2013). It could therefore be the case that, upon listening to these tones for the first time, Dutch listeners interpreted them as prosodic contours, yielding larger repetition suppression in the right IFG. This would be in accordance with lateralization patterns in prosodic processing (Belyk & Brown, 2013; Rota et al., 2009; Witteman, Van Heuven, & Schiller, 2012).

Importantly, repetition suppression in the left IFG, and particularly the POp, correlated positively with tone learning performance, such that individuals who were better learners of tones showed larger repetition suppression to tone in this area even before training. Our findings thus support the hypothesis that variation in sound learning aptitude stems in part from the fact that individuals differ in how efficiently they encode and process non-native sound contrasts. Although all learners improved significantly with training, converging fMRI (pre-training repetition suppression to tone) and behavioral data (pre-training tone identification accuracy) demonstrate that they did not start off at the same level.

Consistent with our repetition suppression (i.e. deactivation) findings, activation in response to non-native sounds in the left IFG has been shown to correlate negatively with sound learning performance (Golestani & Zatorre, 2004; Myers & Swan, 2012). Previous findings have been interpreted in a speculative manner, with accounts alluding to verbal working memory or subvocal rehearsal as the potential underlying mechanisms of left IFG activation patterns. Assuming that they lack clear representations of lexical sounds, less successful learners would rely on encoding any acoustic information available and keeping it online. This would take up more verbal memory resources to support their performance compared to successful learners. Although we cannot completely exclude such an interpretation, it seems unlikely in our case because we did not observe correlation between our behavioral verbal working memory measure and learning performance. Instead, a more favorable interpretation is that the left POp is involved in controlling and deciding on relevant abstract stimulus representations (Hasson, Skipper, Nusbaum, & Small, 2007; Myers et al., 2009; Myers & Swan, 2012) which it unifies with existing phonological knowledge (Hagoort, 2013) thereby guiding learning in sensory encoding areas. This is probably done by means of top-down feedback connections. Less successful learners would accordingly need more top-down feedback than successful learners, since they have not yet built efficient representations of the stimuli to inform perception (see also Golestani & Zatorre, 2004).

As a consequence of learning, repetition suppression to tone was not evident in the whole-brain analysis of post-training data. Providing increased sensitivity, the ROI analysis allowed us to detect repetition suppression to tone after training completion. This effect was there for thalamic and bilateral frontal areas. It was however smaller compared to the pre-training sessions, mainly due to the increase in the BOLD response for tone repetitions rather than decrease for tone changes. Overall, activation was higher in the post- training session along the bilateral IFG, the right STG and thalamus, maybe because the participants had learned to associate the Dutchinese words with meanings over the course of the training. There is strong behavioral evidence (individual variation in learning scores, and tone discrimination and identification scores) suggesting that tone stimuli were perceived differently by different learners, leading to disparate adaptation effects. Successful learners may thus have shown a decrease in activation with tone repetition while less successful learners may have shown an increase in activation. Group effects would then be cancelled out due to individual differences. The ROI analysis revealed that indeed successful learners were more sensitive to the tone repetitions in the left IFG. That is, they showed more repetition suppression to the learning-relevant acoustic dimension in the stimuli in the post-training session.

The differences observed between the pre and the post training fMRI sessions could be due to the emergence of more discreet word representations induced by learning. Our whole brain conjunction analysis on the post- training session revealed that participants showed repetition suppression to any repeated acoustic properties in the stimuli (tone, voice and CVC). This was observed in subcortical and cortical areas known to process acoustic changes (Schönwiesner et al., 2007), including the IC, the thalamus and bilateral IFG. No such overlap was found in the pre-training session. The trials with tone repeats were not always repeating in other dimensions (CVC, voice). This might have resulted in reduction of overall repetition suppression effects. Furthermore, we cannot exclude the possibility that the newly acquired semantic representations of the words might have influenced the brain activity pattern in the post-training session. This could account for the overall higher activation in the post-training session in the bilateral IFG, the right STG and thalamus. It is unlikely, however, that our results could be explained by changes in the awareness of the stimuli, since post-scanning reports indicate that participants were completely unaware of the tone repetition manipulation, and their recall of the presented words and the number of speakers required a lot of effort and was not always successful.

Our functional connectivity analysis revealed an increase in the strength of association between activation in the right HG and the left POp with tone repetition after training, regardless of learning performance. Although it is difficult to make directionality claims, we speculate that this is an increase in feed-forward connectivity from a basic pitch encoding area, such as the right primary auditory cortex, to higher order pitch contour representations in the left frontal cortex. As mentioned earlier, the behavioral results suggest that learning has taken place, as evident from the improvement in discrimination and identification of tone patterns across participants. Thus, in the post-training session, all participants must have improved to some extent in encoding pitch information, which is preferentially engaging the right HG (Luo et al., 2006; McGettigan & Scott, 2012; Warrier et al., 2009; Yisheng Xu et al., 2006). A similar right temporal – left frontal network has been postulated to underlie domain-general pitch processing by Nan and Friederici (2012). They suggest that the right auditory cortex is doing the initial pitch acoustic processing while the left IFG the more cognitive and decision related processing (Nan & Friederici, 2012). The fact that we observe what appears to be feed-forward instead of feed-back connectivity can be attributed to the task participants were performing in the scanner (i.e. the volume change detection task). This required forwarding accurate acoustic information from sensory areas to higher-order representation and decision areas. In this context, feed-back connectivity is rendered unnecessary, which probably explains why the strength of connectivity between these areas did not correlate with learning performance.

The absence of adaptation effects in the temporal lobes, otherwise often reported in auditory fMRI adaptation (Hasson et al., 2006; Hu et al., 2013; Rauschecker, Pringle, & Watkins, 2008), might be due to our design. We used a slow event-related design with a long lag between repetitions (~14s), which may have been too long for more sensory-related repetition suppression effects to arise (Grill-Spector et al., 2006). It is also possible that there was repetition suppression to tone in the primary and secondary auditory cortices, but it might have been sensitive to the number of repetitions. With only four tones available, we could not avoid repeating them multiple times across the experiment. As a consequence of this, activation levels in sensory areas might have reached saturation. Lastly, there is the possibility that these areas showed repetition suppression but that it was not large enough to survive whole brain comparisons. Our ROI analysis, however, argues against this. Myers & Swaan (2012) also did not find changes in STG after categorical phonetic training and attribute this absence to the fact that training was very short. Changes in temporal areas dedicated to more sensory processing may require long-term exposure to new sounds. Given that such changes should occur through top-down feedback from frontal areas, the patterns of IFG activation we report here could be an indication of establishing the first stage of the sound learning process.

We knew from Chandrasekaran et al. (2012) that even basic pitch encoding structures, such as the inferior colliculus, contribute to non-native sound learning. Now we also have evidence that higher-order cortical structures, such as the left IFG, are crucial for learning performance. It is our hope that future studies with longitudinal training paradigms can investigate long-term sound learning and shed more light into the role of fronto-temporal as well as subcortical sound encoding areas in this process.

To conclude, we trained Dutch native speakers in non-native Mandarin tones over five separate sessions. fMRI adaptation data to tones were acquired before and after training to assess tone processing efficiency and how it changes with learning. Participants showed repetition suppression to tones in the bilateral IFG before training. After training, there was no whole brain repetition suppression effect to tone but an increased general sensitivity to any repeated acoustic information. This increased sensitivity could be due to increased feed-forward connectivity between right auditory and left frontal regions. While all participants showed behavioral improvement, they started and ended the training at different levels, with substantial individual variation in their learning scores. Some individuals were thus better than others in learning non-native tones. We attribute their improved learning performance to more efficient processing of tones, as revealed by the correlation between repetition suppression in the left IFG and learning performance. Strikingly, this correlation was there even before training began. This suggests that individual differences in speech learning aptitude reflect differences in neuronal processing efficiency, in particular in the left IFG.

Acknowledgments

We are grateful to Michele Gubian for his assistance in creating the hybrid Dutchinese stimuli. We would also like to thank Pieter van Groenestijn and Lin Wang for their help with the stimulus rating; Wencui Zhou, Hubert F. J. M. Voogd, Pascal de Water, Alina Lartseva, and Paul Gaalman for their technical assistance throughout the experiment; Carlos Gussenhoven, Eelke Visser, Katrien Segaert, Margriet Groen, Frank Eisner, Wencui Zhu, Danchao Cai, and Alex Cristia for helpful discussions.

CHAPTER 5

INDIVIDUAL VARIATION IN NON-NATIVE WORD-LEARNING PERFORMANCE: INSIGHTS FROM CONNECTIVITY IN THE BRAIN AT REST

After: Asaridou, S. S., Fonteijn, H. M., Hagoort, P., & McQueen, J. M. (in preparation). Individual variation in non-native word-learning performance: insights from connectivity in the brain at rest.

Abstract

The current study explored associations between resting-state connectivity in the human brain and listeners' ability to learn non-native language sounds. Since individuals vary greatly in this ability, we expected that individual variation in learning performance would be reflected in resting-state connectivity patterns. We taught Dutch participants the meaning of new words which contained non-native sounds (Mandarin lexical tones). Resting-state fMRI data were acquired before and after five sessions of training and were analyzed using a seed-based correlation approach. We found large individual variation in learning to map non-native sounds to semantic content. The fMRI results showed that post-training resting-state connectivity between areas in the left ventral language-processing pathway correlated negatively with learning attainment. Moreover, pre- to post-training session changes in resting-state connectivity strength between right temporal and frontal areas were negatively correlated with participants' performance. After training, more successful language learners thus appear to employ the left ventral language sound-to-meaning pathway less, and they rely less on a right dorsal temporo-frontal compensatory pathway.

Individual variation in cognitive abilities is a well-documented phenomenon in humans (Thornton & Lukas, 2012). Among the skills in which individuals exhibit striking variation is the ability to learn non-native sounds in their adult life (Chandrasekaran et al., 2010; Golestani & Zatorre, 2009; Hanulíková et al., 2012). These differences in learning aptitude have been associated with differences in the patterns of brain activity in auditory language areas that occur in response to non-native sounds (Golestani & Zatorre, 2004; Ventura-Campos et al., 2013; Wang et al., 2003; Wong, Perrachione, et al., 2007). In the present study we set out to investigate the role of connectivity between these areas at rest. Specifically, we wanted to test whether differences in the strength of the connectivity between areas involved in sound learning are associated with variation in learning performance.

Previous fMRI research on non-native sound learning has shown that training induces changes in the neuronal response to the novel sounds. After receiving extensive training, participants showed increased activity in the bilateral inferior frontal gyri as well as in left parietal areas in response to a trained non-native consonant contrast (Golestani & Zatorre, 2004). As a result of training, non-native sound processing recruited brain areas involved in processing native sounds such as the left insula, and left and right superior temporal gyri STG (Golestani & Zatorre, 2004). When trained in a non-native tone contrast, participants showed overall more activity in the left inferior frontal gyrus (IFG) and bilateral STG in response to tones (Wong, Perrachione, et al., 2007). In addition to these overall effects of training, these studies also report differences in performance across individuals which are associated with differences in brain activity. In particular, activity in the left IFG correlates with the individual’s ability to learn non-native consonant contrasts (Golestani & Zatorre, 2004; Myers & Swan, 2012; Ventura-Campos et al., 2013). Furthermore, activity in response to non-native lexical tones increases in the left superior temporal gyrus differs for successful compared to less successful learners even before training initiation (Wong, Perrachione, et al., 2007). Patterns of brain activity are thus related to how non-native sounds are processed and can be indicative of the individual’s learning abilities.

Although the contribution of specific areas to learning performance investigated so far has offered useful insights into the neural correlates of sound learning, interest is increasingly shifting towards the role of connectivity between areas. A number of studies have focused on the role of white matter structural connectivity in language learning. It has been suggested that the process of mapping sounds to meaning is supported by the ventral language pathway (Hickok & Poeppel, 2007). Evidence for the role of this pathway in word learning comes from a study by Wong et al. (2011). It investigated the role of white matter connectivity and found a positive correlation between non-native word learning performance and Fractional Anisotropy (FA) in a left temporo-parietal cluster. Tractography seeding from this cluster revealed that it is connected to the left IFG via a ventral fiber tract which the authors identify as the Extreme Capsule (EmC) (Wong et al., 2011). Another study, looking at learning the phonological form of words, found that individual differences in learning performance were associated with white matter differences in the dorsal pathway, specifically the arcuate fascicle (López-Barroso et al., 2013). This pathway has been claimed to support matching sounds to articulation (Hickok & Poeppel, 2007). These studies provide valuable insights into the role of connectivity between brain areas by stressing its importance in individual variation in non-native word learning performance.

A more dynamic way of looking at connectivity is by investigating resting-state connectivity or, in other words, temporal correlations in spontaneous, low frequency fluctuations in the blood oxygenation level- dependent signal (BOLD) between brain areas (Biswal et al., 1995). These spontaneous fluctuations are not random and instead reflect connectivity between areas that form stable functional networks very similar to those recruited during structured cognitive activity (Smith et al., 2009). Networks showing consistent connectivity patterns at rest such as the bilateral “auditory network” and the left "frontoparietal network" correspond to areas consistently recruited for speech and language tasks (Smith et al., 2009). Resting-state fMRI can therefore be used in order to study how functional connectivity contributes to individual differences in language learning as well as other cognitive functions (Guerra-Carrillo et al., 2014).

A study by Ventura-Campos et al. (2013) investigated resting-state connectivity before and after participants received multi-session phonetic training in a non-native Hindi dental-retroflex contrast. Participants' learning performance was assessed as the difference in phoneme identification accuracy after as compared to before the training. The results showed that spontaneous fluctuations between left frontal and parietal regions could reliably predict participants' phonetic learning performance. Not only were these areas involved in active processing during the non-native phoneme discrimination task, but their functional connectivity at rest (i.e. while participants were not performing any organized activity in the scanner) was also predictive of participants' improvement in non-native phoneme identification (Ventura-Campos et al., 2013).

Learning to discriminate and identify a non-native phonetic contrast is an important aspect of learning to recognize a second language but it is not all there is to it. The study by Ventura-Campos et al. (2013) thus addresses the role of connectivity in non-native speech learning, but only partially. In natural language situations, an individual will seldom need to identify phonemes in isolation. Instead, a non-native contrast will be encountered embedded in lexical contexts. Moreover, words are always connected to meanings and thus learning to perceive or produce the correct phonemes of words is usually done in the context of learning the words themselves, and their meanings, which ultimately will serve a communicative function.

The present study therefore focuses on the role of resting-state connectivity in learning non-native sounds embedded in words which were learned with meanings. Participants received a lexical training task which indirectly required them to learn to identify non-native sounds. More specifically, they were trained to match auditory-presented words to meaning (pictures of everyday items). The words differed minimally in non-native tone contours modeled after Mandarin tones. In order to map the words to their meanings successfully, participants had to learn to discriminate and identify the non-native tones. We used resting-state fMRI to measure connectivity before and after participants received the lexical training. Our study can be considered complementary to that of Ventura-Campos et al. (2013) as it takes it a step further by looking at the interface between phonetic/phonological, lexical, and semantic processing.

We trained individuals in a suprasegmental contrast. Lexical tones are acoustically different from consonantal contrasts as they rely on pitch differences. These spectral changes recruit right hemisphere auditory cortex alongside left homologue areas, as well as left frontal language areas (Luo et al., 2006; Warrier et al., 2009; Zatorre & Gandour, 2008). As already mentioned, tone learning introduces increased activity in the left IFG and the bilateral Superior Temporal Gyri (STG) (Wong, Perrachione, et al., 2007). In addition, differences in grey matter volume in the left Heschl's Gyrus (HG) are associated with differences in tone learning performance (Wong et al., 2008). These findings illustrate the joint contributions of auditory perception and higher order language processes in tone learning. We therefore chose these frontal and temporal regions as seeds in a Seed Correlation Analysis (SCA) in which we investigated their resting-state connectivity with all other voxels in the brain.

We hypothesized that resting-state connectivity between frontal and temporal regions would be associated with the individual's tone learning performance. These areas are connected anatomically via dorsal and ventral white matter fiber connections (Gierhan, 2013; Hickok & Poeppel, 2007; Saur et al., 2008) which have been shown to contribute to individual variation in language learning abilities (Lopez-Barroso et al., 2011; López-Barroso et al., 2013; Wong et al., 2011). Although resting-state functional connectivity is not necessarily directly mapped to anatomical connectivity, we anticipated the involvement of areas along these pathways. Given the fact that our training paradigm required participants to match non-native sound contrasts to meanings, we expected to find resting-state connectivity differences primarily in the ventral pathway for language in successful compared to less successful learners.

Methods

Subjects

Forty Dutch native speakers took part in the study (15 males, mean age=22.62, SD=3.16). Participants were recruited from the Radboud University Nijmegen and Max Planck Institute for Psycholinguistics databases and had no history of neurological disorders. All participants gave written informed consent prior to the experiment (as approved by the ethics committee, CMO region Arnhem–Nijmegen, The Netherlands) and were compensated monetarily or with course credit for their participation. We excluded one participant from the imaging analyses (due to an incidentally found brain anomaly as assessed by a radiologist).

Stimuli

Participants were trained on 24 Dutch-Chinese hybrid words. These hybrids words (hereafter “Dutchinese”) were based on Dutch monosyllabic pseudowords (e.g. “baaf”). Each of these pseudowords had each of the four Mandarin Chinese tone contours [level Tone 1 (T1), rising Tone 2 (T2), dipping Tone 3 (T3), falling Tone 4 (T4)] superimposed on it (e.g. “baafT1”/ “baafT2”/ “baafT3”/ “baafT4”). The stimuli were created automatically by superimposing the Mandarin pitch contours uttered by four Mandarin native speakers (2 women) on the Dutch pseudowords uttered by four Dutch native speakers (2 women) using the Functional Data Analysis (FDA) method for speech analysis and re-synthesis [http://lands.let.ru.nl/FDA/index.htm (Gubian, 2011.)]. A total of six pseudowords with a Consonant-Vowel-Consonant (CVC) structure were used to create the 24 training words in 6 quadruplets (6 CVCs x 4 tone contours). By using hybrid stimuli we were able to control a number of acoustic variables (e.g. word duration, intensity, vowel length, production rate etc.) which were kept constant within the minimal quadruplets. The non-native pitch contour was therefore the only acoustic information available in the stimuli for participants to distinguish words within each quadruplet. The Dutchinese stimuli were rated by native Mandarin speakers to ensure that the tones were identifiable as the intended Mandarin tones.

Table 1

Transcriptions and meanings of the Hybrid Words Used in the Dutchinese Training

Dutch CVC	Pictures
baaf	duck, foot, bowl, chair
din	sheep, flower, pants, comb
jor	cow, tree, pen, knife
moep	fish, bicycle, glasses, train
nuuk	arm, house, belt, bus
wum	cat, leaf, shirt, door

Dutchinese Training

We used Neurobehavioral Systems Presentation software (www.neurobs.com) for stimulus presentation and response recording. The Dutchinese training (after Chandrasekaran et al., 2010) consisted of five separate training-testing sessions during which participants had to learn to associate 24 auditory-presented Dutchinese words with pictures of everyday items. In the training part, participants were presented with one picture at a time on a computer screen while they heard its Dutchinese name over headphones. During this part, the presentation was blocked per CVC (6 CVC = 6 blocks) and sub-blocked per different speaker in order to facilitate learning. As part of the training, after each block, participants performed a mini-quiz consisting of 16 trials each. They heard one word at a time, while they were presented with the four pictures on the screen, and had to click with the mouse on the correct picture corresponding to the word. Upon clicking one of the pictures, they would hear the word again and get visual feedback on their response (either the printed word “correct” if they were right or the correct picture if they were wrong) (Figure 1). Each training part included a total of 192 training trials (6 CVC x 4 tones x 4 speakers x 2 repetitions). The training data were not analyzed.

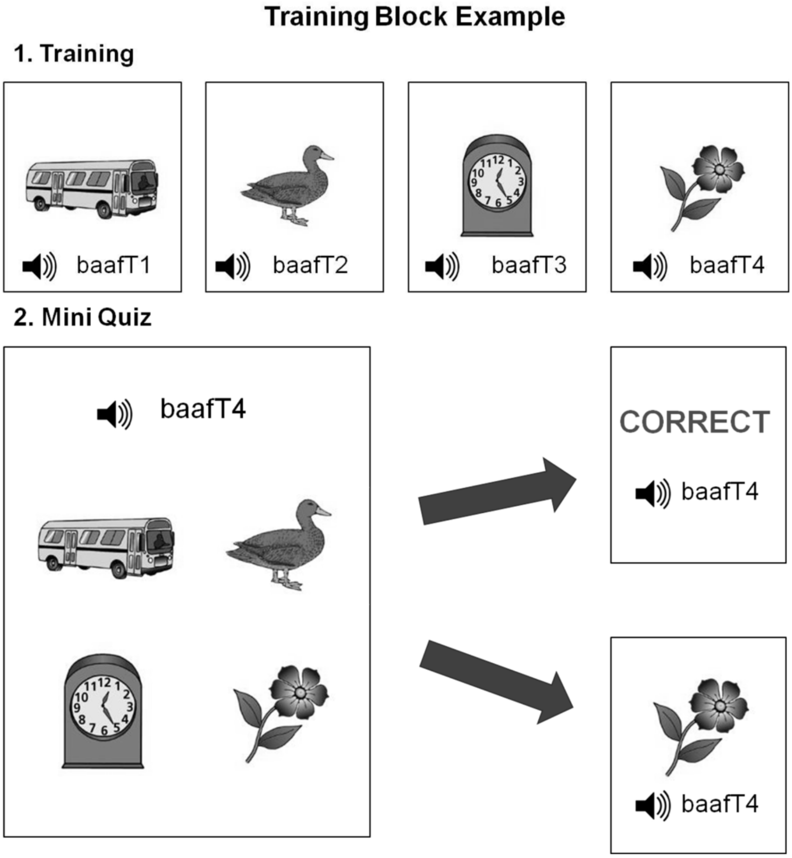


Figure 1. Example of a Dutchinese training block in which the participant is asked to learn the association between words in the minimal quadruplet baafT1, baafT2, baafT3 and baafT4 and their matching pictures.

Within each session, the training phase was always followed by a testing phase. During the latter phase, participants heard one word at a time while they were presented with the full set of twenty-four words presented on the screen and had to click on the matching one. The total number of trials was 96 (6 CVC x 4 tones x 4 speakers) and no feedback was provided throughout this part. In the final session (Session 5), participants performed a generalization test, which was identical to the regular testing phase with the exception that the Dutchinese speakers uttering the words were new (i.e., the other four hybrid speakers). The purpose of using new speakers was to test whether participants had created abstract representations of the words instead of having learned the specific acoustic tokens they were trained on. Participants' response accuracy was recorded (percentage of correct picture-word matches). Following previous studies that used the same type of paradigm (see Chandrasekaran et al., 2012, 2010; Wong et al., 2011), we took accuracy in the final generalization test as participants' final learning attainment. Each training-testing session lasted around 30 minutes in total.

Tone Discrimination and Identification tasks

Participants completed two tone perception tasks prior to training initiation and after training completion (designed after Chandrasekaran et al., 2010). The purpose of these tasks was to ensure that the lexical training indeed trained participants in the non-native tone contrasts instead of just tapping into simple associative learning abilities.

In the tone discrimination task, participants listened to minimal pairs of Dutchinese words and had to report whether or not the words differed in tone. The pairs were CVC words chosen from six minimal tone quadruplets and were different from the ones participants were trained on. All the words were uttered by the same female Dutchinese hybrid speaker so that the only acoustic difference between a pair was the pitch contour. The words were presented using in-house software through headphones with 500 ms ISI and participants were instructed to press one of two buttons on a button box as soon as they had made their same-different decision. The task included eight practice trials with feedback in the beginning and 144 test trials including all possible combinations of tones. Button and trial order were counterbalanced across participants. Response accuracy was recorded.

In the tone identification task, participants listened to single Dutchinese words and had to indicate the direction of the pitch contour in the word. There were three possible directions: upwards (indicated by an upward pointing arrow), downwards (indicated by a downward pointing arrow) and flat (indicated by a horizontal arrow). The words used in this task were different from the ones used in the discrimination and training tasks and consisted of five CVC words uttered by one female and two male speakers. After a fixation cross, the word was presented through headphones together with the three arrows presented on the screen. Participants were instructed to listen carefully and click the button corresponding to the correct arrow. The task included 18 practice trials with feedback and 135 test trials. Response accuracy was recorded.

Control tasks

Since learning abilities may be influenced by general intelligence and memory abilities, we administered two control tasks to assess these abilities in our sample. We used Raven's Advanced Progressive Matrices Test (1998 Edition, set II) to assess non-verbal general intelligence, and the Backward Digit Span subtest adapted from the Dutch version of the Wechsler Adult Intelligence Scale (WAIS) to assess working memory. Participants were also asked to fill out a post-study questionnaire about their language and music background as well as their motivation and the learning strategies used during the training.

Procedure

The experiment consisted of seven testing sessions on seven different days (with maximum three days between sessions): two MRI sessions and five behavioral training sessions (see Figure 2). Resting-state data were acquired at two time points: pre-training (before initiation of training) and post-training (shortly after training completion). For the resting-state fMRI sessions, participants were lying comfortably in the scanner, and were instructed to keep their eyes open and try not to engage in any structured mental operation during scanning. The resting-state protocol lasted 9 minutes and always preceded task-based fMRI sequences (the results of which are reported elsewhere).

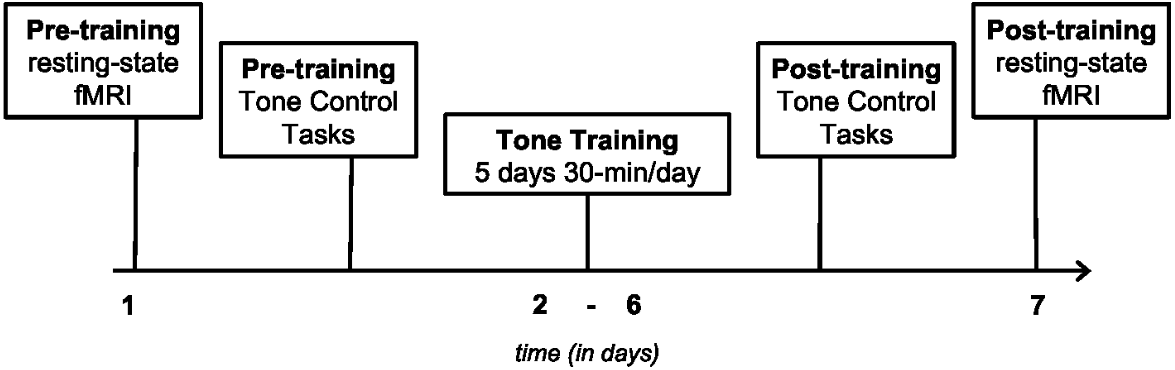


Figure 2. Illustration of the experimental procedure from the Day1 session until the Day7 session. Tone control tasks were administered on Day 2 (the first tone training day) and Day 6 (the last training day).

MR Imaging

Resting-state data were acquired on a Siemens 3T MAGNETOM Trio Tim MR system (Siemens Healthcare, Erlangen, Germany) using a 32-channel head-coil. A multi-echo EPI sequence was used with the following parameters: repetition time (TR) = 2000 ms, echo times (TE): TE = {6.9 ,16.2 ms, 25 ms, 35 ms, 44 ms}, 39 axial slices per volume in interleaved manner and ascending order, 0.51 mm slice gap, 3mm slice thickness, FOV = 224 mm2, matrix size = 64 x 64, and isotropic voxel size of 3.5 x 3.5 x 3.0 mm^3, flip angle= 80°, bandwidth= 2520 Hz/Px, accelerated with GRAPPA parallel imaging (accelerating factor 3). We also acquired a high resolution T1-weighted anatomical image using a magnetization prepared rapid gradient echo sequence (TR: 2.3 s, TE 3.03 ms, 8° flip angle) with 192 slices, 1.0 x 1.0 x 1.0 mm^3 voxel size, 256 mm FOV and matrix size 256 x 256, accelerated with GRAPPA parallel imaging (accelerating factor 2). We used multi-echo fMRI because it reliably increases the temporal signal-to-noise ratio and reduces distortions in ventral regions of the brain (Poser et al., 2006).

Behavioral Analyses

We used the IBM SPSS 19 statistical package to carry out the behavioral data analyses. Participants' response accuracy in matching the Dutchinese words to their corresponding pictures in the testing phase at the end of each day of training was analyzed with a repeated-measures ANOVA, using Session (x 5) as factor and percentage correct as the dependent variable. All post-hoc pairwise comparisons were Bonferroni corrected. Performance in the tone perception tasks was analyzed with paired-sample T-tests, comparing mean response accuracy (percentage correct) before and after training. Pairwise correlations between the final learning score (pictures identified in the generalization test at the end of the last day of training) and the tone perception tasks, as well as the general control tasks, music training duration, and motivation were estimated.

fMRI Analysis

Preprocessing

The five echoes per TR were combined using an in-house toolbox developed in Matlab (MathWorks) implementing the echo-weighting procedure described by Poser et al. (2006). First, all volumes of the first echo volumes were realigned to the first volume of the first echo and all the remaining volumes were subsequently realigned to the first echo and resliced. The first 30 acquired volumes were smoothed with a 3mm Gaussian kernel and were used to calculate the optimal echo-weighting parameters (optimal contrast to noise ratio) for combining the echoes. The resulting weighting parameters were used to combine the echoes in the remaining volumes.

Further image preprocessing was performed in SPM8 (www.fil.ion.ucl.ac.uk) where the mean functional EPI image was co-registered to the T1 weighted anatomical image using normalized mutual information, and the registration parameters were subsequently applied to all the combined functional EPI images. We segmented the T1 image into grey and white matter and cerebrospinal fluid. We used the normalization parameters from segmentation for normalization and transformation of all images into common standard MNI space (2x2x2 voxel size). Finally, all the images were convolved with a Gaussian smoothing kernel of full-width 8mm at half maximum.

Resting-state fMRI statistics

We used a seed-based correlation analysis (Biswal, Yetkin, Haughton, & Hyde, 1995) with *a priori* defined ROIs as seed regions. Seed regions were defined by using the AAL template (Tzourio-Mazoyer et al., 2002) provided by WFU PickAtlas toolbox (Maldjian et al., 2003) and included: the left IFG [the Pars Opercularis (POp), Pars Triangularis (PTr), Pars Orbitalis (POrb)], the Superior Temporal Gyrus (STG) and Heschl's Gyrus (HG) in both hemispheres. The mean BOLD activation time-course of voxels in these seed regions was extracted along with the mean time-course of white matter and cerebrospinal fluid. Since low frequency fluctuations dominate the resting- state signal, Butterworth low-pass filtering (0.005-0.1 Hz) was applied to the data. The first-level GLMs included the mean time-course of the seed region voxels as regressors of interest, and the physiological noise and realignment parameters as nuisance regressors. Linear correlations between seed region voxels and all other voxels in the brain were calculated in SPM8 for each participant and each session (pre- and post-training). The mean connectivity difference between sessions (post-training > pre-training session) was also estimated for each seed region. The connectivity strengths (mean beta coefficients) for each participant in each seed region were averaged across subjects and entered into the second-level random effects analysis. Multiple regressions were performed including the final learning attainment scores of each participant as a covariate in order to investigate how functional connectivity before and after training related to variability in learning performance. Results were corrected for multiple comparisons using family wise error rate (FWE) correction ($p <.05$) (Hayasaka & Nichols, 2003). Four participants couldn't complete the second scanning session due to technical problems and hence the post-training session and session difference analyses were performed on the remaining sample (N=35).

Results

Behavioral Results

The behavioral analysis on participants' learning scores (percentage of pictures correctly identified in the testing phases) revealed a significant effect of Session [F(1.567, 61.127)=99.247, p<.001 (Greenhouse-Geisser corrected), partial η^2=.718]. Participants improved significantly over the course of the five training sessions, as indicated by the post-hoc comparisons (see Figure 3A). Analysis of participants' performance in the tone perception tasks demonstrated that they improved significantly in Pitch Discrimination [t(38)=-4.703, p<.001] and Pitch Identification [t(39)=-4.635, p<.001] after training (see Table 2 for descriptive statistics). Participants' final Dutchinese learning attainment score (generalization in training session 6) correlated significantly with their pre- training Pitch Discrimination and Identification accuracy performance (r=.539, p<.001 and r=.740, p<.001 respectively) but not with general control measures (Backward Digit Span score, Raven's score, music education duration, music education onset and self-reported motivation – see Table 3). Despite the fact that on average participants improved in non-native tone perception, we observed large individual variation in their learning curves (see Figure 3B).

Table 2.

Paired T-tests on Tone Discrimination and Identification Accuracy (percentage correct).

	Pre-training		Post-training			95% CI for Mean			
Accuracy	M	SD	M	SD	n	Difference	r	t	df
Discrimination	93.46	6.09	95.74	5.06	39	-3.26, -1.29	.86*	-4.70*	38
Identification	60.22	19.61	67.65	24.90	39	-10.66, -4.18	.92*	-4.63*	38

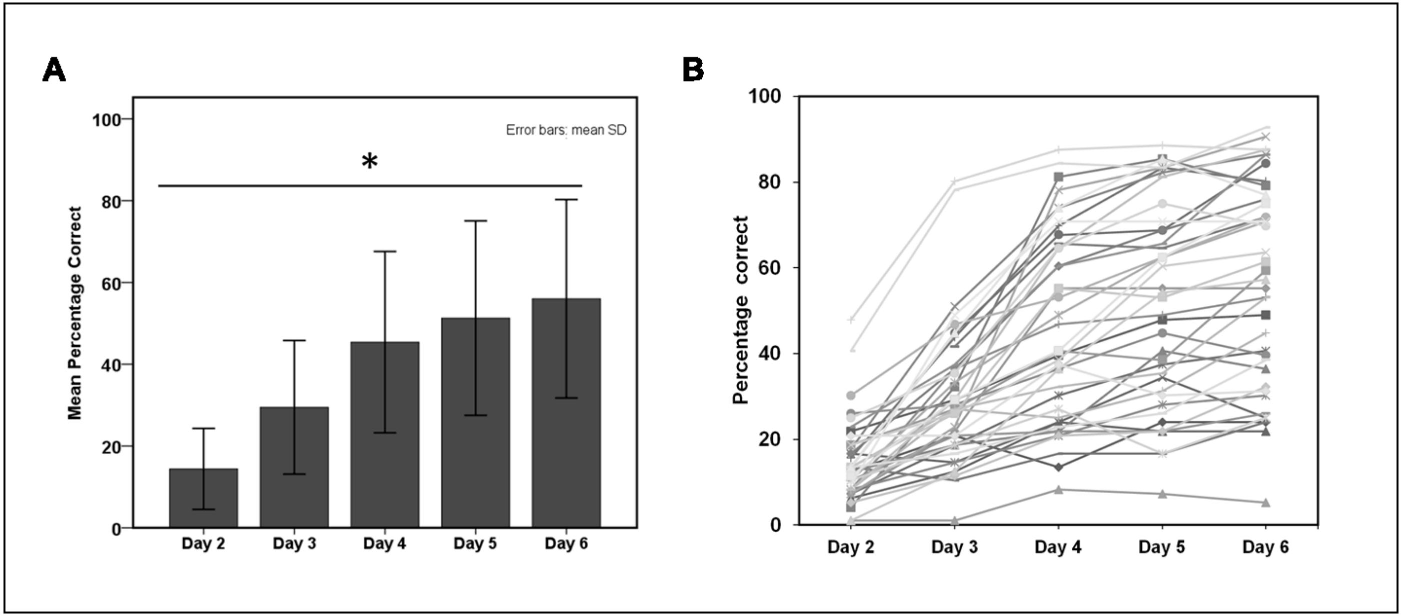
Note. * p<.001

Table 3.

Correlations between Participants' Final Learning Score and Control Measures

Measure	Backward DS	Raven	Music	Onset Music	Motivation
Learning Score	.137	.190	.254	.208	.295
Backward DS		.082	-.068	.090	.042
Raven			-.051	-.299	-.052
Music Edu				.661**	.096
Onset Music					-.082

Note. **p<.001



Note. *p<.005, Bonferroni corrected for multiple comparisons.

Figure 3. Behavioral learning effects. A. Bars representing mean percentage correct word – picture mappings per training session. B. Individual and mean learning scores (percentage of pictures correctly identified from their Dutchinese names) over the five training sessions.

Imaging Results

Main connectivity effects

The connectivity analysis revealed clusters whose activity time-course correlated significantly with our seed regions (see Figures 4, 5). These connectivity patterns were derived while controlling for the effect of the learning covariate (generalization score in the final training session) in the multiple regression models. The general connectivity patterns for each seed in the pre-training and post-training session are summarized in Supplementary Material Tables 1-14

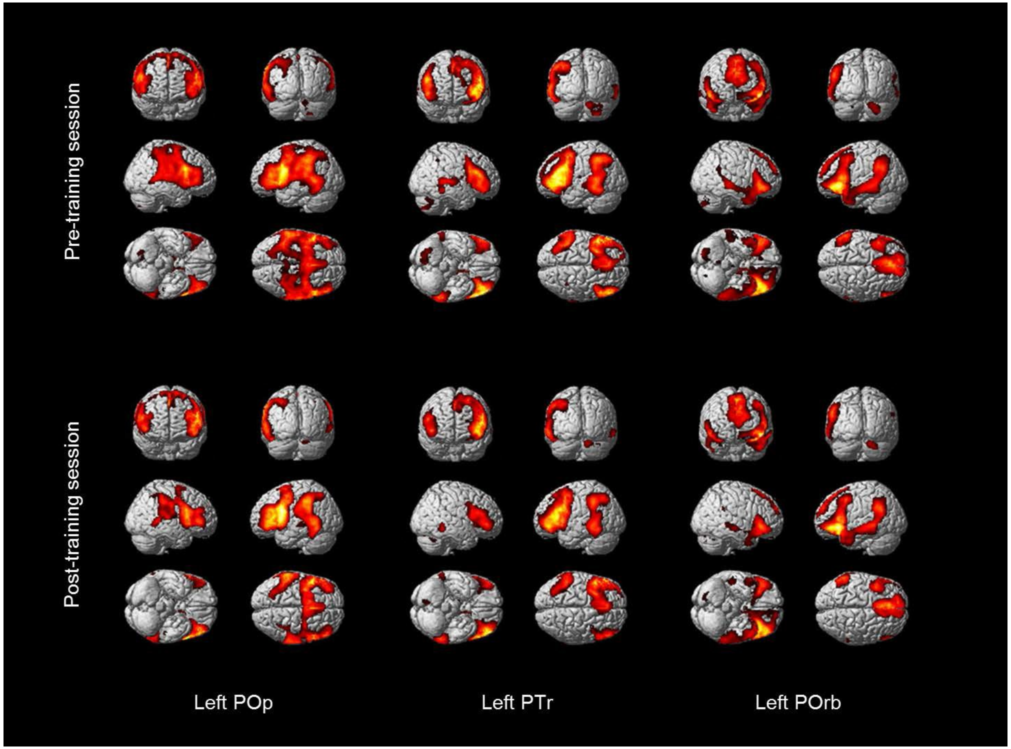


Figure 4. Areas correlating positively with the time-course of BOLD activity in the left Pars Opercularis (POp), the left Pars Triangularis (PTr) and the left Pars Orbitalis (POrb) seed regions in the pre-training and post-training fMRI sessions (p<.05, FEW corrected).

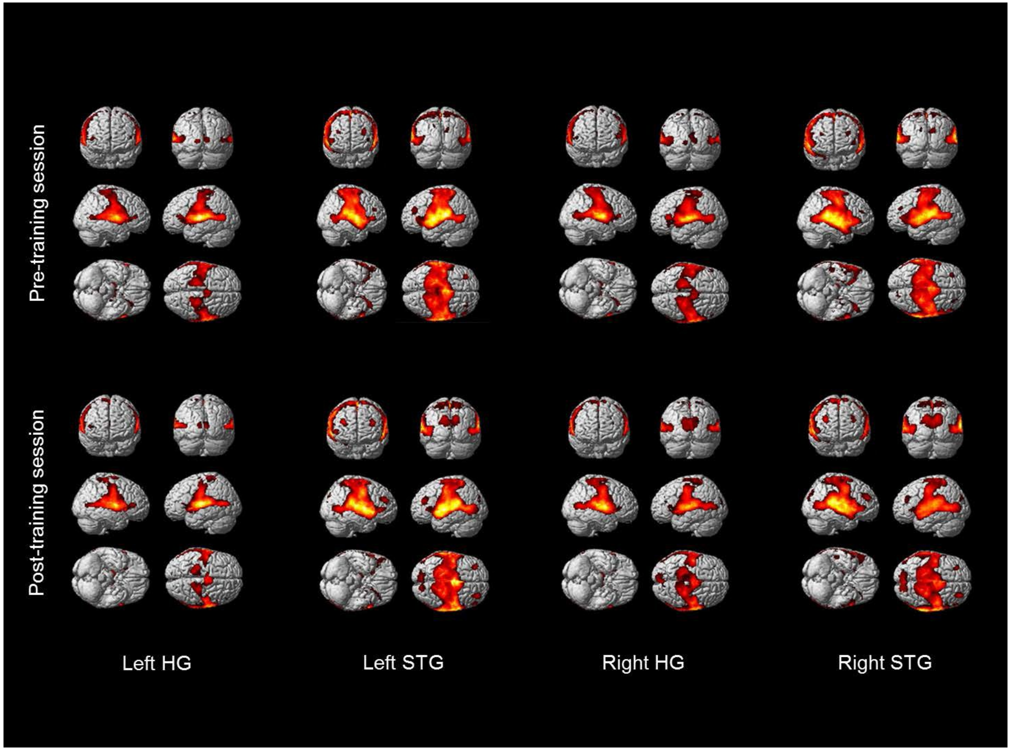


Figure 5. Areas correlating positively with the time-course of BOLD activity in the left and the right Heschl's Gyrus (HG), and the left and the right Superior Temporal Gyrus (STG) seed regions in the pre-training and post-training fMRI sessions (p<.05, FWE corrected).

With respect to the effect of the covariate in explaining variation in connectivity patterns, we found no significant correlation between resting- state connectivity in either the pre-training, or post-training session and learning attainment. At a more lenient threshold of uncorrected p<.001, FWE corrected at the cluster level, however, we found that individual learning attainment correlated negatively with the strength of connectivity in the post- training session between the left POrb and a cluster in the left posterior STG/ Angular Gyrus (see Figure 6).

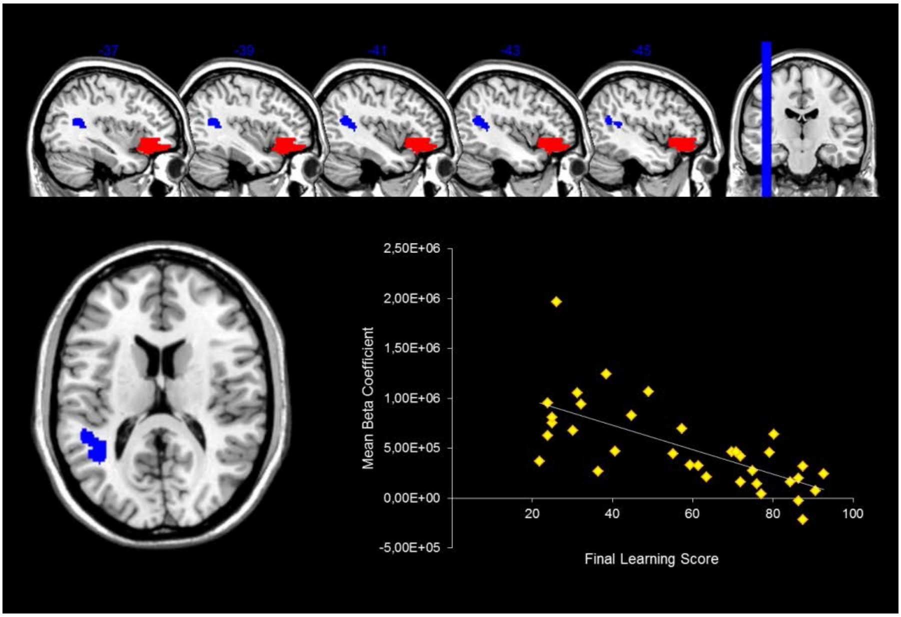


Figure 6. Sagittal multi-slice view of the left Pars Orbitalis seed region (red) and a cluster in the left STG/AG (blue) (also in axial view). The scatter plot illustrates the connectivity strength between the two regions in the post- training fMRI session (mean Beta Coefficient) and its negative correlation with the participants' final learning score (generalization score in the final training session).

Main connectivity differences (post-training > pre-training)

Apart from the general connectivity patterns tested in the pre and post-training sessions, we also ran multiple regressions using the mean difference in connectivity strength over training in each seed region (post-training > pre- training mean beta coefficients) as the dependent measure and with the final learning attainment score as a covariate. No seed region showed changes in connectivity after training at a significance threshold of p<.05, FWE corrected. However, when adopting a more lenient threshold of uncorrected p<.001, FWE corrected at the cluster level, three seed regions showed changes in connectivity in the post compared to the pre-training session. The most significant change in connectivity was observed in the right HG which showed increased connectivity with a cluster in the right Postcentral Gyrus, the right Rolandic Operculum and right STG, in the post-training compared to the pre-training session. The left STG showed increased connectivity with the right primary somatosensory cortex in the Postcentral Gyrus. The right STG showed more connectivity with the bilateral cunei in the occipital lobe after training. Table 4 and Figure 7 summarize the connectivity difference findings. There was no seed region showing decreased connectivity strength with training (i.e. pre-training > post-training).

Table 4.

Areas Showing Significant Connectivity Differences Over Training (post-training > pre training) (uncor. p<.001, FWE cluster corrected).

Seed	Region	Hemisphere	Cluster p(FWE)	Cluster size	T	Z	MNI Coordinates x,y,z {mm}
Left STG							
	Rolandic	R	0.001	440	5.16	4.39	52 -6 18
	Operculum				4.86	4.19	60 -10 24
	Postcentral Gyrus				4.47	3.92	58 0 20
	Postcentral Gyrus						
Right HG							
	Postcentral Gyrus	R	0.000	479	4.76	4.12	60 -14 22
	Rolandic		(4,11E+00)		4.58	4.00	48 -22 24
	Operculum				4.21	3.74	64 -36 18
	Superior Temporal Lobe						
Right STG							
	Cuneus	R	0.001	513	4.82	4.16	12 -78 40
	Cuneus	L			4.39	3.86	-12 -78 34

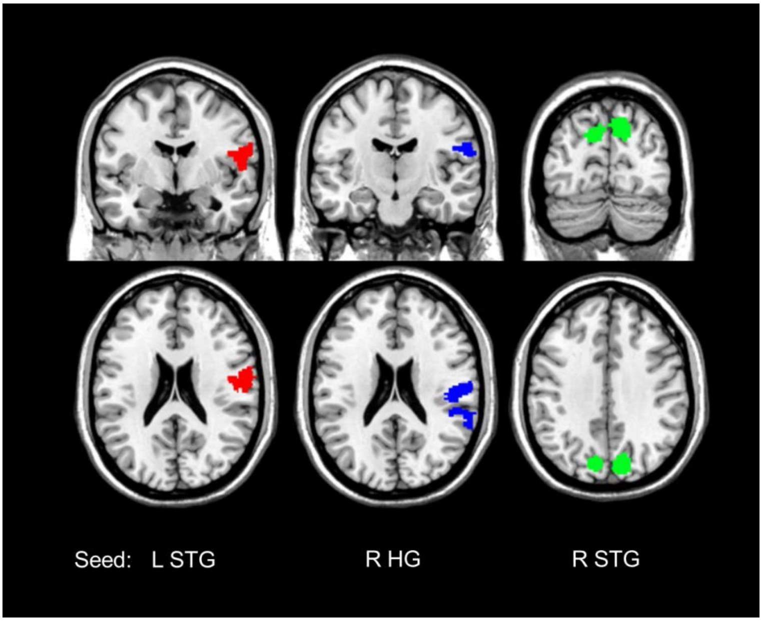


Figure 7. Coronal and axial views of clusters showing significant connectivity differences over training (post-training > pre training) with seed regions in the left STG (red), right HG (blue) and right STG (green) (uncor. p<.001, FWE cluster corrected).

We also tested whether the change in connectivity strength from pre- to post-training (post-training > pre-training) was related to final learning attainment (final learning scores). We found a negative correlation between learning attainment and the difference in connectivity between the right STG and the right MTG, and the difference in connectivity between the right HG and the right POP (see Figure 8) (Table 5 gives the exact MNI coordinates, p values, t and z scores, for the significant clusters).

Table 5.

Areas Showing Negative Correlation with Connectivity Difference in Post>Pre Training Session

Seed	Region	Hemisphere	Cluster p(FWE)	Cluster size	T	Z	MNI Coordinates x,y,z {mm}
Right HG	Pars Opercularis	R	0.047	185	4.89	4.21	60 14 22
	Precentral				4.79	4.14	56 8 28
Right STG	Middle Temporal	R	0.034	222	5.24	4.44	68 -44 -4
	Lobe				4.72	4.09	68 -36 0

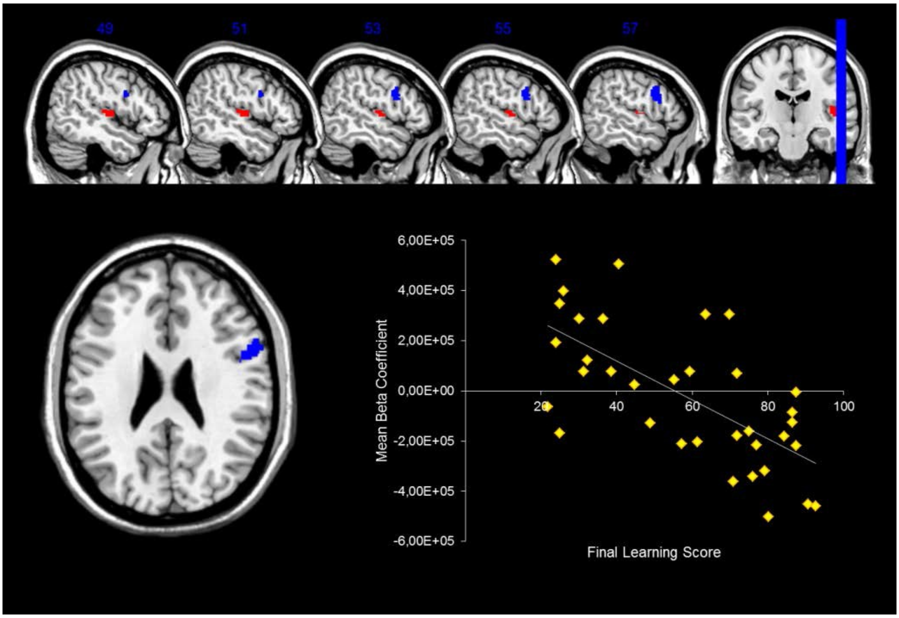


Figure 8. Sagittal multi-slice view of the right Heschl's Gyrus seed region (red) and a cluster in the right Pars Opercularis/ Precentral Gyrus (blue) (also in axial view). The scatter plot illustrates the difference in connectivity strength between the two regions in the post vs. pre-training fMRI session (mean Beta Coefficient) and its negative correlation with the participants' final learning score.

Discussion

In the present study we investigated the contributions of resting-state connectivity to individual variation in the ability to map non-native sounds onto meaning. Although on average all participants improved in matching the non-native words to their meaning, not everyone benefited from training with as many as half of the participants having a poor final learning score (i.e., scoring less than 50% correct on the final training session). This individual variation was reflected in the pattern of spontaneous BOLD fluctuations in the brain, as revealed in resting-state fMRI analyses. There were two main findings. The first one is that participants' final learning scores were associated with reduced resting-state connectivity between the left POrb and the left pSTG/AG in the post-training measurement. That is, after having learned the words, more successful learners showed less connectivity between left frontal and temporo-parietal areas than less successful learners. The second major finding is that changes in connectivity strength in the post vs. pre-training session between the right HG and the right POp/ Precentral gyrus correlated negatively with learning attainment. That is, more successful learners showed decreased connectivity in this right temporo-frontal pathway after training compared to before.

Reduced connectivity between the left POrb and left STG/AG with training in successful learners

The contribution of resting-state connectivity between the left POrb and left STG/ AG areas in word learning performance reported here is consistent with the previous literature on these areas. Both the left POrb and left AG have been repeatedly found to be involved in semantic processing (Binder, Desai, Graves, & Conant, 2009; Hagoort & Indefrey, 2014; Hagoort, 2014; Hope et al., 2014; Price, 2012). A systematic review of PET and fMRI studies on language has shown that the POrb is consistently activated for the retrieval of semantic concepts regardless of language modality (spoken or written) while the posterior STG is activated for auditory processing of words (Price, 2012). Evidence for functional connectivity between the left POrb and left pSTG/AG comes from a study that looked at resting-state data from 970 participants across different institutes (Tomasi & Volkow, 2012). The analysis resulted in the identification of four optimal partitions or modules comprising the language network which were highly reproducible across institutions. Of relevant interest for the present study is the finding that the left POrb is part of the speech comprehension module (or "Wernicke's module") along with other cortical areas that show strong coupling with Wernicke's area (in this study a voxel at the intersection of Brodmann Areas 39, 40 and 20) (Tomasi & Volkow, 2012). The finding is also in agreement with the nature of our training, which required participants to discriminate and identify different pitch contours embedded in novel spoken words in order to access their meaning. Although our results do not allow us to draw any conclusions with regards to the directionality of information flow from one area to the other, they do demonstrate that communication between these areas plays a role when learning to map non-native sounds to meaning.

This resting-state connectivity pattern matches what is known about structural brain connectivity. The left POrb and left pSTG/AG have been described as belonging to the ventral pathway for language, responsible for matching sounds to meaning(Hickok & Poeppel, 2007; Saur et al., 2008). It has been suggested that they are structurally connected via the Extreme Capsule (EmC), a white matter fiber tract which is part of the Inferior Fronto-Occipital Fasciculus (IFOF) (Saur et al., 2008). A previous study that used a very similar sound-to-meaning training paradigm as the one used here has found a positive correlation between word learning performance and Fractional Anisotropy (FA) in a left temporo-parietal cluster very close to our AG cluster (MNI coordinates: −34 −51 24) which is connected to the left IFG via the EmC (Wong et al., 2011). Thus, when participants have to associate the phonological forms of new words with semantic content, they appear to recruit the ventral route and not the dorsal route, which is important when only the phonological forms of words need to be learnt (López-Barroso et al., 2013). Here we report for the first time evidence that functional connectivity in the ventral pathway at rest plays a role in how well an individual can learn to match non-native sounds to meaning. Since resting-state connectivity has been shown to some extent to reflect anatomical connectivity (van den Heuvel, Mandl, Kahn, & Hulshoff Pol, 2009), we may speculate that individual variability in the degree of functional connectivity is indicative of the variability in the underlying white matter connectivity between these areas.

The negative correlation between learning and connectivity in this language pathway is in agreement with Ventura-Campos et al. (2013), who also report that decrease in left fronto-parietal connectivity correlated with learning a phonetic contrast. Despite the fact that we did not find significant connectivity changes induced by training between these areas, the negative correlation between connectivity and learning after training completion can be interpreted as reduced need for top-down control. It can be speculated that less-successful learners needed more guidance from frontal semantic decision areas in order to match the non-native words to their meaning. More research is needed, however, in order to understand the role of spontaneous fluctuations in language learning.

Post vs. pre-training changes in connectivity strength and tone learning

When looking at overall connectivity changes, we observed increase in connectivity strength between the auditory seed regions (the right HG and bilateral STG) and areas belonging to sensory (Postcentral gyrus), auditory (STG), and visual (cunei) resting-state networks (Damoiseaux et al., 2006; Smith et al., 2009). Furthermore, changes in connectivity in the two right hemisphere auditory seed regions (HG and STG) correlated negatively with learning attainment. More specifically, we found that changes in connectivity between the right primary auditory cortex (HG) and the right POp and right Precentral gyrus and in connectivity between the right STG and right MTG correlated negatively with tone learning performance. It is difficult to interpret these results in the absence of an overall significant session difference. However, based on evidence for the engagement of the right IFG in less-successful learners of tone (Wang et al., 2003;Wong, Perrachione, et al., 2007) and its involvement in non-lexical pitch processing (Grimault et al., 2014; Zatorre, Evans, Meyer, & Gjedde, 1992), we hypothesize that less successful learners in our study continued to engage a less effective network more than more successful learners. The fact that we find involvement of the right HG (as opposed to left HG) is not surprising given the fact that we used tones, which preferentially engage the right auditory cortex (Luo et al., 2006; Warrier et al., 2009; Zatorre & Gandour, 2008). Native-like tone processing should in contrast recruit the left IFG and not its right homologue (Nan & Friederici, 2012; Zatorre & Gandour, 2008). Affective prosodic processing has also been shown to engage preferentially the right IFG compared to linguistic prosodic processing which engages primarily the left IFG (Ethofer, Pourtois, & Wildgruber, 2006).

The connectivity pattern found between the right primary auditory cortex and the right POp/ Precentral Gyrus could reflect a right-lateralized dorsal pathway. A similar pathway has been shown to function as a compensation mechanism in young individuals with dyslexia (Hoeft et al., 2011). In particular, activation in the right POp has been found to be positively correlated with reading performance in children with dyslexia, who develop an atypical right AF to compensate for their impoverished phonological/print processing resources (Hoeft et al., 2011). Such correlations between literacy skills in dyslexia and right AF do not hold in adults, who have already learned how to read (Vandermosten et al., 2012) and support the idea that it is mainly recruited during learning. Less successful learners of the non-native pitch patterns in our study relied more on this right temporo- frontal pathway in the post-training compared to the pre-training session.

Given the nature of the training we used, we anticipated that connectivity patterns between auditory and frontal areas would play a role in learning performance. However, the finding that learning attainment is correlated negatively with connectivity changes between the right STG and the right MTG was less expected. According to this finding, successful learners of tone showed decreased connectivity between these areas, again, without an overall main effect of connectivity change in this connection from the pre to the post-training session. Based on evidence from neuropsychological, fMRI and direct cortical stimulation studies, Hickok and Poeppel (2007) have argued that the bilateral superior temporal areas are involved in phonological processing while the bilateral posterior middle temporal gyri are serving as an interface between lexical and semantic processing. The posterior MTG is therefore part of the ventral pathway: it receives phonological input from the STG and passes output to lexical and, further on, to semantic levels of processing (Hickok & Poeppel, 2007). Instead of a bilateral connectivity effect, however, here we report right-lateralized connectivity. This lateralization pattern is again probably related to the acoustic properties of the non-native tones. Interestingly, the right MTG is recruited during emotional word processing (Ethofer et al., 2009) and during the explicit evaluation of affective prosody (Ethofer et al., 2006; Wildgruber et al., 2005). Since affective prosody and emotional expressions rely on non- linguistic acoustic cues, including pitch, it is possible that less successful learners of tone kept processing tones non-linguistically by forwarding phonological or acoustic information from the right STG to the right MTG. Successful learners, in contrast, may have stopped relying on such an ineffective strategy with training. Although this is a plausible interpretation of the association between tone learning performance and right STG - MTG connectivity change, further studies explicitly testing the role of connectivity between these temporal areas in speech processing and non-native speech learning are warranted.

Study limitations

There are several limitations in the interpretability of our study. We report relatively low correlation effects between learning performance and resting-state connectivity. These should be interpreted in the light of our analysis methodology. Instead of using a seed-to-seed correlation analysis, we opted for SCA, a less biased approach that looks at the correlation between the mean time-course of the seed region and the mean time-course of all voxels in the brain. In the absence of a relevant fMRI task, we had to use anatomical ROIs as seed regions instead of functional ROIs; the use of functionally-defined regions could have increased sensitivity in our analysis. In addition, it is possible that the duration of the training may not have been sufficient for major connectivity changes to occur. Our training protocol consisted of five sessions of half an hour duration which was sufficient for individual variation to arise but not for all learners to reach ceiling performance. That would explain the fact that the significant overall changes in connectivity we found with training consisted primarily of increased connectivity between auditory, sensory and visual networks. Increased communication between networks that process sensory input would be a first, basic change induced by our training, which required listening to sounds, matching them to pictures, and pressing keys. The behavioral results indicate that only some participants succeeded in learning the non-native words in such a short period of training. Large individual variation in connectivity may have resulted in the absence of overall connectivity changes, an assumption supported by the fact that connectivity changes correlated with learning performance.

Conclusions

Investigating learning with resting-state fMRI is a slowly but steadily developing line of research (Guerra-Carrillo et al., 2014). Resting-state connectivity is a potentially powerful tool that allows unique insights into the dynamic properties of the human brain (Park & Friston, 2013). Being in its infancy, however, there is still a lot to be learnt about how resting-state connectivity changes with experience and how it contributes to individual variation in language learning performance. Here we investigated its role in individual variation in non-native sound learning performance using a lexical training paradigm. The paradigm required participants to learn to match non- native sounds embedded in words to existing meanings. We found that individuals varied greatly in their learning performance and this variation resonated in the strength of resting-state connectivity patterns between frontal and temporal language areas. Participants' final learning attainment was negatively correlated with post-training connectivity between the left POrb and the left STG/AG, alluding to the ventral language stream responsible for mapping sounds to meaning. Furthermore, increase in the strength of connectivity in the post compared to the pre-training session between the right HG and the right POp/Precentral Gyrus correlated negatively with performance, which possibly points to the recruitment of a compensatory right temporo-frontal connection in less successful learners. In sum, our study provides evidence that resting-state connectivity changes introduced with non-native word training differ for successful compared to less successful learners.

Acknowledgements

We thank Michele Gubian for creating the hybrid Dutch – Mandarin stimuli, Wencui Zhou, Pieter van Groenestijn, Hubert F. J. M. Voogd, Pascal de Water, and Paul Gaalman for their technical assistance, and Atsuko Takashima for her assistance in fMRI data analysis and helpful discussion.

Supplementary Material

Table 1a.

Areas Showing Positive Correlation with the left Pars Opercularis in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Pars Opercularis	L	0.000	54421	0.000	27.02	Inf	-48 8 24
Pars Opercularis				0.000	26.35	Inf	-52 10 16
n.l.				0.000	23.29	Inf	-30 18 0
Cerebellum 8	R	0.000	85	0.000	7.80	5.96	24 -74 -50
Cerebellum 6	R	0.000	413	0.000	7.38	5.75	26 -62 -30
n.l.				0.001	7.13	5.62	18 -66 -36
Cerebellum Crus1				0.001	7.00	5.55	12 -74 -26
Vermis 3	R	0.001	36	0.006	6.33	5.18	0 -38 -12
n.l.	L	0.004	15	0.011	6.13	5.06	-12 -54 -30
Lingual Gyrus	L	0.000	84	0.014	6.03	5.00	-14 -66 0
Lingual Gyrus	R	0.004	16	0.029	5.75	4.83	18 -60 2
Cerebellum Crus1	R	0.034	1	0.035	5.67	4.78	42 -44 -36
Calcarine Sulcus	L	0.027	2	0.037	5.65	4.77	-14 -74 16
n.l.	L	0.034	1	0.042	5.60	4.74	-10 -50 -22

Table 1b.

Areas Showing Negative Correlation with the left Pars Opercularis in the Pre-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
Frontal Medial Orbital L	R	0.000	14485	0.000	12.83	Inf	4 42 -6
Precuneus				0.000	12.26	7.70	12 -40 8
n.l.				0.000	12.15	7.66	36 -48 24
Cerebellum Crus2	L	0.000	662	0.000	9.33	6.65	-46 -72 -38
Cerebellum Crus2				0.000	8.76	6.41	-48 -64 -42
Cerebellum Crus2				0.000	8.67	6.37	-32 -84 -34
n.l.	L	0.000	75	0.000	7.42	5.77	-2 -86 -20
n.l.	R	0.001	33	0.001	7.24	5.68	34 -66 6
Thalamus	R	0.006	12	0.001	6.90	5.50	2 -10 8
n.l.	R	0.006	12	0.003	6.56	5.31	2 -34 -24
Inferior Temporal L	R	0.000	138	0.006	6.38	5.21	60 -10 -26
Middle Temporal L				0.010	6.17	5.09	54 -2 -28
Cerebellum Crus2	R	0.004	15	0.010	6.14	5.07	46 -68 -40
Cerebellum Crus1	R	0.027	2	0.011	6.11	5.05	32 -84 -26
n.l.		0.016	5	0.019	5.91	4.93	0 -56 -50
Cerebellum Crus1	R	0.008	10	0.020	5.90	4.92	36 -82 -24

Cerebellum Crus1	R	0.022	3	0.022	5.85	4.89	46 -74 -24
Cerebellum Crus2	R	0.027	2	0.027	5.78	4.84	34 -80 -36
Precuneus	R	0.022	3	0.030	5.73	4.82	2 -70 56
n.l.		0.012	7	0.030	5.73	4.82	2 2 18
Parahippocampal Gyrus	L	0.034	1	0.032	5.70	4.80	-18 -10 -28
Cerebellum Crus1	R	0.034	1	0.043	5.59	4.73	12 -88 -18
n.l.		0.034	1	0.045	5.58	4.72	12 -36 -46
Vermis 6	R	0.034	1	0.046	5.57	4.72	2 -74 -8

Table 2a.

Areas Showing Positive Correlation with the left Pars Triangularis in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm Region}
Pars Triangularis	L	0.000	15973	0.000	26.02	Inf	-46 20 26
Pars Triangularis	L			0.000	25.50	Inf	-56 28 8
Pars Triangularis	L			0.000	24.37	Inf	-46 28 22
Pars Triangularis	R	0.000	5580	0.000	15.67	Inf	44 28 22
Pars Triangularis	R			0.000	14.44	Inf	54 20 26
Pars Triangularis	R			0.000	14.31	Inf	46 38 14

Inferior Temporal L	L	0.000	7465	0.000	15.30	Inf	-58 -42 -16
Inferior Parietal L	L			0.000	14.42	Inf	-46 -48 44
Mid Temporal L	L			0.000	13.47	Inf	-62 -46 -4
Caudate	R	0.000	3269	0.000	11.66	7.51	12 8 8
Mid Temporal L	R			0.000	9.09	6.55	52 -40 6
n.l.	R			0.000	9.04	6.53	32 -24 8
Cerebellum Crus 1	R	0.000	1118	0.000	9.65	6.78	12 -76 -28
Cerebellum Crus 1	R			0.000	9.51	6.72	34 -66 -32
Cerebellum Crus 1	R			0.000	8.50	6.29	24 -68 -32
n.l.		0.000	68	0.000	7.44	5.78	0 0 28
n.l.	R			0.014	6.02	5.00	6 8 26
Inferior Parietal L	R	0.000	58	0.004	6.48	5.26	36 -48 46
Rolandic Operculum	R	0.002	23	0.008	6.22	5.12	58 -4 14
Heschl's Gyrus	L	0.006	13	0.020	5.88	4.91	-50 -14 8
n.l.	L	0.016	5	0.025	5.81	4.86	-6 -26 -26
Superior Temporal L	R	0.034	1	0.039	5.63	4.75	58 -6 -2
Pars Orbitalis	L	0.034	1	0.046	5.57	4.71	-22 12 -24
Cerebellum Crus 2	L	0.034	1	0.047	5.56	4.71	-12 -76 -30
Superior Temporal L	R	0.034	1	0.048	5.55	4.70	56 -2 -4

Table 2b.

Areas Showing Negative Correlation with the left Pars Triangularis in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Caudate	R	0.000	3559	0.000	10.46	7.09	14 24 -6
n.l.				0.000	10.40	7.07	14 30 2
Olfactory L	R			0.000	9.87	6.87	4 20 -6
Lingual Gyrus	L	0.000	1560	0.000	10.18	6.99	-32 -46 -4
n.l.				0.000	10.16	6.98	-20 -34 12
Hippocampus				0.000	9.97	6.91	-36 -30 -10
n.l.	R	0.000	1639	0.000	9.75	6.82	38 -48 -4
Precuneus				0.000	9.59	6.75	14 -38 8
n.l.				0.000	8.84	6.44	32 -40 6
n.l.	R	0.003	20	0.000	8.05	6.08	2 -34 -24
Parahippocampal Gyrus	R	0.000	62	0.001	7.06	5.59	22 -26 -16
n.l.	R	0.000	47	0.001	6.95	5.53	34 -66 6

Cerebellum Crus 2	L	0.002	25	0.006	6.36	5.20	-48 -62 -42
Vermis 6		0.010	8	0.028	5.75	4.83	2 -72 -6

Table 3a.

Areas Showing Positive Correlation with the left Pars Orbitalis in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Pars Orbitalis	L	0.000	23484	0.000	30.04	Inf	-46 28 -8
Pars Orbitalis				0.000	28.71	Inf	-46 40 -10
Pars Orbitalis				0.000	25.69	Inf	-36 28 -12
Pars Orbitalis	R	0.000	5634	0.000	19.37	Inf	50 38 -10
Pars Triangularis				0.000	18.20	Inf	54 28 0
Pars Orbitalis				0.000	12.35	7.72	30 26 -16
Caudate	R	0.000	535	0.000	12.29	7.71	12 10 6
Cerebellum Crus 1	R	0.000	845	0.000	9.85	6.86	20 -76 -30
Cerebellum Crus 2				0.001	7.24	5.68	24 -82 -42
n.l.	L	0.000	82	0.000	8.48	6.28	-4 -22 -24
Cerebellum 9	R	0.000	44	0.000	7.51	5.82	6 -52 -40
Mid Cingulum	L	0.000	75	0.000	7.41	5.77	-4 -16 42

n.l.	L	0.000	69	0.001	7.07	5.59	-14 -44 -30
n.l.	R	0.001	31	0.004	6.54	5.30	34 -8 -12
Cerebellum Crus 1	L	0.006	12	0.017	5.96	4.96	-20 -76 -32
n.l.	R	0.019	4	0.017	5.94	4.95	18 -52 -34
n.l.	R	0.001	35	0.023	5.83	4.88	32 -24 8
Putamen	R	0.016	5	0.034	5.68	4.79	28 -12 6
Fusiform Gyrus	L	0.027	2	0.042	5.60	4.73	-42 -40 -22
Hippocampus	R	0.034	1	0.042	5.60	4.73	26 -4 -18

Table 3b.

Areas Showing Negative Correlation with the left Pars Orbitalis in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Precuneus	R	0.000	4650	0.000	10.08	6.95	10 -74 48
Precuneus				0.000	9.96	6.90	8 -68 54
Precuneus				0.000	8.95	6.49	16 -64 28
n.l.	L	0.000	2111	0.000	9.31	6.64	-36 -54 -2
n.l.				0.000	9.02	6.52	-30 -66 8
n.l.				0.000	8.97	6.50	-22 -10 30

n.l.	R	0.000	144	0.000	8.03	6.07	16 30 -2
Fusiform Gyrus	R	0.000	104	0.000	7.79	5.96	36 -50 -4
Fusiform Gyrus				0.004	6.47	5.26	36 -40 -10
Cerebellum 9	L	0.000	42	0.000	7.37	5.75	-10 -48 -50
n.l.	R	0.000	78	0.001	7.04	5.57	16 0 28
n.l.	R	0.000	55	0.001	6.91	5.50	20 0 42
Inferior Parietal L	R	0.000	61	0.006	6.32	5.17	46 -38 48
n.l.	R	0.016	5	0.018	5.92	4.93	2 18 4
n.l.	R	0.019	4	0.019	5.91	4.93	10 -20 24
Inferior Parietal L	R	0.014	6	0.020	5.88	4.91	56 -56 -10
Cerebellum 9	R	0.027	2	0.029	5.73	4.82	16 -54 -50
n.l.	R	0.027	2	0.034	5.68	4.79	24 -24 34
Hippocampus	R	0.014	6	0.035	5.67	4.78	30 -36 4
Cerebellum 8	L	0.027	2	0.042	5.59	4.73	-36 -44 -42
n.l.	R	0.034	1	0.049	5.54	4.69	24 -26 28

Table 4a.

Areas Showing Positive Correlation with the left Heschl's Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Heschl's Gyrus	L	0.000	29075	0.000	29.38	Inf	-44 -18 8
Heschl's Gyrus	L			0.000	24.64	Inf	-36 -24 12
Heschl's Gyrus	R			0.000	21.70	Inf	44 -14 6
Calcarine Sulcus	R	0.000	2255	0.000	9.17	6.58	16 -62 8
Lingual Gyrus				0.000	9.00	6.51	18 -56 2
Calcarine Sulcus				0.000	8.69	6.37	4 -70 12
n.l.	L	0.000	359	0.000	8.51	6.29	-14 -56 -26
n.l.	L	0.000	184	0.001	7.00	5.55	-10 36 6
Anterior Cingulate Cortex	R			0.003	6.66	5.37	4 40 10
Anterior Cingulate Cortex	L			0.023	5.87	4.90	-6 44 10
Parahippocampal Gyrus	L	0.004	14	0.003	6.69	5.38	-22 -18 -26
n.l.		0.002	22	0.008	6.28	5.15	0 -8 -16

Mid Cingulum	L	0.007	10	0.016	6.01	4.99	-14 -30 48
n.l.	L	0.015	5	0.017	5.98	4.97	-4 -24 -26
Mid Cingulum	L	0.033	1	0.040	5.64	4.76	-12 -28 46
n.l.	R	0.026	2	0.043	5.62	4.75	54 -74 6
Parahippocampal Gyrus	R	0.033	1	0.050	5.56	4.71	22 -10 -22

Table 4b.

Areas Showing Negative Correlation with the left Heschl's Gyrus in the Pre-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
n.l.	L	0.000	4413	0.000	12.87	Inf	-2 -30 12
n.l	L			0.000	12.75	7.84	-20 26 20
n.l.	R			0.000	10.30	7.03	22 30 16
n.l.	L	0.000	261	0.000	8.75	6.40	-18 -48 24
n.l.				0.000	7.87	5.99	-34 -46 24
n.l.				0.000	7.54	5.83	-26 -54 24
Cerebellum Crus 1	R	0.000	224	0.000	7.89	6.01	50 -62 -32

Cerebellum Crus 2	R			0.001	6.96	5.53	44 -56 -44
Precuneus	L	0.000	77	0.000	7.77	5.95	-2 -66 60
Precuneus				0.001	7.12	5.62	0 -74 52
n.l.	L	0.000	156	0.000	7.70	5.91	-6 -88 -20
Cerebellum Crus 1	R			0.002	6.82	5.45	12 -88 -18
Cerebellum Crus 2	L	0.000	291	0.003	6.62	5.34	-46 -68 -40
Cerebellum Crus 2				0.005	6.46	5.25	-32 -70 -42
Cerebellum Crus 2				0.005	6.45	5.25	-38 -62 -42
Cerebellum Crus 1	L	0.000	47	0.003	6.60	5.33	-30 -86 -26
Cerebellum Crus 2				0.038	5.67	4.78	-36 -84 -34
n.l.	L	0.000	38	0.010	6.17	5.09	-16 32 -6
n.l.				0.021	5.90	4.92	-20 40 -8
Cerebellum Crus 1	R	0.015	5	0.020	5.92	4.93	26 -86 -26
n.l.	L	0.005	13	0.020	5.91	4.93	-24 -20 26
Cerebellum Crus 1	L	0.007	10	0.021	5.89	4.92	-46 -72 -28
Cerebellum Crus 2	L	0.009	8	0.023	5.87	4.90	-28 -80 -46
Cerebellum Crus 2	R	0.007	10	0.027	5.80	4.86	42 -76 -40
n.l.	R	0.026	2	0.027	5.80	4.86	26 -32 30
n.l.	R	0.026	2	0.035	5.70	4.80	24 -16 30
n.l.	R	0.033	1	0.038	5.67	4.78	24 -16 36
Cerebellum Crus 2	L	0.026	2	0.039	5.66	4.77	-16 -88 -34

n.l.	L	0.033	1	0.041	5.64	4.76	-28 -12 24
Cerebellum Crus 2	R	0.033	1	0.042	5.63	4.75	34 -80 -40
Cerebellum Crus 2	R	0.033	1	0.045	5.60	4.74	32 -82 -42
n.l.	R	0.026	2	0.045	5.60	4.74	28 -12 26

Table 5a.

Areas Showing Positive Correlation with the left Superior Temporal Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Superior Temporal L	R	0.000	54148	0.000	26.48	Inf	54 -4 2
Superior Temporal L	L			0.000	25.81	Inf	-58 -4 4
Superior Temporal L	L			0.000	23.76	Inf	-56 -32 14
Cerebellum 6	L	0.000	568	0.000	8.83	6.44	-12 -56 -24
Mid Frontal L	L	0.000	248	0.000	8.33	6.21	-34 40 24
Lingual Gyrus	L	0.000	1035	0.000	8.29	6.19	-20 -60 -4

Fusiform Gyrus				0.001	7.09	5.60	-28 -60 -8
Lingual Gyrus				0.001	6.98	5.54	-20 -78 -8
Mid Frontal L	R	0.000	89	0.005	6.40	5.22	34 40 24
Cuneus	L	0.003	18	0.013	6.06	5.02	-14 -74 38
Fusiform Gyrys	L	0.018	4	0.022	5.87	4.90	-40 -50 -16
n.l.	R	0.026	2	0.026	5.79	4.86	8 -22 -26
Parahippocampal Gyrus	L	0.034	1	0.027	5.78	4.85	-28 -4 -26
Fusiform Gyrus	R	0.015	5	0.028	5.77	4.84	42 -44 -20
n.l.	L	0.034	1	0.042	5.61	4.74	-16 -74 20

Table 5b.

Areas Showing Negative Correlation with the left Superior Temporal Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Cerebellum Crus 2	L	0.000	2703	0.000	12.77	7.84	-46 -72 -38
Cerebellum Crus 2				0.000	10.95	7.27	-34 -84 -38
Cerebellum Crus 1				0.000	9.27	6.62	-42 -80 -34
n.l.	R	0.000	9702	0.000	11.88	7.58	22 34 10
n.l.				0.000	10.99	7.28	22 8 26
n.l.				0.000	10.93	7.26	34 -50 20
Mid Frontal L	L	0.000	177	0.001	7.24	5.68	-24 28 54
Mid Frontal L				0.001	7.08	5.59	-30 20 56
Superior Frontal L				0.001	6.92	5.51	-20 22 60
Superior Frontal L	R	0.000	163	0.002	6.74	5.41	26 30 54
Superior Frontal L				0.006	6.33	5.18	26 38 48
Mid Frontal L				0.020	5.90	4.92	34 26 52
Precuneus	L	0.007	11	0.004	6.52	5.29	0 -74 52

Superior Medial Frontal L	L	0.000	76	0.005	6.41	5.22	0 52 42
Superior Medial Frontal L	R			0.005	6.40	5.22	6 46 48
n.l.				0.008	6.26	5.14	0 62 26
Rectus	R	0.000	66	0.009	6.21	5.11	2 44 -18
Rectus				0.021	5.88	4.91	4 30 -20
Superior Medial Frontal L	L	0.026	2	0.029	5.75	4.83	-4 48 46
Medial Orbital Frontal L	R	0.026	2	0.043	5.60	4.73	8 60 -12
Mid Frontal L	R	0.034	1	0.049	5.55	4.70	38 18 54

Table 6a.

Areas Showing Positive Correlation with the right Heschl's Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Heschl's Gyrus	R	0.000	37044	0.000	31.74	Inf	52 -8 6
Insula				0.000	23.73	Inf	40 -10 6
Heschl's Gyrus				0.000	23.44	Inf	38 -26 12
Calcarine Sulcus	R	0.000	2266	0.000	9.60	6.76	16 -62 10

Lingual Gyrus	R			0.000	9.42	6.69	20 -54 2
Calcarine Sulcus	L			0.000	8.94	6.49	-12 -66 12
Calcarine Sulcus	L	0.000	95	0.002	6.77	5.43	-8 -90 2
n.l.	R	0.025	2	0.021	5.92	4.94	22 -34 -30
n.l.	L	0.033	1	0.022	5.90	4.92	-10 -18 56
n.l.	R	0.025	2	0.025	5.85	4.89	2 -18 -22
n.l.	R	0.025	2	0.025	5.84	4.89	4 -32 -34
Hippocampus	R	0.017	4	0.030	5.78	4.85	20 -8 -20
n.l.	L	0.025	2	0.030	5.77	4.84	-4 -26 -28
Pars Orbitalis	R	0.020	3	0.034	5.72	4.81	26 26 -14
n.l.	R	0.033	1	0.045	5.62	4.75	24 20 -14
n.l.	L	0.025	2	0.047	5.60	4.74	-22 -36 -32

Table 6b.

Areas Showing Negative Correlation with the right Heschl’s Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l.	L	0.000	6446	0.000	14.30	Inf	-2 -24 16
n.l.	R			0.000	13.78	Inf	12 -32 16
n.l.	L			0.000	13.24	Inf	-20 24 20
n.l.	L	0.000	253	0.000	8.76	6.40	-4 -86 -20
Cerebellum Crus 1	R			0.002	6.81	5.45	10 -86 -18
Vermis 6	R			0.010	6.18	5.09	2 -76 -10
Cerebellum Crus 2	L	0.000	887	0.000	8.49	6.29	-42 -60 -40
Cerebellum Crus 2				0.000	8.43	6.26	-30 -70 -40
Cerebellum Crus 1				0.007	6.34	5.18	-30 -86 -28
Precuneus	L	0.000	87	0.000	8.37	6.23	-2 -66 60
Precuneus				0.038	5.69	4.79	-2 -56 66
Cerebellum Crus 1	R	0.000	146	0.000	7.75	5.94	50 -62 -30
Cerebellum Crus 2				0.016	6.02	5.00	48 -62 -42
Cerebellum Crus 1				0.019	5.94	4.95	38 -70 -32
Olfactory L	R	0.000	41	0.002	6.86	5.48	0 20 -4
Cerebellum Crus 2	L	0.033	1	0.033	5.74	4.83	-26 -80 -46

n.l.	R	0.033	1	0.041	5.65	4.77	24 -32 32
n.l.	R	0.033	1	0.044	5.62	4.75	28 -32 30
n.l.	R	0.033	1	0.050	5.58	4.72	26 -26 36

Table 7a.

Areas Showing Positive Correlation with the right Superior Temporal Gyrus in the Pre-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Superior Temporal L	R	0.000	52155	0.000	24.97	Inf	56 -12 -4
Superior Temporal L				0.000	23.27	Inf	52 -36 16
Superior Temporal L				0.000	21.58	Inf	66 -20 6
n.l.	L	0.000	524	0.000	7.68	5.90	-14 -44 -28
n.l.				0.000	7.61	5.87	-12 -56 -26
n.l.				0.002	6.75	5.42	-18 -56 -32
Lingual Gyrus	L	0.000	387	0.000	7.54	5.83	-18 -52 -4
Lingual Gyrus				0.007	6.29	5.15	-20 -78 -8

Fusiform Gyrus				0.012	6.07	5.03	-28 -64 -6
Mid Frontal L	L	0.000	111	0.003	6.66	5.36	-32 40 26
Fusiform Gyrus	L	0.001	38	0.004	6.53	5.29	-40 -50 -16
Fusiform Gyrus	R	0.000	51	0.005	6.43	5.23	42 -44 -20
Cuneus	L	0.000	91	0.006	6.36	5.19	-14 -72 24
Mid Frontal L	R	0.000	57	0.017	5.94	4.95	32 42 24
Mid Frontal L				0.018	5.92	4.93	28 48 20
Inferior Temporal L	R	0.016	5	0.026	5.79	4.85	42 6 -38
n.l.	R	0.022	3	0.032	5.70	4.80	6 -30 -36
Calcarine Sulcus	L	0.022	3	0.048	5.55	4.70	-14 -70 12

Table 7b.

Areas Showing Negative Correlation with the right Superior Temporal Gyrus in the Pre-Training Session

							MNI
Region		Cluster		Peak			Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l.	R	0.000	7857	0.000	11.52	7.46	20 26 14
n.l.	R			0.000	11.11	7.32	0 -32 12
n.l.	L			0.000	11.03	7.29	-24 -2 32
Cerebellum Crus 2	L	0.000	1936	0.000	9.10	6.55	-42 -70 -36
Cerebellum Crus 1	R			0.000	8.69	6.38	46 -74 -24
Cerebellum Crus 1	L			0.000	8.44	6.26	-42 -80 -34
Precuneus	L	0.001	33	0.000	8.21	6.16	-2 -66 60
Cerebellum Crus 2	R	0.027	2	0.041	5.61	4.74	14 -86 -40

Table 8a.

Areas Showing Positive Correlation with the left Pars Opercularis in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Pars Opercularis	L	0.000	28702	0.000	26.92	Inf	-46 10 22
Pars Opercularis				0.000	22.00	Inf	-56 10 8
Pars Opercularis				0.000	19.70	Inf	-48 12 4
Inferior Temporal L	L	0.000	7698	0.000	14.18	Inf	-48 -56 -14
Inferior Parietal L				0.000	14.07	Inf	-54 -36 38
Supramarginal Gyrus				0.000	14.07	Inf	-64 -38 26
n.l.	R	0.000	119	0.001	7.24	5.56	12 -16 -2
Cerebellum 6	R	0.000	197	0.001	7.16	5.52	18 -72 -26
Cerebellum 6				0.004	6.70	5.29	26 -62 -30
0				0.018	6.08	4.95	16 -62 -32
Mid Cingulum	R	0.003	17	0.012	6.25	5.04	14 -28 44
n.l.	L	0.034	1	0.047	5.68	4.71	-12 -24 38

Table 8b.

Areas Showing Negative Correlation with the left Pars Opercularis in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
Hippocampus	R	0.000	13066	0.000	13.80	Inf	36 -30 -10
n.l.	L			0.000	13.77	Inf	-20 -50 24
n.l.	R			0.000	12.55	7.56	20 -46 24
Parahippocampal Gyrus	L	0.000	86	0.000	9.67	6.62	-16 -30 -14
Cerebellum Crus 2	L	0.000	457	0.000	9.13	6.41	-44 -70 -36
Cerebellum Crus 2				0.000	8.72	6.24	-28 -86 -32
Cerebellum Crus 1				0.000	7.59	5.73	-36 -82 -20
n.l.	L	0.000	260	0.000	8.35	6.08	-2 -86 -20
n.l.	R			0.001	7.48	5.68	2 -78 -10
n.l.	L	0.006	12	0.002	7.05	5.47	-30 -72 2
Angular Gyrus	R	0.000	62	0.002	6.96	5.42	48 -62 32
Superior Frontal	R	0.000	67	0.003	6.82	5.35	26 36 48
Mid Cingulum	R	0.002	23	0.005	6.59	5.23	2 -20 34

n.l.	R	0.011	7	0.005	6.58	5.22	2 -8 6
n.l.	R	0.015	5	0.012	6.25	5.04	8 -34 -24
n.l.	R	0.004	16	0.013	6.21	5.02	30 -86 -20
n.l.	L	0.034	1	0.021	6.00	4.90	-30 -70 10
n.l.	L	0.034	1	0.027	5.91	4.85	-30 -66 12
Cerebellum Crus 1	R	0.034	1	0.039	5.76	4.76	50 -66 -26

Table 9a.

Areas Showing Positive Correlation with the left Pars Triangularis in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Pars Triangularis	L	0.000	14721	0.000	25.04	Inf	-44 20 26
Pars Triangularis				0.000	23.42	Inf	-50 28 14
Pars Orbitalis				0.000	23.29	Inf	-48 44 -8
Inferior Parietal L	L	0.000	5991	0.000	15.12	Inf	-50 -44 40
Inferior Temporal L				0.000	14.83	Inf	-58 -52 -14
Middle Temporal L				0.000	14.25	Inf	-60 -46 -6
Pars Orbitalis	R	0.000	4059	0.000	14.29	Inf	46 42 -2
Pars Triangularis				0.000	10.83	7.03	44 28 20

Pars Triangularis				0.000	10.46	6.90	54 20 22
Cerebellum Crus 1	R	0.000	604	0.000	11.02	7.09	14 -76 -26
Cerebellum Crus 1				0.000	9.17	6.42	32 -62 -32
n.l.	R	0.000	896	0.000	9.24	6.45	14 8 4
Pallidum	R			0.000	8.79	6.27	20 2 0
n.l.				0.000	8.68	6.22	20 2 10
Inferior Temporal L	R	0.000	324	0.000	8.55	6.16	64 -44 -10
Middle Temporal L				0.000	7.94	5.89	56 -40 -10
Superior Frontal L	L	0.022	3	0.038	5.76	4.76	-18 56 28
Superior Frontal L	L	0.034	1	0.047	5.67	4.71	-14 52 26

Table 9b.

Areas Showing Negative Correlation with the left Pars Triangularis in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l.	L	0.000	3911	0.000	11.42	7.22	-26 -10 30
n.l.				0.000	10.14	6.79	-26 -44 6
Precuneus				0.000	9.91	6.71	-16 -44 8

Thalamus	R	0.000	2661	0.000	11.15	7.13	14 -34 6
Precuneus				0.000	9.98	6.73	18 -44 10
n.l.				0.000	9.98	6.73	18 -46 24
n.l.		0.000	67	0.004	6.73	5.30	-38 -84 -20
Cerebellum Crus 2	L	0.009	9	0.008	6.39	5.12	-48 -54 -42
n.l.	R	0.002	21	0.017	6.10	4.96	4 2 18
Fusiform Gyrus	R	0.008	10	0.028	5.89	4.84	40 -12 -22

Table 10a.

Areas Showing Positive Correlation with the left Pars Orbitalis in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
Pars Orbitalis	L	0.000	14525	0.000	29.04	Inf	-44 24 -8
Pars Orbitalis				0.000	26.94	Inf	-32 24 -18

Pars Orbitalis				0.000	26.01	Inf	-44 40 -8
Supplementary Motor Area	L	0.000	5730	0.000	19.62	Inf	-4 22 58
Superior Medial Frontal L				0.000	17.20	Inf	-4 42 36
Superior Medial Frontal L				0.000	15.67	Inf	-8 30 48
Inferior Orbital Frontal L	R	0.000	2949	0.000	17.65	Inf	48 28 -6
Insula				0.000	15.65	Inf	32 26 -18
Inferior Temporal L				0.000	7.83	5.85	48 6 -40
Cerebellum Crus 1	R	0.000	633	0.000	11.50	7.24	18 -78 -26
Cerebellum Crus 1				0.000	9.71	6.63	26 -70 -30
Caudate	R	0.000	311	0.000	8.97	6.34	12 10 8
n.l.	R	0.000	561	0.000	8.73	6.24	48 -22 -10
n.l.				0.000	8.34	6.08	46 -30 -6
Superior Temporal L				0.049	5.65	4.69	52 -8 -12
Supramarginal Gyrus	R	0.000	40	0.007	6.43	5.14	58 -44 30
Thalamus	L	0.003	18	0.008	6.38	5.11	-8 -28 2

Mid Cingulum	L	0.006	13	0.019	6.04	4.92	-4 -16 38

Table 10b.

Areas Showing Negative Correlation with the left Pars Orbitalis in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
n.l.	L	0.000	459	0.000	10.74	7.00	-26 -56 16
n.l.				0.002	6.91	5.39	-30 -40 10
n.l.				0.008	6.41	5.13	-28 -54 8
Precuneus	R	0.000	2207	0.000	10.42	6.89	24 -52 24
n.l.				0.000	10.14	6.79	20 -48 16
Lingual Gyrus				0.000	9.65	6.61	6 -42 2
n.l.	L	0.000	572	0.000	8.67	6.21	-18 24 16
n.l.				0.000	8.10	5.97	-22 26 24

n.l.				0.001	7.45	5.67	-26 -14 32
Cerebellum 6	L	0.000	78	0.001	7.37	5.63	-30 -58 -26
n.l.	R	0.000	143	0.001	7.27	5.58	20 30 6
n.l.				0.037	5.77	4.77	20 40 -4
Thalamus	R	0.002	22	0.001	7.13	5.51	18 -28 10
n.l.	R	0.004	15	0.002	6.99	5.44	22 30 20
n.l.	R	0.003	19	0.007	6.47	5.16	22 -38 34
Lingual Gyrus	L	0.005	14	0.007	6.47	5.16	-32 -46 -4
n.l.	R	0.002	22	0.010	6.31	5.07	26 -34 16
n.l.	R	0.002	21	0.010	6.30	5.07	16 -4 32
Cerebellum Crus 1	L	0.014	6	0.015	6.14	4.98	-6 -76 -20
Mid Cingulum	R	0.005	14	0.022	5.98	4.89	6 -38 42
n.l.	L	0.034	1	0.023	5.97	4.88	-24 -38 36
n.l.	R	0.034	1	0.023	5.96	4.88	24 -8 42
Precuneus	L	0.006	13	0.025	5.94	4.86	-8 -66 50
n.l.	R	0.034	1	0.026	5.91	4.85	24 -30 38
Vermis 8		0.027	2	0.034	5.80	4.78	0 -74 -36
n.l.	R	0.034	1	0.037	5.77	4.76	26 -24 36
n.l.	R	0.034	1	0.044	5.70	4.72	20 38 2

Mid Temporal L	R	0.034	1	0.045	5.69	4.72	42 -68 24
n.l.	L	0.034	1	0.048	5.66	4.70	-16 -38 18

Table 11a.

Areas Showing Positive Correlation with the left Heschl’s Gyrus in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
Heschl's Gyrus	L	0.000	26498	0.000	28.86	Inf	-36 -26 10
Heschl's Gyrus	L			0.000	24.39	Inf	-44 -18 10
Insula	R			0.000	23.98	Inf	44 -10 4
Anterior Cingulate Cortex	R	0.000	603	0.000	8.60	6.19	4 28 20
Anterior Cingulate Cortex				0.000	8.06	5.95	6 36 12
Supramarginal Gyrus	R	0.000	1412	0.000	8.39	6.10	10 -6 72
Supramarginal Gyrus				0.000	8.11	5.97	4 -8 60
Supramarginal Gyrus				0.000	8.00	5.93	4 -6 50

Postcentral Gyrus	L	0.000	720	0.001	7.40	5.64	-26 -30 62
Postcentral Gyrus				0.002	7.06	5.47	-22 -36 70
Precuneus				0.002	6.99	5.43	-14 -42 68
Mid Occipital L	L	0.033	1	0.046	5.72	4.73	-42 -88 6
n.l.		0.033	1	0.047	5.71	4.73	0 -4 -20

Table 11b.

Areas Showing Negative Correlation with the left Heschl's Gyrus in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l.	L	0.000	3803	0.000	11.62	7.28	-22 22 24
n.l.	R			0.000	11.12	7.12	24 24 20
n.l.	R			0.000	10.61	6.95	20 14 26
n.l.	L	0.000	202	0.000	8.11	5.97	-22 -46 28
n.l.				0.000	7.97	5.91	-30 -46 26
n.l.				0.014	6.20	5.02	-28 -38 28
n.l.	R	0.000	68	0.001	7.26	5.57	34 -42 24
n.l.				0.003	6.87	5.38	32 -56 18

n.l.	R	0.004	15	0.006	6.54	5.20	26 -24 32
Cerebellum Crus 1	R	0.002	23	0.007	6.47	5.16	50 -64 -30
Cerebellum Crus 1				0.027	5.93	4.86	52 -58 -36
n.l.	R	0.000	94	0.007	6.47	5.16	2 -82 -14
n.l.	L	0.002	21	0.012	6.26	5.05	-16 32 -4
n.l.	L	0.021	3	0.014	6.21	5.02	-28 -2 34
n.l.	L	0.033	1	0.026	5.95	4.87	-24 0 36
Olfactory L	R	0.017	4	0.029	5.91	4.84	0 20 -6
n.l.	L	0.033	1	0.034	5.84	4.81	-26 2 34
n.l.	R	0.033	1	0.042	5.75	4.75	48 -32 -14
n.l.	R	0.033	1	0.047	5.70	4.72	24 -10 36
n.l.	R	0.033	1	0.048	5.69	4.72	22 -12 30

Table 12a.

Areas Showing Positive Correlation with the left Superior Temporal Gyrus in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}

Superior Temporal L	L	0.000	55557	0.000	33.80	Inf	-56 -6 2
Superior Temporal L	R			0.000	26.49	Inf	56 0 2
Rolandic Operculum	L			0.000	23.79	Inf	-56 0 12
Mid Frontal L	L	0.000	358	0.000	11.14	7.13	-32 38 22
Mid Frontal L	R	0.000	351	0.000	8.89	6.31	32 44 24
Parahippocampal Gyrus	L	0.007	10	0.006	6.57	5.22	-18 -38 -4
Parahippocampal Gyrus	L	0.010	7	0.012	6.28	5.06	-22 2 -32
Vermis 3	R	0.025	2	0.026	5.97	4.88	2 -42 -16
Cerebellum 6	R	0.032	1	0.030	5.90	4.84	16 -60 -28
Calcarine Sulcus	L	0.008	9	0.036	5.83	4.80	-2 -94 8

Table 12b.

Areas Showing Negative Correlation with the left Superior Temporal Gyrus in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
n.l.	L	0.000	7872	0.000	15.11	Inf	-30 -46 26

n.l.	R			0.000	13.58	7.82	22 -44 26
n.l.	R			0.000	13.39	7.78	30 -42 24
Cerebellum Crus 1	R	0.000	1914	0.000	11.57	7.26	42 -76 -32
Cerebellum Crus 2	L			0.000	10.77	7.01	-40 -72 -38
Cerebellum Crus 1	R			0.000	10.30	6.84	40 -68 -34
Superior Medial Frontal L	L	0.000	86	0.001	7.47	5.68	0 52 40
Superior Medial Frontal L				0.024	6.00	4.90	0 42 50
Mid Frontal L	L	0.000	127	0.001	7.35	5.62	-34 16 56
Mid Frontal L				0.009	6.42	5.14	-38 22 50
Superior Frontal L	L	0.000	68	0.005	6.60	5.23	-14 38 52
Superior Frontal L				0.049	5.70	4.72	-22 38 48
Superior Medial Frontal L	L	0.002	22	0.005	6.60	5.23	0 66 14
Superior Medial Frontal L				0.023	6.01	4.90	0 62 26
Inferior Temporal L	L	0.001	23	0.008	6.45	5.15	-58 -34 -18
n.l.	R	0.032	1	0.009	6.38	5.11	48 -30 -16
Superior Orbital Frontal L	R	0.032	1	0.017	6.13	4.97	14 46 -18
Superior Orbital Frontal L	R	0.032	1	0.020	6.07	4.94	12 44 -20
Pars Orbitalis	R	0.025	2	0.026	5.97	4.88	46 36 -18
Superior Medial Frontal L	L	0.025	2	0.031	5.89	4.83	0 60 30
Superior Frontal L	L	0.032	1	0.040	5.78	4.77	-16 22 44
Angular Gyrus	R	0.025	2	0.047	5.72	4.74	46 -58 34

Mid Frontal L	L	0.032	1	0.049	5.70	4.72	-24 32 52

Table 13a.

Areas Showing Positive Correlation with the right Heschl's Gyrus in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
Heschl's Gyrus	R	0.000	41543	0.000	26.00	Inf	52 -10 8
Heschl's Gyrus				0.000	24.47	Inf	40 -24 12
Superior Temporal L				0.000	24.42	Inf	52 -22 10
Mid Frontal L	L	0.003	16	0.004	6.78	5.33	-32 36 20
Mid Frontal L	L	0.025	2	0.035	5.86	4.81	-28 40 32

Table 13b.

Areas Showing Negative Correlation with the right Heschl's Gyrus in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l	R	0.000	6075	0.000	13.39	7.78	34 -42 4
n.l				0.000	12.26	7.47	20 -46 28
n.l				0.000	11.70	7.31	20 -38 10
Cerebellum Crus 1	R	0.000	172	0.001	7.16	5.52	38 -70 -32
Cerebellum Crus 2				0.001	7.16	5.52	44 -74 -36
Olfactory L	R	0.002	19	0.003	6.93	5.41	-2 20 -6
Cerebellum Crus 2	L	0.000	157	0.004	6.77	5.32	-42 -72 -36
Cerebellum Crus 1				0.004	6.74	5.31	-32 -74 -28
Cerebellum Crus 2	L	0.000	64	0.006	6.56	5.21	-6 -84 -30
Cerebellum Crus 2				0.009	6.40	5.13	-2 -82 -22
Cerebellum Crus 2	R	0.005	12	0.011	6.32	5.08	36 -62 -40
Cerebellum Crus 2	L	0.016	4	0.033	5.87	4.82	-34 -64 -38

Cerebellum Crus 1	R	0.025	2	0.043	5.77	4.76	50 -62 -32

Table 14a.

Areas Showing Positive Correlation with the right Superior Temporal Gyrus in the Post-Training Session

Region	Hemisphere	Cluster p(FWE-corr)	Cluster size	Peak p(FWE-corr)	T	Z	MNI Coordinates x,y,z {mm}
Superior Temporal L	R	0.000	54420	0.000	24.44	Inf	54 -14 -4
Superior Temporal L				0.000	24.32	Inf	58 -22 12
Superior Temporal L				0.000	22.58	Inf	60 -34 12
Mid Frontal L	R	0.000	373	0.000	9.59	6.58	30 46 22
Mid Frontal L	L	0.000	201	0.000	8.57	6.17	-32 40 22
Mid Frontal L				0.000	7.70	5.78	-30 36 30
Cerebellum 6	L	0.003	17	0.020	6.05	4.93	-10 -62 -20
n.l.				0.045	5.70	4.72	-14 -60 -28
n.l.	L	0.026	2	0.025	5.95	4.87	-16 -54 -32
Cerebellum 4/5	R	0.013	6	0.025	5.94	4.87	14 -54 -18
n.l.	L	0.033	1	0.028	5.90	4.84	-32 0 -26
Superior Occipital L	L	0.033	1	0.046	5.70	4.72	-26 -90 26

Table 14b.

Areas Showing Negative Correlation with the right Superior Temporal Gyrus in the Post-Training Session

Region		Cluster		Peak			MNI Coordinates
	Hemisphere	p(FWE-corr)	Cluster size	p(FWE-corr)	T	Z	x,y,z {mm}
n.l.	R	0.000	7669	0.000	12.67	7.59	30 -52 26
n.l.	R			0.000	12.18	7.45	20 -2 30
n.l.	L			0.000	11.93	7.37	-26 -48 28
Cerebellum Crus 1	R	0.000	1310	0.000	10.03	6.75	48 -68 -32
Cerebellum Crus 1	R			0.000	8.91	6.32	36 -62 -38
Cerebellum Crus 2	L			0.000	8.86	6.30	-2 -82 -28
n.l.	R	0.002	20	0.001	7.46	5.67	28 -40 -40

Footnote. n.l. in Region column: no label available for current MNI coordinates in AAL Atlas. The n.l. clusters usually lie outside grey matter, mainly in ventricles. These can be artifactual maps caused by noise (head motion, non-neural physiological noise) and normalization of individual brains into a common space. L in Region column: Lobe, AAL does not specify gyri and sulci. Inf in Z column: infinite value.

CHAPTER 6

GENERAL DISCUSSION

This thesis has focused on understanding what drives individual variation in sound processing performance, specifically with respect to pitch processing in the domains of speech and music. Although many factors can underlie this variation such as age of commencement (Flege et al., 1999; Habib & Besson, 2009) and motivation (Asmus, 2014; Dörnyei, 2002), two of them were investigated here: the effect of experience and the effect of aptitude and predispositions. The physical sound property studied in all chapters was frequency, with its psychoacoustic correlate pitch. Pitch, along with timbre, plays a central role in both speech and music sounds (Patel, 2008; Zatorre & Gandour, 2008) and thus was used in the present experiments in both contexts (linguistic and musical).

Summary of findings

Chapter 2 reviewed behavioral and neuroimaging studies and demonstrated that experience with pitch in either a linguistic or a musical context can influence pitch processing beyond that specific context. Individuals who have received musical training are better at processing pitch in a linguistic context, such as when it signals lexical tone or sentence intonation, and in domain-neutral pitch processing, such as in the processing of sine-wave sounds (Magne et al., 2006; Moreno et al., 2009; Musacchia et al., 2007; Schön et al., 2004; Wong, Skoe, et al., 2007). Although sparser, evidence also exist for the reverse effect; that is, individuals who have experience with pitch through language, in particular tone language speakers, are better than non-tone language speakers in processing pitch in a musical or domain-neutral context (Bidelman et al., 2011a, 2011b; Chandrasekaran et al., 2007a, 2007b; Pfordresher & Brown, 2009). The review in Chapter 2 proposed that these bidirectional influences occur because speech and music draw on the same cortical and subcortical sound processing resources. These are shaped by experience in either domain, which hence leads to domain-general sound processing advantages.

Following up on this central premise in Chapter 2, the study in Chapter 3 looked at pitch processing in a group of Cantonese-Dutch bilinguals. This group has been exposed to two languages that make different use of linguistic pitch from a very young age onwards, gaining valuable “auditory expertise” with pitch. It was thus expected that Cantonese-Dutch bilinguals would outperform their monolingual Dutch peers in a series of non-linguistic pitch processing tasks. The results revealed that bilingual experience with a tone and a non-tone language influenced the way melodic (pitch) and phonological (vowel) information is processed online in sung speech, with bilinguals processing the two stimulus dimensions more integrally than controls. The study however found no tone experience advantage either in non-linguistic pitch discrimination tasks or in a musical interval learning task. It was thus concluded that although experience with tone has some effect in non-linguistic pitch processing, this effect is rather weak and difficult to elicit.

The large individual variation in performance among the musically naïve monolingual control group, observed in the pitch perception experiments reported in Chapter 3, highlighted the important factor of aptitude and predisposition. Experience is undoubtedly crucial and can influence performance in pitch processing, as supported by longitudinal studies (Hyde et al., 2009; Moreno et al., 2009). However, the variability found in inexperienced individuals indicates that not everyone has the same aptitude or talent with sounds. It has been suggested that differences in aptitude stem from differences among individuals' brain structure and function (Zatorre, 2013). There is great variation in human brain morphology which may result in some individuals having neuroanatomical and/or functional brain characteristics that are advantageous for processing sounds. Chapters 4 and 5 focused on the role of aptitude by investigating the neural correlates of individual variation in sound learning performance.

The experiment in Chapter 4 considered the efficiency with which sounds are processed in the brain and how this efficiency may favor certain individuals over others in learning new linguistic sounds. Dutch participants without any prior experience with tone were trained over five separate sessions to identify non-native (Mandarin) tonal contrasts embedded in words. They were asked to learn to match, in total, twenty-four auditory-presented words to pictures of everyday items. Learning performance was assessed as the accuracy (percentage correct) in a word-picture matching task. Neuronal efficiency was measured with an fMRI-adaptation paradigm as the size of the repetition suppression effect, i.e. the reduction of the

BOLD response to the repeated presentation of the non-native tonal contrast. Repetition suppression to tones was measured at two time-points: before and after training. The results revealed that the better the performance in learning the non-native tonal contrast, the larger the repetition suppression effect to tone in the left Inferior Frontal Gyrus (IFG). This effect was significant even before participants had started the training, arguing in favor of pre-existing differences in sound-processing aptitude. Successful learners processed sounds more efficiently, and appeared to require less guidance from higher order frontal language areas when listening to non-native tones. The fact that the effect was found in the left IFG, and in particular in the left Pars Opercularis (POp), is in agreement with previous studies on phonetic learning (Golestani & Zatorre, 2004; Myers & Swan, 2012) and emphasizes the importance of this area in learning non-native sounds.

The experiment in Chapter 5 was complimentary to the experiment in Chapter 4 as it also looked at variation in the BOLD signal in relation to tone learning performance, only this time the focus was on activity at rest, that is, when the participants (those tested in Chapter 4) were not engaged in any task inside the scanner. Resting-state connectivity was measured before and after the five-session lexical tone training and the connectivity patterns in bilateral auditory areas (Heschl's Gyri and Superior Temporal Gyri) and the left frontal areas (IFG) were analyzed in relation to learning performance. The results revealed that the strength of resting-state connectivity between the left Pars Orbitalis and the left posterior Superior Temporal Gyrus /Angular Gyrus in the post-training session correlated negatively with learning performance. These areas comprise the ventral pathway for language (Hickok & Poeppel, 2007; Saur et al., 2008), which is considered to be important for mapping sounds to meaning (as was required in the learning task). The involvement of functional connectivity along the ventral pathway is in accord with previous findings showing that white matter connectivity in this pathway is positively correlated with word-learning performance (Wong et al., 2011). It is also in agreement with models of speech processing which have suggested a sound-to-meaning role for the ventral pathway (Hickok & Poeppel, 2007; Ueno, Saito, Rogers, & Lambon Ralph, 2011). Furthermore, changes in connectivity strength from the pre-to the post-training session between the right Heschl's Gyrus and the right POp were negatively correlated with participants' learning scores. The same negative correlation was found for the change in connectivity between the right STG and the right Middle Temporal Gyrus (MTG). These rightward connectivity patterns were elicited presumably due to the spectral characteristics in pitch which are often processed in the right hemisphere (Luo et al., 2006; Warrier et al., 2009; Zatorre & Gandour, 2008). Their negative association with learning performance was interpreted as the decreased reliance on inefficient compensatory connections employed for non-linguistic (prosodic and emotional) pitch processing in successful learners. In sum, resting-state connectivity patterns in speech processing areas showed training-induced changes which varied as a function of learning performance.

In contrast to the results in Chapter 4, which demonstrated that the efficiency with which the left IFG processes non-native sounds is predictive of future learning performance, the resting-state connectivity patterns in Chapter 5 did not reveal pre-training differences among learners. In other words, the strength of connectivity between auditory and speech areas in the pre-training session did not correlate with participants' subsequent learning scores. Despite that, differences in aptitude were evident in the post-training session as well as in the connectivity changes from the pre- to the post-training sessions. These results together with findings in Chapter 4 indicate that individual variation in non-native sound learning performance is a consequence of more efficient neuronal processing of sounds as well as more efficient functional connectivity changes between frontal and temporal brain areas.

The effect of aptitude with sounds was tested in this thesis only in the speech domain. It would be of great interest to test sound learning in the linguistic and musical domains in parallel, using the same population, in order to see whether the talented speech learners will be the same as the talented music learners. Although studies on the neural correlates of musical learning variability do exist (Zarate, Delhommeau, Wood, & Zatorre, 2010; Zatorre, Delhommeau, & Zarate, 2012), they focus on tonal discrimination abilities rather than the ability to learn new sound categories, such as musical intervals. The latter ability would be more comparable with the non-native tone training paradigm used in Chapters 4 and 5. Further research comparing learning in both domains is needed in order to establish the degree of overlap and divergence between the neural substrates of musical and speech-sound learning aptitude.

Future studies on sound learning should also address the role of structural connectivity in music vs. speech aptitude and learning. While Chapter 5 explored the important role of functional connectivity, it would be interesting to investigate the underlying white matter fiber connections that enable functional communication between distant brain regions. It is already known that ventral fiber tracts, such as the Extreme Capsule and the Inferior Fronto-Occipital Fasciculus, contribute to the ability to map non-native sounds to meaning (Wong et al., 2011), while dorsal fiber tracts, such as the Arcuate Fasciculus, facilitate learning the phonological form of non-native words (Lopez-Barroso et al., 2013). However, the extent to which these white matter tracts also contribute to musical sound learning is still unknown. Again, studies investigating sound learning in both domains are needed in order to make these cross-domain comparisons.

Another approach to aptitude and predispositions for pitch processing would be to examine the potential contribution of genetic factors. At the population or glossogenetic level, linguistic tone is related to the population frequency of the adaptive haplogroups of two genes: ASPM (Abnormal Spindle-like Microcephaly-Associated) and MCPH1 (Microcephalin) (Dediu & Ladd, 2007). In particular, the frequency of the derived ASPM and MCPH1 allele is higher in populations speaking a non-tonal language, a correlation which cannot be accounted for by other environmental factors (Dediu & Ladd, 2007). Since ASPM and MCPH1 are involved in the adaptive evolution of the human brain (Evans et al., 2005; Mekel-Bobrov et al., 2005) it has been suggested that these alleles may provide an indirect association between brain structure and tone-language use (Dediu & Ladd, 2007). This hypothesis was tested at the individual or ontogenetic level and revealed that individuals who spoke a non-tonal language and had a greater load of the ASPM derived allele showed higher tone identification accuracy (Wong et al., 2012). Based on these findings, it would be interesting for future research to test whether genetic factors, such as the ASPM allele, can also predict pitch-to-meaning performance in non-tone-language speakers.

Weighing the differential contributions of experience and aptitude

The studies reported in this thesis demonstrate that although experience, in the form of musical training or experience with a tone language, can influence pitch processing performance, there are major individual differences in aptitude which may also account for a large proportion of the observed variation in performance. While the contributions of each factor were not directly compared in this thesis, the findings point to a possible interaction between them, such that, for example, individuals with aptitude for sound learning will be more likely to engage in music and/or language learning and will benefit from training faster than their peers.

Self-selection, that is the tendency of an individual to pursue an activity for which he/she is predisposed, has always been a concern in studies on “specialist” groups (i.e., professional musicians, phoneticians, ballet dancers, simultaneous interpreters etc.); the effects of experience cannot be easily disentangled from the effects of aptitude in these groups (Elmer et al., 2011; Gaser & Schlaug, 2003; Golestani et al., 2011). One could speculate that self-selection could have been a reason why the study in Chapter 3 offered relatively poor evidence for enhanced pitch processing in Cantonese-Dutch bilinguals: contrary to musicians, who have chosen to engage in musical training, the participants in Chapter 3’s study did not choose to become bilinguals. The chances of finding a good sound learner in such a population are thus likely to be equal to the chances of finding one in any population. In contrast, it is hypothesized that the chances of finding a good learner are higher in a population of musically educated individuals, due to the interaction between aptitude and experience.

In relation to this point, it is also worth mentioning that performance in the lexical tone training reported in Chapter 4 and Chapter 5 did not correlate with the participants' musical training (neither training duration nor age of music training commencement). This contradicts the theories reviewed in Chapter 2, according to which musical experience should result in an advantage in learning new sound categories (see Patel's Shared Sound Category Learning Mechanism Hypothesis, 2008). It is possible, however, that the participants with musical education tested in Chapters 4 and 5 would have performed worse in the word-learning task had they not received any musical training at all. Furthermore, the experiment did not aim to make a categorical comparison between musicians and non-musicians, so less than half of the participants had musical experience, and none were professional musicians. Previous studies that used the same training paradigm and recruited individuals with up to six years of musical training also did not find a correlation between musical experience and tone-learning performance (Chandrasekaran et al., 2012, 2010). The absence of a musical experience effect could be due to large variation in performance with some musically naïve individuals performing extremely well in learning non-native contrasts, as was observed in the sample recruited for Chapters 4 and 5.

As in every instantiation of the nature vs. nurture debate, the answer lies somewhere in the middle: both aptitude and experience influence sound processing performance and thus explain a large proportion of the individual variation in behavior. It is therefore advisable to consider these factors in interaction when assessing performance. Longitudinal studies would be needed in order to test the aptitude by experience interaction. In such a study, several measurements (e.g., the individuals' tone identification accuracy, repetition suppression to tone, resting-state or white matter connectivity) could be made before and after part of the sample is randomly assigned to extensive musical training. The contribution of each factor could then be tested along with their interaction.

Studying the effect of aptitude and experience is of both theoretical and practical significance. Theoretically, the study of experience offers a window onto the capabilities of human cognition as well as how it is related to brain anatomy and plasticity (for a review see Zatorre, Fields, & Johansen-Berg, 2012). The seminal longitudinal study by Draganski and colleagues (2004), for instance, was one of the first to show that training can alter human brain structure in adulthood. It showed that three months of juggling training induced plastic changes in the grey matter of cortical areas processing visual motion (Draganski et al., 2004). Studies of learning and plasticity provide further insights into the functional role of different brain areas and networks [e.g. seeing white matter increase in fiber tracts underlying the functional working memory network after intense working memory training strengthens knowledge on the role of that tract (Takeuchi et al., 2010)].

Exactly like experience, studying the neural correlates of aptitude or predisposition is informative for gaining better understanding of how cognition is instantiated in the brain (for reviews see Kanai & Rees, 2011; Zatorre, 2013). While individual variation in cognitive performance is often considered "noise", it is exploited in studies that investigate correlations between variation in performance and variation in brain function and structure. For instance, by correlating individual differences in visuospatial attention performance to cortical thickness, a study by Westlye and colleagues elucidated the contributions of different brain areas to different components of the attention network (Westlye, Grydeland, Walhovd, & Fjell, 2011). In the domain of sound learning, individual variation in learning to map sounds to semantic content correlated with white matter integrity along the ventral pathway for language, a result which compliments functional data for the role of this route in speech processing (Wong et al., 2011).

Ultimately, this line of research would find its practical application in improving learning in the classroom as well as in rehabilitation practices. The better our knowledge of the effects of experience and aptitude on learning and cognition, the better informed and adjusted educational policies can be for language or music learners and the better rehabilitation practices can be for individuals with learning disabilities or impairments. With respect to educational policies, awareness of the underlying causes of individual differences in non-native sound learning can lead towards more individualized second language instruction that is most effective for high- versus low-aptitude sound learners. Perrachione and colleagues, for example, found that successful sound learners have higher perceptual abilities and thus benefit more from exposure to highly variable non-native sound input whereas less successful learners have weaker perceptual abilities and benefit more from less variable input (Perrachione, Lee, Ha, & Wong, 2011), a result that has been replicated (Sadakata & McQueen, in press). With respect to clinical practice, increasing understanding of the neural mechanisms involved in sound learning performance can guide rehabilitation approaches. For example, the consistent finding that activation in the left inferior frontal gyrus is predictive of phonetic and perceptual learning performance (see Chapter 4, Eisner, Mcgettigan, Faulkner, Rosen, & Scott, 2010; Golestani & Zatorre, 2004; Myers & Swan, 2012) highlights the importance of higher-order linguistic skills rather than basic acoustic encoding. This relatively abstract knowledge can be potentially useful in guiding rehabilitation towards relevant skills (e.g. training phonological working memory) in clinical populations, such as patients with cochlear implants.

Concluding remarks

This dissertation has demonstrated that experience and aptitude contribute to individual variation in pitch processing performance. The effect of linguistic experience was juxtaposed with the effect of music training in a comprehensive review of the literature on music-speech transfer effects. Theoretical frameworks and experimental evidence pointed to bidirectional influences between the two domains which occur due to shared, domain-general cortical and subcortical pitch processing mechanisms. Speech-to-music transfer effects, that is, the effect of linguistic experience on music processing, are weaker than music-to-speech transfer effects, as also verified in the performance of Cantonese-Dutch bilinguals in music pitch processing tasks. Due to their auditory expertise with pitch in a tone and a non-tone language from an early age, bilinguals were expected to outperform controls with no tone language experience in different levels of pitch processing. Although bilinguals processed melodic and phonological information in sung speech more integrally, they failed to show an advantage in either pitch perception or pitch interval learning tasks. The contribution of aptitude and predisposition was addressed by investigating the neural underpinnings of individual variation in non-native sound learning. Individuals who performed better in learning to match non-native tonal contours to meaning showed more efficient processing of the non-native sounds in the left IFG, even before they started the training. Furthermore, individual variation in sound learning performance was reflected in resting-state functional connectivity patterns, especially in the ventral pathway for language.

In conclusion, this thesis has presented evidence that having “an ear for pitch” in language and music stems from both long-term experience with pitch, which shapes domain-general pitch processing resources, as well as aptitude, in the form of more efficient neuronal processing of pitch and more efficient functional connections in the brain.

REFERENCES

’t Hart, J. (1998). Intonation in Dutch. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 96–111). Cambridge University Press.

Abdul-Kareem, I. a, Stancak, A., Parkes, L. M., & Sluming, V. (2011). Increased gray matter volume of left pars opercularis in male orchestral musicians correlate positively with years of musical performance. *Journal of Magnetic Resonance Imaging : JMRI*, *33*(1), 24–32. doi:10.1002/jmri.22391

Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1515), 285–99. doi:10.1098/rstb.2008.0253

ANSI (1994). *American National Standard Acoustical Terminology. New York: American National Standards Institute.*

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*(June), 321. doi:10.3389/fpsyg.2013.00321

Asmus, E. P. (2014). Sixth Graders’ Achievement Motivation: Their Views of Success and Failure in Music. *Bulletin of the Council for Research in Music Education*, *85*(Late Fall), 1–13.

Bajo, V. M., Nodal, F. R., Moore, D. R., & King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nature Neuroscience*, *13*(2), 253–60. doi:10.1038/nn.2466

Barker, D., Plack, C. J., & Hall, D. a. (2012). Reexamining the evidence for a pitch-sensitive region: a human fMRI study using iterated ripple noise. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(4), 745–53. doi:10.1093/cercor/bhr065

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology* (Vol. 102). Berlin: Mouton de Gruyter.

Belyk, M., & Brown, S. (2013). Perception of affective and linguistic prosody: an ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nst124

Bent, T., Bradlow, A. R., & Wright, B. a. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology. Human Perception and Performance*, *32*(1), 97–103. doi:10.1037/0096-1523.32.1.97

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(7), 1583–96. doi:10.1093/cercor/bhn196

Bermudez, P., & Zatorre, R. J. (2005). Differences in gray matter between musicians and nonmusicians. *Annals of the New York Academy of Sciences*, *1060*(2005), 395–9. doi:10.1196/annals.1360.057

Besson, M., Chobert, J., & Marie, C. (2011a). Language and Music in the Musician Brain. *Language and Linguistics Compass*, *5*, 617–634. doi:10.1111/j.1749-818X.2011.00302.x

Besson, M., Chobert, J., & Marie, C. (2011b). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*(May), 94. doi:10.3389/fpsyg.2011.00094

Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*(1), 3–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9255429

Besson, M., Faita, F., Peretz, I., Bonnel, a.-M., & Requin, J. (1998). Singing in the Brain: Independence of Lyrics and Tunes. *Psychological Science*, *9*(6), 494–498. doi:10.1111/1467-9280.00091

Besson, M., & Schön, D. (2011). What remains of modularity? In P. Rebuschat, M. Rohrmeier, J. A. Hawkings, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 283–291). Oxford: Oxford university press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775. doi:10.1121/1.1332378

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011a). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–34. doi:10.1162/jocn.2009.21362

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011b). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, *77*(1), 1–10. doi:10.1016/j.bandc.2011.07.006

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011c). Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neuroscience Letters*, *503*(3), 203–7. doi:10.1016/j.neulet.2011.08.036

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PloS One*, *8*(4), e60676. doi:10.1371/journal.pone.0060676

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–25. doi:10.1016/j.brainres.2010.07.100

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(12), 2767–96. doi:10.1093/cercor/bhp055

Biswal, B., Yetkin, F. Z., Haughton, V. M., & Hyde, J. S. (1995). Functional Connectivity in the Motor Cortex of Resting Human Brain Using Echo-Planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541. doi:10.1002/mrm.1910340409

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer. Retrieved from http://www.praat.org/

Bonnel, a M., Faita, F., Peretz, I., & Besson, M. (2001). Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Perception & Psychophysics*, *63*(7), 1201–13. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11766944

Bradley, E. (2013). Pitch perception in lexical tone and melody. *Reviews of Research in Human Learning and Music*, (May). doi:10.6022/journal.rrhlm.2013002

Bradley, E. D. (2012). Tone language experience enhances sensitivity to melodic contour, (1).

Brown, R. M., Chen, J. L., Hollinger, A., Penhune, V. B., Palmer, C., & Zatorre, R. J. (2013). Repetition suppression in auditory-motor regions to pitch and temporal structure in music. *Journal of Cognitive Neuroscience*, *25*(2), 313–28. doi:10.1162/jocn_a_00322

Brown, S., Martinez, M. J., & Parsons, L. M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *The European Journal of Neuroscience*, *23*(10), 2791–803. doi:10.1111/j.1460-9568.2006.04785.x

Cambridge Dictionary of American Idioms. (2003). Cambridge: Cambridge University Press.

Cambridge Idioms Dictionary. (2006) (2nd ed.). Cambridge: Cambridge University Press.

Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology*, *13*(1), 61–68. doi:10.1111/j.2044-835X.1995.tb00664.x

Chandrasekaran, B., Kraus, N., & Wong, P. C. M. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, *107*(5), 1325–36. doi:10.1152/jn.00923.2011

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport*, *18*(18), 1963–7. doi:10.1097/WNR.0b013e3282f213c5

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*(1), 148–56. doi:10.1016/j.brainres.2006.10.064

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language*, *108*(1), 1–9. doi:10.1016/j.bandl.2008.02.001

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456–65. doi:10.1121/1.3445785

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover FMRI design. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(4), 958–66. doi:10.1093/cercor/bhs090

Chobert, J., François, C., Velay, J.-L., & Besson, M. (2012). Twelve Months of Active Musical Training in 8- to 10-Year-Old Children Enhances the Preattentive Processing of Syllabic Duration and Voice Onset Time. *Cerebral Cortex (New York, N.Y. : 1991)*, (2009). doi:10.1093/cercor/bhs377

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *The Journal of the Acoustical Society of America*, *131*(6), 4756–69. doi:10.1121/1.4714355

Costa, A., & Sebastián-Gallés, N. (2014). How does the bilingual experience sculpt the brain? *Nature Reviews. Neuroscience*, *15*(5), 336–45. doi:10.1038/nrn3709

Crinion, J. T., Green, D. W., Chung, R., Ali, N., Grogan, A., Price, G. R., … Price, C. J. (2009). Neuroanatomical markers of speaking Chinese. *Human Brain Mapping*, *30*(12), 4108–15. doi:10.1002/hbm.20832

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human primary auditory cortex follows the shape of Heschl's gyrus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(40), 14067–75. doi:10.1523/JNEUROSCI.2000-11.2011

Damoiseaux, J. S., Rombouts, S. a R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(37), 13848–53. doi:10.1073/pnas.0601417103

Dediu, D., & Ladd, D. R. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(26), 10944–9. doi:10.1073/pnas.0610848104

Degé, F., & Schwarzer, G. (2011). The effect of a music program on phonological awareness in preschoolers. *Frontiers in Psychology*, *2*(June), 124. doi:10.3389/fpsyg.2011.00124

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*(1), 46–61. doi:10.1080/09541440802708136

Dörnyei, Z. (2002). Attitudes, Orientations, and Motivations in Language Learning: Advances in Theory, Research, and Applications. *Language Learning*, *53*(1), 3–32. doi:10.1111/1467-9922.53222

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature*, *427*(6972), 311–312. doi:10.1038/427311a

Eisner, F., Mcgettigan, C., Faulkner, A., Rosen, S., & Scott, S. K. (2010). Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. *Journal of Neuroscience*, *30*(21), 7179–7186. doi:10.1523/JNEUROSCI.4040-09.2010.Inferior

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2011). Differential language expertise related to white matter architecture in regions subserving sensory-motor coupling, articulation, and interhemispheric transfer. *Human Brain Mapping*, *32*(12), 2064–74. doi:10.1002/hbm.21169

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex (New York, N.Y. : 1991)*, *22*(3), 650–8. doi:10.1093/cercor/bhr142

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential Influences of Emotion , Task , and Novelty on Brain Regions Underlying the Processing of Speech Melody. *Journal of Cognitive Neuroscience*, *21*(7), 1255–1268. doi:10.1162/jocn.2009.21099

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in Brain Research*, *156*, 345–361. doi:10.1016/S0079-6123(06)56019-4

Evans, P. D., Gilbert, S. L., Mekel-Bobrov, N., Vallender, E. J., Anderson, J. R., Vaez-Azizi, L. M., … Lahn, B. T. (2005). Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science (New York, N.Y.)*, *309*(5741), 1717–20. doi:10.1126/science.1113722

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences*, *1169*, 448–58. doi:10.1111/j.1749-6632.2009.04582.x

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470. doi:10.1006/jpho.1997.0052

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language*, *41*(1), 78–104. doi:10.1006/jmla.1999.2638

Foster, N. E. V, & Zatorre, R. J. (2010). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, *53*(1), 26–36. doi:10.1016/j.neuroimage.2010.06.042

Foxton, J. M., Dean, J. L., Gee, R., Peretz, I., & Griffiths, T. D. (2004). Characterization of deficits in pitch perception underlying “tone deafness”. *Brain : A Journal of Neurology*, *127*(Pt 4), 801–10. doi:10.1093/brain/awh105

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(9), 2038–43. doi:10.1093/cercor/bhs180

Gao, S., Hu, J., Gong, D., Chen, S., Kendrick, K. M., & Yao, D. (2012). Integration of consonant and pitch processing as revealed by the absence of additivity in mismatch negativity. *PloS One*, *7*(5), e38289. doi:10.1371/journal.pone.0038289

Garner, W., & Felfoldy, G. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *241*, 225–241. doi:10.1016/0010-0285(70)90016-2

Garner, W. R. (1983). Asymmetric interactions of stimulus dimensions in perceptual information processing. In T. J. Tighe & B. E. Sheep (Eds.), *Perception, Cognition, and Development: Interactional Analyses* (pp. 1–36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *23*(27), 9240–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14534258

Gierhan, S. M. E. (2013). Connections for auditory language in the human brain. *Brain and Language*, *127*(2), 205–21. doi:10.1016/j.bandl.2012.11.002

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011a). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, *2*(August), 146. doi:10.3389/fpsyg.2011.00146

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011b). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. *Frontiers in Psychology*, *2*(August), 146. doi:10.3389/fpsyg.2011.00146

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(3), 575–82. doi:10.1093/cercor/bhk001

Golestani, N., Price, C. J., & Scott, S. K. (2011). Born with an ear for dialects? Structural plasticity in the expert phonetician brain. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(11), 4213–20. doi:10.1523/JNEUROSCI.3891-10.2011

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage*, *21*(2), 494–506. doi:10.1016/j.neuroimage.2003.09.071

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language*, *109*(2-3), 55–67. doi:10.1016/j.bandl.2008.01.005

Gordon, R. L., Schön, D., Magne, C., Astésano, C., & Besson, M. (2010). Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. *PloS One*, *5*(3), e9889. doi:10.1371/journal.pone.0009889

Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2001). Musical experience and Mandarin tone discrimination and imitation. *Journal of the Acoustical Society of America*, *115*(1), 2545.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Grimault, S., Nolden, S., Lefebvre, C., Vachon, F., Hyde, K., Peretz, I., … Jolicoeur, P. (2014). Brain activity is related to individual differences in the number of items stored in auditory short-term memory for pitch: Evidence from magnetoencephalography. *NeuroImage*. doi:10.1016/j.neuroimage.2014.03.020

Gubian, M. (n.d.). Functional Data Analysis for Phonetics Research, 1–4.

Guerra-Carrillo, B., Mackey, A. P., & Bunge, S. a. (2014). Resting-State fMRI: A Window into Human Brain Plasticity. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, (February). doi:10.1177/1073858414524442

Gutschalk, A., & Uppenkamp, S. (2011). Sustained responses for pitch and vowels map to similar sites in human auditory cortex. *NeuroImage*, *56*(3), 1578–87. doi:10.1016/j.neuroimage.2011.02.026

Habib, M., & Besson, M. (2009). What do music training and musical experience teach us about brain plasticity. *Music Perception*, *26*(3), 279–285. doi:10.1525/MP.2009.26.3.279

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–23. doi:10.1016/j.tics.2005.07.004

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology*, *4*(July), 416. doi:10.3389/fpsyg.2013.00416

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, *28*, 136–141. doi:10.1016/j.conb.2014.07.013

Hagoort, P., & Indefrey, P. (2014). The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, *37*, 347–362. doi:10.1146/annurev-neuro-071013-013847

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual Differences in the Acquisition of a Complex L2 Phonology: A Training Study. *Language Learning*, *62*, 79–109. doi:10.1111/j.1467-9922.2012.00707.x

Hasson, U., Nusbaum, H. C., & Small, S. L. (2006). Repetition suppression for spoken sentences and the effect of task demands. *Journal of Cognitive Neuroscience*, *18*(12), 2013–29. doi:10.1162/jocn.2006.18.12.2013

Hasson, U., Skipper, J. I., Nusbaum, H. C., & Small, S. L. (2007). Abstract coding of audiovisual speech: beyond sensory representation. *Neuron*, *56*(6), 1116–26. doi:10.1016/j.neuron.2007.09.037

Hayasaka, S., & Nichols, T. E. (2003). Validating cluster size inference: random field and permutation methods. *NeuroImage*, *20*(4), 2343–2356. doi:10.1016/j.neuroimage.2003.08.003

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(May), 393–402. doi:10.1038/nrn2113

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., … Gabrieli, J. D. E. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(1), 361–6. doi:10.1073/pnas.1008950108

Hope, T. M. H., Prejawa, S., Parker Jones, ’Ōiwi, Oberhuber, M., Seghier, M. L., Green, D. W., & Price, C. J. (2014). Dissecting the functional anatomy of auditory word repetition. *Frontiers in Human Neuroscience*, *8*(May), 246. doi:10.3389/fnhum.2014.00246

Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin & Review*, *17*(3), 310–6. doi:10.3758/PBR.17.3.310

Hu, X., Ackermann, H., Martin, J. a, Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: behavioural predictors and neural substrates. *Brain and Language*, *127*(3), 366–76. doi:10.1016/j.bandl.2012.11.006

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(10), 3019–25. doi:10.1523/JNEUROSCI.5118-08.2009

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., & Peretz, I. (2007). Cortical thickness in congenital amusia: when less is better than more. *The Journal of*

Neuroscience : The Official Journal of the Society for Neuroscience, *27*(47), 13028–32. doi:10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., Zatorre, R. J., & Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(2), 292–9. doi:10.1093/cercor/bhq094

Iverson, P., Kuhl, P. K., Akahane-yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes, *87*, 47–57. doi:10.1016/S0

Javad, F., Warren, J. D., Micallef, C., Thornton, J. S., Golay, X., Yousry, T., & Mancini, L. (2014). Auditory tracts identified with combined fMRI and diffusion tractography. *NeuroImage*, *84*, 562–74. doi:10.1016/j.neuroimage.2013.09.007

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., Chen, X., & Yang, Y. (2012). Amusia results in abnormal brain activity following inappropriate intonation during speech comprehension. *PloS One*, *7*(7), e41411. doi:10.1371/journal.pone.0041411

Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, *48*(9), 2630–9. doi:10.1016/j.neuropsychologia.2010.05.009

Jones, J. L., Lucker, J., Zalewski, C., & Brewer, C. (2010). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders*, *42*(3), 226–234. doi:10.1016/j.jcomdis.2009.01.001.Phonological

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews Neuroscience*, *12*(4), 231–242. doi:10.1038/nrn3000

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117. doi:10.1016/j.wocn.2005.10.003

Kim, K. H., Relkin, N. R., Lee, K. M., & Hirsch, J. (1997). Distinct cortical areas associated with native and second languages. *Nature*, *388*(6638), 171–4. doi:10.1038/40623

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: vowels sing but consonants speak. *Cognition*, *112*(1), 1–20. doi:10.1016/j.cognition.2009.02.014

Kraus, N., & Banai, K. (2007). Auditory-Processing Malleability: Focus on Language and Music. *Current Directions in Psychological Science*, *16*(2), 105–110. doi:10.1111/j.1467-8721.2007.00485.x

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews. Neuroscience*, *11*(8), 599–605. doi:10.1038/nrn2882

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Research*, *1313*, 124–33. doi:10.1016/j.brainres.2009.11.061

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2012). Experience-dependent plasticity in pitch encoding: from brainstem to auditory cortex. *Neuroreport*, *23*(8), 498–502. doi:10.1097/WNR.0b013e328353764d

Krishnan, A., Gandour, J. T., Smalt, C. J., & Bidelman, G. M. (2010). Language-dependent pitch encoding advantage in the brainstem is not limited to acceleration rates that occur in natural speech. *Brain and Language*, *114*(3), 193–8. doi:10.1016/j.bandl.2010.05.004

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(20), 7877–81. doi:10.1073/pnas.1201575109

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. doi:10.1038/nrn1533

Lee, C.-Y., & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *124*(5), 3235–48. doi:10.1121/1.2990713

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics*, *53*(2), 157–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8433914

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, *38*(2), 202–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16956095

Lidji, P., Jolicoeur, P., Moreau, P., Kolinsky, R., & Peretz, I. (2009). Integrated preattentive processing of vowel and pitch: a mismatch negativity study. *Annals of the New York Academy of Sciences*, *1169*, 481–4. doi:10.1111/j.1749-6632.2009.04770.x

Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in congenital amusia: evidence from Mandarin speakers. *PloS One*, *7*(2), e30374. doi:10.1371/journal.pone.0030374

Liu, F., Patel, A. D., Fourcin, A., & Stewart, L. (2010). Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain : A Journal of Neurology*, *133*(Pt 6), 1682–93. doi:10.1093/brain/awq089

López-Barroso, D., Catani, M., Ripollés, P., Dell'acqua, F., Rodríguez-Fornells, A., & de Diego-Balaguer, R. (2013). Word learning is mediated by the left arcuate fasciculus. *Proceedings of the National Academy of Sciences of the United States of America*. doi:10.1073/pnas.1301696110

Lopez-Barroso, D., de Diego-Balaguer, R., Cunillera, T., Camara, E., Münte, T. F., & Rodriguez-Fornells, A. (2011). Language learning under working memory constraints correlates with microstructural differences in the ventral language pathway. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(12), 2742–50. doi:10.1093/cercor/bhr064

Loui, P., Hohmann, A., & Schlaug, G. (2010). Inducing Disorders in Pitch Perception and Production: a Reverse-Engineering Approach. *Proceedings of Meetings on Acoustics Acoustical Society of America*, *9*(1), 50002. doi:10.1121/1.3431713

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19558–63. doi:10.1073/pnas.0607065104

MacMillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211. doi:10.1162/089892906775783660

Maldjian, J. a., Laurienti, P. J., Kraft, R. a., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, *19*(3), 1233–1239. doi:10.1016/S1053-8119(03)00169-1

Mandell, J., Schulze, K., & Schlaug, G. (2007). Congenital amusia: an auditory-motor feedback disorder? *Restorative Neurology and Neuroscience*, *25*(3-4), 323–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17943009

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–15. doi:10.1162/jocn.2010.21585

Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(4), 447–57. doi:10.1016/j.cortex.2010.11.006

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–63. doi:10.1162/jocn.2007.19.9.1453

Masataka, N. (2007). Music, evolution and language. *Developmental Science*, *10*(1), 35–9. doi:10.1111/j.1467-7687.2007.00561.x

McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends in Cognitive Sciences*, *16*(5), 269–76. doi:10.1016/j.tics.2012.04.006

McMullen, E., & Saffran, J. R. (2004). Music and Language: A Developmental Comparison. *Music Perception*, *21*(3), 289–311. doi:10.1525/mp.2004.21.3.289

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Neurolinguistics: Structural plasticity in the bilingual brain. *Nature*, *431*(7010), 3017. doi:10.1038/nature03016

Mekel-Bobrov, N., Gilbert, S. L., Evans, P. D., Vallender, E. J., Anderson, J. R., Hudson, R. R., … Lahn, B. T. (2005). Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science (New York, N.Y.)*, *309*(5741), 1720–2. doi:10.1126/science.1116815

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. a a, & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: neural and behavioral evidence. *Brain Research*, *1194*, 81–9. doi:10.1016/j.brainres.2007.11.042

Molnar-Szakacs, I., Iacoboni, M., Koski, L., & Mazziotta, J. C. (2005). Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cerebral Cortex (New York, N.Y. : 1991)*, *15*(7), 986–94. doi:10.1093/cercor/bhh199

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2012). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433. doi:10.1177/0956797611416999.Short-Term

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(3), 712–23. doi:10.1093/cercor/bhn120

Mühlau, M., Rauschecker, J. P., Oestreicher, E., Gaser, C., Röttinger, M., Wohlschläger, a M., … Sander, D. (2006). Structural brain changes in tinnitus. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(9), 1283–8. doi:10.1093/cercor/bhj070

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473–478. doi:10.1038/nrn843

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–8. doi:10.1073/pnas.0701498104

Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychological Science*, *20*(7), 895–903. doi:10.1111/j.1467-9280.2009.02380.x

Myers, E. B., & Swan, K. (2012). Effects of category learning on neural sensitivity to non-native phonetic categories. *Journal of Cognitive Neuroscience*, *24*(8), 1695–708. doi:10.1162/jocn_a_00243

Nan, Y., & Friederici, A. D. (2012). Differential roles of right temporal cortex and Broca's area in pitch processing: Evidence from music and Mandarin. *Human Brain Mapping*, *000*. doi:10.1002/hbm.22046

Nan, Y., Friederici, A. D., Shu, H., & Luo, Y.-J. (2009). Dissociable pitch processing mechanisms in lexical and melodic contexts revealed by ERPs. *Brain Research*, *1263*, 104–13. doi:10.1016/j.brainres.2009.01.030

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain : A Journal of Neurology*, *133*(9), 2635–42. doi:10.1093/brain/awq178

Nguyen, S., Tillmann, B., Gosselin, N., & Peretz, I. (2009). Tonal language processing in congenital amusia. *Annals of the New York Academy of Sciences*, *1169*, 490–3. doi:10.1111/j.1749-6632.2009.04855.x

Obleser, J., Zimmermann, J., Van Meter, J., & Rauschecker, J. P. (2007). Multiple stages of auditory speech perception reflected in event-related FMRI. *Cerebral Cortex (New York, N.Y. : 1991)*, *17*(10), 2251–7. doi:10.1093/cercor/bhl133

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch--functional evidence of speech-relevant auditory acuity. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(2), 447–55. doi:10.1093/cercor/bhp113

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5146491

Ott, C. G. M., Langer, N., Oechslin, M. S., Meyer, M., & Jäncke, L. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Frontiers in Psychology*, *2*(September), 195. doi:10.3389/fpsyg.2011.00195

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(45), 14100–7. doi:10.1523/JNEUROSCI.3256-09.2009

Park, H.-J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science (New York, N.Y.)*, *342*(6158), 1238411. doi:10.1126/science.1238411

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford university press.

Patel, A. D. (2011). Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. *Frontiers in Psychology*, *2*(June), 142. doi:10.3389/fpsyg.2011.00142

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications. *Annals of the New York Academy of Sciences*, *1252*, 124–8. doi:10.1111/j.1749-6632.2011.06426.x

Patel, A. D., Xu, Y., & Wang, B. (2010). The role of F0 variation in the intelligibility of Mandarin sentences. In *Proceedings of speech prosody*. Chicago IL USA.

Peretz, I. (2006). The nature of music from a biological perspective. *Cognition*, *100*(1), 1–32. doi:10.1016/j.cognition.2005.11.004

Peretz, I. (2009). MUSIC, LANGUAGE AND MODULARITY FRAMED IN ACTION. *Psychologica Belgica*, *49*, 157–175.

Peretz, I., Brattico, E., Järvenpää, M., & Tervaniemi, M. (2009). The amusic brain: in tune, out of key, and unaware. *Brain : A Journal of Neurology*, *132*(Pt 5), 1277–86. doi:10.1093/brain/awp055

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–91. doi:10.1038/nn1083

Peretz, I., Nguyen, S., & Cummings, S. (2011). Tone language fluency impairs pitch discrimination. *Frontiers in Psychology*, *2*(July), 145. doi:10.3389/fpsyg.2011.00145

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*(1), 461–72. doi:10.1121/1.3593366

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musican pitch in tone language speakers, *71*(6), 1385–1398. doi:10.3758/APP

Plack, C. J., & Oxenham, A. J. (2005). Overview: The present and future of pitch. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 1–6). New York, NY: Springer.

Poser, B. a, Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine : Official Journal* *of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *55*(6), 1227–35. doi:10.1002/mrm.20900

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–47. doi:10.1016/j.neuroimage.2012.04.062

Price, C. J., & Friston, K. J. (2005). Functional ontologies for cognition: The systematic definition of structure and function. *Cognitive Neuropsychology*, *22*(3), 262–75. doi:10.1080/02643290442000095

Rauschecker, A. M., Pringle, A., & Watkins, K. E. (2008). Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Human Brain Mapping*, *29*(11), 1231–42. doi:10.1002/hbm.20460

Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information in speech perception: a cross-linguistic study. *Journal of Phonetics*, *18*, 481–495.

Ressel, V., Pallier, C., Ventura-Campos, N., Díaz, B., Roessler, A., Ávila, C., & Sebastián-Gallés, N. (2012). An effect of bilingualism on the auditory cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(47), 16597–601. doi:10.1523/JNEUROSCI.1996-12.2012

Rodriguez-Fornells, A., De Diego Balaguer, R., & Münte, T. F. (2006). Executive Control in Bilingual Language Processing. *Language Learning*, 133–190. doi:10.1111/j.1467-9922.2006.00359

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. doi:10.1068/p5117

Rota, G., Sitaram, R., Veit, R., Erb, M., Weiskopf, N., Dogil, G., & Birbaumer, N. (2009). Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. *Human Brain Mapping*, *30*(5), 1605–14. doi:10.1002/hbm.20621

Sadakata, M., & McQueen, J. M. (n.d.). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*.

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: a cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10. doi:10.1016/j.actpsy.2011.03.007

Sammler, D., Baird, A., Valabrègue, R., Clément, S., Dupont, S., Belin, P., & Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(10), 3572–8. doi:10.1523/JNEUROSCI.2751-09.2010

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., … Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(46), 18035–40. doi:10.1073/pnas.0805234105

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–4. doi:10.1111/j.0956-7976.2004.00711.x

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, *98*(2), 457–468. doi:10.1037/0022-0663.98.2.457

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–94. doi:10.1038/nn871

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450–61. doi:10.1016/j.neuroimage.2010.02.023

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–9. doi:10.1111/1469-8986.00172.x

Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., & Näätänen, R. (2007). Heschl ' s Gyrus , Posterior Superior Temporal Gyrus , and Mid-Ventrolateral Prefrontal Cortex Have Different Roles in the Detection of Acoustic Changes, 2075–2082. doi:10.1152/jn.01083.2006.

Sebastián-Gallés, N., Soriano-Mas, C., Baus, C., Díaz, B., Ressel, V., Pallier, C., … Pujol, J. (2012). Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics*, *25*(3), 150–162. doi:10.1016/j.jneuroling.2011.11.001

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: a review of fMRI studies. *Neuropsychologia*, *51*(1), 59–66. doi:10.1016/j.neuropsychologia.2012.11.006

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, *124*(2), 128–42. doi:10.1016/j.cognition.2012.05.008

Skoe, E., & Kraus, N. (2011). Human subcortical auditory function provides a new conceptual framework for considering modularity. In P. Rebuschat, M. Rohrmeier, J. A. Hawkings, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 269–282). Oxford: Oxford university press.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter? *Psychological Science*, *17*(8), 675–81. doi:10.1111/j.1467-9280.2006.01765.x

Sluming, V., Matthew, H., & Cezayirli, E. (2002). Voxel-Based Morphometry Reveals Increased Gray Matter Density in Broca ' s Area in Male Symphony Orchestra Musicians, *1622*, 1613–1622.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., … Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13040–5. doi:10.1073/pnas.0905267106

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology : CB*, *19*(6), 498–502. doi:10.1016/j.cub.2009.01.066

Stevens, C. J., Keller, P. E., & Tyler, M. D. (2011). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, *41*(1), 59–74. doi:10.1177/0305735611415749

Stewart, L. (2008). Fractionating the musical mind: insights from congenital amusia. *Current Opinion in Neurobiology*, *18*(2), 127–30. doi:10.1016/j.conb.2008.07.008

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hearing Research*, *261*(1-2), 22–9. doi:10.1016/j.heares.2009.12.021

Strait, D. L., O'Connell, S., Parbery-Clark, A., & Kraus, N. (2013). Musicians' Enhanced Neural Differentiation of Speech Sounds Arises Early in Life: Developmental Evidence from Ages 3 to 30. *Cerebral Cortex (New York, N.Y. : 1991)*. doi:10.1093/cercor/bht103

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language*, *123*(3), 191–201. doi:10.1016/j.bandl.2012.09.001

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., … Kawashima, R. (2010). Training of working memory impacts structural connectivity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(9), 3297–303. doi:10.1523/JNEUROSCI.4611-09.2010

Tervaniemi, M., Jacobsen, T., Röttger, S., Kujala, T., Widmann, a, Vainio, M., … Schröger, E. (2006). Selective tuning of cortical sound-feature processing by language experience. *The European Journal of Neuroscience*, *23*(9), 2538–41. doi:10.1111/j.1460-9568.2006.04752.x

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, a D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *The European Journal of Neuroscience*, *30*(8), 1636–42. doi:10.1111/j.1460-9568.2009.06955.x

Thornton, A., & Lukas, D. (2012). Individual variation in cognitive performance: developmental and evolutionary perspectives. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1603), 2773–83. doi:10.1098/rstb.2012.0214

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex (New York, N.Y. : 1991)*, *23*(2), 249–54. doi:10.1093/cercor/bhs003

Tomasi, D., & Volkow, N. D. (2012). Resting functional connectivity of language networks: characterization and reproducibility. *Molecular Psychiatry*, *17*(8), 841–54. doi:10.1038/mp.2011.177

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–89. doi:10.1006/nimg.2001.0978

Ueno, T., Saito, S., Rogers, T. T., & Lambon Ralph, M. a. (2011). Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. *Neuron*, *72*(2), 385–96. doi:10.1016/j.neuron.2011.09.013

Vaden, K. I., Muftuler, L. T., & Hickok, G. (2010). Phonological repetition-suppression in bilateral superior temporal sulci. *NeuroImage*, *49*(1), 1018–23. doi:10.1016/j.neuroimage.2009.07.063

Van den Heuvel, M. P., Mandl, R. C. W., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human Brain Mapping*, *30*(10), 3127–41. doi:10.1002/hbm.20737

Vandermosten, M., Boets, B., Poelmans, H., Sunaert, S., Wouters, J., & Ghesquière, P. (2012). A tractography study in dyslexia: neuroanatomic correlates of orthographic, phonological and speech processing. *Brain : A Journal of Neurology*, *135*(Pt 3), 935–48. doi:10.1093/brain/awr363

Ventura-Campos, N., Sanjuán, A., González, J., Palomar-García, M.-Á., Rodríguez-Pujadas, A., Sebastián-Gallés, N., … Ávila, C. (2013). Spontaneous brain activity predicts learning ability of foreign sounds. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(22), 9295–305. doi:10.1523/JNEUROSCI.4655-12.2013

Wang, Y., Sereno, J. a, Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of Cognitive Neuroscience*, *15*(7), 1019–27. doi:10.1162/089892903770007407

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(1), 61–9. doi:10.1523/JNEUROSCI.3489-08.2009

Westlye, L. T., Grydeland, H., Walhovd, K. B., & Fjell, A. M. (2011). Associations between regional cortical thickness and attentional networks as measured by the attention network test. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(2), 345–56. doi:10.1093/cercor/bhq101

Wildgruber, D., Riecker, a, Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage*, *24*(4), 1233–41. doi:10.1016/j.neuroimage.2004.10.034

Witteman, J., Van Heuven, V. J. P., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*, *50*(12), 2752–2763. doi:10.1016/j.neuropsychologia.2012.07.026

Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *The*

Journal of Neuroscience : The Official Journal of the Society for Neuroscience, *31*(24), 8780–5. doi:10.1523/JNEUROSCI.0999-11.2011

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PloS One*, *7*(4), e33424. doi:10.1371/journal.pone.0033424

Wong, P. C. M., & Perrachione, T. K. (2007a). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04), 565–585. doi:10.1017/S0142716407070312

Wong, P. C. M., & Perrachione, T. K. (2007b). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04), 565–585. doi:10.1017/S0142716407070312

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping*, *28*(10), 995–1006. doi:10.1002/hbm.20330

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–2. doi:10.1038/nn1872

Wong, P. C. M., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., & Zatorre, R. J. (2008). Volume of left Heschl's Gyrus and linguistic pitch learning. *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(4), 828–36. doi:10.1093/cercor/bhm115

Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4174782

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, *95*(4), 2240–2253. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8201120

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., … Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping*, *27*(2), 173–83. doi:10.1002/hbm.20176

Yan, J., Zhang, Y., & Ehret, G. (2005). Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *Journal of Neurophysiology*, *93*(1), 71–83. doi:10.1152/jn.00348.2004

Zarate, J. M., Delhommeau, K., Wood, S., & Zatorre, R. J. (2010). Vocal accuracy and neural plasticity following micromelody-discrimination training. *PloS One*, *5*(6), e11181. doi:10.1371/journal.pone.0011181

Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science (New York, N.Y.)*, *342*(6158), 585–9. doi:10.1126/science.1238414

Zatorre, R. J., Delhommeau, K., & Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Frontiers in Psychology*, *3*(December), 544. doi:10.3389/fpsyg.2012.00544

Zatorre, R. J., Evans, a C., Meyer, E., & Gjedde, a. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science (New York, N.Y.)*, *256*(5058), 846–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1589767

Zatorre, R. J., Fields, R. D., & Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nature Neuroscience*, *15*(4), 528–36. doi:10.1038/nn.3045

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1087–104. doi:10.1098/rstb.2007.2161

Samenvatting

Dit proefschrift richt zich op het begrijpen van wat de individuele variatie bij de prestaties van geluidsverwerking inspireert, in het bijzonder met betrekking tot het verwerken van toonhoogte in de domeinen van spraak en muziek. Hoewel vele factoren aan deze variatie ten grondslag kunnen liggen, werden twee daarvan hier onderzocht: het effect van de ervaring en het effect van het talent en de aanleg. De fysieke geluidseigenschap, die in alle hoofdstukken onderzocht werd, was frequentie, met zijn psycho-akoestische correlaat toonhoogte. Toonhoogte (pitch) speelt een centrale rol in zowel spraak- als ook muziekklanken en werd dus in de voorliggende experimenten in beide (taalkundige en muzikale) contexten gebruikt. Hoofdstuk 2 beoordeelde gedrags- en neuro-imaging studies en toonde aan dat ervaring met toonhoogte in ofwel een taalkundige of een muzikale context de verwerking van toonhoogte boven de gegeven context uit, kan beïnvloeden. Personen die muzikale opleiding hebben ontvangen zijn beter in het verwerken van toonhoogte in taal. Hoewel ze schaarser zijn, bestaan er toch ook aanwijzingen dat mensen die ervaring hebben met toonhoogte door middel van taal, in het bijzonder toontaal sprekers, beter zijn dan niet-toontaal sprekers in de verwerking van toonhoogte in een muzikale of domein-neutrale context. De beoordeling in hoofdstuk 2 stelde voor dat deze bi-directionele invloeden optreden, omdat spraak en muziek gebruik maken van dezelfde corticale en subcorticale geluidsverwerkings resourcen. Deze worden in beide domeinen door ervaring gevormd, wat dus tot algemene geluidsverwerking voordelen leidt.

Naar aanleiding van het centrale uitgangspunt in hoofdstuk 2 keek de studie in hoofdstuk 3 naar de toonhoogte verwerking in een groep van Cantonees-Nederlandse twee-taligen. Deze groep is vanaf een zeer jonge leeftijd blootgesteld geweest aan twee talen die verschillend gebruik maken van taalkundige toonhoogte en kreeg dus waardevolle "auditieve ervaring" met toonhoogte. Er werd derhalve verwacht dat Cantonees-Nederlandse twee-taligen hun eentalige Nederlandse branchegenoten in een reeks van niet-talige toonhoogte verwerkingsopdrachten zouden overtreffen. Uit de resultaten bleek dat tweetalige ervaring met een toon- en een niet-toontaal invloed heeft op de manier waarop melodische (toonhoogte) en fonologische (klinker) gegevens in gezongen spraak online worden verwerkt, waarbij de twee-taligen de twee stimulus dimensies integraler verwerkten dan controles. De studie vond hoe dan ook, geen toonervarings voordeel, niet in niet-linguïstische toonhoogte discriminatie taken, of in een muzikale interval leertaak. Daarom werd geconcludeerd dat, hoewel de ervaring met toon enig effect op niet-talige toonhoogte verwerking geeft, dit effect vrij zwak en moeilijk op te wekken is.

De grote individuele variatie in prestaties in de muzikaal naïeve eentalige controlegroep, zoals waargenomen in de toonhoogte perceptie experimenten die in Hoofdstuk 3 gerapporteerd zijn, benadrukten de belangrijke factor van het talent en aanleg. Ervaring is ongetwijfeld cruciaal en kan de prestaties in toonhoogte verwerking beïnvloeden, zoals ondersteund door longitudinale studies. De variatie, die bij onervaren mensen gevonden wordt, geeft echter aan dat niet iedereen dezelfde aanleg of talent met geluiden heeft. Het is gesuggereerd, dat verschillen in aptitude afstammen van verschillende hersen structuur en functie, tussen individuen. Er is een grote variatie in de menselijke hersen morfologie die zou kunnen resulteren bij sommige individuen met neuro-anatomische en/of functionele hersen karakteristieken die in het voordeel zijn voor het verwerken van geluid. Hoofdstukken 4 en 5 focusten zich op de rol van aptitude bij het onderzoeken van de neurologische correlaten van individuele variatie in geluidsleerprestatie.

Het experiment in hoofdstuk 4 behelsde de efficiëntie met welke geluiden verwerkt worden in de hersenen en hoe deze efficiëntie meer in het voordeel zou kunnen zijn voor zekere individuen dan anderen met het leren van nieuwe linguïstische geluiden. Nederlandse deelnemers zonder enige voorgaande ervaring met toon werden getraind in vijf afzonderlijke sessies om de niet-oorspronkelijke (Mandarijn) tonale contrasten gelegen in woorden te identificeren. Ze werden gevraagd om te leren om de overeenkomst te vinden van in totaal vierentwintig vocaal gepresenteerde woorden en plaatjes van alledaagse items. Leerprestatie werd vastgesteld als accuratesse (percentage correct) in een woord-afbeelding gelijkenis taak. Neurale efficiëntie werd gemeten met een fMRI-aanpassings paradigma net als het repetition suppression effect, m.a.w. de afname van activiteit door de herhaalde presentatie van de niet-oorspronkelijke tonale contrast. Repetition suppression met tonen werd gemeten op twee verschillende momenten: voor en na de training. De resultaten gaven aan dat des te beter de prestatie in het leren van de niet-oorspronkelijke tonale contrast des te groter het repetition suppression effect naar de toon in de linker Inferior Frontal Gyrus. Dit effect was significant zelfs voordat deelnemers begonnen waren met de training discussiërende in het voordeel van vroeger bestaande verschillen in de aptitude voor het verwerken van geluid. Succesvolle leerlingen verwerkten geluiden efficiënter en bleken minder begeleiding nodig te hebben van hoger orde frontale taalgebieden wanneer er geluisterd wordt naar niet-oorspronkelije tonen.

Het experiment in hoofdstuk 5 was een complement op het experiment van hoofdstuk 4, omdat het ook keek naar variatie in het BOLD signal in relatie tot toon leerprestatie, alleen dit keer lag de focus op rust activiteit, dat betekent, wanneer de deelnemers (degene die getest zijn in hoofdstuk 4) niet bezig waren in enige taak binnen in de scanner. De staat van rust verbinding werd vooraf gemeten en na vijf sessies lexicale toon training en de verbindingspatronen in bilaterale gehoorsgebieden en de linker frontale gebieden werden geanalyseerd in relatie tot de leerprestatie. De resultaten gaven weer dat de kracht van rust-staat verbinding tussen de linker inferieure frontale en de linker posterieure superieure temporale gebieden in de na-training sessie negatief correleerde met leerprestatie. Deze gebieden omvatten de ventrale route voor taal, welke wordt aangezien als belangrijk voor het arrangeren van geluid tot betekenis (zoals werd vereist in de leertaak). De betrekking van functionele verbinding langs de ventral route is in overeenstemming met eerdere bevindingen, welke laten zien dat witte materie verbinding op deze route positief gecorreleerd is met woord leerprestatie.

In contrast tot de resultaten in hoofdstuk4, die demonstreerden dat de efficiëntie met welke de linker IFG niet-oorspronkelijke geluiden verwerkt voorspellend is van toekomstige leerprestaties, gaven de rustende staat verbindings patronen in hoofdstuk 5 geen voortraining verschillen weer tussen leerlingen. Desondanks waren verschillen in aptitude evident in de na-trainnigs sessie, en zo ook in de verbindings veranderingen van de voor tot de na trainings sessies. Deze resultaten tesamen met bevindingen in hoofdstuk 4, geven aan dat individuele variatie in niet-oorspronkelijke geluids leerprestatie, een gevolg is van meer efficiënte neurale verwerking van geluiden, alsmede de meer efficiënte functionele verbindings veranderingen tussen frontale en temporale hersengebieden.

Het effect van aptitude met geluiden was getest in dit proefschrift, allen in het spraak domein. Het zou van grote belangstelling zijn om het leren van geluiden parallel te testen in de linguïstische en muzikale domeinen, gebruikmakende van dezelfde populatie, om te kunnen zien of de getalenteerde spraak leerlingen dezelfden zullen zijn als de getalenteerde muziek leerlingen. Verder onderzoek dat leren in beide domeinen vergelijkt is nodig om de gradatie van overlapping en divergentie te vast te stellen, tussen de neurale substraten van muziek en spraak-geluid lerende aptitude.

De studies gerapporteerd in dit proefschrift demonstreren dat ondanks ervaring, in de vorm van muzikale training of ervaring met een toontaal, invloed kan hebben op toonhoogte verwerkingsprestatie, er grote individuele verschillen in aptitude zijn welke ook verantwoordelijk zouden kunnen zijn voor een groot deel van de waargenomen variatie in prestatie.

Terwijl de bijdragen van elke factor niet direkt vergeleken werdenn in dit proefschrift, wijzen de bevindingen naar een mogelijke interactie tussen hen, zodat, bijvoorbeeld, individuen met aptitude voor het leren van geluiden, waarschijnlijk meer doorgang zullen hebben in muziek en/of taal leren and sneller zullen profiteren van oefenen dan hun peers. Zoals in elke instantiering van de afkomst versus het verzorgings debat, ligt het antwoord ergens in het midden: beide aptitude en ervaring beïnvloeden geluid verwerkingsprestatie, en verklaren zo een groot deel van de individuele variatie in gedrag. Het is daarom aan te raden om deze factoren in interactie te overwegen wanneer prestaties worden geëvalueerd.

Om samen te vatten, dit proefschrift heeft bewijs gepresenteerd dat het hebben van “een oor voor toonhoogte“ in taal en muziek, afstammen van zowel lange termijn ervaring met toonhoogte, welke domein algemene toonhoogte verwerkingsbronnen vormen, alsmede aptitude, in de vorm van meer efficiënte neurale verwerking van toonhoogte en meer efficiënte functionele verbindingen in de hersenen.

Curriculum Vitae

Salomi S. Asaridou completed her undergraduate studies in psychology at the Aristotle University of Thessaloniki in Greece where she graduated with honors. She continued her studies as a Fulbright scholar in the US where she obtained a MSc degree in Neuroscience and Education from Columbia University, New York. In 2010 she was awarded an IMPRS for Language Sciences fellowship by the Max Planck Society to perform her doctoral research at the Max Planck Institute for Psycholinguistics and the Donders Institute for Brain, Cognition and Behaviour in the Netherlands. She is currently a postdoctoral researcher in the Department of Neurology, at the University of California, Irvine.

Publications

Asaridou, S. S., Takashima, A., Dediu, D., Hagoort, P., & McQueen, J. M. (under revision). Repetition suppression in the left Inferior Frontal Gyrus predicts tone learning performance.

Asaridou, S. S., Hagoort, P., & McQueen, J. M. (submitted). Effects of early bilingual experience with a tone and a non-tone language on speech- music integration.

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology, 4*, 321. doi:10.3389/fpsyg.2013.00321

Fessler, D. M., Stieger, S., Asaridou, S. S., Bahia, U., Cravalho, M., de Barros, P., … Voracek, M. (2011). Testing a postulated case of intersexual selection in humans: The role of foot size in judgments of physical attractiveness and age. *Evolution and Human Behavior, 33*(2), 147-164. doi:10.1016/j.evolhumbehav.2011.08.002

Proios, H., Asaridou, S. S., & Brugger, P. (2008). Random Number Generation in Patients with Aphasia: A Test of Executive Functions. *Acta Neuropsychologica, 6*(2), 157-168.

MPI Series in Psycholinguistics

1. *The electrophysiology of speaking. Investigations on the time course of semantic, syntactic and phonological processing*. Miranda van Turennout

2. *The role of the syllable in speech production. Evidence from lexical statistics, metalinguistics, masked priming and electromagnetic midsaggital articulography.* Niels O. Schiller

3. *Lexical access in the production of ellipsis and pronouns*. Bernadette M. Schmitt

4. *The open-/closed-class distinction in spoken-word recognition*. Alette Haveman

5. *The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach*. Kay Behnke

6. *Gesture and speech production*. Jan-Peter de Ruiter

7. *Comparative intonational phonology: English and German*. Esther Grabe

8. *Finiteness in adult and child German*. Ingeborg Lasser

9. *Language input for word discovery*. Joost van de Weijer

10. *Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe.* James Essegbey

11. *Producing past and plural inflections*. Dirk Janssen

12. *Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea*. Anna Margetts

13. *From speech to words*. Arie van der Lugt

14. *Simple and complex verbs in Jaminjung: A study of event categorization in an Australian language*. Eva Schultze-Berndt

15. *Interpreting indefinites: An experimental study of children's language comprehension*. Irene Krämer

16. *Language specific listening: The case of phonetic sequences*. Andrea Weber

17. *Moving eyes and naming objects*. Femke van der Meulen

18. *Analogy in morphology: The selection of linking elements in Dutch compounds*. Andrea Krott

19. *Morphology in speech comprehension*. Kerstin Mauth

20. *Morphological families in the mental lexicon*. Nivja H. de Jong

21. *Fixed expressions and the production of idioms*. Simone A. Sprenger

22. *The grammatical coding of postural semantics in Goemai*. Birgit Hellwig

23. *Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies*. Fermin Moscoso del Prado Martin

24. *Contextual influences on spoken-word processing*. Daniëlle van den Brink

25. *Perceptual relevance of prevoicing in Dutch*. Petra M. van Alphen

26. *Syllables in speech production: Effects of syllable preparation and syllable frequency*. Joana Cholin

27. *Producing complex spoken numerals for time and space*. Marjolein Meeuwissen

28. *Morphology in auditory lexical processing*. Rachèl J.J.K. Kemps

29. *At the same time ...: The expression of simultaneity in learner varieties*. Barbara Schmiedtová

30. *A grammar of Jalonke argument structure*. Friederike Lüpke

31. *Agrammatic comprehension: An electrophysiological approach*. Marlies Wassenaar

32. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Frank Seifart

33. *Prosodically-conditioned detail in the recognition of spoken words*. Anne Pier Salverda

34. *Phonetic and lexical processing in a second language*. Mirjam Broersma

35. *Retrieving semantic and syntactic word properties: ERP studies on the time course in language comprehension*. Oliver Müller

36. *Lexically-guided perceptual learning in speech processing*. Frank Eisner

37. *Sensitivity to detailed acoustic information in word recognition*. Keren B. Shatzman

38. *The relationship between spoken word production and comprehension.* Rebecca Özdemir

39. *Disfluency: Interrupting speech and gesture*. Mandana Seyfeddinipur

40. *The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation*. Christiane Dietrich

41. *Cognitive cladistics and the relativity of spatial cognition*. Daniel B.M. Haun

42. *The acquisition of auditory categories*. Martijn Goudbeek

43. *Affix reduction in spoken Dutch*. Mark Pluymaekers

44. *Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence*. Valesca Kooijman

45. *Space and iconicity in German Sign Language (DGS)*. Pamela Perniss

46. *On the production of morphologically complex words with special attention to effects of frequency*. Heidrun Bien

47. *Crosslinguistic influence in first and second languages: Convergence in speech and gesture*. Amanda Brown

48. *The acquisition of verb compounding in Mandarin Chinese.* Jidong Chen

49. *Phoneme inventories and patterns of speech sound perception*. Anita Wagner

50. *Lexical processing of morphologically complex words: An information-theoretical perspective*. Victor Kuperman

51. *A grammar of Savosavo: A Papuan language of the Solomon Islands*. Claudia Wegener

52. *Prosodic structure in speech perception and production*. Claudia Kuzla

53. *The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension*. Sarah Schimke

54. *Studies on intonation and information structure in child and adult German*. Laura de Ruiter

55. *Processing the fine temporal structure of spoken words*. Eva Reinisch

56. *Semantics and (ir)regular inflection in morphological processing*. Wieke Tabak

57. *Processing strongly reduced forms in casual speech*. Susanne Brouwer

58. *Ambiguous pronoun resolution in L1 and L2 German and Dutch*. Miriam Ellert

59. *Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging.* Ian FitzPatrick

60. *Processing casual speech in native and non-native language*. Annelie Tuinman

61. *Split intransitivity in Rotokas, a Papuan language of Bougainville*. Stuart Robinson

62. *Evidentiality and intersubjectivity in Yurakaré: An interactional account*. Sonja Gipper

63. *The influence of information structure on language comprehension: A neurocognitive perspective*. Lin Wang

64. *The meaning and use of ideophones in Siwu*. Mark Dingemanse

65. *The role of acoustic detail and context in the comprehension of reduced pronunciation variants*. Marco van de Ven

66. *Speech reduction in spontaneous French and Spanish*. Francisco Torreira

67. *The relevance of early word recognition: Insights from the infant brain*. Caroline Junge

68. *Adjusting to different speakers: Extrinsic normalization in vowel perception*. Matthias Sjerps

69. *Structuring language: Contributions to the neurocognition of syntax*. Katrien R. Segaert

70. *Infants' appreciation of others' mental states in prelinguistic communication*. Birgit Knudsen

71. *Gaze behavior in face-to-face-interaction*. Federico Rossano

72. *Sign-spatiality in Kata Kolok: How a village sign language of Bali inscribes its signing space*. Connie de Vos

73. *Who is talking?* Attila Andics

74. *Lexical processing of foreign-accented speech: Rapid and flexible adaptation*. Marijt Witteman

75. *The use of deictic versus representational gestures in infancy*. Daniel Puccini

76. *Territories of knowledge in Japanese conversation*. Kaoru Hayano

77. *Family and neighbourhood relations in the mental lexicon: A cross-language perspective.* Kimberley Mulder

78. *Contributions of executive control to individual differences in word productions*. Zeshu Shao

79. *Hearing and seeing speech: Perceptual adjustments in auditory-visual processing* . Patrick van der Zande

80. *High pitches and thick voices: The role of language in space-pitich association.* Sarah Dolscheid

81. *Seeing what's next: Processing and anticipating language referring to objects*. Joost Rommers

82. *Mental representations and processing of reduced words in casual speech*. Iris Hanique

83. *The many ways listeners adapt to reductions in casual speech*. Katja Poellmann

84. *Contrasting opposite polarity in Germanic and Romance languages: Verum Focus and affirmative particles in native speakers and advanced L2 learners.* Giuseppe Turco

85. *Morphological processing in younger and older people: Evidence for flexible dual-route access*. Jana Reifegerste

86. *Semantic and syntactic constraints on the production of subject-verb agreement*. Alma Veenstra

87. *The acquisition of morphophonological alternations across languages*. Helen Buckler

88. *The evolutionary dynamics of motion event encoding*. Annemarie Verkerk

89. *Rediscovering a forgotten language*. Jiyoun Choi

90. *The road to native listening: Language-general perception, language-specific input.* Sho Tsuji

91. *Infants' understanding of communication as participants and observers*. Gudmundur Bjarki Thorgrímsson

92. *Information structure in Avatime*. Saskia van Putten

93. *Switch reference in Whitesands.* Jeremy Hammond

94. Machine learning for gesture recognition from videos. Binyam Gebrekidan Gebre

95. *Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish.* Beyza Sumer

96. *An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music.* Salomi S. Asaridou