Based on results evidencing music training-related advantages on speech processing, perceptive and cognitive functions, we examine whether music training facilitates novel word learning throughout the lifespan. We show that musically-trained children and young professional musicians outperform controls in a series of experiments, with faster brain plasticity and stronger functional connectivity, as measured by electroencephalography. By contrast, advantages for old adult musicians are less clear-cut, suggesting a limited impact of music training to counteract cognitive decline. Finally, young musicians show better long-term memory for novel words, which possibly contributes, along with better auditory perception and attention, to their advantage in word learning. By showing transfer effects from music training to semantic processing and long-term memory, results reveal the importance of domain-general cognitive functions and open new perspectives for education and rehabilitation.

Aix*Marseille université
UMR 7309 · CNRS · AMU PAROLE ET LANGAGE
CNRS
BLRI Brain & Language Research Institute
Laboratoire de Neurosciences Cognitives

FROM AUDITORY PERCEPTION TO MEMORY: Musicianship as a window into novel word learning | Eva Dittinger

FROM AUDITORY PERCEPTION TO MEMORY:
Musicianship as a window into novel word learning

Eva Dittinger

From Auditory Perception to Memory:

Musicianship as a Window into Novel Word Learning

Cover by Tincho Schmidt

Printed and bound by Duplica Print (www.duplicaprint.com)

The research reported in this thesis was carried out within the Laboratoire de Neurosciences cognitives (LNC) in the Fédération 3C of the CNRS and Aix-Marseille University, the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002). It has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). Eva Dittinger was supported by a PhD grant from the Labex BLRI.

UNIVERSITE D'AIX-MARSEILLE

Ecole Doctorale Science de la Vie et de la Santé (ED 62)

Laboratoire de Neurosciences Cognitives
& Laboratoire Parole et Langage

Thèse présentée pour obtenir le grade universitaire de docteur
Discipline: Neurosciences

Eva Maria DITTINGER

"From Auditory Perception to Memory: Musicianship as a Window into Novel Word Learning"

"De la perception auditive à la mémoire: la musicalité comme fenêtre sur l'apprentissage de nouveaux mots"

Soutenue le 18/12/2018 devant le jury :

Minna HUOTILAINEN	Department of Education, Cognitive Brain Research Unit (CBRU), University of Helsinki, Helsinki, Finland
Martin MEYER	Division Neuropsychology, Institute of Psychology, University of Zürich, Zürich, Switzerland
Cheryl FRENCK-MESTRE	CNRS & Aix-Marseille Université, Laboratoire Parole et Langage, Aix-en-Provence, France
Corine ASTESANO	Université de Toulouse, Unité de Recherche Interdisplinaire Octogone-Lordat, Toulouse, France
Mariapaola D'IMPERIO	CNRS & Aix-Marseille Université, Laboratoire Parole et Langage, Aix-en-Provence, France
Mireille BESSON	CNRS & Aix-Marseille Université, Laboratoire de Neurosciences Cognitives, Marseille, France

"Music gives a soul to the universe, wings to the mind,

flight to the imagination and life to everything."

Plato

ABSTRACT

Intensive music training positively impacts auditory perception, as well as cognitive functions such as auditory attention, working and verbal memory, executive functions, and general intelligence. Music training-related advantages also transfer to several levels of language processing. However, whether music training influences the semantic level of speech processing has never been investigated.

The aim of the present work was twofold: first, we evaluated whether music training facilitated the learning of novel words. Considering that, in our interconnected world, novel words need to be acquired at all ages, we aimed at understanding how potential transfer effects from music training to word learning changed throughout the lifespan. Having shown a positive impact of musicianship on word learning, the second goal was to better understand the origin of this effect and to determine whether musicians showed better long-term memory for newly learned words than non-musicians.

Related to our first goal, participants (school-aged children, younger and older adults, with and without music training) performed a series of experiments: they categorized novel words with native or non-native initial phonemes, learned the meaning of these words through picture-word associations, and were tested for successful word learning (matching task) and meaning integration into semantic networks (semantic task). Young professional musicians discriminated native from non-native phonemes better than non-musicians with larger differences in brain activity, suggesting that fine auditory perception is an important first step for better word learning. Moreover, both musically-trained children and young adult musicians outperformed controls without music training, with faster developing N400 components during the learning phase and larger N400s to unexpected than to expected words (N400 effects) during the test phase. By contrast, the influence of music training in older adults was not as clear-cut, possibly showing a limited impact of music training in counteracting the deleterious effects of aging. Finally, for young adults, we also analyzed functional connectivity in the ventral and dorsal streams during word learning. Musicians showed increased left-hemispheric theta connectivity in both streams, suggesting that music training modified the functional architecture of the networks involved in semantic processing and in sensorimotor integration.

Related to our second goal, we modified our original design to neutralize the contribution of auditory attention and perception. Under these conditions, non-musicians performed as well as musicians (young adults) when tested immediately after learning. When retested one month later, however, musicians showed better long-term memory for novel words and larger N400 effects than controls. These results reveal that, along with better auditory perception and attention, enhanced long-term memory in musicians is possibly one of the driving factors contributing to the musician’s advantage in word learning. Theoretically, demonstrating transfer effects from music training to semantic processing and long-term memory further reveal the importance of domain-general cognitive functions, and challenge modular views of the anatomo-functional organization of brain functions.

Taken together, these results open new perspectives for education in showing that early music training can facilitate foreign language learning. They also offer new possibilities for the rehabilitation of people with language impairments through a musical route.

Keywords:

musical expertise, brain plasticity, novel word learning, semantic processing,

long-term memory, EEG, N400, transfer effects

RÉSUMÉ

L’entrainement musical intensif améliore la perception auditive, ainsi que l'attention auditive, la mémoire de travail et la mémoire verbale, les fonctions exécutives et l'intelligence générale. Les avantages liés à la formation musicale transfèrent également à différents niveaux du traitement linguistique. Cependant, l’impact de la formation musicale au niveau sémantique n’a jamais été étudié.

Le but de ce travail est double : premièrement, déterminer si la formation musicale facilite l'apprentissage de nouveaux mots. Considérant que, dans notre monde interconnecté, de nouveaux mots doivent être acquis à tous les âges, nous avons cherché à comprendre comment les effets potentiels du transfert de la formation musicale à l'apprentissage des mots changent au long de la vie. Ayant démontré un impact positif, le deuxième but était de mieux comprendre l’origine de cet effet et de déterminer si les musiciens ont une meilleure mémoire à long terme pour les mots nouveaux que les non-musiciens.

En lien avec notre premier but, les participants (enfants d’âge scolaire, adultes jeunes et plus âgés, avec ou sans formation musicale) ont effectué une série d’expériences : ils catégorisaient des nouveaux mots avec des phonèmes initiaux natifs ou non, ils apprenaient le sens des nouveaux mots grâce à des associations images-mots, et ils étaient testés pour l’apprentissage des mots et leur intégration dans les réseaux sémantiques. Les jeunes musiciens professionnels distinguent mieux les phonèmes non-natifs que les non-musiciens, avec des différences plus importantes dans l'activité cérébrale, suggérant qu’une perception auditive fine est une première étape importante pour un meilleur apprentissage des mots. De plus, comparés aux enfants « musiciens » et les jeunes adultes musiciens montrent un développement plus rapide de la composante N400 pendant la phase d'apprentissage, un meilleur niveau de performance et des N400s plus amples pour les mots inattendus que pour les mots attendus (effets N400) pendant la phase de test. En revanche, les résultats chez les adultes âgés n’étaient pas aussi clairs, suggérant que la formation musicale a un impact limité pour contrecarrer les effets délétères du vieillissement cognitif. Enfin, les jeunes musiciens montrent une augmentation de connectivité fonctionnelle dans les voies ventrales et dorsales au cours de l'apprentissage des mots, ce qui suggère que la formation musicale modifie l'architecture fonctionnelle des réseaux impliqués dans le traitement sémantique et l'intégration sensorimotrice.

En lien avec notre deuxième but, nous avons modifié notre protocole expérimental afin de neutraliser la contribution de l'attention et de la perception auditives. Dans ces conditions, les non-musiciens sont aussi performants que les musiciens (jeunes adultes) lorsqu'ils sont testés immédiatement après avoir appris. Mais un mois plus tard, les musiciens ont une meilleure mémoire à long terme des mots nouveaux et des effets N400 plus amples que les participants de contrôle. Ces résultats révèlent, qu'ajoutée à une meilleure perception et attention auditives, l'amélioration de la mémoire à long terme chez les musiciens serait l'un des facteurs déterminant l'avantage du musicien dans l'apprentissage des nouveaux mots. Théoriquement, montrer des effets de transfert de la formation musicale vers le niveau sémantique et la mémoire à long-terme révèle l'importance des fonctions cognitives générales et remet en question les vues modulaires de l'organisation anatomo-fonctionnelle des fonctions cérébrales.

Pris dans leur ensemble, ces résultats ouvrent de nouvelles perspectives d'éducation en montrant qu'une formation musicale précoce peut faciliter l'apprentissage des langues étrangères. Ils offrent également de nouvelles possibilités de rééducation, par un parcours musical, aux personnes ayant des troubles du langage.

Mots clés :

Expertise musicale, plasticité cérébrale, apprentissage de nouveaux mots,

traitement sémantique, mémoire à long terme, EEG, N400, effets de transfert

ACKNOWLEDGEMENTS

Having started my educational journey into the direction of molecular biotechnology several years ago, it seems unbelievable that this same journey first, is arriving to an end, and second, ends with a PhD in cognitive neuroscience. I cannot help thinking of all the amazing moments that have accompanied me throughout this years and all the coincidences that have brought me to this point. I call it “journey”, since most of my energy was (and is still) driven by curiosity to discover new countries, new people, new ways of thinking. On this journey my path crossed many people who enriched me with ideas and curiosities, and taught me to develop a grand passion for learning new languages. Having played myself an instrument from childhood on, it may not be surprising that the possibility to do a PhD on transfer effects from music training to foreign language learning inspired myself much more than many other topics in the world of research. Before this journey, in which I have learnt and grown so much, both personally and professionally, comes to an end, I will try to put in words how thankful I am for the continuous support from several people, without whom I would not have been able to take and bring to an end this fantastic journey.

First of all, my great thanks to Mireille and Mariapaola, for supervising me during more than four years of research. Looking back, it seems incredible that you, Mireille, welcomed me in your team and trusted me to succeed in this project even though I had no formal education in cognitive neuroscience. Without your confidence, as well as your time to accompany me during these years I would have not been able to do this thesis. Moreover, your energy to take me or to send me abroad, made me get to know several internationally respected researchers, students and several new angles of this world. Thanks for your continuous motivation, and for having contributed to make these years not only fruitful but also a lot of fun. Thanks, Mariapaola, for your similar trust in me succeeding in this project and for accepting me as your PhD student. Without your motivation and patience to discuss basic linguistic issues, many questions and discussions that came up during these last years would have been more difficult to be answered.

That I succeeded in running my experiments, and in analyzing and publishing my data, is in no small part also thanks to our international collaborator, Stefan Elmer. Whenever suggestions in electrophysiological brain data analysis or more dynamic views on experimental data or research in general were needed, he was very generous with his time, energy and positivity. Many thanks for that.

Afterwards, thanks to all the members of our research group, Jean-Luc Velay, Marieke Longcamp, Jérémy Danna, Mylène Barbaroux, Sarah Palmis, Elie Fabiani, and Lauriane Veron-Delor, for their patience with my initially poor knowledge of French, for their crash course on French culture, for smiling moments and last but not least for having been a great audience for many presentations and discussions. I am furthermore deeply indebted to all the students who invested energy in the framework of my thesis, Mylène Barbaroux, Benjamin Furnari, Elizaveta Parfenova, Johanna Scherer and Betina Korka, and without any doubt also to all the coauthors of my publications. Thanks to all of you!

Moreover, great thanks to all the members of both labs that I am affiliated to, the Laboratoire de Neurosciences Cognitives in Marseille and the Laboratoire Parole et Langage in Aix-en-Provence, for continuous support, fruitful discussions and warm celebrations. Additionally, thanks to the collective Auditory Research Group in Zürich for having me hosted and supported for two months. In this context, special thanks to Stefan Elmer and Lutz Jäncke. They both warmly welcomed me to Zürich and were available for suggestions and discussions whenever needed.

I am furthermore deeply grateful to the Brain and Language Research Institute for having provided me with the financial means to do my research. Additionally, the Institute provided a social and scientific network facilitating excellent research. I greatly enjoyed my time as a member of the Institute, and this network would not exist without the social initiative and energy of Philippe Blache. Great thanks to you! Moreover, I really appreciate having met Nadéra Bureau, whose support was essential for several administrative questions and for all my wonderful experiences and conferences abroad. Her effective work within the Institute deserves big recognition.

Thanks also to the numerous group of people that contributed to the research reported in this thesis: a total of more than 150 participants engaged in the novel word learning tasks of Chapters 3 to 6. Their anonymity should not be confused with lack of importance.

Finally, thanks to James McQueen, Kristof Strijkers, Leonardo Lancia and Christina Tremblay for their suggestions and feedbacks during my “Comité de suivi de these”, and thanks to all the members of this jury for taking time and energy to evaluate the work presented here.

Additionally, it should not be forgotten that the path to a PhD begins much earlier than the actual PhD: I am grateful to John Trichereau in Vienna, Sylvia Krobitsch in Berlin and Giorgio Inghirami in Turin for having supported my first steps on the path to becoming a researcher.

Beyond that, I am also deeply grateful to my parents, Gertraud and Erich, to my sister Michaela and to my whole family. All of them continuously supported me to take the journey I wanted to, and believed in that my decisions were the right ones and would make me happy. Furthermore, thanks to my friends, back in Vienna, in Berlin and in Turin. These last four years were also a time of great emotional challenges, accompanied by motivation and positivity, but also by doubt and frustration. All of them motivated me in one way or the other to continue this journey, and without their emotional support despite their geographical distance, I could not have brought together the will to finish this PhD. It is impossible to name them all, but the names which come to mind spontaneously are Caro and Gabriel, Monika and Philipp, Mirjam and Romy, Tine, Hestie, Sylvi, Annika, Claudio, Simone, Marco, Rosina and Davide, Federica, and Roberto. Thanks again! Great thanks also to all the people I met in Marseille, Michel, Jean-Claude, Gerome, Mathieu, Agnès and Nicolas, Christina, Logan, Elena, Isabell, Lionel, Kelly, Florent and Manuela, who provided regular (or less regular) social scaffold that picked me up when the PhD got me down. Special thanks to Stéphanie for her recharging yoga classes and her success to contemporary free my mind from research questions. I am immensely grateful to have met them all, and to have shared my immersion into French culture with them.

Last but definitely not least, deepest thanks to my companion Matteo, for his infinite patience and positivity, and for his taking wonderful care of our son Kimo whenever I needed time to work on this thesis.

While I am the sole PhD author and I alone take responsibility of its content, I could not have done it without you all. Great thanks to all of you for unwaveringly believing in me completing this PhD, and for accompanying me on this journey: feel proud of what I achieved with your generous help.

Table of contents

Chapter 1

INTRODUCTION

1. 1. The musician brain as a model of brain plasticity

1.1.1. Music in the brain: an introduction

The boom of investigating the effects of music on the brain possibly started 25 years ago with the so-called Mozart-effect[1]. This effect was described as a temporary improvement of visuo-spatial abilities after listening to sonatas from Mozart (Rauscher, Shaw, & Ky, 1993). While these results were not very convincing, scientifically speaking, since they were based on an experimental design not fully controlled for other influences, the reaction of the media was enormous and encouraged the public belief that "music makes intelligent". Since then, hundreds of studies have tried to evidence the positive effect of music listening and music training on cognitive abilities, as well as on the anatomical and functional brain organization. However, what is it that makes cognitive neuroscientists, psychologists, music scientists as well as musicians themselves so interested in studying music?

Musicality is an essential part of human life and society, and humans all over the world comprehend, make and share music. And by no way, music is only a perceptive and cognitive capacity, but it is rather viewed as a collection of forms of human actions (i.e., a physical process): musicians' bodies interact with instruments to create vibrations that travel through the air (Bishop & Goebl, 2017). Before going into detail of how studying music may contribute to better understand the brain, two main points should be made: first, there is no single region for music in the brain (as there is no single musical experience). Rather, there are distributed areas of the brain forming networks to give rise to the different aspects of music. And second, music offers an appropriate test case for hypotheses related to how brain areas and networks enable human behavior. In fact, by studying music (but not only), cognitive neuroscience shifted in recent years from "modularity" (Fodor, 1983), a focus on identifying individual brain areas that subserve specific cognitive functions, to a more network-based approach.

Coming back to music: when music reaches the ear, variable and largely-distributed brain areas are activated. Initially, the sound reaches the auditory brainstem at the cochlear nucleus and travels up the subcortical pathway to the thalamus. Then, the signal reaches the superior

---

[1] This is not to say that philosophers or scientists were not interested in studying musical sounds beforehand. In fact, the first studies date back to Pythagoras around 570-500 before Christ. However, scientific studies relating musical experience to the brain and body using modern empirical methods have rapidly increased during these past 25 years.

temporal lobe with the primary auditory cortex (Heschl's gyrus) and the secondary auditory cortex (superior temporal plane). Finally, the sounds travel beyond auditory cortices, and several brain areas seem to be involved during music processing: the premotor cortex for action selection such as identification of rhythmic sequences (Janata & Grafton, 2003), the occipital lobe during musical imagery (Herholz, Halpern, & Zatorre, 2012), superior parietal regions and the supramarginal gyrus in the parietal lobe for musical learning and memory (Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003), and the prefrontal cortex for tracking tonal structures (Janata et al., 2002) or jazz improvisation (Limb & Braun, 2008). In sum, music requires multiple functions and engages distributed brain networks (Loui & Przysinda, 2017, Figure 1), thereby supplying neuroscientists with rich possibilities of how to test hypotheses regarding regional and/or network views of the human brain.

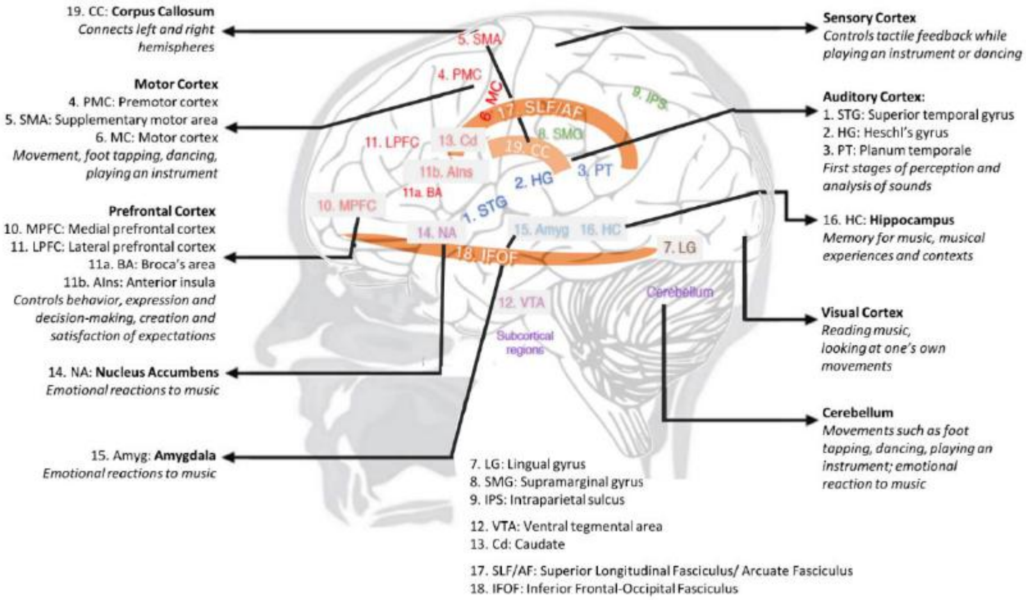


Figure 1. Some of the brain regions involved in music perception and cognition. Regions shaded in grey represent mesial structures that are rather seen in cross-section views than in this present surface view of the brain. Regions shaded in orange represent white matter pathways. Colors of text represent general areas of the brain: Blue = temporal lobe; red = frontal lobe; brown = occipital lobe; green = parietal lobe; pink and violet = subcortical structures. (Adapted from Loui & Przysinda, 2017 and Levitin, 2006)

1.1.2. Musical expertise: learning, practice and neuroplasticity

Although music perception and the positive effects of music listening on the brain are more than fascinating to me, this thesis focuses on the influence of regular music training on the brain rather than on "music" itself. Thus, each time I speak about musicians I refer to instrumentalists who acquired their musical skills through years of classical music training, excluding jazz musicians as well as vocalists.

This is where neuroplasticity comes into play: research in the last century has provided us with important evidence about the malleability of our brain as a function of our environment. In fact, the structure, function and connectivity of our nervous system can be reorganized in order to make proper responses to intrinsic and extrinsic stimuli (Bayona, Bitensky, & Teasell, 2005; Cramer et al., 2011). In other words, the ability to learn a new skill is thought to be determined by the capacity to create new functional connections allowing the brain to integrate new materials and activities (Bayona et al., 2005). This capacity is called neuroplasticity (or brain plasticity). Potentially any training or experience can induce neural changes, and neuroplasticity has been demonstrated in several domains and brain regions, both at the structural and functional neural levels (Erickson et al., 2007). Consequently, also behavioral alterations, that require modifications in the underlying neural circuits, are possible (Kolb & Muhammad, 2014).

Recently, Zatorre and colleagues (Zatorre, Fields, & Johansen-Berg, 2012) summarized several different plastic mechanisms linking the molecular/cellular alterations with their translation into grey and white matter changes observed by using Magnetic Resonance Imaging (MRI) methods[2]. Such grey matter changes can be caused by several events: first, axon sprouting. Second, neurogenesis or creation of new neurons. Third, gliogenesis, a process that has been reported to be increased in response to learning and experience. Forth, changes in dendritic organization, as well as creation of new dendritic spines and synapses. Fifth, vascular changes, in form of angiogenesis or accompanying neurogenesis. Finally, also some signaling molecules such as BDNF ("Brain-Derived Neurotrophic Factor") can promote plastic cellular changes, affecting neurons, glial cells or blood vessels (Kolb & Muhammad, 2014; Zatorre et al., 2012). White matter microstructure alterations, as measured by diffusion imaging technologies, include

---

[2] For more details on the MRI methodology I refer the reader to Chapter 2.

axon branching and myelination of unmyelinated axons, changes in axonal diameter and fiber organization, changes in astrocyte morphology or number, and angiogenesis (see Figure 2).

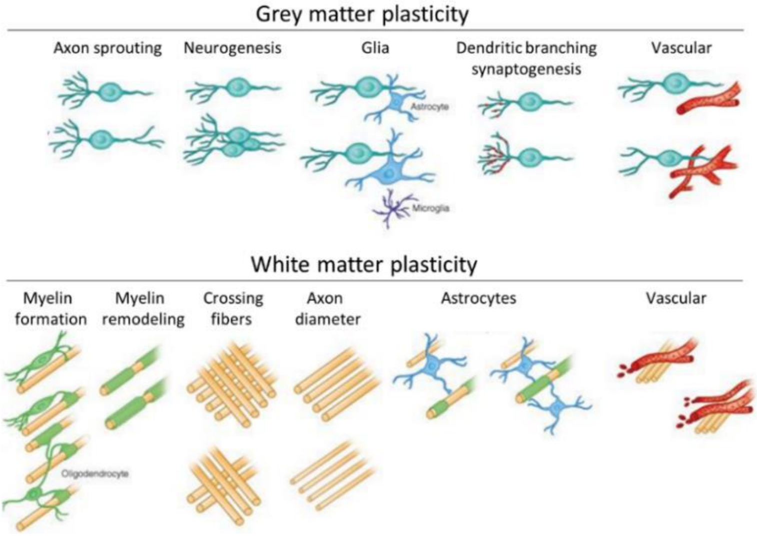


Figure 2. Cellular and molecular mechanisms of gray and white matter plasticity. (Adapted from Zatorre et al., 2012)

Interestingly, although manifestations of neuroplasticity have been described to be present over the course of the lifetime, mechanisms for this adaptive capacity tend to decrease in older adults (Freitas, Farzan, & Pascual-Leone, 2013), thereby raising the question of whether there could be sensitive periods across the lifespan. In other words, do experiences we have in life have different effects depending on the developmental state of our brain? In fact, for several brain functions and skills, the existence of a critical period during normal development in which the effect of experience on the brain is particularly strong (Hensch, 2005; Huttenlocher, 2003) and can contribute to long-lasting changes in brain structure and function as well as in behavior (Penhune, 2011) have been proposed. This question is still a debated issue, but in the last decades the concepts of neural plasticity, neurogenesis and brain repair have been redefined and have resulted in a more dynamic view of the adult brain (DeFelipe, 2006), shedding doubts on the existence of a sensitive time window, at least for some perceptive functions and cognitive abilities such as language and music learning in adults (Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009).

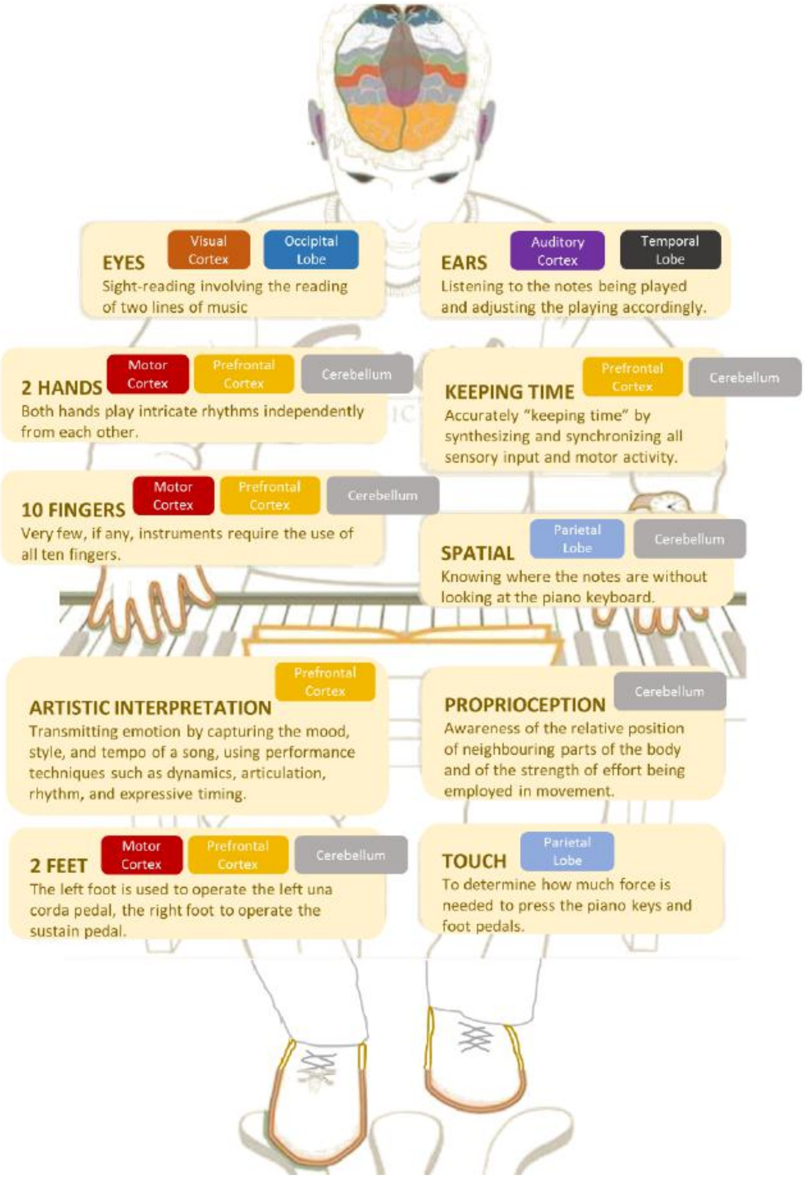


Figure 3. Multimodality of piano playing. (Adapted from http://www.leadingnotestudios.com)

Turning to music training, it is clear that playing an instrument at a professional level is arguably among the most complex human activities. For example, a pianist has to produce (and coordinate the production) of up to 1800 notes per minute (Münte, Altenmüller, & Jäncke, 2002), and the multimodality of such a production requires fine motor skills, limb and bimanual coordination, auditory and somatosensory perception, audio-motor integration, and several cognitive functions such as attention, executive functions and memory, thereby engaging distributed brain networks (Figure 3). Moreover, music training includes a great amount of emotional content that transforms musical practice in an enjoyable and motivational experience (Koelsch, 2010; Pantev & Herholz, 2011; Schmithorst & Wilke, 2002; Zatorre, Chen, & Penhune, 2007). Thus, it may not come as a surprise that regular training with music is linked to widespread anatomical and functional changes in the brain based on neuroplasticity, not only during childhood but throughout the lifespan.

1.1.3. Training-related brain plasticity

Numerous studies have tried to investigate whether professional music training has an influence on the perceptive, cognitive and motor functions recruited when playing an instrument. When exploring the link between musical expertise and the brain, cross-sectional and longitudinal approaches can be used. In cross-sectional studies, levels of performance and/or brain correlates of professional musicians and non-musicians in a variety of tasks are compared to investigate potential associations between music training and the brain. However, only longitudinal studies can provide evidence regarding the causality of effects from music training onto the brain: in this case, non-musicians are assigned randomly to a music education program or to a non-musical control activity (e.g., painting) for several weeks or years. Pre- and post-training data are then compared between groups. Importantly, since longitudinal studies are costly in terms of time and financial means, it is essential to first demonstrate between-group differences with cross-sectional designs before running longitudinal studies to demonstrate causality. Reporting all cross-sectional and longitudinal studies aiming at understanding music-related influences on the brain is clearly impossible: here I report only few studies highlighting the importance of studying professional musicians to investigate plastic changes in the human brain. I refer interested readers to more complete reviews (Herholz & Zatorre, 2012; Jäncke, 2009; Wan & Schlaug, 2010) or to several chapters within the recently published book “The Routledge Companion to Music Cognition” (Edited by Ashley & Timmers, Routledge, 2017).

One of the first pioneering study in the field of the neuroscience of music was conducted by Schlaug and colleagues (Schlaug, Jäncke, Huang, & Steinmetz, 1995) who demonstrated that professional musicians had a larger anterior half of the corpus callosum (see Figure 4) compared to non-musician controls. Interestingly, these differences were driven by musicians who had begun their training before the age of 7 years. This is an interesting issue: professional musicians are typically characterized by the fact that they start music training at an early age and that this training continues throughout their entire life, with several hours per day of playing their instrument. In fact, the early onset of training continues to be an important parameter when studying music-related brain changes: for example, in a study with pianists using Diffusion Tensor Imaging (DTI)[3], Bengtsson and colleagues (Bengtsson et al., 2005) found that several white matter tracts correlated with the estimated amount of musical practice during childhood, although the estimated number of practicing hours was lower in this period than in adolescence and adulthood. Similarly, in a more recent DTI study, Steele and colleagues (Steele, Bailey, Zatorre, & Penhune, 2013) found that early-trained musicians (age of start before 7 years) outperformed late-trained musicians (age of start after 7 years) in a visuo-motor synchronization task, although early- and late-trained groups did not differ in their years of musical experience, years of formal training or hours of current practice. The authors also observed a negative correlation between the age of onset of music training and fractional anisotropy[4] in the posterior mid-body of the corpus callosum (Figure 4), so that the earlier the onset of musical practice, the better the white-matter microstructure in that region. The authors interpreted these findings as evidence for a sensitive period and hypothesized that training during this period may induce changes in white matter connectivity that serve as a scaffold on which later training can build (Steele et al., 2013).

Not only age of onset, but also the enormous intensity of music training in professional musicians has been taken into consideration in several studies: for example, Ericsson and colleagues (Ericsson, Krampe, & Tesch-romer, 1993) demonstrated that excellent music playing is positively correlated with the number of training hours. Specifically, they computed the accumulated number of training years until the 18th birthday of their violin players, and obtained 7400 training hours for the best violinists (nominated by music professors of the Music Academy of West Berlin to have the potential for careers as international soloists), compared to 5300 training hours for the very good ones. Since all musicians had graduated from a prestigious music

---

[3] DTI is a magnetic resonance imaging method that measures the diffusion patterns of molecules in biological tissue thereby providing information on the microscopic structure of that tissue. It is commonly used to image neuronal fiber tracts.

[4] Fractional anisotropy describes the degree of anisotropy of a diffusion process: it can have a value between zero and one, where zero means that diffusion is unrestricted (or equally restricted) in all directions, and one means that diffusion occurs only along one axis and is fully restricted along all other directions. In DTI it is thought to reflect fiber density, axonal diameter, and myelination in white matter.

school and therefore obtained a similar quality of musical education, the authors concluded that the amount of practice was one of the most important factors influencing musical expertise.

Regarding plasticity of the auditory cortex (Figure 4), one of the first cross-sectional studies was provided by Pantev and colleagues (Pantev et al., 1998). In this study, the authors computed equivalent current dipoles[5] from evoked magnetic fields[6] in response to piano tones and to pure tones of equal fundamental frequency and loudness. In musicians, responses to piano tones were about 25 percent larger than in non-musicians, thereby evidencing increased receptive functions in response to isolated tones in musicians. Moreover, Schneider and colleagues (Schneider et al., 2002) showed structural and functional plasticity of the primary auditory cortex in professional musicians compared to non-musicians: musicians showed twice as large activity in response to pure tones 20-30 ms after stimulus onset (corresponding to the earliest evoked activity in the primary auditory cortex) compared to controls, as well as on average 130 percent enhancement in grey matter volume in the bilateral Heschl's gyrus (Figure 4). Interestingly, both measures were also highly correlated with musical aptitude, which led the authors to conclude that both the neurophysiology as well as the morphology of the Heschl's gyrus played an essential role in the processing of music. Similarly, Gaser and Schlaug (Gaser & Schlaug, 2003) showed increased grey matter density in professional musicians in brain regions such as the inferior temporal gyrus, precentral gyrus, inferior frontal gyrus, as well as in the left Heschl's gyrus and cerebellum (Figure 4).

Regarding plasticity of the motor cortex, for example, Elbert and colleagues (Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995) revealed that the cortical representation of the digits of the left hand of string players (who typically practice with the left hand controlling the manipulation of the strings and thereby experiencing enhanced sensory motor stimulation) was larger than in controls. No such differences were observed for the representations of the right hand that manipulates the bow, involving less distal dexterity. Moreover, the cortical reorganization in the representation of the fingering digits was correlated with the age of onset of music training, leading the authors to suggest that the representation of different parts of the

---

[5] Calculating a set of equivalent current dipoles is a source-modeling technique for magnetoencephalography, allowing to gain information about the location of relevant brain activity based on the registered data.

[6] Evoked magnetic fields are part of the magnetoencephalogram. They are brain signals that can be recorded after multiple repetition of a similar sensory stimulation. For more details on the methodology of magnetoencephalography, I refer the reader to Chapter 2.

body in the primary somatosensory cortex (Figure 4) of humans depends on changes made to conform to the current needs of each individual.

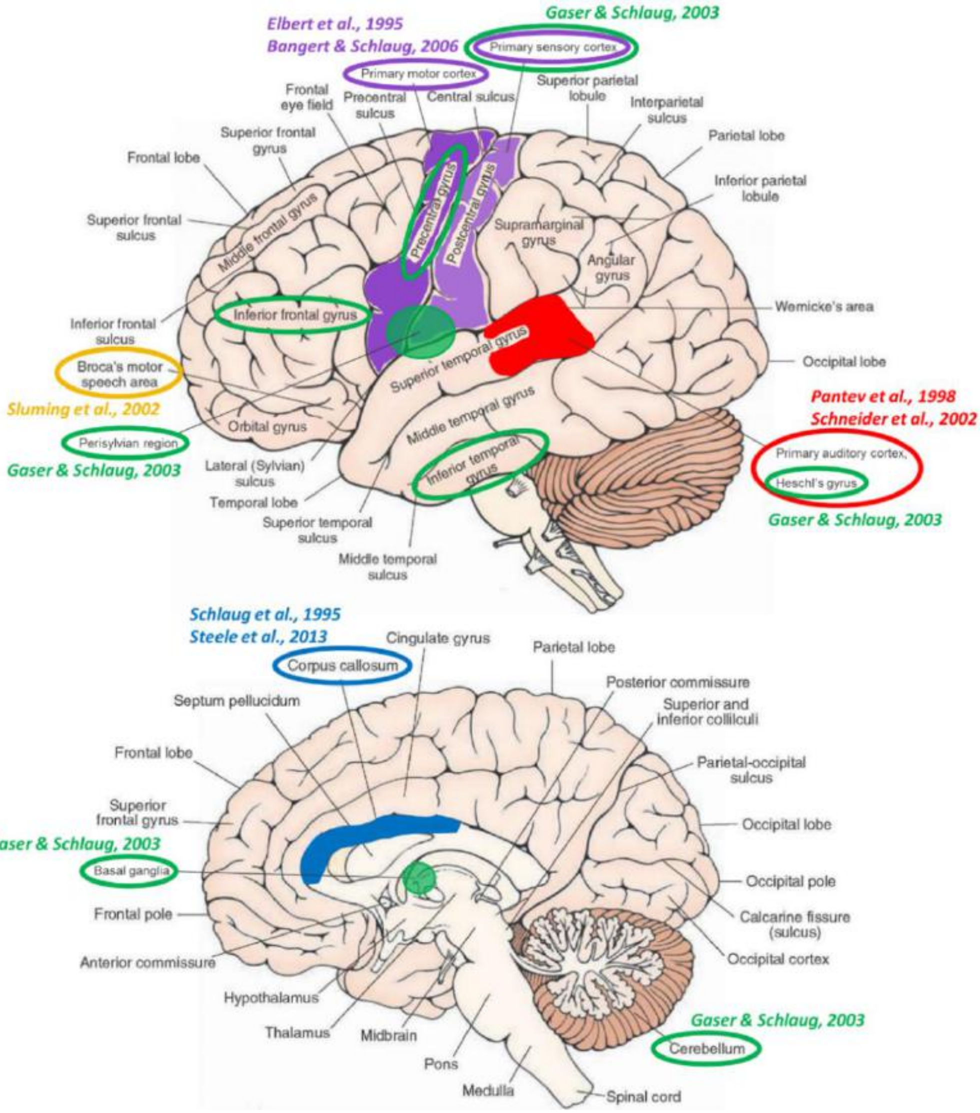


Figure 4. Lateral view of the cerebral cortex showing the principal gyri and sulci (above) and midsagittal view of the brain (below). Brain areas underlying music training-related brain plasticity, as evidenced by exemplary studies, are marked. (Adapted from http://what-when-how.com/neuroscience/overview-of-the-central-nervous-system-gross-anatomy-of-the-brain-part-1/)

Similarly, Bangert and Schlaug (Bangert & Schlaug, 2006) also reported specific changes in the anatomical features in musicians as a function of the instrument played: while pianists (who typically practice with both hands) were characterized by larger hand motor areas in both hemispheres, violinists (practicing with the left hand) showed such an effect mainly over the right hemisphere, controlling the left hand. Furthermore, by using a voxel-based morphometry approach[7], Gaser and Schlaug (2003) evidenced increased grey matter volume in musicians in the motor network, including the bilateral primary sensorimotor regions, the left basal ganglia, the bilateral cerebellum, as well as the left posterior perisylvian region (Figure 4). Interestingly, the authors also found a strong association between these structural differences and musical status as well as practice intensity, thereby supporting the view that practice has an impact on brain anatomy.

Finally, Sluming and colleagues (Sluming et al., 2002) also reported anatomical differences in Broca's area[8] (Figure 4) between musicians and non-musicians. Specifically, musicians showed increased grey matter volume in the left inferior frontal gyrus compared to non-musicians. Interestingly, these between-group differences increased with age: the musicians' grey matter density in the frontal cortex remained stable with age, while non-musician controls showed significant volume reductions.

Having reviewed some of the numerous studies reporting functional and/or structural differences between musicians and non-musicians, a general finding of these studies is that nearly all of the brain areas involved in the control of musical expertise (motor cortex, auditory cortex, frontal and parietal regions, cerebellum and other areas) show specific anatomical and functional plasticity in professional musicians, thereby supporting the idea of increased brain plasticity possibly driven by musical training. These plasticity effects on brain structure and on functional patterns of activation are translated into benefits for music- and non-music-related activities at the behavioral level (Besson, Chobert, & Marie, 2011; Kraus & Chandrasekaran, 2010; Moreno & Bidelman, 2014; Patel, 2014): for example, long-term musical training has been reported to enhance (i) auditory analysis, encoding of complex sounds and auditory discrimination (Bidelman, Gandour, & Krishnan, 2011; Bidelman & Krishnan, 2010; Chartrand & Belin, 2006; Parbery-Clark, Skoe, Lam, & Kraus, 2009; Zendel & Alain, 2009); (ii) speech perception and verbal intelligence (described in more detail in the following part of this chapter);

---

[7] A statistical method for revealing differences in brain anatomy between groups, without the need to focus on target structures.

[8] Broca's area is typically defined in terms of the pars opercularis and pars triangularis of the inferior frontal gyrus. It was named after Pierre Paul Broca and originally linked to language processing, based on two patients who lost the ability to speak after injury to this brain region. Nowadays, the anatomical and functional specificity of Broca is a broadly discussed issue (see Chapter 1.2.3 of this thesis or e.g., Tremblay & Dick, 2016).

(iii) attentional processing, general intelligence, memory and executive functions (Besson et al., 2011; Kraus & Chandrasekaran, 2010; Moreno et al., 2011; Pallesen et al., 2010; Schellenberg, 2004; for more details see Chapter 1.2.2. of this thesis, and for a recent review see Benz, Sellaro, Hommel, & Colzato, 2016), as well as (v) motor function and synchronization ability (Baer et al., 2015; Bailey & Penhune, 2010). However, there are also studies that fail to evidence far transfer effects from music training to other cognitive or academic abilities. Interested readers are referred to a recent meta-analysis by Sala and Gobet (Sala & Gobet, 2017) that failed to support the hypothesis that music skill transfers to cognitive or academic skills in the general population of children and young adolescents.

1. 2. Music and speech

Speech processing is another fascinating research area. Below, I consider some general components and information about speech. However, for a more complete introduction on the cortical networks involved in speech processing, as well as for the perceptual and cognitive demands underlying speech learning, I refer the reader to our recent book chapter that can be found in the appendix of this thesis (Elmer, Dittinger, & Besson, in press).

1.2.1. Commonalities between music and speech processing

Speech processing involves similar functions as described above for music processing: perceptive and cognitive functions, fine motor control and organization of movements, as well as auditory-motor integration that allows information transfer between interconnected brain regions (Tillmann, 2012; Zatorre et al., 2007). Moreover, music and speech are both structured and organized along several dimensions: the basic elements are notes for music and phonemes for language and they can be bound together using well-established harmonic or syntactic rules (Besson & Schön, 2003). Both music and language are based on similar acoustic parameters (i.e., frequency, duration, intensity and timbre), that define the structural organization and that can lead to similarities and differences in music and speech (Patel, 2008). They are described in detail below.

a. Acoustic parameters

Results of recent studies opened the possibility that a potential driver of the link between music training and language skills is the enhanced sensitivity of musicians to prosodic cues of speech. Prosody is a feature supplying both linguistic (intention, focus, discourse meaning) and non-linguistic (emotional) information (D'Imperio, 2011; Gandour & Krishnan, 2016), and it comprises many different aspects: accent, stress pattern, prosodic boundaries, intonation and rhythm (Magne, Schön, & Besson, 2003). Like music, prosody is expressed through a combination of variations in frequency, duration, intensity and timbre (i.e., spectral characteristics; D'Imperio, Dittinger, & Besson, 2016). These variations correspond to the four acoustic parameters described in detail in the following.[9]

Frequency: is defined as the number of periodic variations per second of the sound wave, measured in Hertz (Hz). Its perceptive correlate is pitch. For pure tones this terminology does not pose any problem, but for complex sounds frequency generally describes the lowest frequency of resonance or fundamental frequency (F0). Multiple frequencies of F0 are called harmonics. For speech, the F0 is related to the vibration velocity of the vocal cords and for music it depends on the instrument played: the faster the vibration, the higher the sound.

Duration: is defined as the length of time during which the sound vibrations persist, measured in milliseconds (ms). The two main duration-related parameters are rhythm and meter: rhythm corresponds to the pattern of regular or irregular pulses, and meter refers to regularly recurring patterns and accents such as measures and beats.

Intensity: is defined as the amplitude of the acoustic pressure, measured in decibel (dB). It allows to distinguish a loud from a soft sound. The perceived intensity is called loudness.

Timbre: is also called "the color of sound" and allows to distinguish two sounds with the same frequency, duration and intensity. It is a multidimensional parameter that results from the interaction of spectral and temporal variations related to the harmonics of the fundamental frequency, and of the temporal characteristics of the onset and offset of the sound. Based on these interactions, the quality of a given sound of a certain instrument or voice emerges.

---

[9] For a more detailed introduction on prosodic processing and on the overlap between speech and music, I suggest readers with knowledge of French our recently published chapter included in the appendix of this thesis (D'Imperio, Dittinger, & Besson, 2016).

b. Musical and linguistic structures

For music, three levels of organization have been described: rhythm, melody and harmony. Rhythm corresponds to the subdivision of sounds in time and it organizes the temporality of sounds. Variations of rhythm can be created by the variation of note durations and musical accents. The level of melody corresponds to the sequence of notes that shape a musical phrase: it is structured by the frequency of the notes, and corresponds to the horizontal component of music. The vertical aspect of music is given by the level of harmony that describes the simultaneous emission of several sounds and their relation. Harmony is structured by intervals and chords. Finally, a “fourth level of organization” should be mentioned that is debated in the literature and that corresponds to musical interpretation: musical signification explores expressive motivations behind musical structures and the ways in which listeners and performers interpret musical works and practices. It includes a wide variety of repertoire (e.g., early, recent, popular, and world music), of approaches (e.g., historical, philosophical, cultural, critical, semiotic, and cognitive), and of themes (e.g., gesture/movement, drama/narrativity, gender/sexuality, intertextuality).

For language, four levels of organization can be considered: first, the level of phonology with the phoneme constituting the smallest sound unit of speech and allowing to modify the meaning of a word. This level includes segmental (phonemes and syllables) as well as supra-segmental cues (prosodic structures). Second, the level referring to the combination of phonemes to create words. Third, the lexico-semantic level allowing to access the meaning of a word. Words are combined following syntactic rules that define the structural relations between words to create sentences. Thus, the meaning of a sentence results from the semantic relations between the words as well as from the syntactic organization of the words. Fourth, the organization of sentences at the discourse level and is referred to as the pragmatic level.

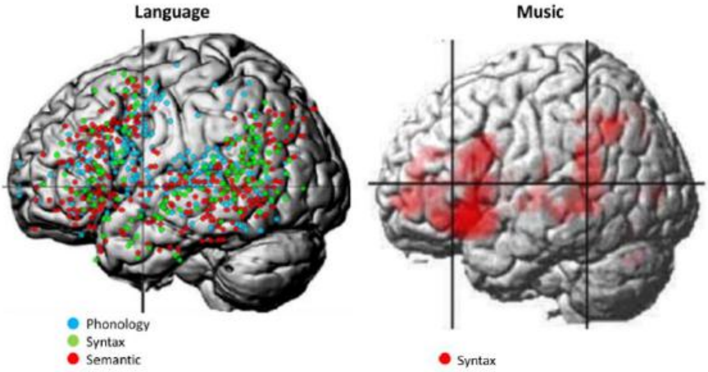


Figure 5.
Activated brain regions during the processing of syntax in language (left, green dots) and in music (right, red). (Adapted from Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005 and Vigneau et al., 2006)

Based on the possibility to similarly structure music and language (starting from single units that are combined into larger units, see Figure 5), several studies have used neuroimaging methods to compare the processing of music and speech at different levels of organization.

1.2.2. Transfer effects from music training to speech processing

Transfer effects (also called transfers of training effects) are defined as the influence of knowledge or ability acquired in one domain (in our case music) on the knowledge or acquisition of knowledge in another domain (in our case speech)[10]. The relation between music training and speech has been well-documented and results of many experiments have demonstrated the positive influence of music training on speech perception (Besson et al., 2011; Kraus & Chandrasekaran, 2010). For instance, there is clear evidence that music training influences the segmental processing of speech sounds (consonants, vowels and syllables; Bidelman & Krishnan, 2010; Chobert, Francois, Velay, & Besson, 2014; Elmer, Meyer, & Jancke, 2012; Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007), as well as the supra-segmental level (couple of syllables, words, and sentences; Besson et al., 2011; Lima & Castro, 2011; Schön, Magne, & Besson, 2004). Bidelman and collaborators demonstrated that younger (Bidelman, Weiss, Moreno, & Alain, 2014) and older musicians (Bidelman & Alain, 2015) were faster and showed steeper boundaries between phonetic categories in a vowel categorization task than non-musicians. Moreover, Bidelman and colleagues (Bidelman et al., 2014) also showed an influence of music training on categorical perception of speech sounds (on a /u/ to /a/ continuum) at the cortical level: the P200 component of Event-Related Potentials (ERPs)[11] showed increased amplitudes in musicians, and these neural enhancements were also correlated with their years of formal music training. The authors concluded, that increased auditory sensitivity, as reflected by increased P200 amplitudes, may thus be one of the driving forces behind enhanced categorical perception and enhanced speech processing in musicians.

Importantly, transfer from music to speech is by no means limited to low-level speech processing, but extends to higher-level speech processing. For example, based on the idea that both music and language are structured sequences of events that unfold in time, several studies

---

[10] Here I focus on transfer effects from music to speech processing. For reversed transfer effects (from speech to music processing) the interested reader may consult the review by Asaridou & Mc Queen (2013).

[11] See Chapter 2 for an introduction on the ERPs methodology and ERP components.

have investigated the influence of music training on syntactic processing (Fitzroy & Sanders, 2012; Jentschke & Koelsch, 2009). Jentschke & Koelsch (2009) evaluated ERPs to violations of linguistic and musical syntax in musically-trained and untrained children (ten- to eleven-year-olds). Results showed that both types of violations elicited larger ERP modulations in the musically-trained group than in the control group, thereby reflecting more strongly developed and sustained syntax processing as a function of music training.

The literature evidencing transfer effects from music training to speech perception is large, and the studies mentioned above are only few examples. Interested readers will find a more complete review of this topic in a chapter that we wrote recently and that is added as an appendix to this thesis (Besson, Barbaroux, & Dittinger, 2017). The main question addressed here is how to explain transfer effects from music to speech processing. In this context, I would like to introduce two main interpretations that have been proposed in the literature and to which I will come back several times within this thesis: the so-called cascade and multi-dimensional interpretations. Following the cascade interpretation, transfer effects arise because, as described above, speech and music are auditory signals relying on the same acoustic parameters. Consequently, if music training facilitates the perception of acoustic structures of sounds not only in music but also in speech, perceptive advantages in musicians should facilitate higher levels of speech processing.

The multi-dimensional interpretation refers to the multi-dimensionality of music training, including auditory and visual perception (the notes on the score), visuo-auditory-motor integration (transforming visual notes into sounds through movements), motor control (adapting posture and fine distal movements), as well as attentional functions (focusing attention on one own instrument and dividing attention between the different instruments of the orchestra and the conductor), mnemonic functions (playing music by heart), and executive control (switching between visual and auditory codes; inhibiting a movement to play at the right moment and updating information).

Several studies point to advantages of professional musicians when these functions are called into play: for example, Strait and colleagues (Strait, Slater, O'Connell, & Kraus, 2015) provided evidence for enhanced selective auditory attention in young musicians compared to non-musicians, reflected by less variable cortical auditory-evoked responses to attended than to ignored speech streams. Moreover, the authors compared preschoolers, school-aged children and young adults, and suggested that music training could support the maturation of auditory attention during developmental years, based on reduced auditory-evoked response variability during selective auditory attention in school-aged children with music training and adult musicians, but not in preschoolers.

Regarding memory functions, George and Coch (George & Coch, 2011) demonstrated that professional music training was related to improvements in auditory and visual working memory: behaviorally, musicians outperformed controls on standardized subtests of visual, phonological, and executive memory. Moreover, this behavioral advantage was accompanied by shorter latencies and larger amplitudes of P300 components of ERPs in both auditory and visual oddball tasks in musicians compared to non-musicians, thereby suggesting faster and less effortful updating of working memory. Similarly, Zuk and colleagues (Zuk, Benjamin, Kenyon, & Gaab, 2014) tested adult musicians and musically-trained children on several measures of executive functions, including working memory, cognitive flexibility, and verbal fluency. They showed enhanced performance as well as increased brain activation in musically-trained participants, and concluded that musical training was linked to enhanced cognitive skills. These few examples (see Besson et al., 2017 for a more detailed review) may point to the interest of considering a more multi-dimensional interpretation and to the possibility that musicians process speech more easily not only based on better auditory perception but also based on better processing of the multiple cognitive components important for speech perception.

1.2.3. Neural networks of music and speech: domain-general versus domain-specific brain regions

Although there is accumulating evidence suggesting that music and speech processing overlap in terms of brain regions (Jantzen, Large, & Magne, 2016; Koelsch et al., 2002; Tillmann, Janata, & Bharucha, 2003; see Figure 6), such a spatial overlap of functional activation does not necessarily imply that music and language are processed by the same underlying neuronal networks.

In fact, the question of whether music and speech rely on shared or separated brain regions and/or neurons, is a controversial and hotly debated issue. To better understand this issue, I would like to mention the example of syntactic processing and Broca's area. For over a century, Broca's area was considered as specifically involved in the processing of linguistic syntactic structures (Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006). However, in the years 2000, results of several studies demonstrated that Broca's area was also activated when processing musical structures (Koelsch et al., 2002; Maess, Koelsch, Gunter, & Friederici, 2001), thereby providing evidence of shared neural substrates for music and language. These results were further supported by those of recent studies by Abrams and colleagues (Abrams et al., 2011) and Rogalsky and colleagues (Rogalsky, Rong, Saberi, & Hickok, 2011) showing that similar activations of frontal and temporal regions in both hemispheres are associated with temporal violations in linguistic sentences and melodies. However, when using multivariate pattern analysis[12], the same authors showed that the two types of stimuli elicited spatially distinct activity. Thus, the authors concluded that temporal structure is encoded differently within the two domains and that distinct cortical networks are activated.

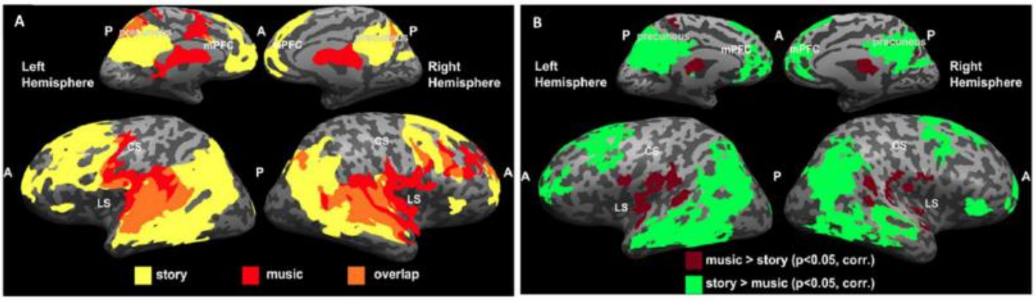


Figure 6. (A) Brain areas responding to intact story (yellow), intact music (red) or both (orange). (B) Significant differences between intact music and intact story conditions. (Figure adapted from Farbood, Heeger, Marcus, Hasson, & Lerner, 2015)

This example shows that one of the challenges for answering the question of shared versus distinct networks involved in music and language is related to the method chosen for analysis. However, characteristics of the stimuli and the task at hand, or even outdated terminology, may work against the establishing of a conjoint view across studies: related to the example of Broca's area it is worth mentioning a recent review by Tremblay and Dick (Tremblay & Dick, 2016) with the provocative title "Broca and Wernicke are dead". In this review, the authors take the current

---

[12] Multivariate pattern analysis (MVPA) is considered to be a more fine-grained approach, in which data from individual voxels within a region are jointly analyzed as distributed patterns of activity. Thus, it investigates the varying brain states that a cortical field or system can produce, and increases the amount of information that can be decoded from brain activity in contrast to simpler univariate measures that indicate the extent to which a cortical field or system is globally engaged. MVPA is frequently applied in clinical research, with two main goals: first, to understand underlying mechanisms and patterns in the brain that characterize a disease. Second, to develop sensitive and specific image-based biomarkers for disease diagnosis, the prediction of disease progression, or prediction of treatment response (Haxby, 2012; Norman, Polyn, Detre, & Haxby, 2006)

terminology of Broca's and Wernicke's area as an example to illustrate problems that arise when using historical terminology based on the classical ''Wernicke-Lichtheim-Geschwind" model, in spite of fast advancement of research in the field of cognitive neuroscience and neuropsychology. In fact, the authors demonstrated that there is no consistent anatomical definition of Broca's and Wernicke's areas when comparing different studies (see Figure 7). The authors point to the (necessarily) resulting difficulty in finding a consensus on shared versus distinct neuronal networks for music and speech processing. For more details on this controversial issue, interested readers may consult our chapter on music and language processing, and/or our review on domain-general versus domain-specific brain regions in the appendix of this thesis (Besson et al., 2017 and Besson, Dittinger, & Barbaroux, in press, respectively).

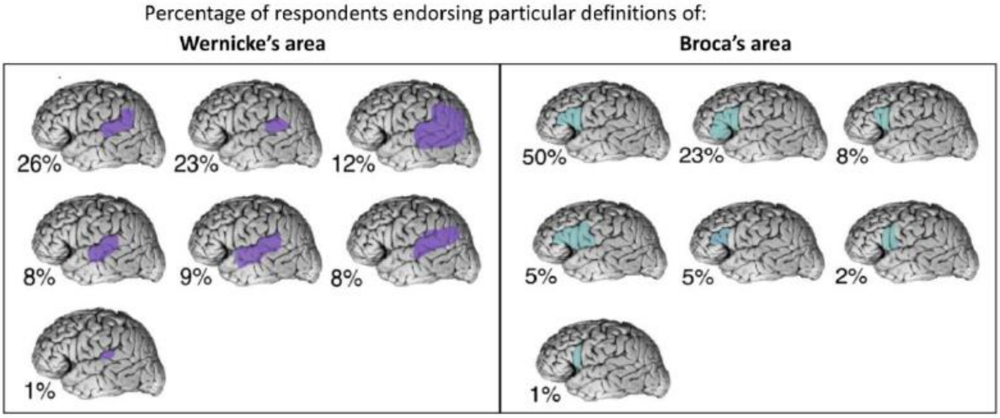


Figure 7. Anatomical definitions of Wernicke's (left) and Broca's area (right), and the percentages of respondents to the survey endorsing each definition. (Figure adapted from Tremblay & Dick, 2016)

1. 3. Language acquisition and novel word learning

Acquiring vocabulary is one of the main cornerstones of human language learning, and we are continuously encountering novel words during our lifespan. In fact, studies suggest that between the age of 2 and 20 years humans learn between 6 and 25 words per day (Nagy & Anderson, 1984), and that the vocabulary size of high school graduates lies between 40.000 and 100.000 words (Nagy, Anderson, & Herman, 1987). Importantly, it has been pointed out that efficient and accurate verbal comprehension skills depend on knowing many words in a language

(Anderson & Freebody, 1981; Sternberg & Powell, 1983), and that the underlying meaning of these words can be easily and reliably accessed (Perfetti, 1985).

In the mother tongue, language acquisition has traditionally been considered an implicit process, and infants become proficient in a very short period of time (Rodríguez-Fornells et al., 2009). In this vein, children can learn novel words and their meanings based on few exposures, and usually without explicit feedback from their parents (Brown & Hanlon, 1970; Marcus, 1993). To better understand the enormous rate at which children gain vocabulary, Carey and Bartlett (Carey & Bartlett, 1978) tested three-year-olds on a novel word learning task. Specifically, children were presented with two trays, a red and an olive one, and were asked to bring the “chromium” tray, not the red one. By contrasting "chromium" with "red" all children were able to infer that “chromium” was a color and to identify its intended referent. Moreover, when tested after a one-week retention delay, more than half of the children displayed some type of comprehension about the word chromium. Carey and Bartlett called the process, allowing children to learn a new concept based on a single exposure to a given unit of information, “fast mapping”, and this term has become central to developmental psychology’s studies about how words are learned. However, it is important to point out that for the authors, the demonstration of fast mapping was noteworthy not because children appropriately determined that “chromium” was a color word (which could have been inferred without any learning). Rather, the authors focused on what they called “extended mapping”, a process based on the fact that after very few exposures children were able to create a new lexical entry and to change their interpretation of how the color space is lexicalized (i.e., to bring into alignment the lexical entry “chromium” and its position in a color space that had not contained that region as a separate category before).

Independent of the open question if one exposure is sufficient to learn a novel word (I will come back to this issue below when considering ERP studies), the capacity of children to learn very fast is quite striking, since novel word learning is a complex problem to solve: when children acquire new vocabulary, they are usually presented with multiple novel words and multiple meanings. Therefore, one could ask how human beings assign correct meaning to an unknown word. Studies have evidenced that the speech signal first is segmented into differentiated units to identify the novel word from fluent speech, and that then meaning is attached (i.e., word-to-world mapping; Rodríguez-Fornells et al., 2009). Speech segmentation describes the ability to identify the boundaries between words, syllables, or phonemes in continuous speech, and has been suggested to rely (at least in part) on a domain-general learning mechanism, that is statistical learning. Generally speaking, statistical learning refers to the fact that learners are able to detect the different transitional probabilities within a language after being sufficiently exposed to it: for example, in speech, the probability of syllable B happening given syllable A will be higher for syllables that follow one another within a word than for those at word boundaries (e.g., in "pretty music", the probability of "ty" given "pre" is higher than that of "mu" given "ty"). Saffran and colleagues (Saffran, Aslin, & Newport, 1996) showed that eight-month-old infants were already able to extract word units from fluent speech, and more recent research has shown that this ability is also present in neonates (Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), as well as in human adults (Cunillera et al., 2009; López-Barroso et al., 2015; McNealy, Mazziotta, & Dapretto, 2006; Schön & François, 2011). Then, once the novel word has been extracted, the proper meaning can be assigned to it.

However, we constantly learn novel words first, not only in our mother tongue but also in second or third languages, and second, not only during infancy but also during adulthood, potentially extending to advanced ages. In this respect it needs to be made clear that processes involved in foreign word learning may differ from processes involved in novel word learning in the mother tongue: while in the mother tongue novel words need to be mapped onto novel meanings (word-to-world mapping), novel words of foreign languages are most often mapped onto already known referents. In fact, previous research has shown that in early learning stages, associations are created between novel words of foreign languages and their corresponding translations in the mother tongue (word-to-word mapping). By contrast, only as learners become more proficient, direct associations are formed between these novel words and their corresponding meanings (word-to-world mappings; Doughty & Long, 2003). Furthermore, second language learning is usually considered to be less successful than the acquisition of the mother tongue, and to be a challenging and effortful process for many learners (Rodríguez-Fornells et al., 2009). However, it is difficult to explain why both lead to different proficiency levels. For instance, children can generate full sentences at the age of three years, while adults are able to learn sentences in a new language within some minutes. Such comparisons might be relevant and odd at the same time as the higher learning ability of adults may be the result of higher cognitive ability, but surprisingly this has only limited positive effects on the adults' ability to learn foreign languages: while infants are virtually open to new sound combinations of all languages, adult learners will not acquire novel languages as successfully as infants do (Kuhl, 2004). Such controversies have led to important assumptions in the field of language acquisition research, as for example the suggestion that language acquisition processes do not exclusively rely on higher cognitive abilities such as intelligence, but also on other factors such as, for example, on critical periods during which specific abilities are easily taken up, or individual differences.

1.3.1. Neural correlates of novel word learning: ERP studies

Over the years, a great amount of scientific research has been devoted to study how children and adults are able to quickly learn novel words and to better understand the neural underpinnings of novel word learning: along with behavioral measures, that can reveal whether word meaning has been learned, several studies have indicated the benefits of using ERPs in word learning experiments to examine additional information on the temporal dynamics of learning (Key, Molfese, & Ratajczak, 2006; McCandliss, Posner, & Givón, 1997; Perfetti, Wlotko, & Hart, 2005), on word recognition (Curran, 1999; Finnigan, Humphreys, Dennis, & Geffen, 2002; Wilding & Rugg, 1997) as well as on meaning processing (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Kutas & Hillyard, 1980). In fact, fast plastic brain changes underlying word encoding and retrieval are reflected by changes in the amplitude and/or latency of several ERP components: first, one ERP component of interest reflecting meaning retrieval and integration processes is the N400 component. Second, a group of ERP components (frontal N400, FN400, and P600) that are thought to index so-called old/new effects[13] and that have been implicated in recognition processes should be mentioned related to novel word learning. Please consider that a more general view on these ERP components can be found in Chapter 2.

a. Meaning processing: N400 component

The N400 component has been implicated in many tasks requiring semantic processing, and its amplitude reflects meaning congruence between a word and its previous context,

---

[13] The classical old/new effect refers to the finding of several studies in recognition memory, in which the presentation of old/repeated items elicits more positive-going ERPs than the presentation of new/unrepeated items. While some basic information on the components reflecting the old/new effect is reported below, readers seeking more information on this effect may refer to the review by Michael Rugg (Rugg, 1995).

.

thereby evoking a so-called N400 effect with larger amplitudes for incongruent compared to congruent trials (see Figure 8A). It has been implicated in both sentence context presentations (Kutas & Hillyard, 1980) and single word semantic priming presentations (Bentin, McCarthy, & Wood, 1985). Regarding novel word learning, several studies have taken advantage of the N400 component to examine if novel words have been learned and integrated into pre-existing semantic networks. For example, Perfetti and colleagues (Perfetti et al., 2005) reported N400 effects related to meaning acquisition: in this study, adult participants learned 60 novel rare words through definitions (e.g., “clowder” was the word for “a collection or group of cats”), and following word training, participants were presented with the trained words followed by meaning probes that were related (e.g., clowder – cats) or unrelated (clowder - evening) to the words’ meanings. They observed a robust N400 effect with larger amplitudes for unrelated meaning probes than for related meaning probes, thereby indicating that participants had learned the meaning of words. Similarly, Mestres-Missé and colleagues (Mestres-Missé, Rodriguez-Fornells, & Münte, 2007) reported N400 effects for words that were learned in high constraint sentences (e.g., “Mario always forgets where he leaves the lankey”), and that were tested by meaning probes related (e.g., lankey – car) or unrelated (e.g., lankey – house) in meaning to the learned words.

Finally, N400 effects not only depict markers of novel word learning in adults, but also in infants: for example, Friedrich and colleagues (Friedrich, Wilhelm, Born, & Friederici, 2015) tested nine- to sixteen-month-old infants on the learning of novel words (e.g., “Beker”) through picture-word associations. Initially, the authors showed N400 effects only to the specific word meanings trained, but after a consolidation break (i.e., one and a half hours of sleep) also to novel category exemplars. In sum, by summarizing only few results of the numerous experiments that have been conducted in infants and adults, the N400 seems to reliably index that meaning has been attached to the novel words.

Importantly, the N400 is not only elicited once novel words have been learned, but it is also one of the most eligible components to follow the learning dynamics during meaning acquisition: in fact, when meaningless items acquire meaning, the amplitude of the N400 increases. Considering the same study as already mentioned above (Perfetti et al., 2005), the authors were able to show N400 enhancements after only 45 minutes of learning the meaning of the novel words. Importantly, this increase was larger for skilled compared to less skilled comprehenders, thereby pointing to a relationship between N400 amplitude and higher word learning abilities. Similarly, McLaughlin and colleagues (McLaughlin, Osterhout, & Kim, 2004) evidenced increased N400 in native English speakers after 14 hours of learning the meaning of novel French words. Interestingly (and possibly contrary to what one would imagine), recent studies have also evidenced, that the learning of novel words can be very fast: for example, Friedrich and Friederici (Friedrich & Friederici, 2008) explored the brain activity related to novel word learning through picture-word mappings in fourteen-month-old infants, and revealed N400 increases after only four presentations, thereby suggesting the very fast formation of associative memory structures.

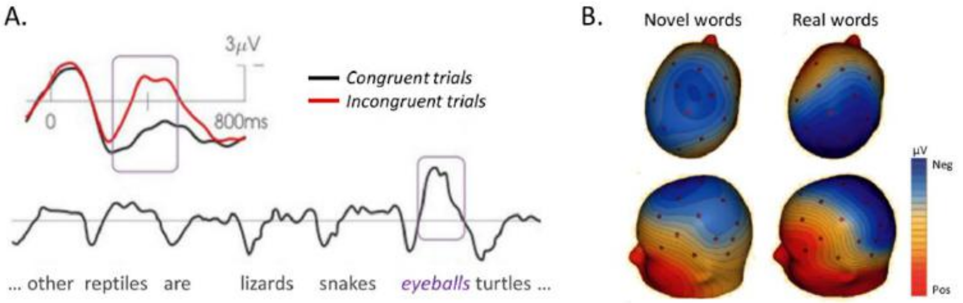


Figure 8. (A) The N400 component is larger to incongruent compared to congruent trials. (Adapted from Kutas & Federmeier, 2011) (B) The N400 effect shows a right parietal scalp distribution in the real-word condition, and a fronto-central maximum in the novel-word condition. (Adapted from Mestres-Missé et al., 2007)

Similar results were obtained with adult participants: in a study by Mestres-Missé and colleagues (Mestres-Missé et al., 2007) adult participants learned the meaning of novel words after only 3 presentations. Moreover, it has also been suggested that if a novel word is presented in a strongly constrained and meaningful context, even a single exposure can be sufficient to build up initial word representations (Batterink & Neville, 2011; Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2010). Importantly, although these studies evidence that novel words have been acquired from context after only a single exposure, the question of whether participants have learned the meaning of these novel words remains debated: for example, in Borovsky et al. (2010) adult participants learned the word “marf” by reading the sentence “He tried to put the pieces of the broken plate back together with MARF”. Subsequently, participants were tested on plausible (e.g., “he needed the MARF”) or implausible sentences (e.g., “she drove the MARF”), and were asked whether the novel word was used appropriately in the sentence or not. As illustrated by this example, participants may have responded correctly to these sentences without necessarily knowing but simply guessing the meaning of the novel word (in this case “glue” was the meaning for “marf”). Thus, concluding whether novel words can be learned through fast mapping, and whether the N400 increase observed in the Borovsky et al. (2010) study is related to meaning acquisition or not, is a difficult issue that still needs to be solved. In fact, although the initial building up of word representations may be very fast, several studies have pointed to the importance of an incubation-like period and/or further exposures for consolidation and integration of novel words into existing lexical networks (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Dumay & Gaskell, 2007; Tamminen, Lambon Ralph, & Lewis, 2013).

In sum, the N400 can be taken as a reliable index of word learning in infants and in adults, reflecting the formation of semantic representations (at least in combination with behavioral responses). Importantly, it needs to be noted that, compared to the typical N400 component that is elicited by known words and that is typically larger over centro-parietal scalp sites with a slightly right-hemispheric predominance (Kutas, Van Petten, & Besson, 1988; Figure 8B "real words"), this fast increase in N400 amplitude reflecting meaning acquisition is typically largest over fronto-central brain regions (Borovsky et al., 2010; Mestres-Missé et al., 2007; Figure 8B "novel words"). In fact, this frontal N400 distribution is also interesting for the following two reasons: first, previous results have evidenced that prefrontal and temporal brain regions are associated with the maintenance of novel information in working or short-term memory and the formation of new associations (Hagoort, 2014) and/or with the initial building-up of word representations in episodic memory (Rodríguez-Fornells et al., 2009; Wagner et al., 1998). Second, it may relate to the FN400 of the old/new effect, which I will detail in the following.

b. Familiarity and recollection: FN400 and P600 components

The old/new effect includes two topographically distinct ERP components that have been implicated as markers of familiarity and recognition (Rugg & Curran, 2007): the FN400 and the P600. Generally speaking both components show more positive amplitudes to trained (i.e., old) words than to untrained (i.e., new) words. The FN400 has been hypothesized to index familiarity, not only at an orthographic level, but also at a semantic level, as reflected by old/new effects elicited for new items that share semantic features with the old studied items. Moreover, Finnigan and colleagues (Finnigan et al., 2002) observed that the FN400 old/new effect is also modulated by repetition and memory strength. In this study, words were presented once (weak condition) or three times (strong condition) during the study phase. ERPs evoked by words in the strong condition elicited a greater reduction in FN400 amplitude (i.e., greater positivity) than words in the weak condition. Thus, in novel word learning experiments, the FN400 may index the dissociation between unfamiliar (new) items and familiar (old) items, and the dissociation in memory strength as a consequence of repetition. Importantly, although I don't know any study simultaneously investigating the frontally-distributed N400 elicited during word meaning acquisition mentioned above, and the FN400 related to old/new effects, I may speculate on their relationship: in fact, it is conceivable that the frontal N400 first increases (i.e., gains negativity) when a novel word acquires meaning, and that then the same component decreases again (i.e., gains positivity, typically called FN400) once the word has been learned, based on repetition effects and memory strength. In the following (i.e., in the result and discussion sections), I will refer to this component as "frontal N400", without aiming at resolving the question whether these two N400s refer to independent components or not. Finally, in this vein it is also interesting to mention a recent study by François and colleagues (François, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017): in this study, adult participants were asked to perform a speech segmentation and a word-to-picture association task. Results showed the appearance of an early FN400 (around 200 ms) as well as a more delayed FN400 (around 400 ms). The authors concluded that the frontally-distributed N400 not only reflects the building-up of novel word meaning, but also processes involved in speech segmentation, leading them to suggest that speech segmentation and meaning acquisition may take place in parallel and act in synergy to enhance novel word learning. Thus, although it is an open issue if the frontal N400 and/or FN400 should be considered as a "specific" marker of novel word acquisition, there is no doubt that this frontally-distributed negativity depicts a component that can reveal interesting information related to speech processing.

The second component related to old/new effects, the P600, also shows greater positive amplitudes for old items than for new items (Curran, 1999; Rugg & Curran, 2007), and has been suggested to be modulated by recollection processes: for example, Rugg and Curran (Rugg & Curran, 2007) reported that P600 amplitudes index the recognition of recollected words that have been recently learned, rather than a familiarity-driven recognition process that may be also modulated by the presence of semantic knowledge. Furthermore, the P600 old/new effect has also been directly observed in word learning studies. Perfetti et al. (2005) exposed learners to the form and meaning of rare unknown words, and reported a P600 after word presentation that separated learned words from untrained and familiar words. Similarly, Balass and colleagues (Balass, Nelson, & Perfetti, 2010) exposed learners to unknown words and replicated the P600 old/new effects by showing a difference between trained (old) words and untrained (new) words. Moreover, the authors also found that the depth of learning affected the subsequent recognition of the trained words as reflected by P600 amplitude: in fact, the "deeper" training condition in which participants learned the spelling of the word as well as its meaning elicited significantly larger P600 components compared to the conditions in which they only heard the word and saw its spelling or in which they heard the word and learned its meaning. In sum, P600 effects during word learning have been related to the activation of meaning-related information during learning, and the P600 may depict a potential component of interest for investigating neural correlates of novel word learning. However, the P600 has also been suggested to be especially relevant for examining the effects of high versus low constraint contexts (i.e., the quantity of meaning that can be attributed to a novel word; Mestres-Missé et al., 2007; Nessler, Mecklinger, & Penney, 2001). Since my experimental designs rely on novel word learning through picture-word associations rather than from context, and since the P600 has been also related to a number of other processes not directly related to novel word learning (see Chapter 2), I may point out already here that analyses of P600 modulations were not of main interest for this thesis.

1.3.2. The dual stream model of language processing & brain areas of novel word learning

Our understanding of the functional anatomy of language has advanced enormously since the first studies of Broca and Wernicke almost 150 years ago (Broca, 1863; Wernicke, 1881). In recent years, the dual-stream model of language processing is the predominant view regarding the cortical organization of language. From a functional point of view, this model describes a left-lateralized dorsal stream, responsible of mapping acoustic speech signals to their articulatory-based representations (i.e., sound-to-articulation mapping or auditory-motor integration), and a bilateral ventral stream controlling the mapping of sounds to its corresponding meaning (sound-to-meaning mapping). Anatomically, the dorsal stream involves the parieto-temporal boundary and posterior frontal regions, including the inferior frontal gyrus (IFG), Broca's area and premotor regions. The ventral stream encompasses the middle and superior portions of the temporal lobe, as well as the inferior temporal cortex and the anterior temporal pole (see Figure 9).

Several authors have proposed detailed information and interpretations of this dual-stream model (Hickok & Poeppel, 2000, 2004, 2007; Rauschecker & Scott, 2009; Rodríguez-Fornells et al., 2009; Saur et al., 2008), and describing them here is beyond the scope of this thesis. However, a short overview of these models is included in our recent written book chapter and interested readers may refer to it for a broader view on the dual-stream model (Elmer et al., in press).

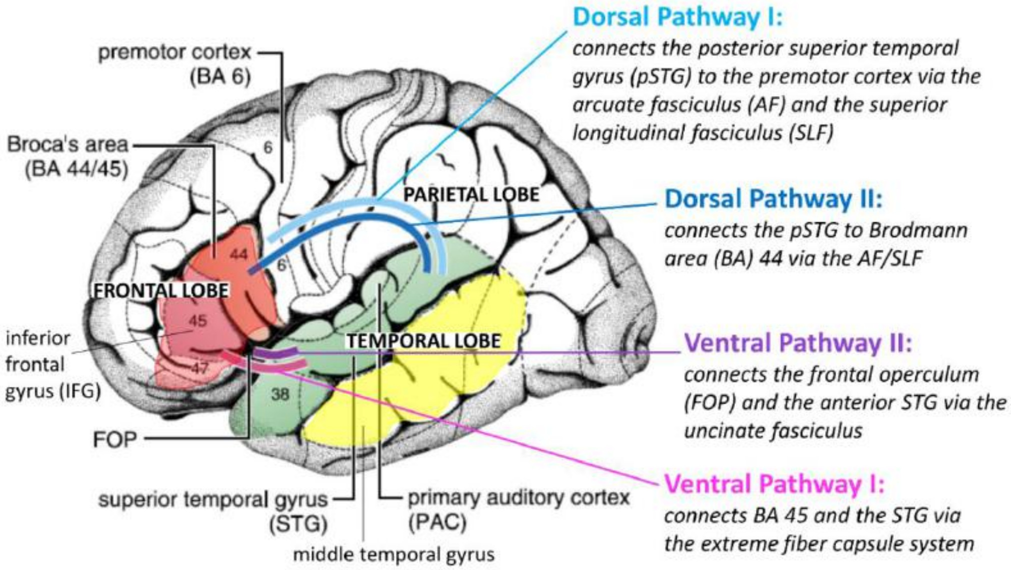


Figure 9. Dual stream model of speech processing. (Adapted from Friederici, 2011)

What should be mentioned here is that, although the dual-stream model has been very successful and there is clear evidence to support its existence, it may not tell the whole story about the neural correlates of language processing. Language involves an incredible amount of mechanisms, each of them sub-serving very different processes (e.g., semantics, syntax, phonology, etc.) and these processes may tap areas outside this "classical language network" (Fitch, 2010; Kraus & White-Schwoch, 2015; Lieberman, 2016). In fact, a recent review on studies of speech perception, spoken language and reading (Price, 2012) showed that there is a variety of cortical regions that respond to language. Moreover, a recent study by Huth and colleagues (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016) is revealing in this regard: in this study the authors collected fMRI data while participants listened to narrative stories, and, by using voxel-wise modeling[14], were able to create a semantic brain atlas[15] (Figure 10). Interestingly, this map evidences extensive cortical areas within temporal, parietal and prefrontal cortices in both hemispheres, making it hard to believe that semantic processing relies on brain areas that are language-specific and that are exclusively located along the two processing streams.

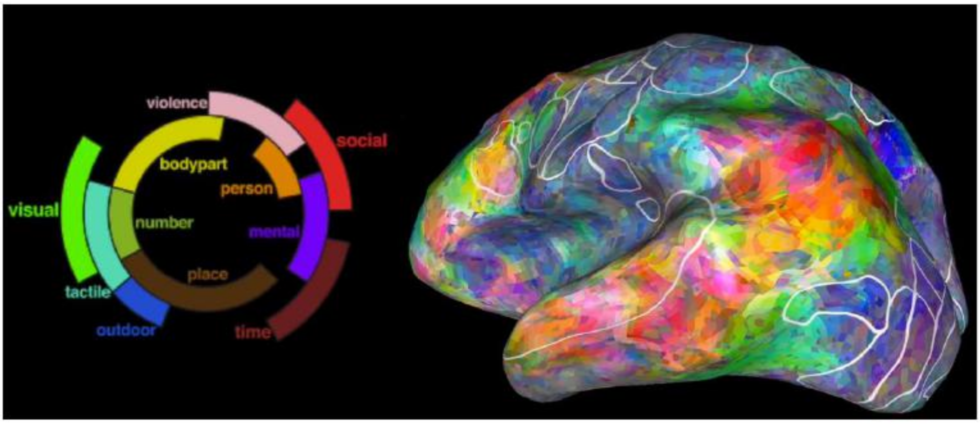


Figure 10. 3D semantic atlas using voxel-based mapping and RGB coloring, a method to encode tri-variate imaging data, where each parameter map is assigned to one of the basic display colors red, green, blue. (Adapted from Huth et al., 2016)

Independently of whether the dual-stream model is realistic or not, both processing streams are involved in novel word learning. Besides supporting auditory-motor integration, recent results suggest a role for the arcuate fasciculus (one of the white fiber bundles of the dorsal stream) in verbal recall (Catani et al., 2007), the acquisition of literacy (Thiebaut de Schotten, Cohen, Amemiya, Braga, & Dehaene, 2014), and also novel word learning in adults (López-Barroso et al., 2013) as well as in infants (François et al., 2016). Moreover, the anterior part of the IFG has been associated with semantic control (Bokde, Tagamets, Friedman, &

---

[14] Voxel-wise modeling refers to the approach of fitting linearized models, embodying different feature spaces, to every cortical voxel within a scanning window. It is typically estimated and validated independently for each subject, and has been shown to represent a highly effective approach for modelling responses to complex natural stimuli, while at the same time, minimizing both Type I and Type II errors (Huth, Nishimoto, Vu, & Gallant, 2012; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009)

[15] The semantic brain atlas describes the distribution of semantically selective functional areas in the human cerebral cortex. It is based on the Bayesian algorithm PrAGMATiC. This algorithm represents patterns of functional tuning (recovered by voxel-wise modelling) as a dense map of functionally homogeneous brain areas, while respecting individual differences in anatomical and functional anatomy (Huth, de Heer, Griffiths, Theunissen, & Gallant, 2016).

Horwitz, 2001; Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001), and activity in the inferior parietal lobe has been observed during the learning (Breitenstein et al., 2005) or during the retrieval of previously learned novel words (Cornelissen et al., 2004; Hultén, Laaksonen, Vihla, Laine, & Salmelin, 2010).

Regarding brain areas along the ventral stream, it has been suggested that they have an important role in the emergence of novel conceptual structures (Rodríguez-Fornells et al., 2009; Saur et al., 2008). Furthermore, the middle temporal gyrus has been related to the storage of long-term conceptual knowledge (Martin & Chao, 2001), to lexical-semantic processes (Baumgaertner, Weiller, & Büchel, 2002; Ferstl & von Cramon, 2001; Indefrey & Levelt, 2004) and word recognition (Binder et al., 1997; Friederici, 2012; Hickok & Poeppel, 2007; Saur et al., 2008), as well as the activation of visual forms and word meanings (Hagoort et al., 1999; Howard et al., 1992). Finally, also the anterior temporal lobes seem to be associated with semantic representations (DeWitt & Rauschecker, 2012; Friederici, 2012; for a review, see Patterson, Nestor, & Rogers, 2007). White matter connections along the ventral stream and their relationship to language processing are still a matter of debate, but accumulating evidence has related them to different aspects of lexical and semantic processing (Agosta et al., 2010; Catani & Mesulam, 2008; Friederici, 2015; Harvey, Wei, Ellmore, Hamilton, & Schnur, 2013).

Recent theories have also suggested that the hippocampus and the subcortical medial temporal lobe (MTL) possibly play a critical role in the process of word learning (Davis & Gaskell, 2009; Mestres-Missé et al., 2008; Rodríguez-Fornells et al., 2009; Takashima, Bakker, van Hell, Janzen, & McQueen, 2017; Ullman, 2001): for example, the parahippocampal gyrus has been linked to comprehension and production of meaningful speech (Awad, Warren, Scott, Turkheimer, & Wise, 2007) and to listening to and reading of meaningful text (Lindenberg & Scheef, 2007). Moreover, Ullman (2001) suggested that novel word learning is subserved by the hippocampus and the MTL, with anterior inferior frontal regions mediating the retrieval of declarative memories (i.e., words and meanings). Davis and Gaskell (2009) suggested that word learning occurs in two differentiated stages: first, isolated representations of novel words trigger a rapid initial acquisition process which is sustained by MTL structures. Second, stable cortical representations are achieved by offline consolidation processes (Davis & Gaskell, 2009).

In this vein, a recent study by Takashima and colleagues (Takashima et al., 2017), testing whether the neural representations of novel words change as a function of consolidation, is revealing: by observing brain activation patterns directly after learning and after a one-week-delay, the authors were able to show that both episodic (hippocampus-dependent) and semantic (dependent on distributed neocortical areas) memory systems were recruited during recognition of the novel words. Specifically, while the hippocampal episodic memory network was more strongly involved on the first day, neocortical activation increased (especially in the left IFG) with lexical integration after the one-week-delay. Interestingly, the authors also compared words that were learned with meanings (meaningful words) and without meanings (form-only words), and evidenced that the extent to which the two systems were involved not only changed as a function of time but also as a function of the amount of associated information: in fact, meaningful words elicited larger hippocampal activation, neocortical activation as well as inter-cortical functional connectivity compared to form-only words after the one-week-delay. These results suggest, that meaningful words are possibly remembered better than form-only words maybe because their retrieval can benefit more from these two complementary memory systems.

Finally, Rodriguez-Fornells et al. (2009) suggested the existence of three learning interfaces that work together during novel word learning: first, an audio-premotor interface (part of the dorsal pathway) fundamental for initial learning of new phonological word forms. Second, a meaning integration interface (part of the ventral pathway) involved in inferring meaning using both internal and external cues. Third, an episodic-lexical interface responsible for fast-mapping new words onto specific MTL dependent contexts and for controlling the long-term consolidation of the newly learned words (Rodríguez-Fornells et al., 2009). Importantly, the authors also suggested that the interaction of these three systems can be modulated by other areas (e.g., prefrontal cortex, basal ganglia) that govern other high level cognitive functions. Considering the above-reported studies and the numerous brain areas that seem to be involved in speech processing and novel word learning, brings support to the view that language is not processed independently from other cognitive functions. Future studies will (hopefully) shed more light on this issue that seems to be not only a hot but also a complex topic.

1. 4. Novel word learning and music training: aims and hypotheses

Based on the background that word acquisition is influenced by higher (and possibly domain-general) cognitive abilities and individual differences, it is time for music training to come into play. As suggested above, musicians constitute a wonderful model to study language processing based on the multidimensional nature of music training that boosts perceptive, cognitive and motor functions. Moreover, above I also described evidence that training-related advantages are not limited to the domain of music but also transfer to the domain of speech. However, whether such transfer effects extend to the semantic processing of speech, and to novel word learning, has not been investigated before. This idea rests upon the fact, that novel word learning (as music training) is a multidimensional task with strong parallels between novel word learning and music training regarding the recruitment of perceptive and higher-order cognitive abilities. Let us imagine, how a learner typically proceeds when it comes to learn novel words: the learner has to focus attention on the acoustic stream to discriminate spectral and temporal phonetic contrasts, to build new phonological representations, to associate these representations with their corresponding lexical meaning and to store these novel associations by recruiting working, short-term, episodic, and semantic memory processes. Thus, if music training translates into improved auditory perception and attention together with enhanced working and verbal memory, it should be easier for musically-trained children as well as for adult musicians to learn novel words. This general hypothesis is at the heart of my thesis and details are described below.

Before going into more details, it is important to point out that the word learning designs used in this thesis should be considered related to foreign language learning, rather than in relation with research on the acquisition of novel words in the mother tongue for the following three reasons: first, transfer effects from music training to speech are at the heart of my thesis, but it was not possible to test infants (i.e., the probably most adequate population when trying to understand first language acquisition but who cannot be trained with music). Here, I focused on novel word learning across the lifespan starting at the age of 8 years, thereby allowing a minimum of 3-4 years of regular music training for the youngest children, and until 65 years.

Second, the novel words used in my experimental designs are natural words from Thai and Finnish, including non-native phonology for French speakers. Thus, these designs are ecologically valid considering foreign language learning.

Third, in the experimental designs, participants learned novel words through picture-word associations. While such a procedure has been frequently used in children (Friedrich & Friederici, 2008; Friedrich et al., 2015; Torkildsen et al., 2009) and adults (Breitenstein et al., 2005; Dobel et al., 2010; Dobel, Lagemann, & Zwitserlood, 2009), infants learn most vocabulary implicitly from context, without any explanation (see for example Batterink & Neville, 2011; Borovsky et al., 2012, 2010 for the use of similar procedures in adults). Note, that word learning from context differs notably from associative learning: it strongly depends on how constraining a given sentence or definition is, resulting in more incidental vocabulary learning. Moreover, when learned from context, the acquisition of word meaning may be more incremental and may need several exposures to capture the correct meaning. Previous studies conducted with children and with adults have demonstrated that the knowledge gained about a previously unknown word after a few contextual exposures is rarely complete (i.e., decontextualized), but rather partial (Daneman & Green, 1986; Nagy, Herman, & Anderson, 1985; Schwanenflugel, Stahl, & McFalls, 1997; Swanborn & de Glopper, 2002). By contrast, novel words presented with a picture depicting its meaning can be explicitly encoded right from the first exposure, and such direct memorizing of vocabulary is frequently applied in foreign language learning classes. In this context, it should be underlined that most people are faced with the challenge of learning at least one foreign language. Understanding the mechanisms of foreign novel word learning at the brain level, not only during childhood but also during adulthood, is without any doubt of great importance for improving foreign language teaching and learning.

With this purpose in mind, children with and without music training, as well as professional musicians and non-musician young adults were tested on a first series of tasks ("APPSENS") including a phonological categorization task, a word learning phase as well as a test phase. Behavioral measures and ERPs during phonological categorization, novel word learning and novel word retrieval, as well as functional connectivity during word learning and retrieval were investigated. Subsequently a second series of tasks ("MUMEMO") was designed to better understand results from the first series of experiments. Again, behavioral and electrophysiological data were analyzed. In the following I will give a quick overview of what is described in detail in Chapters 3 to 6:

First, music training and non-native phoneme perception (Chapter 3):
I asked the question whether professional musicians would process non-native phonemes more efficiently than non-musician adults. This is an important issue: phoneme perception and the development of phoneme categories is one of the building blocks for higher level speech processing such as novel word learning. Non-native phonemes and contrasts (i.e., the difference between two phonemes) may be very challenging to learn. Therefore, showing that music training facilitates their perception has strong implications for foreign language learning.

Second, music training and novel word learning across the lifespan (Chapter 4):
Children, as well as younger and older adults, continually confront a changing world and need to adapt by acquiring new skills and information, as for example novel words. While understanding the neural underpinnings of word learning in infants and young adults is a hot topic, not much is known about these learning mechanisms in school-aged children and how they change with age in older adults. By using a very similar experimental design in all three populations, not only could I investigate differences between musically trained and control participants for each age group, but I was also able to compare processes involved in novel word learning from childhood to late adulthood. Such an approach is also interesting in regard to transfer effects from music to semantic processing that have not yet been investigated in these populations.

Third, functional connectivity during novel word learning (Chapter 5):
In this chapter the focus shifts from ERPs to functional connectivity in the dorsal and ventral streams during novel word learning in young adults, and how connectivity patterns differ between professional musicians and non-musicians.

Fourth, music training, novel word learning and memory (Chapter 6):
The second novel word learning design is introduced with the aim of better understanding the origin of the musicians' advantage during novel word learning by reducing the weight of perceptive functions during semantic integration. In addition, we asked whether musicians would remember the newly learned words better than non-musicians. While we typically describe the acquisition of novel words as a learning process, it is clear that it is also a memory process (i.e., remembering words is as important as learning them; Wojcik, 2013). In this vein, several studies have evidenced that a brain structure central to memory processes, the hippocampus, plays a crucial role in learning new words during adulthood (Davis & Gaskell, 2009; Mestres-Missé et al., 2008; Rodríguez-Fornells et al., 2009; Takashima et al., 2017). Based on the

extending literature evidencing increased working- and short-term memory functions in professional musicians (George & Coch, 2011; Ho, Cheung, & Chan, 2003), and based on the fact that typical word learning studies do not investigate memory (compare Wojcik, 2013), I tried to close this gap by taking advantage of professional musicians to better understand how cognitive functions such as memory relate to novel word learning, and how they contribute to the musicians' advantage to learn and remember novel words demonstrated in previous studies presented in Chapter 4.

Finally, the findings are summarized and discussed in Chapter 7, and future directions are suggested for investigating transfer effects from music training to semantic processing.

To conclude: in this thesis I chose to study novel word learning from the perspective of transfer effects from music training. I would like to make clear that at no point in this document will I claim that music training is the only or most efficient way to improve foreign language learning. I will instead provide, hopefully, compelling evidence suggesting that musicians are a wonderful model to better understand the fascinating puzzle of mechanisms underlying language learning in a wide sense, and that the hours of music training at school in early years may be worth to be extended rather than to be reduced.

Chapter 2

METHODOLOGICAL BACKGROUND

2. 1. Electroencephalogram

2.1.1. History

The electroencephalogram (EEG) is one of the oldest methods of brain imaging. The first work using this method go back to 1790, when Luigi Galvani studied the relationship between electricity and the nervous system in the frog. By conducting an experiment allowing to observe muscular contractions of the frog's leg after the electrical stimulation of the muscle's nerve, he evidenced a phenomenon that he called "animal electricity". This was the first demonstration of neural conduction. In 1875, Caton reported the first observations of brain electrical activity from electrodes placed on the exposed brain in rabbits and monkeys. Half a century later, in 1928, Berger, a German neurologist, considered as the pioneer of human electrophysiology, placed electrodes on the scalp of a trepanned patient and observed rhythmic fluctuations of electrical activity in the human brain: Berger noted that the alpha rhythm (between 8 and 12 Hz) disappeared when the participant opened his eyes and payed attention to the context. Finally, in 1951 Dawson developed a method for summing up and averaging the EEG signal, based on stimulation onset, thereby allowing to isolate variations of electrical potential directly related to stimulus presentation: the Event-Related Potentials (ERPs).

2.1.2. Physical principals

The adult brain is composed of around 100 milliards of neurons, and single neurons are connected through their dendrites (information transport to the cell body) and axons (information transport from the cell body). These connections build up a huge network: for succeeding in different tasks, thousands of neurons need to be wired together and function at the same time. Along the axons, electrical activity allows communication between different neurons, and this activity can be measured with the EEG method. To do so, electrodes are placed on the surface of the scalp of participants that allow measuring small electrical fluctuations that mainly originate from the activity of pyramidal cells of the cortex. Pyramidal cells are multipolar neurons and show dendrites that are organized in parallel macro-columns perpendicular to the cortical surface. If a pyramidal neuron receives an afferent stimulation, the ionic channels on the surface of the cellular membrane open and allow the moving of electrically charged particles at the intracellular level (primary current) and at the extracellular level (volume current). However, only the synchronization and the sum of the activity of several thousands or millions of pyramidal neurons can generate a global volume current that can be captured at the surface of the scalp.

2.1.3. Material and Registering

Data presented in this thesis were registered with Biosemi amplifiers. Biosemi is a recording system that uses electrodes with pre-amplificators allowing the direct amplification of the electrical signal directly at the place of recording, thereby improving the signal-to-noise ratio (SNR).

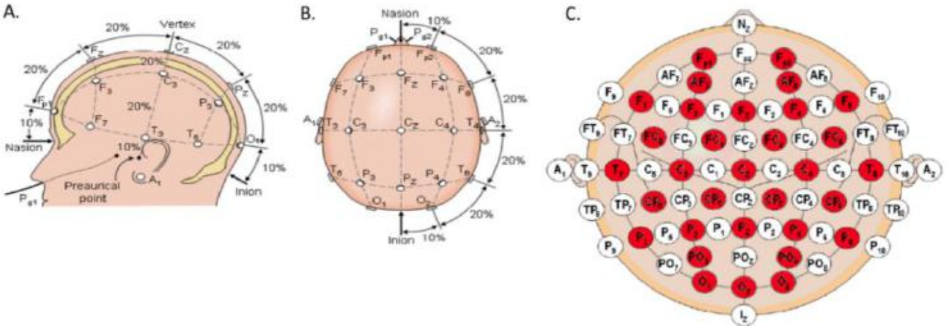


Figure 11. Electrode positions according to the international standard system 10/20 (Jasper, 1958). Electrodes in red circles correspond to the 32 electrode positions on the electrocap used for the studies of this thesis.

In all the experiments, EEG was registered continuously from 32 electrodes that were placed according to the international standard system 10/20 (Jasper, 1958). This system defines the specific spatial position of each electrode based on two anatomic reference points of the scalp: the nasion that is located between the two eyes, and the inion that corresponds to the most prominent projecting point of the occipital bone at the base of the skull. Based on these two anatomic references, the electrodes' positions along the median and transverse planes are calculated respecting 10 % or 20 % intervals between the electrodes (see Figure 11). The electrodes' names are defined by the cerebral regions [frontal (F), central (C), temporal (T), parietal (P) or occipital (O)] and their laterality (odd numbers for the left hemisphere and even numbers for the right hemisphere).

Differences of potentials are measured between two electrodes, an active electrode and a reference electrode. In the Biosemi electrocap, the by-default reference electrode is "Common

Mode Sense" (CMS) that is placed next to Cz. However, due to its central position, this electrode cannot be used as the reference for analyzing the data. The reference needs to be placed at a position that is not too sensitive to the sources that generate the brain activity. Several positions have been discussed in the literature, and the position of the reference is an important choice related to the components of interest. Three solutions are typically used in research today: first, the average activity across all the electrodes, second, an electrode on the tip of the nose, and third, two electrodes on the left and right mastoids (behind the ears). This last solution is the one that I used for my experiments and all data were re-referenced offline to the average activity of the left and right mastoids in order to avoid lateralization biases towards the left or right hemisphere.

2.1.4. Data pre-processing: Filtering and Artefact rejection

Although the registered EEG signal is amplified, its amplitude is very small [of the order of microvolts (µV)], and the SNR is weak. For improving this ratio, three main solutions are possible: first, working in an environment that is not influenced by electrical activity from the ambient, like for example a Faraday cage. Second, recording EEG with pre-amplified electrodes. And the third one is related to signal processing methods, in particular filtering and artefact rejection that (to a certain degree) eliminate noise from the data.

Filtering is used for frequency bands in which the noise is highest, that is high frequencies related to muscular movements and frequencies overlapping with ambient currents (50 or 60 Hz). Low frequency filtering is also applied, but this needs to be done with caution since it can change the characteristics of certain ERP components. For the experiments that I conducted during my thesis, I used a filter between 0.1 and 30 Hz.

Artefacts are non-permanent disturbances in the EEG signal, with larger amplitudes than the average noise. They can be of electrical or physiological origin (eye movements, cardiac and muscular activity) and eliminating these artefacts is very important. Considering eye blinks and horizontal eye movements: the eye-ball acts as a dipole and whenever the eye is moving, electric field fluctuations are generated that interfere with the EEG recording. Although participants are typically asked to control their eye movements by, for example, fixing a point on the screen, blinks and eye movements are nevertheless often present in the recordings. Therefore, recording periods (trials) that include artefacts need to be eliminated before further data processing, typically by defining an amplitude threshold (in µV): if the value of the signal exceeds the threshold (e.g., +/- 75 µV), the trial is rejected. In case the number of artefacts is too high (more than 15 % of all trials), one may consider not to include this dataset into further analyses. Eye movements can also be corrected by applying mathematical algorithms, or by the substraction of an estimated value of the disturbing phenomenon adjusted to the signal by linear regression (Gratton, Coles, & Donchin, 1983). Another method, called "Independent Component Analyses" (ICA; Hyvärinen & Oja, 2000), can be applied that decomposes the EEG signal into independent components (elementary vectors) ordered as a function of their frequency of appearance in the recorded signal. Therefore, by identifying the components that show the typical shape of an eye blink and that are distributed correspondingly on the topographic distribution maps (i.e., frontally), the components related to eye movements can be eliminated and the signal is recomputed based on the remaining sources. This type of correction for eye movements is very satisfying for keeping a maximum number of trials per condition (without rejection of trials) or for experiments in which controlling eye movements is problematic (e.g., with children or patients). However, since components that correspond to eye movements, and that are eliminated, possibly also include physiological data, the ICA algorithm introduces certain changes in the signal that may need to be considered.

2.1.5. Event-Related Potentials (ERPs)

a. How are ERPs generated?

After the pre-processing steps described above, trials are summed up and averaged to obtain the ERP and to reduce the SNR. As mentioned above, this method has been established by Dawson based on mathematical and physical concepts of signal modeling and noise, that is of stable or of stochastic phenomena. For stable phenomena, the brain electrical activity generated by perceptive and cognitive processes in response to a stimulus is supposed to be relatively similar each time the same stimulus is repeated. Thus, the signal is stable throughout time if the EEG is synchronized to stimulus onset. By contrast, stochastic phenomena activity is typically not synchronized to stimulus onset. Therefore, stable variations of the electrical activity related to stimulus presentation will emerge from the EEG, compared to random variations related to noise that will be eliminated (see Figure 12). Importantly, the number of trials required to produce a relatively noise-free ERP is highly variable and depends on the size of the ERP component of interest (i.e., SNR): while for large ERP components (e.g., N400, P300) in general a minimum of about 20-30 trials is necessary to obtain a good SNR, this number can increase to several hundreds or thousands of trials for smaller components (e.g., brainstem responses, P50).

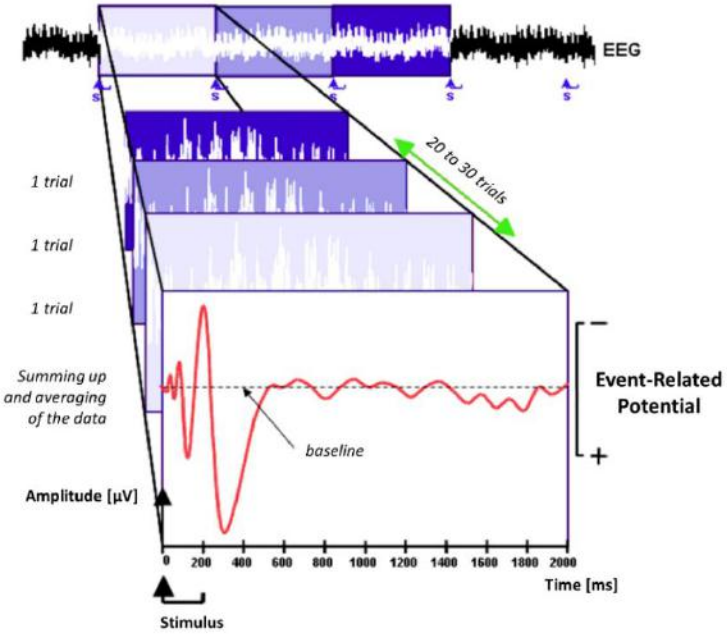


Figure 12. Event-Related Potential (ERP; red trace) in response to the presentation of an auditory stimulus. Several trials need to be averaged to obtain a sufficient signal-to-noise ratio allowing to see the ERP.

b. ERP components

ERPs allow to track the brain electrical activity with a temporal precision that depends upon the sampling rate [generally of the order of milliseconds (ms)]. Several ERP components are characterized by their polarity, latency, topographic distribution and functional significance. Since it is of main interest for my work, I will focus on previous literature related to ERP components in the auditory modality.

b.1. Short-latency responses

The earliest responses, within the first 10 ms after stimulus presentation, arise from brainstem structures and are called Auditory Brainstem Responses [ABRs; also called Brainstem Auditory Evoked Responses (BAERs) or Brainstem Auditory Evoked Potentials (BAEPs)]. ABRs are characterized by very small amplitudes (of approximately one µV), due to which 500 or more repetitions of the auditory stimulus are required in order to average out the random background electrical activity. In humans, Jewett and colleagues (Jewett, Romano, & Williston, 1970) were the first to describe ABRs with latencies as short as 1 - 4 ms, evoked by sounds and detected at the vertex. One year later, by using a linearized analogue averager and a slow repetition rate, Jewett and Williston (Jewett & Williston, 1971) were able to demonstrate that ABRs first, can be recorded from the vertex in response to auditory “clicks” as a series of 7 positive waves (labelled I through VII) and second, are very consistent from run to run in the same as well as across participants. Importantly, since this study showed the possibility to investigate the functional integrity of subcortical streams of the auditory system in a non-invasive (i.e., electrodes at the surface of the scalp) and objective way (i.e., without the active involvement of the participants), this method gained a critical role for diagnosis and in clinical research.

Following the initial works of Jewett and colleagues (Jewett et al., 1970; Jewett & Williston, 1971) there has been a large number of studies on ABRs, also focusing on more complex sounds than clicks: tone-bursts (hybrids between a click and a pure sound), pure sounds, verbal sounds (such as consonants, vowels, syllables or words), musical sounds (notes or short melodies), and even environmental sounds (Krishnan, 2002; Musacchia et al., 2007; Strait, Kraus, Skoe, & Ashley, 2009). Responses to these more complex sounds arise from the inferior colliculus and are called complex ABRs (cABRs). Interestingly, cABRs resemble the complex stimulus structure that evokes it in showing a precise timing relation of its peaks to the evoking stimulus. Numerous studies on cABRs have enriched our knowledge of temporal and spectral encoding of sounds at the level of the brainstem, but since reviewing this literature is beyond the scope of my thesis, I would refer interested readers to, for example, Kraus (2011).

b.2. Mid-latency responses

A second group of responses, from 10 - 50 ms post-stimulus onset, are known as the middle-latency responses [MLRs; also called middle latency auditory evoked potential (MLAEP)], and are thought to arise from the medial geniculate nucleus of the thalamus and from the primary auditory cortex. The MLR has been described for the first time by Goldstein and Rodman (Goldstein & Rodman, 1967) and is made up of a set of positive ("P") and negative ("N") waves: the first negative wave is called Na, followed by the positive wave Pa and by the Nb, Pb and, sometimes, Nc and Pc waves. However, analyses typically focus on Na–Pa peak-to-peak amplitude since it is the most robust one. MLRs are influenced by maturity-related factors, by gender (with higher latencies for males and higher amplitudes for females), as well as by sleep (Hall, 1992). Thus, although MLRs are a promising methodology for the identification of alterations in the central auditory nervous system and its morphology is clinically important, the large variability in latency and amplitude values among subjects makes it difficult to establish values of normality.

b.3. Long-latency responses

The longer-latency components (>50 ms post-stimulus) are also called cortical evoked potentials or Event-Related Potentials (ERPs) that represent cortical processing stages, and consist of a series of positive and negative voltage deflections. Their polarity is referred to by a letter (N for negative-going and P for positive-going components), followed by a number that indicates either the component's latency (in ms) or ordinal position in the waveform (Figure 13).

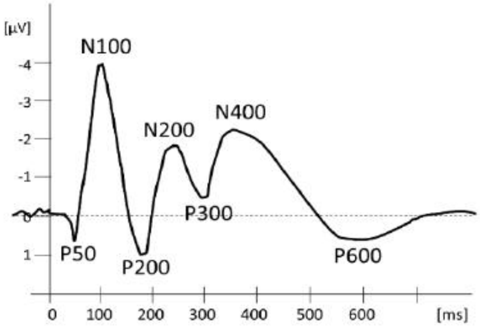


Figure 13.
Long-latency responses of Event-Related Potentials

b.3.1. P50-N100-P200

The P50 component is a positive-going waveform peaking approximately 50 ms after stimulus presentation. Typically, the P50 is used to measure sensory gating or the reduced neurophysiological response to repeated stimuli, as tested in the "paired click" paradigm (Freedman, Adler, Waldo, Pachtman, & Franks, 1983): one auditory click is presented, followed by a second click approximately 500 ms after the first one. Since the second click is redundant, it will result in decreased P50 amplitude (typically for around 80 %) compared to the first one.

The N100 component is a negative-going waveform with maximum amplitudes around 100 ms after stimulus onset over fronto-central brain regions (Näätänen & Picton, 1987). N100 amplitude is sensitive to different acoustic parameters, as well as stimulus rise time, intensity and frequency: typically, the N100 is larger to sounds with short onset times (such as piano sounds or consonants with short onset times such as “p” or “k”) compared to sounds with long onset times (such as violin sounds or vowels; Loveless & Brunia, 1990). Similarly, the higher the intensity of a sound, the larger the N100 (Picton, Goodman, & Bryce, 1970). Importantly, N100 amplitude is also sensitive to attention: the N100 is larger when participants pay attention to sounds compared to when they listen passively. This effect was described for the first time by Hillyard and colleagues in 1973, in a dichotic listening task: the participants’ attention was directed to one ear at a time and the main result showed that the N100 component of the attended ear was enhanced compared to that of the unattended ear (Hillyard, Hink, Schwent, & Picton, 1973).

The P200 component is a positive-going waveform, following the N100, with maximum amplitude typically observed 200 ms after stimulus presentation over fronto-central as well as parieto-occipital regions. Similar to the N100, the P200 component is also sensitive to the acoustic parameters of the stimuli with larger amplitude to sounds with short onset times (Shahin, Roberts, Pantev, Trainor, & Ross, 2005) and when participants pay attention to the sounds (Hillyard, Woldorff, Mangun, & Hansen, 1987; Snyder, Alain, & Picton, 2006).

b.3.2. N200

The N200 is a negative-going wave typically developing between 200 and 350 ms after stimulus onset with largest amplitude over central scalp sites. Since the N200 is typically evoked before the motor response, it has been linked to processes of stimulus identification and discrimination (Hoffman, 1990). Interestingly, N200 latency allows to measure the time that is needed for the detection and discrimination of a stimulus (Ritter, Simson, Vaughan, & Friedman, 1979), as well as the difficulty of discrimination (Ford, Roth, & Kopell, 1976). Three distinct N200 components have been identified: first, the N2a that is elicited in a task-independent manner (i.e., regardless of attention) in auditory oddball paradigms[16] and that reflects the automatic processing of the difference between the deviating stimulus and the sensory-memory representation of the standard stimulus. This auditory N2a is probably better known as the Mismatch Negativity (MMN). Näätänen and colleagues (Näätänen, Gaillard, & Mäntysalo, 1978)

---

[16] The oddball paradigm describes an experimental design in which a sequence of repetitive stimuli is infrequently interrupted by a deviant (i.e., “oddball” stimulus) and in which the participant’s reaction to the deviant is recorded.

demonstrated the temporal dissociation between the N2a and the N100 component: they showed larger N200 components to infrequent unattended changes compared to standard stimuli during a listening task. The delayed latency of this effect (i.e., 200-300 ms after stimulus presentation) compared to the attention-dependent N100 effect described by Hillyard and colleagues (Hillyard et al., 1973) was considered to be related to the rate of stimulus presentation: while Hillyard et al used very short inter-stimulus-intervals (ISIs) to avoid “unnecessary processing” (p. 322, Näätänen et al., 1978) and to speed up stimulus processing, Näätänen and colleagues used constant ISIs of 800 ms allowing slower processing.

Second, the N2b that corresponds to voluntary processing and that is elicited by controlled processes when participants selectively attend to deviations in oddball paradigms. Moreover, in contrast to the N2a, the N2b is not only elicited in auditory tasks and has also been found to reflect alterations in orthography, phonology, and semantics (Patel & Azzam, 2005). For example, Connolly and colleagues (Connolly, Phillips, Stewart, & Brake, 1992) revealed larger N200s to terminal words of spoken low than high constraint sentences with no influence of auditory masking on N200 latency. The authors interpreted this result as showing that the N200 was related to a phonological process that assessed whether the initial sounds of a word fit the sentential context. Specifically, on the basis of the preceding context a semantic expectation of a particular word is formed, and if the initial phonemes of the perceived word do not match the initial phonemes of the expected word (as in low constraint sentences), an N200 is elicited, whose amplitude is larger than if the initial phonemes of the perceived word matches the initial phonemes of the expected word (i.e., high constraint sentence). Moreover, the authors interpreted that the N200 reflecting phonological processing was a functionally distinct component compared to the N400 (a component that typically reflects conceptual processing, details are described below).

Two years later, Connolly and Phillips (Connolly & Phillips, 1994) evidenced that the N200 might reflect some kind of “pre-processing phase” related to contextual influences. In this study, participants were presented with medium to high constraining context sentences that ended with a semantically correct word or an anomalous word beginning with a similar or different phoneme compared to the expected word. An early negativity (between 270 and 300 ms) was elicited only by anomalous words with onset phonemes that did not match the expected word. This response to the discrepancy between initial phonemes of the expected and the actually perceived word was called the “phonological mismatch negativity” (PMN) and was taken to reflect the formation of specific phonological expectations during sentence processing. Several years later, van den Brink and colleagues (van den Brink, Brown, & Hagoort, 2001) investigated the time course of contextual effects on spoken-word recognition to provide more insight into early (i.e., N200) versus late (i.e., N400) context effects. The authors used a design very similar to Connolly and Phillips (1994), and revealed congruency effects on two negative deflections, one peaking at 200 ms and the other at 400 ms, supporting the idea that the N200 and N400 reflect two distinct effects involved in spoken-word recognition. Importantly, their results showed that sentence contexts have an early influence on the N200 component.

Third, early studies also identified an N2c: unlike the N2a and N2b, its latency covaries with reaction times and its scalp distribution is modality-specific, with posterior distributions for visual stimuli and fronto-central distributions in the auditory modality. This N2c is interpreted as reflecting stimulus classification during physical and semantic discriminations (Ritter, Simson, & Vaughan, 1983).

Nowadays, the subdivision of the N200 into distinct components is still a matter of debate and generally N200s are divided based on their scalp distribution with anterior N200 components involved in cognitive control, novelty, and sequential matching and with posterior N200s belonging to the family of attention-related components obtained in oddball paradigms (for review see Folstein & Van Petten, 2008).

b.3.3. N400

For already more than 35 years researchers have tried to understand how cognitive processes, such as processing the meaning of words, are reflected in the ERPs. Kutas and Hillyard were the first to show that a negative-going waveform typically elicited between 300 and 600 ms post-stimulus presentation, the N400 component, was larger to words that were incongruent compared to congruent relative to the sentence context in which they were presented (Kutas & Hillyard, 1980). The authors suggested two possible interpretations: first, the N400 could possibly depict a continuation of the sustained negative potential (Contingent Negative Variation, CNV) that developed across the course of the sentence that had to be read. Specifically, the participant could anticipate a further word following the incongruent word to resolve the ambiguity, and this expectation could prolong the CNV. Second, the so-called N400 effect (i.e., the difference between N400 amplitude to incongruent minus congruent words) could reflect the greater difficulty to integrate the meaning of an incongruent word related to the sentence context and the "re-processing" when trying to extract meaning from senseless sentences.

Three years later, in 1983, Kutas and Hillyard showed that the N400 was also elicited by incongruent words in intermediate sentence positions, thereby ruling out the possibility that the N400 reflects a continuation of the CNV. They also found that the N400 component was independent from syntactic processing[17], and largest over centro-parietal scalp sites[18] (Kutas & Hillyard, 1983). Moreover, by using a "cloze procedure" and including high or low probable words in high or low constrained contexts, the authors evidenced, in 1984, that N400 amplitude reflects the extent to which a word is semantically primed: its amplitude increases progressively to words of decreasing cloze probability (Kutas & Hillyard, 1984).

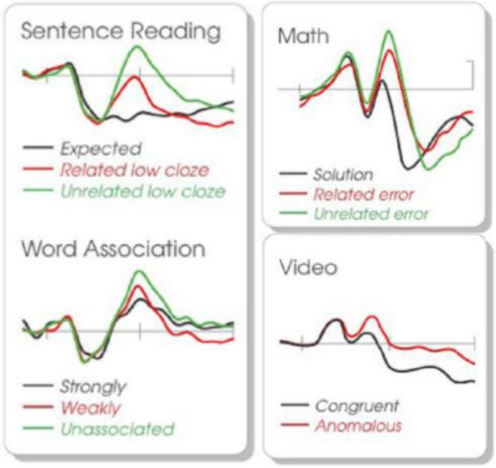


Figure 14.
Examples of stimulus types in different conditions that elicit N400s, with different amplitudes as a function of the stimulus' relatedness. (Adapted from Kutas & Federmeier, 2011)

Following these original investigations, the N400 effect was reproduced many times in both the visual and auditory modalities. Interestingly, although the N400 was long considered as

---

[17] Note that recently studies showed that this is not necessarily the case, since syntactic constructions may have meaning (Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; Magne, Besson, & Robert, 2014). Whether semantic and syntactic processing are two independent processes or not is still an open issue and hotly debated.

[18] As already mentioned in Chapter 1, an N400 may also be observed over frontal scalp sites, related to word meaning acquisition (François, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007), and/or familiarity-based recognition (the FN400; Rugg & Curran, 2007). Whether the FN400 is functionally distinct from the N400 indexing semantic processing described here, remains an open issue, with studies against (Voss & Federmeier, 2011) and studies in favor (Bridger, Bader, Kriukova, Unger, & Mecklinger, 2012) of such functional distinction.

a “language” measure, numerous studies evidenced the impressive range of stimulus types that can elicit the N400s, such as words, pseudowords, drawings, photos, faces, objects as well as mathematical symbols (see Figure 14, and for a review see Kutas & Federmeier, 2011). This current interpretation of the N400, that it is also sensitive to non-linguistic manipulations (and therefore domain-general), and that it is effective for probing semantic memory and determining how the neurocognitive system dynamically uses bottom-up and top-down information to make sense of the world, needs to be underlined: in fact (and related to this thesis), it gives us a unique possibility to study similarities and differences between meaning processing of speech and music, and transfer effects from one domain to the other.

Related to the understanding of similarities versus differences between music and speech processing, it is interesting to ask if N400s can be elicited when processing the meaning of musical concepts. A first study by Besson and Macar (Besson & Macar, 1987) did not evidence such N400 effects during music processing: participants were tested on linguistic sentences as well as on musical phrases from well-known melodies. In both conditions, incongruent stimuli were used in the final position to test for the occurrence of N400 components that were elicited only by the presentation of incongruent words. Incongruent notes (i.e., wrong notes) elicited late positive components, thereby evidencing that different processes are involved when processing violations in language and music. The authors interpreted the lack of N400 effect in music as showing that participants possibly relied upon melody recognition from memory: since melodies were very simple, from the children repertoire, participants possibly recognized them as soon as the first notes were presented, without engaging in more elaborate processing based on specific musical knowledge (comparable to sentence processing). Insofar as wrong notes were surprising, they elicited P300-like positivities rather than N400 components.

By contrast, Koelsch and colleagues (Koelsch et al., 2004), used a typical priming protocol with either linguistic or musical phrases as primes and words as targets. The target word was processed differently as a function of its conceptual relation with the prime. The authors interpreted this result as showing that an N400 was not specific to the processing of linguistic stimuli, and that processing of meaning in music shared, at least in part, cognitive processing with processing of meaning in speech. However, these results only indirectly argued for conceptual processing during music processing because words were used as targets and musical excerpts were not tested as targets. In a follow-up study using musical excerpts as targets (with words as primes), Daltrozzo and Schön (Daltrozzo & Schön, 2009) showed a conceptual relatedness effect during music processing, very similar in latency and topography to word targets. These results point to overlapping processes in music and language and confirm the presence of an N400 component to musical stimuli.

In sum, the view of the N400 as a marker of semantic processing shifted towards a view of the N400 as reflecting more broadly concept processing across different interacting cognitive domains: in fact, the N400, rather than reflecting the activation of a word's meaning, may be more accurately described as reflecting the activity in a multimodal long-term memory system that is dynamically created and context dependent (Kutas & Federmeier, 2011).

b.3.4. P300, P600, Late Positive Component

The N200 and N400 components described above are frequently followed by positivities of large amplitude with frontal or centro-parietal distributions. Their latency is thought to index classification speed (Kutas, McCarthy, & Donchin, 1977) with maximum amplitude around 300 ms (e.g., for simple sounds) and 600 ms (e.g., for words) after stimulus presentation, resulting in different names (and distinct components with eventually different functional significances): P300, P600, or more generally "Late Positive Component" (LPC). The P300 component was described for the first time by Sutton and colleagues (Sutton, Braren, Zubin, & John, 1965) in response to the presentation of deviant stimuli in a typical oddball paradigm. Subsequently, Squires and colleagues (Squires, Squires, & Hillyard, 1975) showed that the P300 possibly comprises two components: the P3a and P3b. The P3a shows a fronto-central distribution and reflects the automatic orientation of attention to a surprising event (Courchesne, Hillyard, & Galambos, 1975), processes involved in stimulus evaluation (Donchin, 1981), as well as the processing of novelty (Polich, 2003). By contrast, when the event is expected in an active manner or is relevant for the task at hand, a P3b component is elicited (Polich, 1989). The P3b is centro-parietally distributed and interpreted to reflect stimulus categorization processes, decision making and degree of confidence in decision making, as well as cognitive workload (Donchin, 1981; for a more recent review see Polich, 2007). Moreover, the P3b has also been related to subsequent memory storage: in the field of explicit recognition memory, this ERP component was rather called LPC (for a review see Friedman & Johnson, 2000), and is thought to be linked to consolidation processes and context updating (Donchin & Coles, 1988). In addition, LPC amplitude is sensitive to levels of processing manipulation and is larger for more deeply encoded items (Paller, Kutas, & McIsaac, 1995), suggesting that LPC amplitude is also closely related to recollection, and strategic and effortful information retrieval (Rugg & Curran, 2007; Swaab, Brown, & Hagoort, 1998; Van Petten, Kutas, Kluender, Mitchiner, & McIsaac, 1991).

Finally, as mentioned above, late positivities in the domain of language processing have also been referred to as P600 components, typically elicited by sentences with syntactic violations (Hagoort, Brown, & Groothusen, 1993) or in sentences with complex or non-preferred syntactic structure (Osterhout & Holcomb, 1992). The question of whether the P600 is specific to syntactic processing has long been debated. Nowadays, researches consider the P600 either as a specific component reflecting interpretative and integrative brain processes (Brouwer, Fitz, & Hoeks, 2012) or as belonging to the P300 family of components (Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014).

2.1.6. Functional Connectivity

a. Neuronal oscillations

Neurons do not function in isolation, but they are embedded in networks that are rhythmically activated and inhibited (Buzsáki, 2006). This rhythmicity is reflected in oscillations[19] of the extracellular field potential that can be registered with EEG. The frequency of such neural oscillations depend on various time constants and network properties and may range from slow activity with oscillation periods of several seconds, to fast activity with one cycle lasting a few milliseconds. The following oscillations can be captured with EEG: delta (0-3 Hz), theta (3-7 Hz), alpha (8-12 Hz), beta (12-28 Hz) and gamma (>28 Hz). Oscillations of different frequencies can occur at the same time in the same brain regions (Buzsáki & Draguhn, 2004). As a function of the anatomic origin (and often also of the experimental context), each frequency band corresponds to different types of activity or cognitive states (Engel, Fries, & Singer, 2001): delta oscillations can be observed during profound sleep, and theta oscillations may be observed during states of drowsiness (daytime sleepiness, meditation ...) and memorization. Alpha waves correspond to a diffuse awake state (i.e., at rest with eyes closed and relaxed), and seem to play

---

[19] An oscillation is a variation of a mechanical, electrical magnitude, characterized by a periodic change of direction around a position of stable equilibrium, successively passing through a maximum and a minimum value. The oscillations can be regular (periodic) or decreasing (damped). The cycle of an oscillation is the time elapsed between two successive passages by the equilibrium position. The frequency of an oscillation is the number of cycles per second and is expressed in Hertz (Hz).

an important role in inhibition processes. Beta oscillations can be observed during active or intense concentration periods as well as during emotional states. Finally, gamma oscillations can reflect several brain functions, especially related to attention and perceptual binding.

b. Phase synchronization

As mentioned earlier, groups of neurons communicate across brain networks. This may be possible through, for example, a process called phase synchronization. Phase synchronization is a fundamental neural mechanism that supports neural communication, neural plasticity, as well as cognitive processes (Engel et al., 2001). Two brain regions are characterized by phase synchronization when the oscillatory phases in both regions are correlated (see Figure 15A).

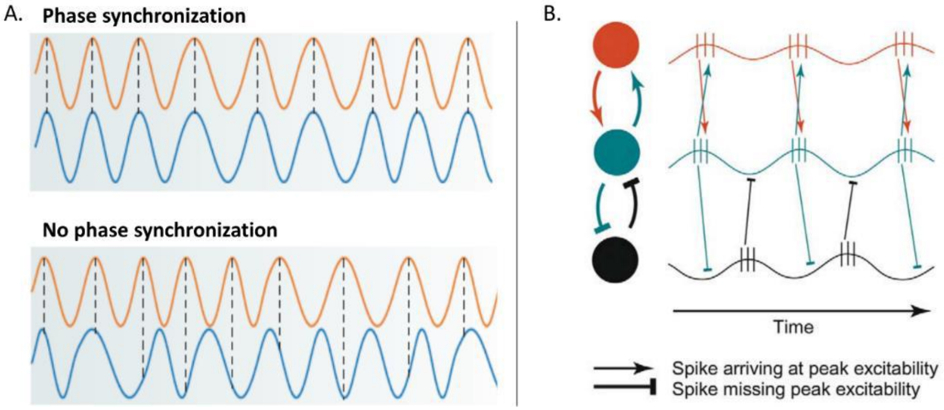


Figure 15. (A) Neural oscillations may show stable phase relationships (above), or variable phase relationships (below). (Adapted from Fell & Axmacher, 2011) (B) Neuronal communication. Spikes that arrive at excitability peaks of the receiving neuronal group have pointed arrowheads, spikes that miss excitability peaks have blunt arrowheads. Red and blue neuronal groups undergo coherent excitability fluctuations. Thus, their communication is effective. By contrast, the black neuronal group undergoes excitability fluctuations that are not coherent with the blue neuronal group. Communication between these two neuronal groups is therefore prohibited. (Adapted from Fries, 2005)

In networks of synchronized neurons, the oscillatory phase is defined as the angle that corresponds to the momentary deflection of an oscillation and is taken to reflect the degree of excitability of the neurons. Thus, the role of neural oscillations during communication can be thought of as follows: oscillations reflect fluctuations in neural excitability, and these fluctuations not only affect neural output but also neural input (Volgushev, Chistiakova, & Singer, 1998). If action potentials generated during the depolarized phase of an oscillation are transferred from one brain area to another one and reach the target area during the depolarized phase of an oscillation in that area, they are likely to also trigger action potentials in that area. By contrast, if there is no phase synchronization (or phase synchronization with inappropriate phase lag), the action potential will arrive at a phase in which sensitivity to neural input is non-optimal and communication is likely to be blocked. In sum, only phase-locked neuronal groups can communicate effectively, because their communication windows for input and for output are open at the same time (Fries, 2005; see Figure 15B).

Interestingly, neural oscillations are not only phase synchronized across different brain regions, but low-frequency oscillations may also synchronize with high-frequency oscillations, an effect known as "m:n phase synchronization" and related to memory processes (Fell & Axmacher, 2011). In such a case, not the entire cycle but individual phases of the single cycle of an oscillation with lower frequency (e.g., theta oscillation of 5 Hz) is synchronized with a specific phase of an oscillation with a higher frequency (e.g. gamma oscillation of 40 Hz in a 1:8 phase synchronization).

c. Functional connectivity

Functional connectivity is defined as the correlations between spatially remote neurophysiological events (compared to effective connectivity that describes the influence of one neuronal system on another, based on computational modelling; Gerstein & Perkel, 1969; Park & Friston, 2013). Thus, functional connectivity provides a characterization of functional interactions, through the observed correlations. Importantly, since it is based on correlation analyses, it cannot infer causality (i.e., it cannot infer that one event that followed another one was necessarily a consequence of the first event), and it cannot reveal information on how these correlations are mediated.

In my thesis, functional connectivity was evaluated using the current implementation of the sLORETA software (i.e., standardized LOw REsolution brain electromagnetic Tomography; Pascual-Marqui et al., 2011). EEG data are re-referenced to an average reference[20], and

---

[20] Using an average reference for inverse solution methods of EEG data (i.e., methods to reconstruct the intracranial sources of surface-recorded data) is an important issue because of the biophysical assumption of quasi-stationarity predicting that the net source activity at each instant in time within the brain sums to zero. Since the reference electrode adds a constant potential value to the value recorded at each electrode and instant in time, a "re-centering" of the data (i.e., a removal of this constant value) is necessary before applying an inverse solution to avoid violating this assumption. Mathematically this is equivalent to calculating the average reference of the surface-recorded EEG (Pascual-Marqui & Lehmann, 1993).

computations are made within a realistic head model (Fuchs, Kastner, Wagner, Hawes, & Ebersole, 2002) relying on the Montreal Neurological Institute 152 templates (Mazziotta et al., 2001). The 3D solution space is restricted to cortical grey matter, as determined by the probabilistic Talairach atlas (Lancaster et al., 2000), and the intracranial volume is partitioned in 6239 voxels at 5 mm spatial resolution. To determine functional connectivity, spatially remote regions of interest (ROIs) are selected, and, as recommended by sLORETA, the method of using a single voxel at the centroid of each ROI was selected for connectivity calculation for the following reason: source calculation by sLORETA is based on the assumption that the smoothest of all possible activation distributions is the most plausible one, and this is supported by neurophysiological data demonstrating that neighboring neuronal populations present highly correlated activity (Michel, Koenig, Brandeis, Gianotti, & Wackermann, 2009). Thus, signals of spatially adjacent voxels of neighboring ROIs are highly correlated, inducing larger connectivity, which might not be physiological in nature. By taking the single centroid voxel of each ROI, that is an accurate representative for activity within the ROI (Jäncke & Langer, 2011), such contamination for connectivity estimation can be minimized.

Finally, different measures of synchronization are available to compute functional connectivity, and one frequently used example is lagged coherence. Lagged coherence is a measure of variability between two oscillatory time series in specific frequency bands (Lehmann, Faber, Gianotti, Kochi, & Pascual-Marqui, 2006; Thatcher, 2012), based on the estimation of the linear dependency of two signals in a frequency band. Thus, it refers to the equivalent in the frequency domain, as correlation measures in the temporal domain: coherence values can vary between 0 and 1, and a perfectly linear relation between two signals in a certain frequency band would be equal to 1. By contrast, the complete absence of a linear dependency is represented by 0. Importantly, in sLORETA lagged coherence analyses are independent from volume conduction since the instantaneous zero-lag contribution has been excluded mathematically (Nolte et al., 2004; Stam & van Straaten, 2012). This is an important issue when analyzing functional connectivity based on EEG data, since zero-lag connectivity in a given frequency band is often due to non-physiological effects or intrinsic physical artifacts (e.g., volume conduction and low spatial resolution).

2.1.7. Advantages & Limitations

Practical advantages of EEG are related to its non-invasiveness, with no real safety restrictions, as well as to its low cost. Moreover, the EEG equipment can be easily transported, thereby allowing experiments in more natural environments (e.g., in schools). Furthermore, theoretical advantages include the ability to capture electrical fluctuations of large groups of neurons and brain activity as it unfolds in real time at the level of milliseconds. By contrast, limitations are linked to its poor spatial resolution since EEG recordings are obtained from the surface of the head. By putting lots of electrodes all over the scalp (e.g., 64, 128 or 256 electrodes), it is possible to get some idea of where effects are strongest, but this cannot tell where in the brain the signal originates. Interestingly, there are modern analyzing techniques (e.g., Backus-Gilbert; weighted resolution optimization, WROP; low resolution brain electromagnetic tomography, LORETA; bidomain model; S-MAP; for a review see Grech et al., 2008) trying to solve this so-called inverse problem for EEG data. These signal processing techniques allow to estimate the current sources inside the brain that best fits the measurements of the actual EEG data registered from specified positions of electrodes on the scalp (the more electrodes, the better), thereby revealing interesting information. However, it needs to be kept in mind that the accuracy with which a source can be located is affected by a number of factors, including head-modelling errors, source-modelling errors and EEG noise (instrumental or biological).

2.2. Magnetoencephalogram

Magnetoencephalography (MEG) is another online measure of brain activity, and some characteristics are very similar to EEG: MEG is characterized by an excellent temporal but by a limited spatial resolution[21]. However, there are important differences compared to EEG: a practical advantage of the MEG is that no electrodes need to be attached to the scalp and the subject's head is simply positioned close to the surrounding coils. The magnetic field emerges from the brain through the skull and the scalp without any distortion allowing better spatial

---

[21] This is a debated issue. The spatial resolution of MEG is better compared to EEG, mainly because magnetic fields are not deformed by the tissues they have to pass through and the underlying activity is therefore more easily localizable. However, compared to techniques with high spatial resolution (MRI) the resolution of MEG remains limited. Moreover, only signals from sources not too deep in the brain can be recorded using MEG.

resolution with MEG than with EEG. Moreover, MEG differs from EEG related to its sensitivity to the spatial orientation of the dipoles. Brain currents result of the activity of groups of neurons with dendrites perpendicular to the cortical surface (radial source). However, only currents that have a component tangential to the surface produce a magnetic field that can be recorded by MEG. Thus, the MEG signal is more sensitive to tangential dipoles (i.e., originated in the sulci but not on the top of the cortical gyri) than to radial dipoles. Partly for this reason, MEG and EEG "see" different types of activity, with MEG being sensitive to activity in fewer brain regions.

2.3. Structural and functional magnetic resonance imaging

Structural Magnetic Resonance Imaging (MRI) allows two- or three-dimensional views of the brain, and functional MRI (fMRI) is used for the precise localization of brain areas with higher blood flow in response to different perceptive, cognitive or motoric activities. These two methods are based on Nuclear Magnetic Resonance (NMR) that allows to analyze different tissues with excellent spatial resolution.

Participants are placed in a strong magnetic field (varying between 1.5 and 7 Tesla as a function of the machine) that acts on the magnetic properties of certain atom nuclei, mainly hydrogen nuclei (or protons), that are highly present in biological tissue. To obtain the NMR signal two steps are necessary: first, the magnetic field induces the alignment of the proton spins (i.e., the vector corresponding to the axe of rotation) of the hydrogen atom in a parallel or antiparallel position relative to the magnetic field. These two positions correspond to different, not equivalent energy levels. Less energy is needed for moving a proton into a parallel than into an antiparallel position, resulting in more protons in parallel positions. Second, energy is applied in form of radio frequency waves, whose frequency corresponds to the resonance frequency of the protons. Thus, the protons enter into resonance and align themselves in an antiparallel position. As soon as the emission of radio frequency waves is stopped, the protons fall back to their initial position (i.e., parallel to the magnetic field) and release their energy in form of a weakened sinus curve: the MRI signal. This signal differs as a function of the water composition of the tissue and provides information related to the anatomy of the tissue.

By contrast, the fMRI signal relies on the so-called BOLD effect (Blood Oxygenation Level Dependent effect; Kwong et al., 1992; Ogawa, Lee, Kay, & Tank, 1990). This effect is based on the magnetization properties of the hemoglobin of the red blood cells that can be present in two forms: oxyhemoglobin or deoxyhemoglobin. The oxyhemoglobin is contained in oxygenated red blood cells of the lungs and is diamagnetic, which means it hardly influences the local magnetic field. By contrast, the deoxyhemoglobin is found in deoxygenated red blood cells of the tissues and is paramagnetic, thereby influencing the local magnetic field and reducing the NMR signal. If a brain area is activated, the local blood flow and thus the oxygen concentration goes up, resulting in higher oxyhemoglobin than deoxyhemoglobin levels. Therefore, the local variation of magnetic susceptibility is low and the NMR signal is strong (the BOLD effect). Since neurovascular coupling takes five to six seconds for reaching its maximum, the temporal resolution of the fMRI is low and not comparable to EEG and MEG methods.

In conclusion, advantages of MRI and fMRI relate to the fact that, if done correctly, these methods are noninvasive, and allow good localization of activation by producing high-resolution, three-dimensional images (without using radiation like X-rays or computed tomography). Thus, it is frequently applied also for clinical work. By contrast, limiting inconveniences are connected to the fact that the participant needs to remain completely immobile throughout the whole session, since MRI is very susceptible to movement artifacts. Moreover, it is not suitable for claustrophobic people, and, since no metal-based equipment can be around the machine, not accessible for people with metallic implants. Furthermore, the production of the electromagnetic pulses by the gradients produces intensive noise (around 130 dB), so that experiments including auditory stimuli may be challenging. Finally, it should be kept in mind that the BOLD signal is not a direct measure of neural activity, but “only” a measure of oxygen consumption and blood flow in the brain, and that it is a relatively expensive method.

2.4. Transcranial magnetic stimulation

Transcranial magnetic stimulation, unlike the other methods, is an experimental manipulation method rather than a recording technique: it’s a magnetic method to non-invasively stimulate neurons in the brain and induce reversible, immediate or long-term changes in brain activity. In practice, small regions of the brain are stimulated with a magnetic field generator (“coil”) that is placed near the head of the participant. This coil is connected to a pulse generator that delivers electric current to the coil, which in turn produces small electric currents in the brain region just under the coil via electromagnetic induction: the magnetic field passes unimpeded through the skull (generally it reaches no more than 5 cm depth into the brain), and causes a change in the transmembrane current of nearby neurons, thereby leading either to depolarization (i.e., activation) or to hyperpolarization (i.e., inhibition). Consequently, actions potentials can be captured with EEG or MEG. By using different frequencies, intensities and durations of stimulation fluctuations in brain activity are assessed, or temporary changes in brain function are induced.

Non-invasive brain stimulation has mainly two areas of functionality: first, by allowing to modify human brain activity, the potential importance of TMS is to study the consequences of an induced perturbation of a specific brain region on subsequent behavioral performance, subjective experience, or brain imaging and electrophysiological measures. This allows TMS to be used both experimentally as a mean of exploring neural function and clinically as a diagnostic and therapeutic tool. Importantly, its great strength is related to the fact that it allows brain/behavior relationships to be established causally (rather than just as a correlation as in the case of brain imaging). Second, because it can modulate brain function, it has the potential of acting as a treatment for neuropsychiatric diseases (e.g., depression, schizophrenia). Finally, it should be noted that, although the occurrence of seizures, discomfort, pain, or cognitive changes is very rare, few cases have been reported that depict the risk and disadvantage of TMS.

Chapter 3

MUSIC TRAINING AND NON-NATIVE PHONEME PERCEPTION

Dittinger, E., D'Imperio, M., & Besson, M. (2018) Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *European Journal of Neuroscience*. doi: 10.1111/ejn.13939. [Epub ahead of print]

Received: 29 September 2017 | Revised: 24 January 2018 | Accepted: 16 April 2018

DOI: 10.1111/ejn.13939

RESEARCH REPORT

WILEY EJN European Journal of Neuroscience FENS

Enhanced neural and behavioural processing of a nonnative phonemic contrast in professional musicians

Eva Dittinger[1,2,3] | **Mariapaola D'Imperio**[2,4] | **Mireille Besson**[1]

[1]CNRS & Aix-Marseille Université, Laboratoire de Neurosciences Cognitives (LNC, UMR 7291), Marseille, France

[2]CNRS & Aix-Marseille Université, Laboratoire Parole et Langage (LPL, UMR 7309), Aix-en-Provence, France

[3]Brain and Language Research Institute (BLRI), Aix-en-Provence, France

[4]Institut Universitaire de France (IUF), Paris, France

Correspondence
Eva Dittinger, Laboratory of Cognitive Neuroscience (LNC - UMR 7291), Centre Saint-Charles, Pole 3C, Case C, 3 place Victor Hugo, 13331 Marseille Cedex 3, France.
Email: eva.dittinger@blri.fr

Funding information
Institut Convergence ILCB, Grant/Award Number: ANR-16-CONV-0002; Labex BLRI, Grant/Award Number: ANR-11-LABX-0036

Abstract

Based on growing evidence suggesting that professional music training facilitates foreign language perception and learning, we examined the impact of musical expertise on the categorisation of syllables including phonemes that did (/p/, /b/) or did not (/p^h/) belong to the French repertoire by analysing both behaviour (error rates and reaction times) and Event-Related brain Potentials (N200 and P300 components). Professional musicians and nonmusicians categorised syllables either as /ba/ or /pa/ (voicing task), or as /pa/ or /p^{h}a/ with /p^h/ being a nonnative phoneme for French speakers (aspiration task). In line with our hypotheses, results showed that musicians outperformed nonmusicians in the aspiration task but not in the voicing task. Moreover, the difference between the native (/p/) and the nonnative phoneme (/p^h/), as reflected in N200 and P300 amplitudes, was larger in musicians than in nonmusicians in the aspiration task but not in the voicing task. These results show that behaviour and brain activity associated to nonnative phoneme perception are influenced by musical expertise and that these effects are task-dependent. The implications of these findings for current models of phoneme perception and for understanding the qualitative and quantitative differences found on the N200 and P300 components are discussed.

KEYWORDS

event-related potentials, music training, N200, P300, syllabic categorisation

1 | INTRODUCTION

As Patricia Kuhl once mentioned "infants are born citizen of the world," a beautiful manner to express that infants can learn any human language. Independently of how speech sounds differ from one language to the other, infants are able discriminating them at early phases of their development. However, this ability decreases by the time they are 12 months old (Werker & Tees, 1984), in a process implying a "tuning-out" of phonetic contrasts, reflected by less sensitivity to the difference between two phonemes that are not relevant for their phonological system or, seen the other way around, a perceptual attunement to their native speech. Nevertheless, nonnative speech sound contrasts remain somewhat discriminable in adult life, with performance levels that vary according not only to the specific acoustic-perceptual properties of the sound contrast under investigation but also to individual differences and training (Sadakata & McQueen, 2013, 2014).

Abbreviations: ANOVA, analysis of variance; ASP, automatic selective perception; EEG, electroencephalogram; ERPs, event-related potentials; ERRs, error rates; ICA, independent component analyses; MBEA, Montreal battery for the evaluation of amusia; MUS, musicians; NEPSY, a developmental NEuroPSYchological assessment; NM, nonmusicians; PAM, perceptual assimilation model; PM47, Progressive Matrices; RTs, reaction times; VOT, voice onset time; WISC, Wechsler Intelligence Scale for Children.

Edited by Sophie Molholm. Reviewed by Paula Virtala, University of Helsinki, Finland; and Ana Francisco, Albert Einstein College of Medicine, USA

In this experiment, we were particularly interested in studying the impact of musical expertise on native and nonnative phoneme perception, using both behavioural and electrophysiological measures. The study presented here is part of an on-going project that aims at better understanding the influence of music training on novel word learning. To this aim, participants performed a series of tasks, including syllabic categorisation tasks, novel word learning as well as matching and semantic tasks (described in Dittinger et al., 2016). More important, in Dittinger et al. (2016) data from the categorisation tasks were averaged across all syllables because the main focus was on novel word learning in professional musicians and in nonmusicians, independently of the syllables' phonetic characteristics. By contrast, here we compare native and nonnative phoneme perception in musicians and in nonmusicians by focusing on two syllabic categorisation tasks relying on the discrimination of Voice Onset Time (VOT): one based on voicing (categorising syllables as /ba/ or /pa/, with /b/ and /p/ being native phonemes for French speakers) and the other one based on aspiration (categorising syllables as /pa/ or /p^{h}a/, with /p^h/ being a nonnative phoneme for French speakers). VOT is a relevant acoustic cue for speech perception that allows to differentiate phonemic categories such as /ba/ and / pa/, for instance. It is interesting that VOT differs for the same phonemic category across languages (e.g., the French and English /p/ have allophones with VOT values that are quite different), which may be very challenging when learning a foreign language, especially if, as mentioned above, nonnative VOT contrasts have been "tuned-out" so that they are less salient to adult participants (Werker & Tees, 1984).

It is interesting to note that the impact of musical expertise on speech perception has become a hot topic in the past 15 years (Asaridou & McQueen, 2013; Besson, Barbaroux, & Dittinger, 2017; Besson, Chobert, & Marie, 2011; Slevc, 2012). Becoming a professional musician requires hours of training with music instrument(s) (more than 10,000 hr; Krampe & Ericsson, 1996) and such an intensive training has profound consequences on the anatomo-functional organisation of the brain (Jäncke, 2009; Münte, Altenmüller, & Jäncke, 2002). There is also growing evidence based both on behavioural and brain data evidencing that musicians are more sensitive than nonmusicians to several aspects of speech processing. For instance, musical expertise is known to improve speech perception at the segmental level (consonants and vowels; Chobert, Francois, Velay, & Besson, 2014; Elmer, Meyer, & Jancke, 2012; Kühnis, Elmer, Meyer, & Jäncke, 2013; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Marie, Magne, & Besson, 2011; Musacchia, Sams, Skoe, & Kraus, 2007; Sadakata & Sekiyama, 2011; Tervaniemi et al., 2009; Wong, Skoe, Russo, Dees, & Kraus, 2007) as well as at the supra-segmental level (words and sentences; Besson et al., 2011; Lima & Castro, 2011; Schön, Magne, & Besson, 2004).

The electroencephalogram (EEG) and the Event-Related Potentials (ERPs) methods have been used in numerous experiments because of their excellent temporal resolution that allows to follow the processes of interest with a precision of the order of milliseconds. This is very useful when studying the fast processes involved in speech perception and comprehension. Directly related to the present study, Bidelman and collaborators (Bidelman & Alain, 2015; Bidelman, Weiss, Moreno, & Alain, 2014) used the ERPs method to examine the influence of musical expertise on categorical perception, one of the most fundamental aspects of speech perception (Toscano, McMurray, Dennhardt, & Luck, 2010). Results showed that both younger and older musicians were faster at categorising vowels on a five-step vowel continuum from /u/ to /a/ than nonmusicians. Moreover, in young adults (Bidelman et al., 2014), the P200 component of the ERPs, peaking around 175–200 ms, was larger in musicians than in nonmusicians, possibly reflecting the positive influence of long-term auditory training on vowel perception (Seppänen, Hämäläinen, Pesonen, & Tervaniemi, 2012; Shahin, Bosnyak, Trainor, & Roberts, 2003). It is interesting that P200 amplitude and brain activity following the P200 (between 250 and 300 ms) also varied as a function of whether ambiguous stimuli at the categorical boundary were heard as /u/ or /a/ (Bidelman, Moreno, & Alain, 2013). The finding that the same ambiguous stimulus-evoked different brain activity depending upon how it was heard was taken as evidence that ERP components after 200 ms poststimulus onset no longer reflect processing of the acoustic properties of the signal but rather the perceived phonemic category of the stimulus (but see Toscano et al., 2010, for the influence of acoustic differences on later ERP components).

Based on Bidelman et al.' results (2013, 2014), we focused on ERP components that followed the P200 and specifically on the N200 and P300 components that are of main interest here for the following reasons. The N200 component of the ERPs is elicited in various discrimination and categorisation tasks (Friedrich & Friederici, 2008; Ritter et al., 1984), including phonological categorisation (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992). It is important that the fronto-centrally distributed N200 component is often associated with a later positivity of the P300-family that, among different processes, reflects memory updating when unexpected, unfamiliar events are presented, decision-related processes and the degree of confidence that participants have in their response (Donchin & Coles, 1988; Polich, 2007).

To investigate the impact of musical expertise on native and nonnative phoneme perception, we considered two models, the Perceptual Assimilation Model (PAM; Best & Tyler, 2007) and the Automatic Selective Perception model (ASP; Strange, 2011) that allow to make specific predictions regarding the effects of context on syllabic categorisation. PAM is based on the idea that the process of perceptual attunement to native categories is accompanied by the tuning-out of categories that are not used to create meaning contrasts (that are hence "nonphonemic" in the language in question) and that are consequently assimilated to the existing ones (e.g., because they share articulatory events at a distal level, according to a direct-realist approach). An important consequence of this theory is that "If a given nonnative phonological contrast is perceptually assimilated as phonetically similar to a phonological distinction, that is, a contrast, in the native language, then it should be easily discriminated" (Tyler, Nil, Best, Faber, & Levitt, 2014, p. 6). Therefore, success in categorisation of nonnative sounds depends on how easily they can be assimilated (i.e., considered as part of the same native category) or not to existing categories.

In the present experiment, we asked participants to categorise syllables that started with one of three phonemes of Thai, /b/, /p/ and /p^h/ either based on an existing, similar French native contrast (/ba/ vs. /pa/) in the voicing task or based on a nonnative contrast (/pa/ or /p^{h}a/) in the aspiration task. In the voicing task, we expected exemplars of the /b/ and /p/ phonemes (that both exist in the French repertoire) to be readily categorised into their respective native phonemic categories /ba/ or /pa/. In addition, and in line with PAM, we predicted that the /p^h/ Thai phoneme (that does not exist in French) would be perceptually assimilated to the French phoneme /p/ and, consequently, that it would also be easily discriminable from /b/. In other words, based on PAM, we expected the /p^h/ phoneme to be assimilated to /p/ in the voicing task with no or small differences between /p/ and /p^h/ in behaviour (i.e., error rates and reaction times) and in the electrophysiological responses (i.e., N200 and P300; Bidelman et al., 2013; Connolly et al., 1992; Connolly & Phillips, 1994). By contrast, as /p^h/ cannot be assimilated to the /p/ category in the aspiration task, we expected differences between /p/ and /p^h/ to be found both in behaviour and in ERP measures (N200 and P300 amplitudes). In other words, we predicted that the same nonnative phoneme /p^h/ would be processed differently depending on whether it can be assimilated to /p/ or not and that this difference would be reflected both in behavioural and electrophysiological measures.

The ASP model (Strange, 2011) allowed us to go a little further in the processing mechanisms involved in the two tasks. In this model, Strange (2011) argues that if phonemes belong to different phonemic categories, as /b/ and /p/ in the voicing task, discrimination is based on fast, automatic, phonological selective perceptual routines. However, if phonemes differ by subtle phonetic details, as /p/ vs. /p^h/ in the aspiration task, with /p^h/ possibly being perceived as an allophone of /p/ by French speakers, discrimination would be based on phonetic selective processing routines that require focused attention and enhanced cognitive resources. Of interest here was to determine whether the potential involvement of two different selective perceptual routines would translate into different patterns of ERP responses. In a specific manner, we predicted that if phonological routines are involved in the voicing task, and if both /p/ and /p^h/ are assigned to the same phonemic category, both phonemes should elicit N200 components of similar amplitude. By contrast, if phonetic routines are involved in the aspiration task and if /p/ is assigned to a preexisting phonemic category but /p^h/ is not, the N200 component should be larger to /p/ than to /p^h/ and the P300, that is taken to reflect the processing of an unfamiliar, unexpected phoneme and decision-related processes should be larger to /p^h/ than to /p/.

At last, specific predictions can be made regarding the impact of musical expertise on native and nonnative phoneme categorisation. Previous results in the literature showed increased auditory sensitivity (Bidelman & Krishnan, 2010; Marie, Delogu et al., 2011, Marie, Magne, et al., 2011; Moreno et al., 2009; Sadakata & Sekiyama, 2011), as well as enhanced selective attention in musicians compared to nonmusicians (Strait, Slater, O'Connell, & Kraus, 2015; Wang, Ossher, & Reuter-Lorenz, 2015). More important, it was also found that musicians outperform nonmusicians when the acoustic/phonetic difference between two stimuli is difficult to perceive but not when it is easy (Elmer et al., 2012; Schön et al., 2004). Thus, based on these results, we predicted larger between-group differences in behaviour and ERPs in the aspiration task that is most difficult and possibly requires attentional and cognitively-demanding phonetic routines, than in the voicing task that is likely to rely on fast automatic phonological routines available to both musicians and nonmusicians.

In sum, we predicted that the differences in behaviour and brain activity between the native phoneme /p/ and the nonnative phoneme /p^h/ would depend upon musical expertise and upon the task at hand (voicing vs. aspiration tasks). Showing that music training can boost nonnative phoneme perception has potentially strong implications for foreign language learning.

2 | MATERIALS AND METHODS

2.1 | Participants

A total of 31 participants with 16 professional musicians (MUS, 8 women) and 15 controls without formal music education (nonmusicians, NM, 8 women) but involved in a regular leisure activity (e.g. sports, dance, theatre) were paid to participate in the experimental session. The two groups did not differ in age (MUS: mean age = 25.4 years, age range = 19–30, $SD = 3.9$; NM: mean age = 25.7 years, age range = 19–33, $SD = 4.8$; $F_{(1,29)} = 0.03$, $p = 0.85$). All participants were native French speakers, had comparable education (university degree) and socio-economic background (criteria of the National Institute of Statistics and Economic Studies; MUS: 4.4; NM: 4.9; $F_{(1,29)} = 1.45$, $p = 0.24$), and reported no past or current audiological or neurological deficits. MUS practiced their instruments for an average of 17 years (range = 11–24, $SD = 4.1$) and the musician group included three pianists, two accordionists, four violinists, one cellist, two guitarists, one hornist, one tubist, and one flutist. None of the participants was bilingual, but all spoke English as a second language and most participants (except for 1 MUS and 3 NM) had a rudimentary knowledge of a third language that was neither tonal nor quantitative. This study was approved by the local ethic committee of Aix-Marseille University and was conducted in agreement with the Helsinki declaration for human participants. All participants gave their informed consent before enrolling in the experiment and they received monetary compensation for their participation.

2.2 | Screening measures

2.2.1 | Cognitive ability

Standardised psychometric tests were used to examine short-term- and working memory (forward and reverse Digit Span, WISC-IV: Wechsler Intelligence Scale for Children; Wechsler, 2003), visual attention (NEPSY: a developmental NEuroPSYchological assessment; Korkman, Kirk, & Kemp, 1998) and nonverbal general intelligence (progressive matrices, PM47; Raven, Corporation, & Lewis, 1962).

2.2.2 | Musical aptitude

Participants performed two musicality tests [adapted from the Montreal Battery for the Evaluation of Amusia (MBEA); Peretz, Champod, & Hyde, 2003] consisting in judging whether pairs of piano melodies were same or different, based either on melodic or on rhythmic information.

2.3 | Experimental stimuli

Nine syllables were presented: /ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/, /p^{h}a:1/, /ba:0/, /pa:0/, /p^{h}a:0/ with 1 referring to low tone (F0 = 175 Hz), 0 to mid-tone (F0 = 218 Hz), the colon to long vowel duration (530 ms), no colon to short vowel duration (260 ms), and p^h to aspiration. As we were specifically interested in how the native /p/ and nonnative /p^h/ phonemes were processed in the two categorisation tasks, the different types of /p/ (i.e., /pa1/, /pa:1/, /pa:0/) and /p^h/ (i.e., /p^{h}a1/, /p^{h}a:1/, /p^{h}a:0/) were averaged together leading to 30 trials per phoneme in each task. The VOT for /b/ (/ba1/, /ba:1/ and /ba:0/) was of −144 ms, for /p/ (/pa1/, /pa:1/ and /pa:0/) of 3 ms and for /p^h/ (/p^{h}a1/, /p^{h}a:1/ and /p^{h}a:0/) of 77 ms. All nine syllables were recorded by a female Thai-French early bilingual, ensuring that all syllables were produced naturally. For each syllable, 5 versions were digitally recorded to reproduce natural speech variability. Sound pressure level was normalised across all syllables to a mean level of 70 dB using the Praat software (Boersma & Weenink, 2011).

2.4 | Experimental tasks

Participants were tested individually in a quiet experimental room (i.e., Faraday cage), where they sat in a comfortable chair at about 1 m from a computer screen. Auditory stimuli were presented through HiFi headphones (Sennheiser, HD590) at 70 dB sound pressure level. Stimuli presentation, as well as the collection of behavioural data, were controlled by the "Presentation" software (NeuroBehavioral Systems, Version 11.0). Participants performed two different categorisation tasks based on VOT: the voicing task, in which they had to categorise the syllables as /ba/ (including /ba1/, /ba:1/ and /ba:0/) or /pa/ (including /pa1/, /pa:1/, /pa:0/, /p^{h}a1/, /p^{h}a:1/ and /p^{h}a:0), and the aspiration task, in which they had to categorise the syllables as /pa/ (including /pa1/, /pa:1/, /pa:0/) or /p^{h}a/ (including /p^{h}a1/, /p^{h}a:1/ and /p^{h}a:0). In the voicing task, participants pressed one response key if they heard a /ba/ and the other key if they heard a /pa/. In the aspiration task, they pressed one response key if they heard a /pa/ and the other key if they heard a /p^{h}a/. The side of response keys was balanced across participants who were asked to respond as quickly and accurately as possible. No feedback was given. In both tasks, each syllable was presented 10 times in a pseudo-randomised order with the constraints of no immediate repetition of the same syllable, and no more than four successive same responses. The inter-stimulus-interval was 1500 ms, so that the voicing task lasted for about 2.3 min and the aspiration task lasted for about 1.5 min. Task order was counter-balanced across participants.

2.5 | EEG data acquisition

The electroencephalogram was continuously recorded at a sampling rate of 512 Hz with a band-pass filter of 0–102.4 Hz using a Biosemi amplifier system (Amsterdam, BioSemi Active 2) with 32 active Ag/Cl electrodes (Biosemi Pintype) located at standard positions according to the international 10/20 System (Jasper, 1958). The electro-oculogram was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Two additional electrodes were placed on the left and right mastoids. Electrode impedance was kept below 5 kΩ. EEG data were analysed using the Brain Vision Analyzer software (Version 1.05.0005 & Version 2.1.0; Brain Products, Gmbh). All data were re-referenced off-line to the averaged left and right mastoids, filtered with a bandpass filter from 0.1 to 30 Hz (slope of 12 dB/oct), and independent component analysis (ICA) and inverse ICA were used to identify and remove components associated with vertical and horizontal ocular movements. At last, DC-detrend and removal of artifacts above a gradient criterion of 10 µV/ms or a max-min criterion of 100 µV over the entire epoch were applied automatically. For each participant, ERPs were time-locked to syllable onset, segmented into 1200 ms epochs, including a 200 ms baseline and averaged for each phoneme (i.e., /b/, /p/ and /p^h/ for the voicing task, /p/ and /p^h/ for the aspiration task). Individual averages were then averaged together to obtain the grand average across all participants.

2.6 | Statistical analyses

Repeated measures Analyses of Variance (ANOVAs) were used to analyse error rates (ERRs) and reaction times (RTs) in the categorisation tasks (factors are specified below), and post hoc Tukey tests (reducing the probability of Type I errors) were used to determine the origin of significant main effects and interactions. Based on ERRs, three outliers (2 MUS and 1 NM, ± 2 *SD* away from the mean) were excluded from further analyses, resulting in 14 MUS and 14 NM. For ERPs, only trials for correct responses were analysed for nine representative electrodes (i.e., F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). One participant had on average 25 accepted trials (range 22–29) per condition and task. As for behaviour, ANOVAs were computed (factors are specified below) and post hoc Tukey tests were used to determine the origin of significant main effects and interactions. Based on typical latency bands used in the literature and on visual inspection of the data, mean amplitude values were computed (using Brain Analyzer 1 software) within 100 ms time windows centred on the ERP components of interest: N200 (250–350 ms) and P300 (350–450 ms). At last, post hoc correlation analyses (i.e., Pearson, one-tailed) were computed between the number of years and the age of onset of music training, on one side, and behavioural data (i.e., error rates) as well as N200 and P300 amplitudes in the voicing and aspiration tasks, on the other side. Correlations were corrected for multiple comparisons using the Bonferroni procedure (Bonferroni-corrected p-value for 6 tests = 0.009).

3 | RESULTS

3.1 | Screening measures

Results of univariate ANOVAs showed no significant between-group differences regarding general reasoning abilities (i.e., progressive matrices, PM47; $F_{(1,\ 26)} = 2.10$, $p = 0.16$) and working memory (reverse digit span: $F_{(1,\ 26)} = 2.77$, $p = 0.11$), but a trend towards better auditory/verbal short-term memory (direct digit span: $F_{(1,\ 26)} = 3.73$, $p = 0.07$) and visual attention ($F_{(1,\ 26)} = 3.64$, $p = 0.07$) in musicians than in nonmusicians.

3.2 | Musical aptitude

Results of an ANOVA including Group (MUS vs. NM) as between-subjects factor and Task (Melody vs. Rhythm) as within-subjects factor showed that MUS made fewer errors (6.6%) than NM (17.1%, main effect of Group: $F_{(1,\ 26)} = 12.02$, $p = 0.002$). The main effect of Task and the Group × Task interaction were not significant.

3.3 | Behaviour

For ERRs, results of an ANOVA including Group (MUS vs. NM) as a between-subjects factor, and Task (Voicing vs. Aspiration) and Phoneme (/p/ vs. /p^h/) as within-subjects factors showed that MUS (8.6%) made overall fewer errors compared to NM (17.4%; main effect of Group: $F_{(1,26)} = 11.15$, $p = 0.003$). Moreover, while musicians outperformed nonmusicians in the aspiration task (MUS: 6.6%; NM: 22.7%; Tukey, $p < 0.001$), no between-group difference was found in the voicing task (MUS: 10.6%; NM: 12.1%; Tukey, $p = 0.97$; Group × Task interaction: $F_{(1,26)} = 13.68$, $p = 0.001$; see Figure 1). At last, all participants made fewer errors to /p^h/ (2.6%) than to /p/ (20.1%) in the voicing task (Tukey, $p < 0.001$) and more errors to /p^h/ (19.5%) than to /p/ (9.8%) in the aspiration task (Tukey, $p < 0.001$; Phoneme × Task interaction: $F_{(1,26)} = 105.39$, $p < 0.001$). No main effect of Task was obtained.

For RTs, results of an ANOVA including the same factors as for ERRs showed that MUS (574 ms) were not faster than NM (622 ms; no main effect of Group), but with a trend for the influence of music training to be larger in the aspiration task (MUS: 578 ms; NM: 657 ms) than in the voicing task (MUS: 571 ms; NM: 588 ms; Group × Task interaction: $F_{(1,26)} = 3.89$, $p = 0.06$; see Figure 1). In addition, all participants were faster to /p^h/ (562 ms) than to /p/ (597 ms) in the voicing task (Tukey, $p = 0.002$), with no differences in the aspiration task (/p^h/: 626 ms, /p/: 609 ms; Tukey, $p = 0.25$; Phoneme × Task interaction: $F_{(1,26)} = 17.54$, $p < 0.001$). At last, RTs were faster in the voicing task (580 ms) than in the aspiration task (617 ms; main effect of Task: $F_{(1,26)} = 5.89$, $p = 0.02$).

3.4 | Electrophysiological data

Each ERP component was analysed separately using ANOVAs that included Group (MUS vs. NM) as a between-subjects factor, and Task (Voicing vs. Aspiration) and Phoneme (/p/ vs. /p^h/) as within-subjects factors together with Laterality [(Left (F3, C3, P3) vs. Midline (Fz, Cz, Pz) vs. Right (F4, C4, P4)] and Anterior/Posterior dimension

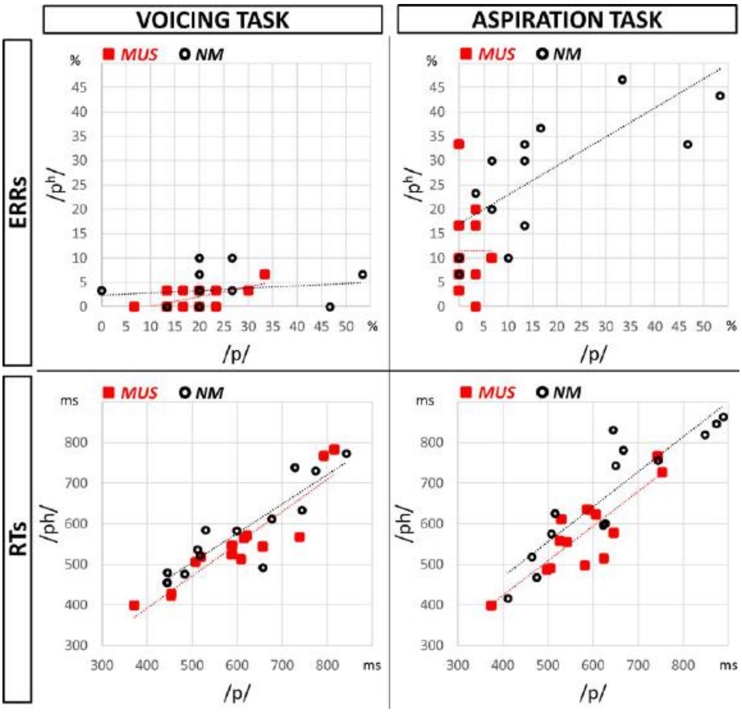


FIGURE 1 Scatterplots of paired observations (/p/ and /p^h/) are illustrated for error rates (ERRs, percentages of error, first row) and reaction times (RTs, milliseconds, second row) separately for the voicing (left) and aspiration tasks (right). Musicians (MUS) are illustrated with squares and nonmusicians (NM) with circles. Trend curves are shown with dotted lines

[(Frontal (F3, Fz, F4) vs. Central (C3, Cz, C4) vs. Parietal (P3, Pz, P4)]. To simplify results presentation, Laterality or Anterior/Posterior effects are only reported when they interacted with the factors of main interest (Group, Task or Phoneme).

N200 (250–350 ms) results showed smaller N200 in MUS (0.50 μV, $SD = 0.94$) than in NM (−2.09 μV, $SD = 0.93$; main effect of Group: $F_{(1,26)} = 6.28$, $p = 0.02$). Moreover, both the Group × Phoneme × Anterior/Posterior and the Group × Task × Phoneme × Laterality × Anterior/Posterior interactions were significant ($F_{(2,52)} = 5.17$, $p = 0.009$, and $F_{(4,104)} = 3.02$, $p = 0.02$, respectively). Separate ANOVAs (2 Groups × 2 Phonemes × 3 Laterality × 3 Anterior/Posterior) were computed for each task. In the voicing task, the N200 was larger in amplitude to /p/ (−1.50 μV, $SD = 2.08$) than to /p^h/ (−0.10 μV, $SD = 2.14$; main effect of Phoneme: $F_{(1,26)} = 9.16$, $p = 0.006$) with no significant group differences (no Group × Phoneme interaction; see Figures 2a and 3a). By contrast, in the aspiration task, only MUS showed significantly larger N200 to /p/ (0.05 μV, $SD = 2.41$) than to /p^h/ (3.19 μV, $SD = 2.71$; Tukey, $p < 0.001$) with no difference between /p/ (−1.57 μV, $SD = 2.41$) and /p^h/ in NM (−.63 μV, $SD = 2.71$; Tukey, $p = 0.51$; Group x Phoneme interaction: $F_{(1,26)} = 5.34$, $p = 0.03$; see Figures 2b and 3a).

P300 (350–450 ms) results showed larger P300 in MUS (1.53 μV, $SD = 0.83$) than in NM (−1.25 μV, $SD = 0.83$; main effect of Group: $F_{(1,26)} = 5.61$, $p = 0.03$). This between-group difference was significant over central (MUS: 1.75 μV, $SD = 2.99$; NM: −1.47 μV, $SD = 3.00$; Tukey, $p = 0.05$) and parietal sites (MUS: 3.40 μV, $SD = 3.23$; NM: −0.20 μV, $SD = 3.20$; Tukey, $p = 0.01$) but not over frontal sites (MUS: −0.55 μV, $SD = 2.92$; NM: −2.07 μV, $SD = 2.90$; Tukey, $p = 0.82$; Group × Anterior/Posterior interaction: $F_{(2,52)} = 4.63$, $p = 0.01$). Moreover, MUS showed significantly smaller P300 to /p/ (0.89 μV, $SD = 3.75$) than to /p^h/ (2.17 μV, $SD = 3.47$; Tukey, $p = 0.01$) with no difference between /p/ (−1.15 μV, $SD = 3.75$) and /p^h/ in NM (−1.35 μV, $SD = 3.47$; Tukey, $p = 0.95$; Group × Phoneme interaction: $F_{(1,26)} = 7.69$, $p = 0.01$). At last, both the Group x Phoneme × Anterior/Posterior and the Group × Task × Phoneme × Laterality × Anterior/Posterior interactions were significant ($F_{(2,52)} = 5.97$, $p = 0.005$, and $F_{(4,104)} = 3.83$, $p = 0.006$, respectively). Separate ANOVAs (2 Groups × 2 Phonemes × 3 Laterality × 3 Anterior/Posterior) were computed for each task. In the voicing task, the P300 was smaller in amplitude to /p/ (−0.90 μV, $SD = 2.39$) than to /p^h/ (−0.02 μV, $SD = 2.34$; main effect of Phoneme: $F_{(1,26)} = 6.58$, $p = 0.02$) with no significant group difference (no Group × Phoneme interaction; see Figures 2a and 3b). By contrast, in the aspiration task, only MUS showed significantly smaller P300 over centro-parietal sites to /p/ (central: 1.81 μV,

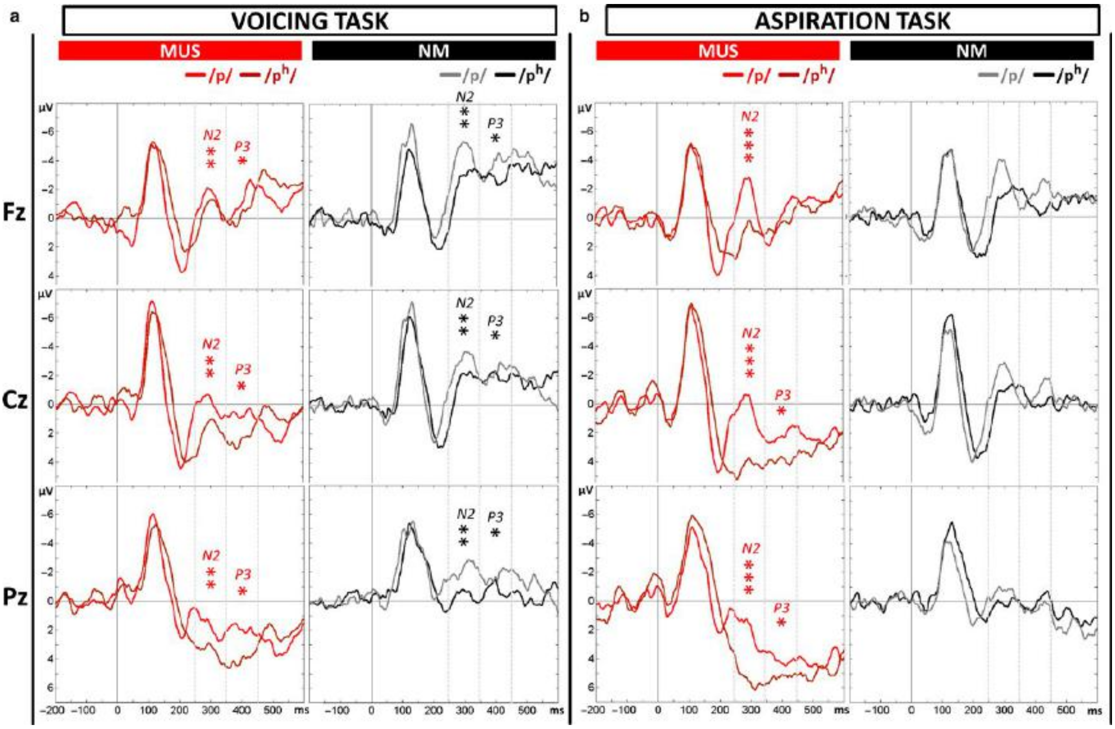


FIGURE 2 ERPs recorded at frontal (Fz), central (Cz) and parietal sites (Pz) are shown for the voicing task (a, left) and aspiration task (b, right). Averages across syllables starting with /p/ or /p^h/ are overlapped for musicians (MUS) and for nonmusicians (NM). Time in milliseconds is in abscissa and the amplitude of the effects in microvolt is in ordinate. Time zero corresponds to syllable onset and negativity is plotted upwards. Latency windows for statistical analyses are indicated with dotted lines and the level of significance is represented by stars with $^*p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$

$SD = 1.55$, parietal: 3.43 µV, $SD = 1.75$) than to /p^h/ (central: 3.09 µV, $SD = 1.34$; parietal: 4.78, $SD = 1.43$; Tukey, $p = 0.03$ and $p = 0.02$, respectively) with no difference between /p/ (central: -0.56 µV, $SD = 1.55$; parietal: 0.83, $SD = 1.75$) and /p^h/ in NM (central: -1.36 µV, $SD = 1.34$; parietal: -0.33, $SD = 1.43$; Group × Phoneme × Anterior/Posterior interaction: $F_{(2,52)} = 4.38$, $p = 0.02$; see Figures 2b and 3b).

3.5 | Correlations between musical expertise, behavioural data and brain activity

No significant correlations were obtained when considering age of onset of music training, or the behavioural data. However, the number of years of music training was significantly correlated with N200 amplitude to /p^h/ syllables in both the voicing and aspiration tasks ($r_{(14)} = 0.67$, $p = 0.004$ and $r_{(14)} = 0.65$, $p = 0.005$, respectively; one-tailed, Bonferroni-corrected). In both cases, N200 decreased in amplitude with increasing numbers of training years (see Figure 4).

4 | DISCUSSION

By comparing French professional musicians and nonmusicians, and by analysing both behaviour and brain electrical activity, the aim of this study was to examine both the impact of musical expertise and of the task at hand on the categorisation of phonemes that did (/p/) or did not (/p^h/) belong to the French repertoire. In particular, our main hypothesis was that the influence of musical expertise would be larger in the aspiration task that required fine discrimination between a native (/p/) and a nonnative (/p^h/) phoneme than in the voicing task that required to discriminate native /b/ and /p/ phonemes, with /p^h/ most likely assimilated to /p/.

Both behavioural and electrophysiological results were in line with these hypotheses. The between-group differences on the error rates were significant in the aspiration task but not in the voicing task with a similar trend for RTs. By showing that the discrimination of native and nonnative phonemes in the aspiration task is enhanced in musicians compared to nonmusicians, these

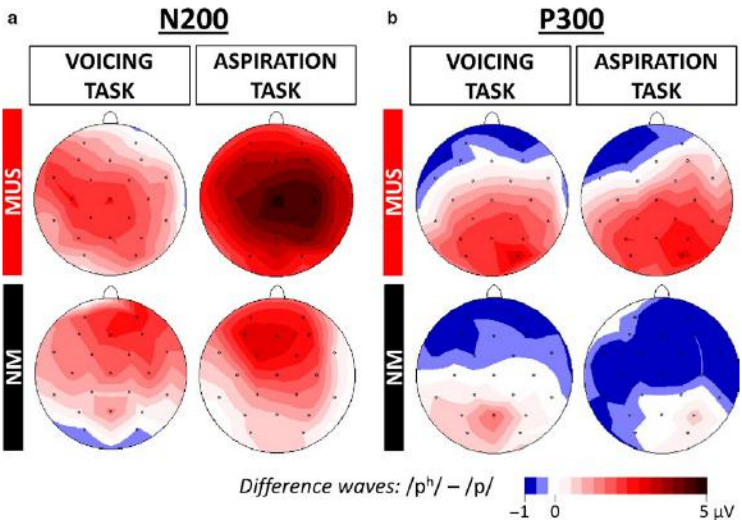


FIGURE 3 Topographic voltage distribution maps of the difference between the two syllables (/pʰ/ minus /p/) are illustrated for (a) the N200 and (b) the P300 separately for the voicing and aspiration tasks, as well as for musicians (MUS) and nonmusicians (NM). Voltage values are scaled from −1 to +5 µV

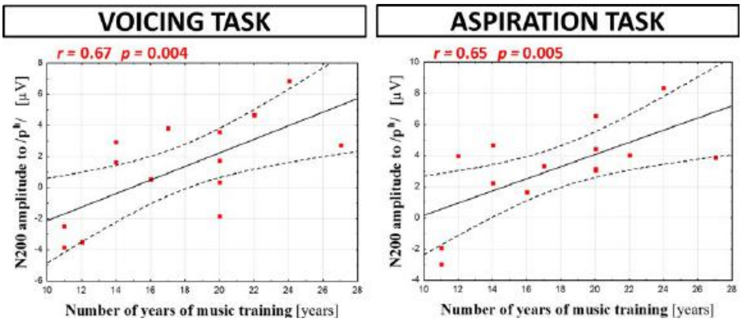


FIGURE 4 Correlations between the number of years of music training and N200 amplitude to /pʰ/ are illustrated for professional musicians in the voicing (left) and aspiration tasks (right). Dotted lines represent the 95% confidence interval of the correlation line

results extend previous findings in the literature showing that musical expertise improves consonant and vowel perception (Chobert et al., 2014; Elmer et al., 2012; Kühnis et al., 2013; Marie, Delogu et al., 2011; Marie, Magne, et al., 2011 Musacchia et al., 2007; Sadakata & Sekiyama, 2011; Tervaniemi et al., 2009; Wong et al., 2007), processing of linguistic prosody (Magne, Schön, & Besson, 2006; Schön et al., 2004) and of emotional prosody (Lima & Castro, 2011) as well as categorical perception of vowels (Bidelman & Alain, 2015; Bidelman et al., 2014). These results are also in line with recent findings showing that, compared to nonmusicians, French musicians may compensate for the lack of experience with English phonemes (e.g., [ðæ]) by strengthening the encoding of acoustic information at the subcortical level (Intartaglia, White-Schwoch, Kraus, & Schön, 2017). At last, the finding of no between-group differences when participants were asked to discriminate two native phonemes in the voicing task is also in line with previous results showing no effect of musical expertise when the discrimination contrast is easy to perceive and does not require fine-grained auditory perception (Elmer et al., 2012; Schön et al., 2004).

By showing that the musicians' advantage depends upon the task at hand, our results support the ASP model proposed by Strange (2011) following which different selective processing routines are called into play depending upon the experimental context and the task to be performed on the stimuli. Thus, musicians outperformed nonmusicians in the aspiration task that, following the ASP model, required costly, effortful phonetic routines to discriminate a nonnative /pʰ/ from a native /p/ phoneme. By contrast, no between-group differences were found in the voicing task that required phonological routines that are automatically called into play, with no difference between musicians and nonmusicians. This is not to say that differences between musicians and nonmusicians cannot be found when automatic, preattentive processing is involved, as demonstrated in many experiments using the Mismatch Negativity (Kujala, Tervaniemi, & Schröger, 2007). However, when the task is very easy because it requires categorisation of native phonemes in their respective categories, nonmusicians are expected to perform comparably to musicians.

It is interesting that analyses of the brain electrical activity in musicians and in nonmusicians allow us to go one step further in the interpretation of the results. Let us first consider results in the voicing task. In this task and in line with the predictions derived from PAM (Best & Tyler, 2007) and from the ASP model (Strange, 2011), N200 components were qualitatively similar for both the native /p/ and the nonnative /p^h/ phonemes in both musicians and nonmusicians with no significant between-group differences (see Figures 2 and 3). This is taken as evidence that qualitatively similar phonological routines were automatically involved to categorise the two phonemes. However, results in the voicing task also revealed quantitative differences with smaller N200 amplitude and lower error rate to /p^h/ than to /p/. These results seem to go against the predictions from PAM (Best & Tyler, 2007) of no or small differences between /p/ and /p^h/; rather they show differences in N200 amplitude that may reflect the fact that /p/ and /p^h/ are processed differently. A possible interpretation of these results is linked to the relative position of /p/ and /p^h/ compared to /b/ on the voicing continuum. Because /b/ ($VOT = -144$ ms) and /p^h/ ($VOT = 77$ ms) were located at extreme positions of the continuum, they were easily discriminable from each other. By contrast, /p/ ($VOT = 3$ ms) was more difficult to discriminate because of its intermediate VOT value between /b/ and /p^h/. In other words, these results suggest that the inclusion of nonnative phonemes is shifting the category boundary between /b/ and /p/ (in French typically around 0 ms) towards more positive values (VOT for /p^h/: 77 ms) so that the native /p/ is more difficult to categorise than the nonnative /p^h/ phoneme. Such an interpretation would also be in line with the Auditory Enhancement Theory (Diehl & Kluender, 1989; Kluender, 1994) following which the redundancy of the cues to perceived contrasts predicts that certain features will enhance the perceptibility of others. In our case, spread glottis which is a feature used for the production of aspirated sounds (in our case /p^h/) is a secondary enhancing cue to voiceless sounds (in our case /p/), because opening at the glottal level makes it impossible for the vocal folds to vibrate. This additional cue possibly facilitated the categorisation of /p^h/ compared to /p/. Insofar as the amplitude of the N200 component is smaller to items that are easier to categorise (Patel & Azzam, 2005), this interpretation also accounts for the smaller N200 to /p^h/ than to /p/ phonemes.

Turning to the aspiration task and in contrast to the voicing task, qualitatively different processes, as reflected in the ERPs, seem to be involved for musicians and nonmusicians. The N200 was larger to the native phoneme /p/ than to the nonnative /p^h/ phoneme in musicians but not in nonmusicians (see Figures 2 and 3). In line with the ASP model (Strange, 2011), we interpret this finding as showing that the native /p/ phoneme is categorised within a native phonemic category using phonological routines, while the nonnative /p^h/ phoneme is discriminated from /p/ using phonetic routines but is not categorised into a phonemic category. For nonmusicians, results are quite different from musicians, with no significant difference in N200 amplitude to both /p/ and /p^h/ (as in the voicing task). In line with the higher error rates in nonmusicians than in musicians, these findings support the interpretation that nonmusicians are less sensitive than musicians to the acoustic difference between allophones of /p/ and /p^h/ that were used as stimuli in our task.

Note, however, that another interpretation of the results in the aspiration task needs to be considered in terms of categorisation and decision-related processes. Based on their enhanced auditory perception (e.g., Micheyl, Delhommeau, Perrot, & Oxenham, 2006), musicians possibly discriminated nonnative /p^h/ and native /p/ phonemes more easily and with higher confidence in their judgement than nonmusicians, as reflected by larger P300 components to /p^h/ than to /p/ in musicians but not in nonmusicians (see Figure 2). As P300 components over centro-parietal sites started to develop in the same latency band as the N200 (250–350 ms) and lasted until 450 ms poststimulus onset, the two components possibly overlapped, thereby reducing N200 amplitude. This interpretation is also in line with the correlation results showing that the N200 to /p^h/ decreased in amplitude with increasing numbers of training years (see Figure 4). These results are taken to show that long-term music training enhances the ability to discriminate native from nonnative phonemes in adult life.

In sum, the reduction of N200 amplitude to /p^h/ compared to /p/ over centro-parietal sites in musicians would reflect the easiness of categorisation and decision-related processes based on better auditory perception (Bidelman et al., 2014; Musacchia et al., 2007; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Wong et al., 2007), rather than the impossibility to categorise the nonnative /p^h/ into a preexisting phonemic category, as proposed above. While these two interpretations are difficult to disentangle, the important point remains that musical expertise clearly influenced the categorisation of nonnative phonemes in the aspiration task (larger P300 over parietal sites to /p^h/ than to /p/ in musicians but not in nonmusicians, see Figure 2) with smaller between-group differences in the voicing task (larger P300 to /p^h/ than to /p/ in both groups).

At last, it should be mentioned that psychometric measurements revealed a trend towards enhanced auditory/verbal short-term memory in musicians ($p = 0.07$). Such a result is in line with previous literature showing an influence of musical expertise on mnemonic functions (George & Coch, 2011; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011). As P300 has been related to cognitive workload (Polich, 2007), memory-related processes (Friedman & Johnson, 2000) and memory updating (Donchin & Coles, 1988), it is possible that the P300 results reflect between-group differences in short-term memory. However, two points argue against such an interpretation in the present experiment: first, the categorisation tasks used here do not require a strong memory component but rather accurate auditory perception and focused attention. Thus, enhanced P300 amplitude may rather reflect enhanced selective attention in musicians compared to nonmusicians (Strait et al., 2015; Wang et al., 2015). Second, if short-term memory influenced the musicians' level of performance, this effect should have been found in both tasks and for both phonemes, which was not the case.

4.1 | Limitations

One limitation of the present study is that the effects of the task discussed above are confounded with the general effects of selective attention. In other words, it is difficult to determine whether the results are best explained by the involvement of different modes of perception, phonological in the voicing task and phonetic in the aspiration task, or by the focus of attention on different aspects of the stimuli. In the voicing task, participants were asked to categorise syllables as /ba/ or /pa/ so that attention was oriented towards the discrimination of a native contrast. By contrast, in the aspiration task, participants were asked to categorise the syllables as /pa/ or as /p^{h}a/ so that attention was oriented towards discriminating a native /p/ from a nonnative /p^h/ phoneme. This issue is also important in view of psychometric measurements that showed a trend towards better attentional function in musicians than in nonmusicians. Future studies should aim at disentangling phonological and phonetic processing differences in musicians and nonmusicians from general attention effects by directly comparing the processing of native and nonnative phonemes in the two groups under passive and attentive listening conditions.

Another limitation is linked to the relatively low number of participants in each group ($N = 14$). However, it is difficult to find professional musicians matching our restrictive criteria (e.g. age, sex, music training years, number of languages spoken, etc.) who agree to devote a few hours of their precious time to participate in experimental research. Nevertheless, as the size of the between-condition and between-group differences were large and significant, we feel confident that our findings are reliable (e.g., Friston, 2012).

5 | CONCLUSIONS

Taken together, both behavioural and electrophysiological results concur in showing that musicians are more sensitive than nonmusicians to the difference between native and nonnative phonemes but that this between-group difference depends upon the task at hand. These results provide partial support to PAM (Best & Tyler, 2007; Tyler et al., 2014). On the one hand, N200 amplitude was not significantly different for native /p/ and nonnative /p^h/ phonemes in the voicing task, thereby supporting the hypothesis that /p^h/ was assimilated to the native /p/ category. On the other hand, smaller N200 amplitude for /p^h/ than for /p/ possibly reflected the fact that the nonnative phoneme was not completely assimilated to the native /p/ category. These results also provide support for the ASP model (Strange, 2011), with a phonological mode of perception involved in the voicing task and a phonetic mode required in the aspiration task. In addition, these results suggest that increased auditory sensitivity may be one of the driving forces behind enhanced phonemic perception in musicians. They are in line with previous work showing that musical expertise facilitates the perception and pronunciation of foreign linguistic sounds (Delogu, Lampis, & Belardinelli, 2010; Gottfried, 2007; Kühnis et al., 2013; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010; Sadakata & Sekiyama, 2011; Slevc & Miyake, 2006) and possibly the learning of foreign languages (Chobert & Besson, 2013; Dittinger, Chobert, Ziegler, & Besson, 2017; Dittinger et al., 2016). Taken together, these results have potentially strong societal implications for education because fine perception of nonnative phoneme is an important first step in foreign language learning. These findings may contribute to foster the coupling between music training and second language learning in elementary schools. At last, they point to the interest of measuring both the level of performance in specific tasks and the accompanying changes in brain electrical activity that provided complementary information highlighting both qualitative and quantitative differences.

ACKNOWLEDGEMENTS

We would like to thank all the participants, Chotiga Pattamadilok for registering the auditory stimuli, and Mylène Barbaroux for help in analysing the data. This work, carried out within the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). ED is supported by a doctoral fellowship from the BLRI. No private or commercial sponsor supported this research.

CONFLICT OF INTERESTS

The authors declare no competing financial interests.

DATA ACCESSIBILITY

Data used in this research were collected under provision of informed consent of the participants. Access to the data will be granted in line with that consent, subject to approval by the university ethics board and under a formal Data Sharing Agreement. Auditory stimuli can be accessed by directly contacting the first author.

AUTHOR CONTRIBUTIONS

ED and MB designed the research. ED collected and analysed the data and contributed to the manuscript. MB supervised the research and wrote the manuscript. MDI helped for data interpretation and contributed to the manuscript.

ORCID

Eva Dittinger http://orcid.org/0000-0002-5121-096X

REFERENCES

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. https://doi.org/10.3389/fpsyg.2013.00321

Besson, M., Barbaroux, M., & Dittinger, E. (2017). Music in the brain: Music and language processing. In R. Ashley & R. Timmers (Eds.), *The Routledge Companion to Music Cognition* (pp. 37–48). New York, NY: Routledge/Taylor & Francis.

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

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Second Language (Ed.), *Speech Learning: The Role of Language Experience in Speech Perception and Production*. Amsterdam: John Benjamins.

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *Journal of Neuroscience*, *35*, 1240–1249. https://doi.org/10.1523/JNEUROSCI.3292-14.2015

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125. https://doi.org/10.1016/j.brainres.2010.07.100

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, *79*, 201–212. https://doi.org/10.1016/j.neuroimage.2013.04.093

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*, 2662–2673. https://doi.org/10.1111/ejn.12627

Boersma, P., & Weenink, D. (2011) “Praat: Doing phonetics by computer”, http://www.praat.org

Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences*, *3*, 923–940. https://doi.org/10.3390/brainsci3020923

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). 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*, *24*, 956–967. https://doi.org/10.1093/cercor/bhs377

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256–266. https://doi.org/10.1162/jocn.1994.6.3.256

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*, 1–18. https://doi.org/10.1016/0093-934X(92)90018-A

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*, 46–61. https://doi.org/10.1080/09541440802708136

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology*, *1*, 121–144. https://doi.org/10.1207/s15326969eco0102_2

Dittinger, E., Barbaroux, M., D’Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, *28*, 1584–1602. https://doi.org/10.1162/jocn_a_00997

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience*, *11*, 233. https://doi.org/10.3389/fnhum.2017.00233

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*, 357–374. https://doi.org/10.1017/S0140525X00058027

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, *22*, 650–658. https://doi.org/10.1093/cercor/bhr142

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: A selective review. *Microscopy Research and Technique*, *51*, 6–28. https://doi.org/10.1002/(ISSN)1097-0029

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport*, *19*, 1757–1761. https://doi.org/10.1097/WNR.0b013e328318f014

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *NeuroImage*, *61*, 1300–1310. https://doi.org/10.1016/j.neuroimage.2012.04.018

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*, 1083–1094. https://doi.org/10.1016/j.neuropsychologia.2011.02.001

Gottfried, T. L. (2007). Music and language learning: Effects of musical training on learning L2 speech contrasts. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 221–237). Amsterdam, the Netherlands: John Benjamins.

Intartaglia, B., White-Schwoch, T., Kraus, N., & Schön, D. (2017). Music training enhances the automatic neural processing of foreign speech sounds. *Scientific Reports*, *7*, 12631. https://doi.org/10.1038/s41598-017-12575-1

Jäncke, L. (2009). The plastic human brain. *Restorative Neurology and Neuroscience*, *27*, 521–538. https://doi.org/10.3233/RNN-2009-0519

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 173–217). San Diego, CA: Academic Press.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio: The Psychological Corporation.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: Deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology: General*, *125*, 331–359. https://doi.org/10.1037/0096-3445.125.4.331

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, *51*, 1608–1618. https://doi.org/10.1016/j.neuropsychologia.2013.04.007

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology*, *74*, 1–19. https://doi.org/10.1016/j.biopsycho.2006.06.001

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*, 1021–1031. https://doi.org/10.1037/a0024521

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*, 199–211. https://doi.org/10.1162/jocn.2006.18.2.199

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*, 2701–2715. https://doi.org/10.1162/jocn.2010.21585

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, *23*, 294–305. https://doi.org/10.1162/jocn.2010.21413

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36–47. https://doi.org/10.1016/j.heares.2006.05.004

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*, 56–60. https://doi.org/10.1016/j.lindif.2009.11.003

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*, *19*, 712–723. https://doi.org/10.1093/cercor/bhn120

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*, 473–478. https://doi.org/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*, *104*, 15894–15898. https://doi.org/10.1073/pnas.0701498104

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *Internation Journal of Medical Sciences*, *2*, 147. https://doi.org/10.7150/ijms.2.147

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75. https://doi.org/10.1196/annals.1284.006

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*, 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Raven, J., Corporation, P., & Lewis, H. (1962). *Coloured Progressive Matrices: Sets A, AB, B*. London: Oxford Psychologist Press.

Ritter, W., Ford, J. M., Gaillard, A. W., Harter, M. R., Kutas, M., Näätänen, R., … Rohrbaugh, J. (1984). Cognition and event-related potentials. *Annals of the New York Academy of Sciences*, *425*, 24–38. https://doi.org/10.1111/j.1749-6632.1984.tb23521.x

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *Journal of the Acoustical Society of America*, *134*, 1324–1335. https://doi.org/10.1121/1.4812767

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*, 1318. https://doi.org/10.3389/fpsyg.2014.01318

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

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349. https://doi.org/10.1111/1469-8986.00172.x

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*, 771–783. https://doi.org/10.1002/hbm.21060

Seppänen, M., Hämäläinen, J., Pesonen, A.-K., & Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of n1 and p2 source activation for unattended sounds. *Frontiers in Human Neuroscience*, *6*, 43. https://doi.org/10.3389/fnhum.2012.00043

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, *23*, 5545–5552. https://doi.org/10.1523/JNEUROSCI.23-13-05545.2003

Slevc, L. R. (2012). Language and music: Sound, structure, and meaning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*, 483–492. https://doi.org/10.1002/wcs.1186

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency does musical ability matter? *Psychological Science*, *17*, 675–681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

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*, 22–29. https://doi.org/10.1016/j.heares.2009.12.021

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104. https://doi.org/10.1016/j.dcn.2015.01.001

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*, 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

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. *European Journal of Neuroscience*, *30*, 1636–1642. https://doi.org/10.1111/j.1460-9568.2009.06955.x

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science*, *21*, 1532–1540. https://doi.org/10.1177/0956797610384142

Tyler, M. D., Nil, M. D., Best, C. T., Faber, A., & Levitt, A. G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, *71*, 4–21. https://doi.org/10.1159/000356237

Wang, X., Ossher, L., & Reuter-Lorenz, P. A. (2015). Examining the relationship between skilled music training and attention. *Consciousness and Cognition*, *36*, 169–179. https://doi.org/10.1016/j.concog.2015.06.014

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children*. San Antonio: The Psychological Corporation (WISC-IV).

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422. https://doi.org/10.1038/nn1872

How to cite this article: Dittinger E, D’Imperio M, Besson M. Enhanced neural and behavioural processing of a nonnative phonemic contrast in professional musicians. *Eur J Neurosci*. 2018;00:1–13. https://doi.org/10.1111/ejn.13939

Chapter 4

MUSIC TRAINING AND NOVEL WORD LEARNING ACROSS THE LIFESPAN

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and word meaning acquisition: from faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, 28, 1584-1602.

Dittinger, E., Chobert, J., Ziegler, J., & Besson, M. (2017). Fast brain plasticity during word learning in musically trained children. *Frontiers in Human Neuroscience*, 11, 233.

Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (submitted). Testing Relationships between Music Training and Word Learning across the Lifespan by using Cross-Sectional and Comparative Electrophysiological Approaches in Children, young Adults, and older Adults.

Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations

Eva Dittinger[1], **Mylène Barbaroux**[1], **Mariapaola D'Imperio**[1], **Lutz Jäncke**[2], **Stefan Elmer**[2*], **and Mireille Besson**[1*]

Abstract

■ On the basis of previous results showing that music training positively influences different aspects of speech perception and cognition, the aim of this series of experiments was to test the hypothesis that adult professional musicians would learn the meaning of novel words through picture–word associations more efficiently than controls without music training (i.e., fewer errors and faster RTs). We also expected musicians to show faster changes in brain electrical activity than controls, in particular regarding the N400 component that develops with word learning. In line with these hypotheses, musicians outperformed controls in the most difficult semantic task. Moreover, although a frontally distributed N400 component developed in both groups of participants after only a few minutes of novel word learning, in musicians this frontal distribution rapidly shifted to parietal scalp sites, as typically found for the N400 elicited by known words. Finally, musicians showed evidence for better long-term memory for novel words 5 months after the main experimental session. Results are discussed in terms of cascading effects from enhanced perception to memory as well as in terms of multifaceted improvements of cognitive processing due to music training. To our knowledge, this is the first report showing that music training influences semantic aspects of language processing in adults. These results open new perspectives for education in showing that early music training can facilitate later foreign language learning. Moreover, the design used in the present experiment can help to specify the stages of word learning that are impaired in children and adults with word learning difficulties. ■

INTRODUCTION

The aim of the present experiment was to examine the influence of music training on word learning using both behavioral and electrophysiological measures. On the basis of the evidence reviewed below, we tested the hypothesis that musicians would be more efficient at word learning than nonmusicians and that the differences would be reflected in their pattern of brain waves. There is strong evidence from previous cross-sectional studies comparing adult musicians and nonmusicians that long-term music training promotes brain plasticity (Münte, Altenmüller, & Jäncke, 2002) in modifying the functional (Schneider et al., 2002; Pantev et al., 1998) and structural (Elmer, Hänggi, Meyer, & Jäncke, 2013; Gaser & Schlaug, 2003; Schneider et al., 2002) architecture of the auditory pathway. Results of longitudinal studies, mostly in children, showed that music training can be the cause of the observed effects (François, Chobert, Besson, & Schön, 2013; Strait, Parbery-Clark, O'Connell, & Kraus, 2013; Chobert, François, Velay, & Besson, 2012; Moreno et al., 2011; Hyde et al., 2009; Moreno et al., 2009). Most importantly for the present purposes, there is also evidence that music training improves different aspects of speech processing (for review, see Asaridou & McQueen, 2013; Besson, Chobert, & Marie, 2011; Kraus & Chandrasekaran, 2010). These transfer effects possibly arise because speech and music are auditory signals relying on similar acoustic cues (i.e., duration, frequency, intensity, and timbre) and because they share, at least in part, common neuronal substrates for auditory perception (Peretz, Vuvan, Lagrois, & Armony, 2015; Jäncke, 2009) and for higher-order cognitive processing (Rogalsky, Rong, Saberi, & Hickok, 2011; Patel, 2008; Maess, Koelsch, Gunter, & Friederici, 2001). For instance, music training facilitates the processing of a variety of segmental (Bidelman, Weiss, Moreno, & Alain, 2014; Kühnis, Elmer, & Jäncke, 2014; Elmer, Meyer, & Jäncke, 2012; Chobert, Marie, François, Schön, & Besson, 2011; Musacchia, Sams, Skoe, & Kraus, 2007) and suprasegmental speech attributes (Marie, Delogu, Lampis, Olivetti Belardinelli, & Besson, 2011; Wong & Perrachione, 2007) within native (Schön, Magne, & Besson, 2004) and nonnative languages (Marques, Moreno, Castro, & Besson, 2007). Moreover, both musically trained children (Jentschke & Koelsch, 2009) and adults (Fitzroy & Sanders, 2013) are more sensitive to violations of linguistic and music syntax than participants without music training. Perhaps most importantly, recent results also showed that

---

[1]CNRS and Université Aix Marseille, [2]University of Zurich
*Shared last authorship.

Journal of Cognitive Neuroscience 28:10, pp. 1584–1602
doi:10.1162/jocn_a_00997

long-term music training positively improves cognitive functions such as auditory attention (Strait, Slater, O'Connell, & Kraus, 2015), visual attention (Wang, Ossher, & Reuter-Lorenz, 2015), working and verbal memory (George & Coch, 2011; Ho, Cheung, & Chan, 2003), executive functions (Zuk, Benjamin, Kenyon, & Gaab, 2014; Moreno et al., 2011; Pallesen et al., 2010), and general intelligence (Schellenberg, 2004). These findings are not surprising insofar as playing an instrument at a professional level is a multidimensional task that, together with specific motor abilities, requires acute auditory perception and focused attention, code switching between the visual information on the score and the corresponding sounds, as well as the ability to maintain auditory information in short- and long-term memory. Taken together, these results are in line with dynamic models of human cognition (Friederici & Singer, 2015; Hagoort, 2014) positing that language—and possibly music—are processed in interaction with other cognitive functions.

Similar to playing music, word learning is also a multidimensional task requiring both perceptive and higher-order cognitive abilities. Let us take the example of Thai. Thai is a tonal and a quantitative language that mainly comprises monosyllabic words (as Mandarin Chinese and other tonal languages) and in which both tonal (i.e., five tones) and vowel length contrasts are linguistically relevant for understanding word meaning (e.g., /pa1/ low tone with a short vowel means "to find" and /pa:1/ low tone with a long vowel means "forest"; Gandour et al., 2002). Thus, when it comes to learn novel words in Thai, the learner has to focus attention on the acoustic stream to discriminate spectral and temporal phonetic contrasts and to build new phonological representations that can then be associated with lexical meaning by recruiting working, short-term, episodic, and semantic memory processes. Thus, if music skills translate into improved auditory perception and attention together with enhanced working and verbal memory, it should be easier for musicians to learn a language such as Thai.

The ERP method is one of the most eligible methods to capture the fast temporal dynamics of word learning and to examine brain plasticity, as reflected by changes in the amplitude and/or latency of ERP components during learning. Previous results in adults have shown that the N400, a negative-going component that typically develops between 300 and 600 msec poststimulus onset (Kutas & Hillyard, 1980), increases in amplitude when meaningless items acquired meaning. Specifically, results showed N400 enhancements in native English speakers after 14 hr of learning the meaning of novel French words (McLaughlin, Osterhout, & Kim, 2004) and after 45 min of learning the meaning of rare words (e.g., "clowder"; Perfetti, Wlotko, & Hart, 2005). Moreover, if a novel word (Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2010; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007) or pseudoword (Batterink & Neville, 2011) is presented in a strongly constrained and meaningful context, even a single exposure can be sufficient to build up initial word representations, an effect referred to as "fast mapping" (Carey, 1978). An incubation-like period and further exposures are then required for consolidation and integration into existing lexical networks (Dumay & Gaskell, 2007). Thus, the N400 is taken as a reliable index of word learning, reflecting the formation of semantic representations.

Note though that the N400 component at the core of the above-mentioned experiments clearly showed a more frontal scalp distribution (Borovsky et al., 2010; Mestres-Missé et al., 2007) than the centroparietal N400 typically elicited by already known words (Kutas, Van Petten, & Besson, 1988). This frontal N400 distribution is compatible with results showing that prefrontal and temporal brain regions are associated with the maintenance of novel information in working or short-term memory and the formation of new associations (Hagoort, 2014) and/or with the initial building-up of word representations in episodic memory (Rodriguez-Fornells, Cunillera, Mestres-Missé, & De Diego-Balaguer, 2009; Wagner et al., 1998).

As mentioned above, most studies of music-to-language transfer effects have focused on segmental, suprasegmental, and syntactic processing levels. On the basis of the results on word learning reviewed above, this study aimed at going a step further so as to determine whether professional music training would also influence the semantic level of processing, most often considered as language-specific (but see Koelsch et al., 2004) by facilitating the learning process of novel word meaning. The general hypothesis was that the optimization of perceptual and cognitive functions in professional musicians would positively influence the speed and quality of word learning as reflected by a behavioral advantage for musicians (i.e., lower error rates [ERRs] and faster RTs). Moreover, based on the ERPs and word learning literature (Borovsky et al., 2010, 2012; Batterink & Neville, 2011; Mestres-Missé et al., 2007; Perfetti et al., 2005; McLaughlin et al., 2004), we expected a frontally distributed N400 component to develop in all participants during the early stages of novel word learning. However, if the perceptual and cognitive computations involved in word learning were facilitated in musicians, the development of the N400 component should be faster in musicians than in controls in the learning phase. By contrast, we expected the N400 to show a centroparietal distribution when novel word learning was consolidated.

To test these general hypotheses, we used an ecologically valid experimental design inspired by Wong and Perrachione (2007) and based on a series of four experiments that comprised several tasks performed during the main experimental session (see Figure 1A–E). First, to further test the hypothesis of improved auditory speech discrimination in musicians compared with controls, participants performed a phonological categorization task at the beginning and at the end of the main experimental session (see Figure 1A and E). On the basis of previous results showing that musicians are advantaged when

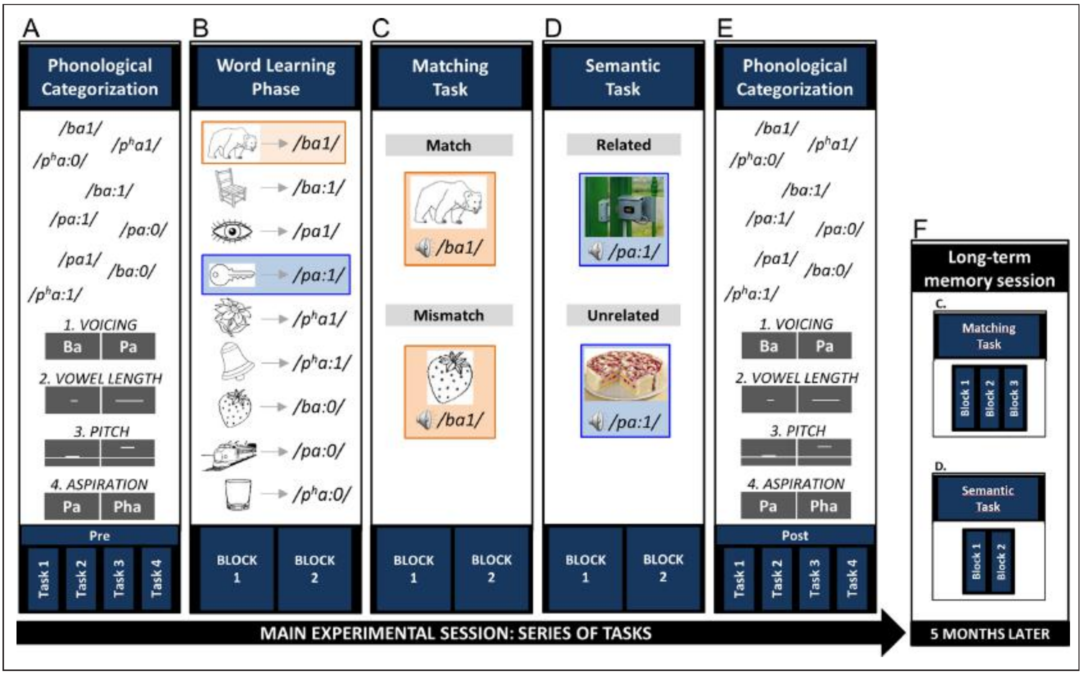


Figure 1. Experimental design. Participants performed a series of tasks in the main experimental session (A–E): First, in the phonological categorization task (A), nine natural Thai monosyllabic words had to be categorized based on voicing (Task 1), vowel length (Task 2), pitch (Task 3), or aspiration contrasts (Task 4). Second, in the word learning phase (B), each word was paired with its respective picture. This phase included two separate blocks of trials. Third, in the matching task (C), the words were presented with one of the pictures, either matching or mismatching the previously learned associations. This task included two separate blocks of trials. Fourth, in the semantic task (D), the words were presented with novel pictures that were either semantically related or unrelated to the novel words. Again, this task included two separate blocks of trials. Fifth, participants did again the four tasks of the phonological categorization task (E). Finally, participants came back 5 months after the main session to perform again the matching and semantic tasks (F).

the discrimination is most difficult (Diamond, 2013; Schön et al., 2004), we expected musicians to outperform controls in identifying phonemic contrasts that are not relevant for lexical discrimination in French. Moreover, based on previous literature reporting that the N100 component reflects encoding of auditory cues in the auditory-related cortex (Kühnis et al., 2014) and is influenced by auditory attention and perceptual learning (Seppänen, Hämäläinen, Pesonen, & Tervaniemi, 2012; Woldorff & Hillyard, 1991), we expected this behavioral advantage to be accompanied by an increased N100 amplitude in musicians.

Second, participants were asked to learn the meaning of the novel words through picture–word associations (see Figure 1B), a design that has often been used in word learning experiments in children (Friedrich & Friederici, 2008; Torkildsen et al., 2008) and in adults (Dobel, Lagemann, & Zwitserlood, 2009). No behavioral response was required during this word learning phase, but ERPs were recorded to test the main hypothesis that frontally distributed N400 components would develop in both groups of participants (François et al., 2013; Borovsky et al., 2010; Rodriguez-Fornells et al., 2009) but with faster temporal dynamics in musicians than in controls.

Third, to test for the efficacy of the learning phase, participants performed a matching task and were asked to decide whether a picture–word pair matched or mismatched the previously learned pairs (see Figure 1C). Fourth, an important aspect was to determine whether word learning was specific to the picture–word pairs learned during the word learning phase or whether the meaning of the newly learned words was already integrated into semantic networks so that priming effects generalized to new pictures. To this end, participants performed a semantic task during which novel pictures that had not been seen in the previous tasks were presented. They were asked to decide whether the picture and the word were semantically related or unrelated (see Figure 1D). In both the matching and the semantic tasks and in both groups of participants, we predicted that N400 amplitudes would be larger for mismatch and semantically unrelated words than for match and semantically related words (i.e., the typical N400 effect; Kutas & Hillyard, 1980), thereby showing that participants had learned the meaning of the novel words. Moreover, if the novel words' meanings were already integrated into existing semantic networks at the end of the word learning phase (Borovsky et al., 2012;

Batterink & Neville, 2011), we expected the N400 effect (mismatching–matching and unrelated–related words) in the matching and semantic tasks to show the centroparietal distribution typically found for already known words (Kutas et al., 1988). Of main interest was to specify the spatiotemporal dynamics of the N400 effect in musicians and in nonmusicians. Finally, if we were to find that music training influenced word learning, then we expected musical ability to be positively correlated with word learning efficacy, as reflected by behavioral measures and/or the N400 effect in the matching and semantic tasks.

Finally, a subset of the participants was behaviorally retested after 5 months (see Figure 1F) in the matching and semantic tasks to evaluate the maintenance of novel words in long-term memory. It was of interest to determine whether the behavioral advantages of musicians in a variety of cognitive domains, as reviewed above, extend to long-term memory. To the best of our knowledge, this aspect has not yet been investigated.

In summary, this experimental design is relevant for specifying whether music expertise influences the semantic level of speech processing, an issue that, to our knowledge, has not been addressed before. By analyzing ERPs, we aimed at better understanding the dynamics of word learning, how fast semantic processes develop, and whether and how the N400 is influenced by music training. Showing that long-term music training with an early start (as it is most often the case in professional musicians) may facilitate foreign language learning later in life should add evidence to the claim that music training has important societal consequences for education (Besson et al., 2011; Kraus & Chandrasekaran, 2010). Finally, this experimental design is of potential interest for clinical research: using several different tasks that call upon several perceptual and cognitive functions (phonological categorization, formation of picture–word associations, maintaining these associations in short-term and long-term memory and generalization of learning effects) within the same patient may help specify the processing stages that are deficient in adults or children with language learning disorders.

METHODS

Participants

A total of 30 participants with 15 professional musicians (MUS, eight women) and 15 controls without formal music training (nonmusicians, NM, eight women) but involved in a regular leisure activity (e.g., sports, dance, theater) were paid to participate in the experimental session lasting for 2.5 hr (including the application of the Electrocap, psychometric measurements, and experimental tasks). The two groups did not differ in age (MUS: mean age = 25.1 years, age range = 19–30, *SD* = 3.9; NM: mean age = 25.7 years, age range = 19–33, *SD* = 4.8; $F(1, 28) = 0.02, p = .68$). All participants were native French speakers, had comparable education levels (university degree) and socioeconomic background (criteria of the National Institute of Statistics and Economic Studies; MUS: 4.4; NM: 4.9; $F(1, 28) = 1.45, p = .24$), and reported no past or current audiological or neurological deficits. MUS practiced their instruments for an average of 17 years (range = 11–24, *SD* = 4.1) and the musician group included three pianists, two accordionists, four violinists, one cellist, two guitarists, one hornist, one tubist, and one flautist. None of the participants was bilingual, but all spoke English as a second language and most participants (except for 1 MUS and 3 NM) had a rudimentary knowledge of a third language that was neither tonal nor quantitative. The study was conducted in accordance with the Helsinki declaration, and all participants gave their informed consent before enrolling in the experiment.

Screening Measures

Cognitive Ability

Standardized psychometric tests were used to examine short-term and working memory (forward and reverse Digit Span, WISC-IV; Wechsler, 2003), visual attention (NEPSY from Korkman, Kirk, & Kemp, 1998) and nonverbal general intelligence (progressive matrices, PM47; Raven, Corporation, & Lewis, 1962).

Musical Aptitude

Participants performed two musicality tests (adapted from the MBEA battery; Peretz, Champod, & Hyde, 2003) consisting in judging whether pairs of piano melodies were same or different, based either on melodic or on rhythmic information.

Experimental Stimuli

Auditory Stimuli

Nine natural Thai monosyllabic words were selected for the experiment: /ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/, /p^{h}a:1/, /ba:0/, /pa:0/, /p^{h}a:0/.[1] These words varied in vowel duration, with short (/ba1/, /pa1/ and /p^{h}a1/; 261 msec on average) and long vowels (/ba:1/, /pa:1/, /p^{h}a:1/, /ba:0/, /pa:0/ and /p^{h}a:0/; 531 msec on average), and in fundamental frequency, with low-tone (/ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/ and /p^{h}a:1/; F0 = 175 Hz on average) and midtone vowels (/ba:0/, /pa:0/ and /p^{h}a:0/; F0 = 218 Hz on average). Furthermore, words contained voicing contrasts (/ba1/, /ba:1/ and /ba:0/, VOT = −144 msec vs. /pa1/, /pa:1/ and /pa:0/, VOT = 3 msec) as well as aspiration contrasts (/pa1/, /pa:1/ and /pa:0/, VOT = 3 msec vs. /p^{h}a1/, /p^{h}a:1/ and /p^{h}a:0/, VOT = 77 msec).[2] Stimuli were recorded by a female Thai–French bilingual, ensuring that all words were produced naturally. For each word, five versions were digitally recorded to reproduce natural speech variability. Sound pressure level was normalized across all words to a mean level of 70 dB by using Praat software (Boersma & Weenink, 2011).

Visual Stimuli

For the learning phase, nine pictures representing familiar objects (i.e., bear, flower, key, chair, bell, eye, strawberry, train, glass) were selected based on the standardized set of 260 pictures (that are matched for name and image agreement, familiarity, and visual complexity) built by Snodgrass and Vanderwart (1980).[3] The same pictures as in the learning phase were then presented in the matching task. For the semantic task, 54 new pictures that the participants had not seen before in the experiment and that were semantically related or unrelated to the meaning of the newly learned words were chosen from the Internet by two of the authors (ED and MB). Students from our university (n = 60; age range = 19–25 years) were asked to rate the semantic relatedness between new and old pictures (that is, those previously presented during the word learning phase). Half of the presented pairs were semantically related and the other half were semantically unrelated, and this was confirmed by the students' ratings.

Experimental Tasks

Participants were tested individually in a quiet experimental room (i.e., Faraday cage), where they sat in a comfortable chair at about 1 m from a computer screen. Auditory stimuli were presented through HiFi headphones (HD590, Sennheiser Electronic GmBH, Wedemark, Germany) at 70-dB sound pressure level. Visual and auditory stimuli presentation as well as the collection of behavioral data were controlled by Presentation software (Version 11.0, Neurobehavioral Systems, Berkeley CA).

Main Experimental Session (See Figure 1A–E)

Phonological categorization task. At the beginning and at the end of the experiment, participants performed four different phonological tasks that lasted for 2.3 min each. All nine Thai monosyllabic words were presented in each task, but participants were asked to categorize them based upon different features in each task: (1) the voicing contrast (e.g., /ba1/ vs. /pa1/), (2) the vowel length (e.g., short: /ba1/ vs. long /ba:1/), (3) pitch (e.g., low: /pa:1/ vs. high: /ba:0/), and (4) the aspiration contrast (e.g., /pa1/ vs. /pha1/; see Figure 1A and E). For each task, the contrast was visually represented on the left (e.g., "short" with a short line) and right (e.g., "long" with a long line) half of the screen and participants had to press one of two response buttons according to the correct side (e.g., left one for short and right one for long vowels), as quickly and accurately as possible. Each word was presented 10 times in a pseudorandomized order with the constraints of no immediate repetition of the same word and no more than four successive same responses. Task order and response side were counterbalanced across participants.

Word learning phase. Participants were asked to learn the meaning of each word previously presented in the phonological categorization task using picture–word associations. For instance, a drawing of a bear was followed by the auditory presentation of the word /ba1/, and thus, /ba1/ was the word for bear in our "foreign" language (see Figure 1B). Each of the nine picture–word pairs was presented 20 times, resulting in 180 trials that were pseudorandomly presented (i.e., no immediate repetition of the same association) in two blocks of 3 min each. The picture was presented first and then followed after 750 msec by one of the nine words. Total trial duration was 2000 msec. Two different lists were built, so that across participants different pictures were associated with different words. No behavioral response was required from the participants but they were told that subsequent tests would evaluate whether they learned the meaning of the novel words.

Matching task. One of the nine pictures was presented, followed after 750 msec by an auditory word that matched or mismatched the associations previously learned in the word learning phase. For instance, whereas the drawing of a bear followed by /ba1/ (i.e., bear) was a match, the drawing of a strawberry followed by /ba1/ was a mismatch (see Figure 1C). Participants were asked to press one of two response keys accordingly, as quickly and accurately as possible. Response hand was counterbalanced across participants. At the end of the trial, a row of XXXX appeared on the screen, and participants were asked to blink during this time period (1000 msec; total trial duration: 3750 msec) to minimize eye movement artifacts during word presentation. Each word was presented 20 times, half in match condition and half in mismatch condition. The total of 180 trials was pseudorandomly presented (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 5.6 min each.

Semantic task. One of the new pictures was presented, followed after 1500 msec by an auditory word that was semantically related or unrelated. For instance, although the picture of a lock was semantically related to the previously learned word /pa:1/ (i.e., "key"), the picture of a strawberry cake was semantically unrelated to /pa:1/ (see Figure 1D). Participants were asked to press one of two response keys accordingly, as quickly and accurately as possible. A familiarization task including four trials was administrated before starting the task. Response hand was counterbalanced across participants. At the end of the trial, a row of XXXX appeared on the screen, and participants were asked to blink during this time period (1000 msec; total trial duration = 4500 msec). Each word was presented 12 times, but none of the new pictures were repeated, so that on each trial the word was associated with a different related or unrelated picture. Half of the picture–word pairs were semantically related, and half were semantically unrelated. A total of 108 trials was presented pseudorandomly (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 4 min each.

Long-term Memory Session (See Figure 1F)

To test for long-term memory effects on behavior (i.e., ERRs and RTs), participants performed again the matching (always administered first) and semantic tasks 5 months after the main experimental session (no ERPs were recorded). Because of a dropout rate of 33%, only 10 participants were retested in each group. In the matching task, a total of 270 trials were presented within three blocks. In the semantic task, a total of 216 trials were presented in two blocks (with a short pause within each block).

EEG Data Acquisition

The EEG was continuously recorded at a sampling rate of 512 Hz with a band-pass filter of 0–102.4 Hz by using a Biosemi amplifier system (BioSemi Active 2, Amsterdam, The Netherlands) with 32 active Ag/Cl electrodes (Biosemi Pintype) located at standard positions according to the International 10/20 System (Jasper, 1958). The EOG was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi and from an electrode beneath the right eye. Two additional electrodes were placed on the left and right mastoids. Electrode impedance was kept below 5 kΩ. EEG data were analyzed using Brain Vision Analyzer software (Version 1.05.0005 & Version 2.1.0; Brain Products, München, Germany). All data were re-referenced offline to the averaged left and right mastoids, filtered with a bandpass filter from 1 to 30 Hz (slope of 24 dB/oct), and independent component analysis and inverse independent component analysis were used to identify and remove components associated with vertical and horizontal ocular movements. Finally, DC-detrend and removal of artifacts above a gradient criterion of 10 μV/msec or a max–min criterion of 100 μV over the entire epoch were applied automatically. For each participant, ERPs were time-locked to word onset, segmented into 2700 msec epochs, including a 200-msec baseline and averaged within each condition. Individual averages were then averaged together to obtain the grand average across all participants.

Statistical Analyses

ANOVAs were computed using Statistica software (Version 12.0, StatSoft, Inc., Tulsa, OK). For ERRs and RTs, ANOVAs always included Group (MUS vs. NM) as between-subject factor as well as specific factors for each task. As the phonological categorization task was the only task that was performed both at the beginning and at the end of the experiment, Order (preexperiment vs. postexperiment) was included as a within-subject factor together with Task (Voicing vs. Vowel length vs. Pitch vs. Aspiration). The matching and semantic tasks were only performed once but each experiment was divided into two blocks of trials so that factors were Block (1 vs. 2) and Condition (match vs. mismatch or related vs. unrelated). To further increase clarity, factors were again specified at the beginning of each task in the Results section. On the basis of ERRs, four outliers (2 MUS and 2 NM, ±2 *SD* away from the mean) were excluded from further analyses.

For ERPs, we analyzed the early stages of auditory processing in the phonological categorization task using N100 peak amplitude measures. By contrast, during the word learning phase, as well as in the matching and semantic tasks, we focused on semantic processing and we analyzed the mean amplitude of the N400 component. Effects on the N200 were also analyzed using mean amplitudes. Because ERPs were only analyzed for correct responses and because the ERP traces of the four outliers that were eliminated from behavioral analyses were similar to the grand average in each group, all participants (i.e., 15 MUS and 15 NM) were included in the ERP analyses. ANOVAs always included Group (MUS vs. NM) as a between-subject factor and Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; right: F4, C4, P4) and Anterior/Posterior (frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4) as within-subject factors, together with specific factors for each task. As for behavior, for the phonological categorization task, these factors were Order (pre vs. post) and Task (Voicing vs. Vowel length vs. Pitch vs. Aspiration). For the matching and semantic tasks, factors were Block (1 vs. 2) and Condition (match vs. mismatch or related vs. unrelated). Post hoc Tukey tests (reducing the probability of Type I errors) were used to determine the origin of significant main effects and interactions. To simplify results presentation, we only report significant results related to our hypotheses (full statistical results can be seen in Table 1). Finally, correlation analyses (Pearson's coefficient) were computed between error rates in the musicality task with error rates or N400 effects in the semantic task. General linear models (including Group as a categorical factor, error rates in the musicality task as a continuous factor, and error rates or N400 effects in the semantic task as a dependent factor) were used to test whether the differences between the slopes and intercepts of the two groups were significant.

RESULTS

Results are presented first for the screening measures of cognitive ability and musical aptitude and second for the experimental tasks. For each experimental task, behavioral data are presented first followed by the ERPs data, except for the word learning phase in which no behavioral data were recorded and for the long-term memory tasks in which no ERPs were recorded. Finally, for ERPs data (except in the

Table 1. Results of ANOVAs on the ERPs Data in Different Tasks of the Main Experimental Session

	Word Learning Phase						*Matching Task*						*Semantic Task*					
	ALL		*MUS*		*NM*		*ALL*		*MUS*		*NM*		*ALL*		*MUS*		*NM*	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
N400 (340–550 msec)																		
G	2.72	.11	–	–	–	–	3.85	.06	–	–	–	–	0.93	.34	–	–	–	–
G × B	0.12	.73	–	–	–	–	3.23	.08	–	–	–	–	0.02	.89	–	–	–	–
G × L	0.41	.67	–	–	–	–	0.21	.82	–	–	–	–	0.21	.81	–	–	–	–
G × R	0.13	.88	–	–	–	–	0.75	.48	–	–	–	–	*3.14*	*.05*	–	–	–	–
G × B × L	*4.65*	*.01*	–	–	–	–	0.07	.93	–	–	–	–	0.37	.69	–	–	–	–
G × B × R	1.87	.16	–	–	–	–	2.40	.10	–	–	–	–	0.01	.99	–	–	–	–
G × L × R	0.67	.62	–	–	–	–	1.27	.29	–	–	–	–	2.23	.07	–	–	–	–
G × B × L × R	0.82	.52	–	–	–	–	0.40	.81	–	–	–	–	0.60	.67	–	–	–	–
G × C	–	–	–	–	–	–	1.61	.21	–	–	–	–	1.06	.31	–	–	–	–
G × B × C	–	–	–	–	–	–	0.42	.52	–	–	–	–	0.84	.37	–	–	–	–
G × C × L	–	–	–	–	–	–	1.25	.29	–	–	–	–	0.21	.81	–	–	–	–
G × C × R	–	–	–	–	–	–	*3.14*	*.05*	–	–	–	–	0.04	.96	–	–	–	–
G × B × C × L	–	–	–	–	–	–	0.83	.44	–	–	–	–	0.01	.99	–	–	–	–
G × B × C × R	–	–	–	–	–	–	1.63	.21	–	–	–	–	0.15	.86	–	–	–	–
G × C × L × R	–	–	–	–	–	–	1.82	.13	–	–	–	–	1.19	.32	–	–	–	–
G × B × C × L × R	–	–	–	–	–	–	1.97	.43	–	–	–	–	0.52	.72	–	–	–	–
B	*4.25*	*.05*	1.57	.23	2.72	.12	3.13	.09	<.001	.99	7.68	.02	1.69	.20	0.63	.44	1.11	.31
B × L	*7.05*	*.002*	*16.38*	*<.001*	0.10	.91	0.52	.60	0.28	.76	0.31	.74	0.81	.45	0.25	.78	0.91	.42
B × R	1.32	.27	*3.29*	*.05*	0.12	.89	1.46	.24	0.08	.93	5.08	.01	0.36	.70	0.10	.90	0.32	.73
B × L × R	0.42	.80	0.68	.61	0.54	.71	1.06	.38	0.84	.50	0.54	.71	0.65	.63	0.71	.59	0.51	.73
C	–	–	–	–	–	–	0.01	.95	0.53	.48	1.38	.26	0.10	.75	0.65	.43	0.42	.53
C × B	–	–	–	–	–	–	1.17	.29	0.07	.79	2.06	.17	1.46	.24	0.04	.85	2.43	.14
C × L	–	–	–	–	–	–	*9.72*	*<.001*	*5.36*	*.01*	*5.58*	*.009*	1.41	.25	0.19	.83	2.25	.12
C × R	–	–	–	–	–	–	*27.80*	*<.001*	*28.34*	*<.001*	*6.38*	*.005*	*19.18*	*<.001*	*8.98*	*<.001*	*10.22*	*<.001*
C × B × L	–	–	–	–	–	–	0.47	.63	1.02	.38	0.39	.68	0.45	.64	0.43	.65	0.13	.88
C × B × R	–	–	–	–	–	–	2.43	.10	2.99	.07	0.09	.91	1.12	.33	1.36	.27	0.18	.84
C × B × L × R	–	–	–	–	–	–	1.29	.28	1.11	.36	1.15	.34	0.59	.67	0.57	.69	0.54	.71
N200 (230–340 msec)																		
G	1.24	.28	–	–	–	–	*5.56*	*.03*	–	–	–	–	1.49	.23	–	–	–	–
G × B	0.13	.73	–	–	–	–	0.04	.84	–	–	–	–	*7.26*	*.01*	–	–	–	–
G × L	0.60	.55	–	–	–	–	0.27	.77	–	–	–	–	2.41	.10	–	–	–	–
G × R	0.04	.96	–	–	–	–	0.87	.43	–	–	–	–	2.56	.09	–	–	–	–
G × B × L	1.05	.36	–	–	–	–	0.03	.97	–	–	–	–	0.29	.75	–	–	–	–
G × B × R	0.70	.50	–	–	–	–	0.69	.51	–	–	–	–	0.85	.43	–	–	–	–
G × L × R	0.99	.42	–	–	–	–	1.55	.19	–	–	–	–	1.76	.14	–	–	–	–
G × B × L × R	0.68	.61	–	–	–	–	0.20	.94	–	–	–	–	0.69	.60	–	–	–	–

Table 1. (*continued*)

	Word Learning Phase						*Matching Task*						*Semantic Task*					
	ALL		*MUS*		*NM*		*ALL*		*MUS*		*NM*		*ALL*		*MUS*		*NM*	
	F	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*	*F*	*p*
G × C	–	–	–	–	–	–	2.19	.15	–	–	–	–	1.49	.23	–	–	–	–
G × B × C	–	–	–	–	–	–	0.28	.60	–	–	–	–	0.004	.95	–	–	–	–
G × C × L	–	–	–	–	–	–	1.85	.17	–	–	–	–	0.64	.53	–	–	–	–
G × C × R	–	–	–	–	–	–	1.41	.25	–	–	–	–	0.88	.42	–	–	–	–
G × B × C × L	–	–	–	–	–	–	0.14	.87	–	–	–	–	0.37	.70	–	–	–	–
G × B × C × R	–	–	–	–	–	–	0.31	.74	–	–	–	–	0.50	.61	–	–	–	–
G × C × L × R	–	–	–	–	–	–	1.51	.21	–	–	–	–	2.19	.08	–	–	–	–
G × B × C × L × R	–	–	–	–	–	–	1.99	.10	–	–	–	–	0.69	.60	–	–	–	–
B	*11.61*	*.002*	*6.50*	*.02*	*5.11*	*.04*	2.28	.14	0.81	.38	1.55	.23	0.19	.67	*5.36*	*.04*	2.35	.15
B × L	*4.44*	*.02*	*4.14*	*.03*	0.82	.45	1.62	.21	0.77	.47	0.90	.42	1.24	.30	0.78	.47	0.75	.48
B × R	0.93	.40	1.25	.30	0.47	.63	0.30	.74	0.20	.82	1.33	.28	0.26	.77	0.58	.57	0.52	.60
B × L × R	0.51	.73	0.26	.91	1.33	.27	0.71	.58	0.37	.83	0.62	.65	0.41	.80	1.40	.25	0.09	.99
C	–	–	–	–	–	–	*16.53*	*<.001*	*15.22*	*.002*	3.38	.09	*4.16*	*.05*	*4.87*	*.05*	0.37	.55
C × B	–	–	–	–	–	–	0.72	.40	0.93	.35	0.05	.82	*4.32*	*.05*	2.00	.18	2.33	.15
C × L	–	–	–	–	–	–	3.02	.06	4.04	.03	1.18	.32	1.59	.21	1.60	.22	0.22	.80
C × R	–	–	–	–	–	–	*20.06*	*<.001*	*15.12*	*<.001*	*6.12*	*.006*	*6.83*	*.002*	*5.05*	*.01*	1.89	.17
C × B × L	–	–	–	–	–	–	0.21	.81	0.20	.82	0.16	.85	0.28	.76	0.57	.57	0.10	.90
C × B × R	–	–	–	–	–	–	1.34	.27	1.13	.34	0.26	.77	*5.25*	*.008*	*4.97*	*.01*	1.17	.33
C × B × L × R	–	–	–	–	–	–	*4.25*	*.003*	*3.98*	*.007*	1.88	.13	1.53	.20	1.24	.30	0.94	.45

Although the Condition × Anterior/posterior interactions are significant in musicians (MUS) and in nonmusicians (NM) for the N400 component, the effects are reversed in both groups (typical N400 effect over parietal sites in MUS, inversed N400 effect over frontal sites in NM). Significant effects are printed in italics, and exact levels of significance are indicated except when the p values are inferior to .001 (<.001). G = Group; B = Block; C = Condition; L = Laterality; R = Anterior/Posterior.

phonological categorization task where the N100 component is of main interest), analysis of the N400 component is presented first, followed by analyses of the N200.

Screening Measures

Cognitive Ability

Psychometric data were evaluated by means of univariate ANOVAs. Results showed no significant Group differences regarding general reasoning abilities (i.e., progressive matrices, PM47; $F(1, 28) = 1.37$, $p = .25$), verbal working memory (reverse digit span; $F(1, 28) = 2.88$, $p = .10$), nor visual attention ($F(1, 28) = 3.17$, $p = .09$). By contrast, MUS (mean = 7.6, $SD = 0.30$) showed better short-term memory abilities than NM (mean = 6.7, $SD = 0.30$; direct digit span; $F(1, 28) = 5.53$, $p = .03$).

Musical Aptitude

A 2 × 2 ANOVA (i.e., 2 Groups × 2 Tasks) showed that MUS made fewer errors (6.7%, $SD = 2.0$) than NM (17.6%, $SD = 2.0$; main effect of Group: $F(1, 28) = 14.71$, $p < .001$), and all participants performed better on the rhythmic (9.8%, $SD = 1.6$) than on the melodic task (14.4%, $SD = 2.0$; main effect of Task: $F(1, 28) = 4.19$, $p = .05$) with no Group × Task interaction.

Experimental Tasks

Phonological Categorization Task

Behavioral data. Results of 2 × 2 × 4 ANOVAs [i.e., 2 Groups (MUS vs. NM) × 2 Orders (pre vs. post) × 4 Tasks (Voicing vs. Vowel length vs. Pitch vs. Aspiration)] showed that MUS (6.6%, $SD = 2.2$) made overall fewer errors compared with NM (19.1%, $SD = 2.2$; main effect of Group: $F(1, 24) = 16.29$, $p < .001$; Figure 2A). The influence of music training was largest in the pitch (MUS: 4.1%, $SD = 4.8$; NM: 25.8%, $SD = 4.9$; Tukey, $p < .001$) and aspiration tasks (MUS: 7.6%, $SD = 5.5$; NM: 28.7%, $SD = 5.5$; Tukey, $p < .001$; Group × Task interaction: $F(3, 72) = 11.82$, $p < .001$). Finally, only NM improved their level of performance from pre to post in the pitch task (Group × Task × Order interaction: $F(3, 72) = 3.31$,

p = .03; NM: pre: 30.3%, SD = 4.1, post: 21.3%, SD = 3.3, Tukey, p = .02; MUS: pre: 4.5%, SD = 4.1, post: 3.7%, SD = 3.3, Tukey, p = .99). Analyses of RTs did not reveal significant Group differences (Figure 2B).

Electrophysiological data. N100 amplitudes were evaluated by means of a 2 × 2 × 4 × 3 × 3 ANOVA (i.e., 2 Groups × 2 Orders × 4 Tasks × 3 Laterality positions [left vs. midline vs. right] × 3 Anterior/Posterior positions [frontal vs. central vs. parietal]). Results revealed a significant Group × Task × Laterality interaction effect (F(6, 168) = 3.19, p = .005). Separate ANOVAs for each group showed that only MUS were characterized by a larger N100 in the aspiration task (MUS: −6.50 µV, SD = 2.59 and NM: −5.57 µV, SD = 1.81) compared with the other three tasks over the left hemisphere and at midline electrodes (Task × Laterality interaction: MUS: −5.47 µV, SD = 1.78; F(6, 84) = 3.13, p = .008, Tukey, both ps < .001; NM: −5.01 µV, SD = 1.74; F < 1; Figure 3A and B).

Word Learning Phase

Electrophysiological data. The N400 as well as the N200 were evaluated by means of 2 × 2 × 3 × 3 ANOVAs (i.e., 2 Groups × 2 Blocks [1 vs. 2] × 3 Laterality × 3 Anterior/Posterior positions).

For all participants and in line with previous results, the N400 component was larger over frontal (−2.63 µV, SD = 0.87) and central (−2.52 µV, SD = 0.76) sites compared with parietal sites (−1.29 µV, SD = 0.66; Tukey, both ps < .001; main effect of Anterior/Posterior: F(2, 56) = 25.25, p < .001). In addition, the Group × Block × Laterality interaction effect was significant (F(2, 56) = 4.65,

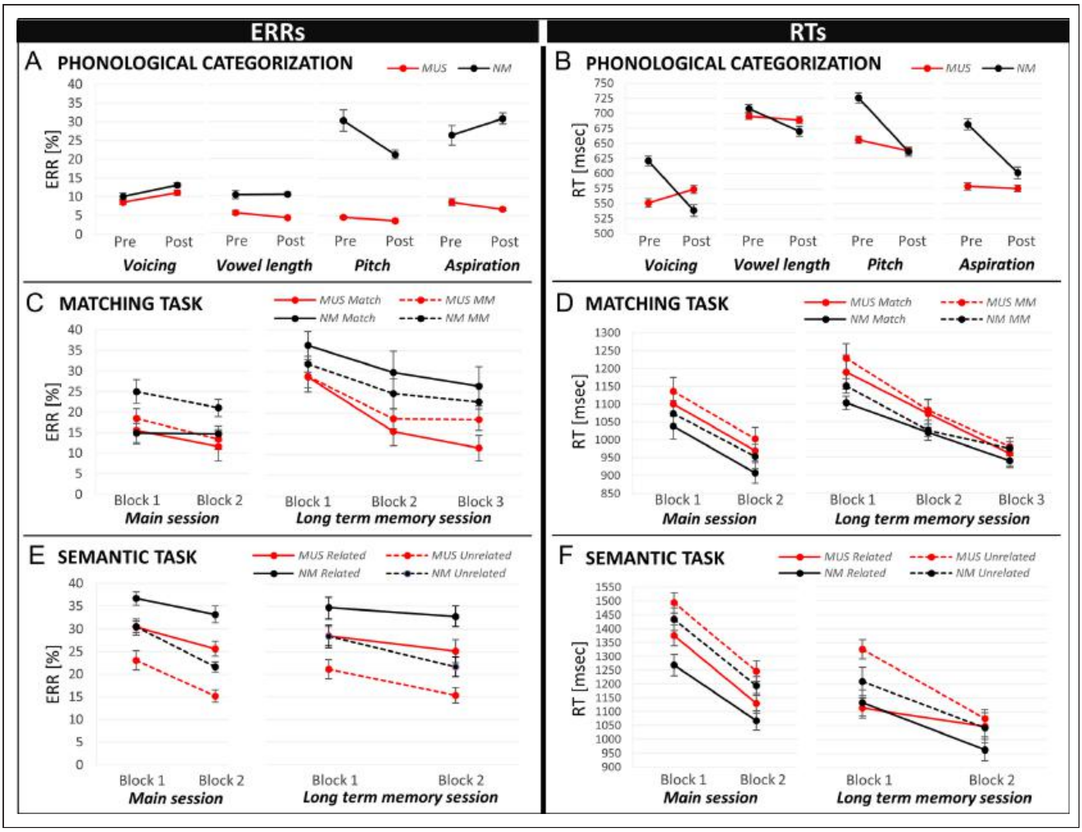


Figure 2. Percentages of errors (ERRs) and RTs in the different tasks are shown for musicians (MUS) in red and for nonmusician controls (NM) in black. For the phonological categorization task (A and B), results are illustrated for the premeasurement (beginning) and postmeasurement (end of main experimental session), separately for each task (voicing, vowel length, pitch, and aspiration). For the matching task (C and D), results for Match (solid lines) and Mismatch (MM: dotted lines) words are illustrated in the two blocks of the main experimental session and in the three blocks of the long-term memory session. For the semantic task (E and F), results for semantically Related (solid lines) and Unrelated (dotted lines) words are illustrated in the two blocks of the main experimental session and of the long-term memory session.

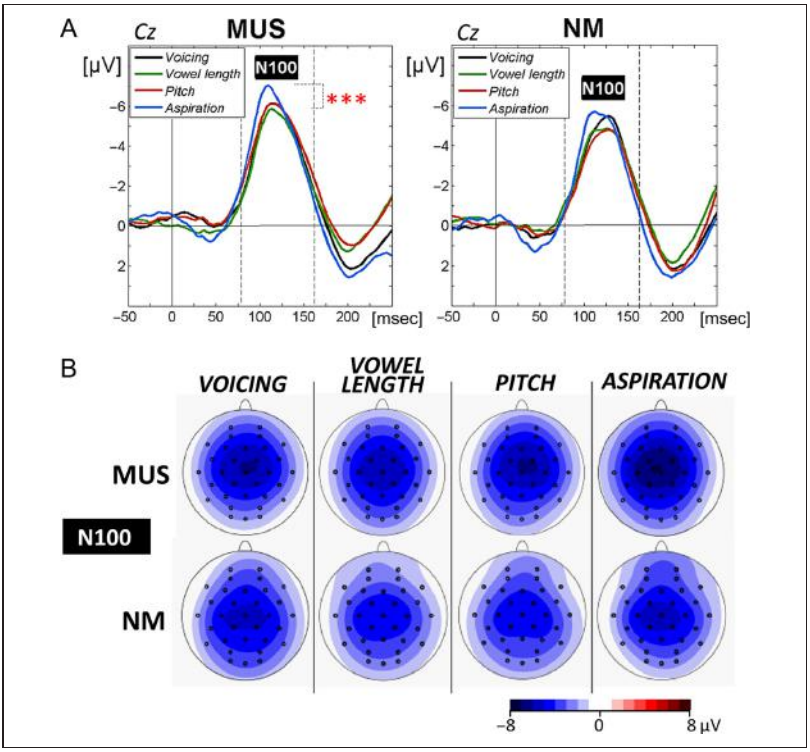


Figure 3. Phonological categorization. (A) N100 components at the Central (Cz) electrode are compared between tasks for musicians (MUS) and for nonmusician controls (NM). In this and subsequent figures, time in milliseconds is in abscissa and the amplitude of the effects in microvolt is in ordinate. Time zero corresponds to word onset and negativity is plotted upwards. Latency windows for statistical analyses are indicated with gray dotted lines and the level of significance is represented by stars with $*p < .05$, $**p < .01$, and $***p < .001$ (red stars for MUS and black stars for NM). (B) Topographic voltage distribution maps illustrate the N100s to the words separately for each task and for MUS and NM. Voltage values are scaled from −8 to +8 µV.

$p = .01$). Separate group analyses showed that only MUS showed significantly increased amplitudes from Block 1 to Block 2. This effect was localized over the left hemisphere and midline electrodes (MUS: Block 1: −1.54 µV, *SD* = 0.76 and Block 2: −2.16 µV, *SD* = 0.79; Block × Laterality interaction: $F(2, 28) = 16.38$, $p < .001$; Tukey, both *p*s < .001 and NM: Block 1: −2.34 µV, *SD* = 1.36 and Block 2: −2.91 µV, *SD* = 1.38; main effect of Block: $F(1, 14) = 2.72$, $p = .12$; Figure 4A and B).

Analyses of the N200 component did not reveal significant Group differences (main effect of Group: $F(1, 28) = 1.24$, $p = .28$), but all participants showed significantly increased amplitudes from Block 1 (−1.13 µV, *SD* = 1.22) to Block 2 (−1.79 µV, *SD* = 0.99; main effect of Block: $F(1, 28) = 11.61$, $p = .002$; Figure 4A and B).

Matching Task

Behavioral data. Results of three-way ANOVAs (i.e., 2 Groups × 2 Blocks × 2 Conditions [match vs. mismatch]) showed that ERRs did not significantly differ between the two groups (main effect of Group: $F(1, 24) = 2.19$, $p = .15$). However, all participants committed overall fewer errors for match (14.2%, *SD* = 2.6) compared with mismatch words (19.4%, *SD* = 2.2) and fewer errors in Block 2 (15.2%, *SD* = 2.1) than in Block 1 (18.5%, *SD* = 2.2; main effect of Condition: $F(1, 24) = 7.68$, $p = .01$; main effect of Block: $F(1, 24) = 9.27$ $p = .006$; Figure 2C). In line with ERRs, analyses of RTs did not reveal between-group differences (main effect of Group: $F < 1$), but overall faster RTs for match (1041 msec, *SD* = 64) than for mismatch words (1080 msec, *SD* = 70; main effect of Condition: $F(1, 24) = 5.90$, $p = .02$) and faster RTs in Block 2 (994 msec, *SD* = 61) than in Block 1 (1128 msec, *SD* = 73; main effect of Block: $F(1, 24) = 60.45$, $p < .001$; Figure 2D).

Electrophysiological data. The N400 as well as the N200 component were evaluated by means of 2 × 2 × 2 × 3 × 3 ANOVAs (2 Groups × 2 Blocks × 2 Conditions × 3 Laterality × 3 Anterior/Posterior positions).

Analysis of the N400 revealed a significant Group × Condition × Anterior/Posterior interaction effect ($F(2, 56) = 3.14$, $p = .05$). Results of separate group analyses showed larger N400 amplitudes in MUS for mismatch (−0.10 µV, *SD* = 1.82) compared with match words over centroparietal regions (0.67 µV, *SD* = 1.58; Condition × Anterior/Posterior interaction: $F(2, 28) = 28.34$, $p < .001$; Tukey, central: $p = .02$; parietal: $p < .001$). The opposite pattern was found in NM with larger N400 for match (−1.34 µV, *SD* = 1.42) than for mismatch words over frontocentral sites (−0.89 μV, *SD* = 1.53; Condition × Anterior/Posterior interaction: *F*(2, 28) = 6.38, *p* = .005; Tukey, frontal: *p* = .001; central: *p* = .03; Figure 5A and B).

The N200 amplitude was overall smaller in MUS (−0.15 μV, *SD* = 0.60) compared with NM (−2.14 μV, *SD* = 0.60; main effect of Group: *F*(1, 28) = 5.56, *p* = .03) and the N200 effect (i.e., mismatch minus match words) was more widely distributed in MUS compared with NM (Group × Block × Condition × Laterality × Anterior/Posterior interaction: *F*(4, 112) = 1.99, *p* = .10; Figure 5A and B). MUS showed larger N200 amplitudes for mismatch (P4: −1.98 μV, *SD* = 0.76) than for match words (P4: 0.01 μV, *SD* = 0.68) over centroparietal scalp sites with largest differences over midline and right hemisphere (Condition × Anterior/Posterior interaction: *F*(2, 28) = 15.12, *p* < .001; Condition × Laterality interaction: *F*(2, 28) = 4.04, *p* = .03). In addition, the N200 effect was larger in Block 2 (P4: −2.22 μV, *SD* = 0.50) than in Block 1 (P4: −1.66 μV, *SD* = 0.63) over midline and right centroparietal sites (Condition × Block × Laterality × Anterior/Posterior interaction: *F*(4, 56) = 3.98, *p* = .007). NM also showed an N200 effect that was localized over parietal sites (−1.06 μV, *SD* = 1.41; Condition × Anterior/Posterior interaction: *F*(2, 28) = 6.12, *p* = .006).

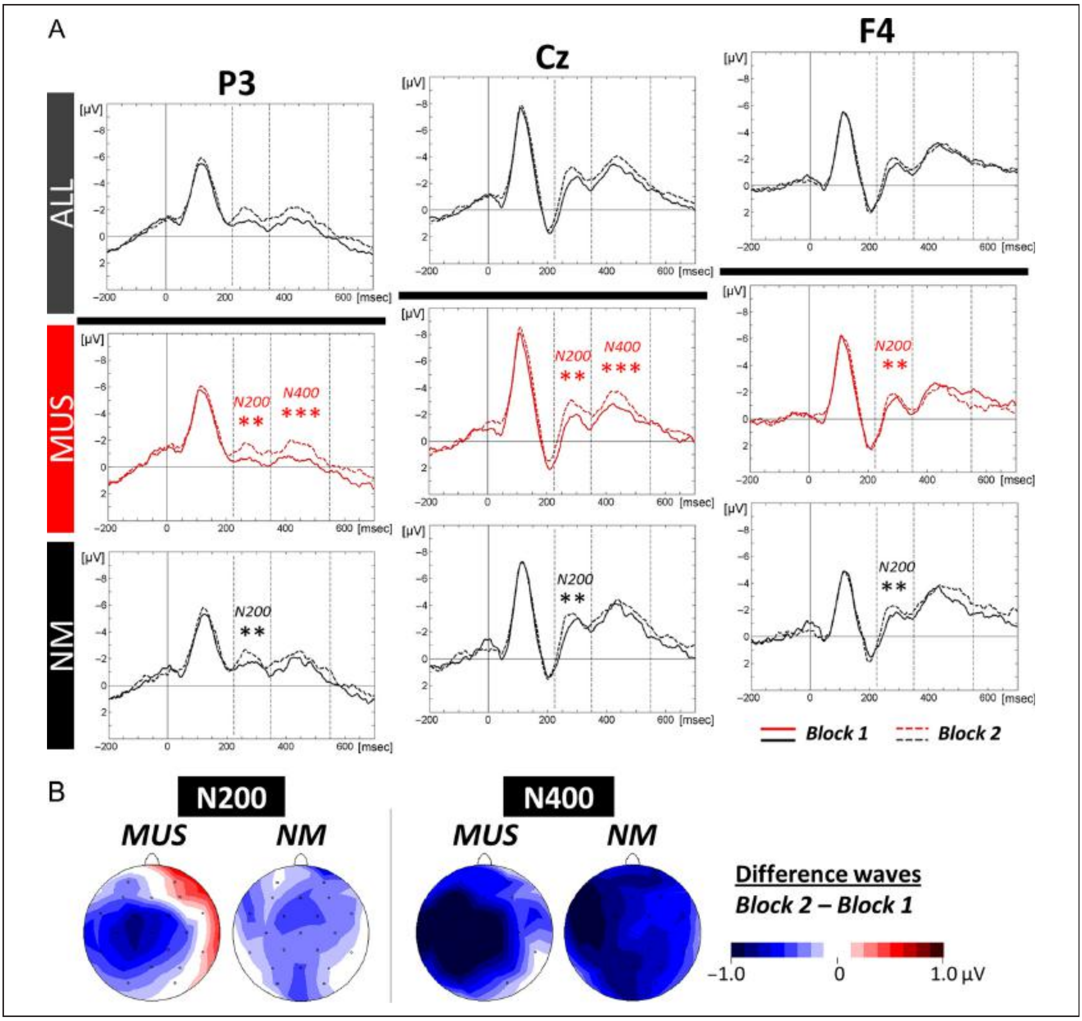


Figure 4. Word learning phase. (A) ERPs recorded at left parietal (P3), central (Cz), and right frontal sites (F4) in Block 1 (solid line) and Block 2 (dotted line) are overlapped for all participants (ALL: black lines, top), and separately below for musicians (MUS: red lines) and nonmusician controls (NM: black lines). (B) Topographic voltage distribution maps of the differences between the two blocks (Block 2 minus Block 1) are illustrated for the ERP components of interest (N200, N400), separately for MUS and for NM. Voltage values are scaled from −1.0 to +1.0 μV.

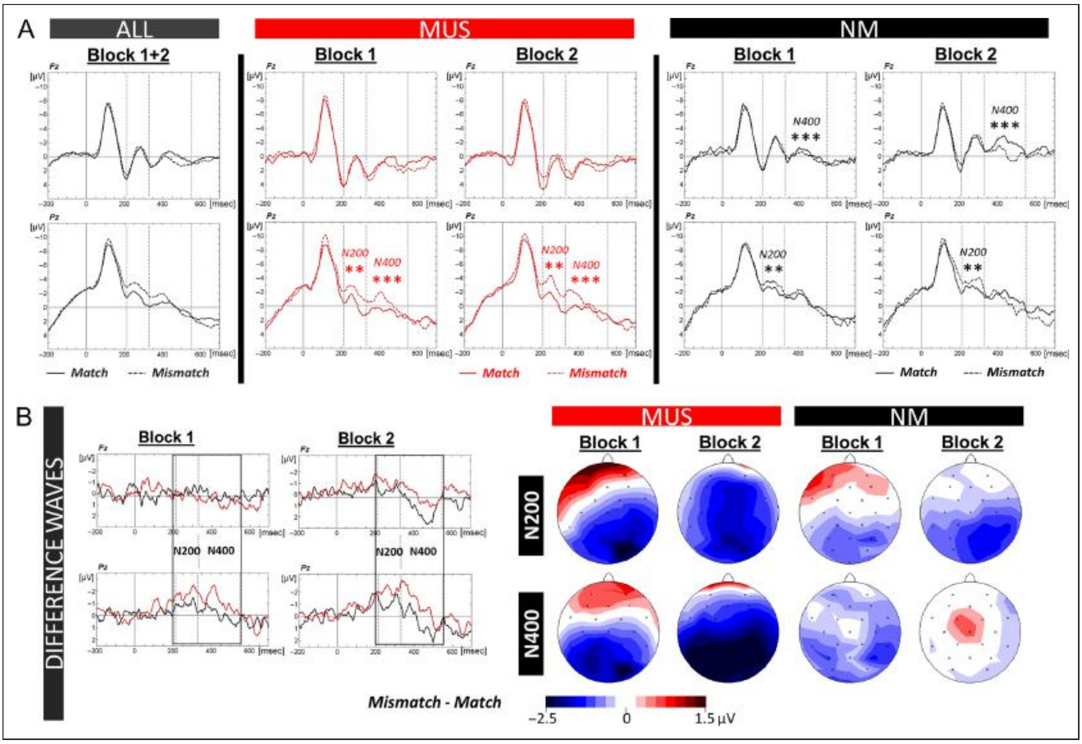


Figure 5. Matching task. (A) Left: ERPs recorded at frontal (Fz) and parietal (Pz) sites are overlapped for Match (solid lines) and Mismatch (dotted lines) words for all participants across the two blocks of trials (ALL: black lines). Central and right: ERPs are presented separately for Block 1 and Block 2 and for musicians (MUS: red lines) and nonmusician controls (NM: black lines). (B) Difference waves (Mismatch minus Match) are overlapped for MUS (red) and NM (black) separately for Block 1 and for Block 2 at Fz and Pz. Topographic voltage distribution maps of the Mismatch minus Match differences are illustrated for N200 and N400 components separately and for MUS and NM in Block 1 and Block 2. Voltage values are scaled from −2.5 to +1.5 μV.

Semantic Task

Behavioral data. Results of three-way ANOVAs (i.e., 2 Groups × 2 Blocks × 2 Conditions [related vs. unrelated]) showed that MUS (23.6%, *SD* = 2.0) made overall fewer errors than NM (30.5%, *SD* = 2.0; main effect of Group: *F*(1, 24) = 5.82, *p* = .02), and all participants made fewer errors for unrelated (22.6%, *SD* = 2.6) than for related words (31.5%, *SD* = 2.9; main effect of Condition: *F*(1, 24) = 11.24, *p* = .003; Figure 2E). Moreover, all participants made fewer errors in Block 2 (30.2%, *SD* = 2.3) than in Block 1 (23.9%, *SD* = 2.4; main effect of Block: *F*(1, 24) = 12.37, *p* = .002). RTs were faster for related (1210 msec, *SD* = 72) than for unrelated words (1342 msec, *SD* = 71; main effect of Condition: *F*(1, 24) = 41.32, *p* < .001) and faster in Block 2 (1159 msec, *SD* = 70) than in Block 1 (1393 msec, *SD* = 75; main effect of Block: *F*(1, 24) = 88.92, *p* < .001), with no between-group differences (main effect of Group: *F* < 1; Figure 2F).

Electrophysiological data. The N400 as well as the N200 component were evaluated by means of 2 × 2 × 2 × 3 × 3 ANOVAs (i.e., 2 Groups × 2 Blocks × 2 Conditions × 3 Laterality × 3 Anterior/Posterior positions).

N400 analyses revealed a significant Group × Anterior/Posterior interaction effect (*F*(2, 56) = 3.14, *p* = .05). As typically reported in the literature (Kutas et al., 1988), the N400 was larger for semantically unrelated (2.17 μV, *SD* = 1.93) compared with related words (3.29 μV, *SD* = 1.66) over parietal sites in MUS (Condition × Anterior/Posterior interaction: *F*(2, 28) = 8.98, *p* < .001; Tukey, parietal: *p* < .001). By contrast, a reversed N400 effect was found in NM with larger N400 for related (−2.09 μV, *SD* = 1.60) than for unrelated words (−1.19 μV, *SD* = 1.06) over frontal sites (Condition × Anterior/Posterior; *F*(2, 28) = 10.22, *p* < .001; Tukey, frontal: NM: *p* = .002, Figure 6A and B).

The N200 amplitude was larger in Block 2 than in Block 1 in MUS only (Group × Block: (*F*(1, 28) = 7.26, *p* = .01; MUS: Block 1: 1.58 μV, *SD* = 2.62 and Block 2: 0.89 μV, *SD* = 2.68; *F*(1, 14) = 5.35, *p* = .04 and NM: Block 1: −0.11 μV, *SD* = 3.15 and Block 2: 0.39 μV, *SD* = 2.58; *F*(1, 14) = 2.35, *p* = .15; Figure 6A and B). In addition, the N200 was also larger for unrelated than for related words in MUS but not in NM (main effect of

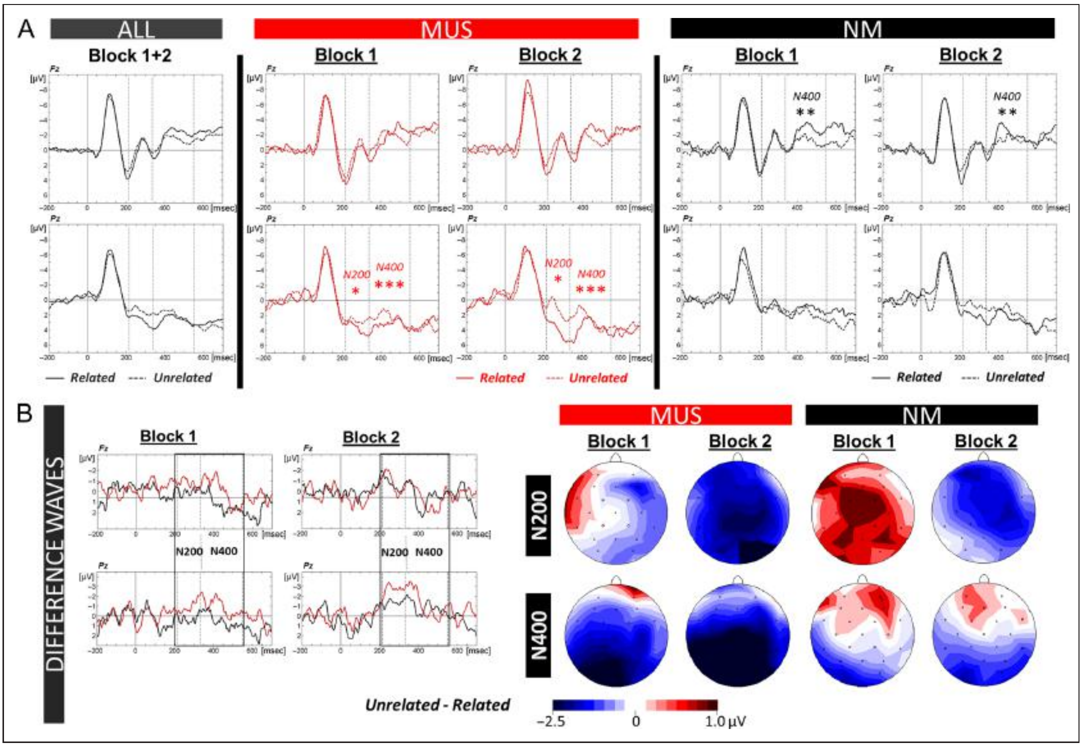


Figure 6. Semantic task. (A) Left: ERPs recorded at frontal (Fz) and parietal sites (Pz) are overlapped for semantically Related (solid lines) and Unrelated (dotted lines) words for all participants across the two blocks of trials (ALL: black lines). Central and right panels: ERPs are presented separately for Block 1 and Block 2 and for musicians (MUS: red lines) and nonmusician controls (NM: black lines). (B) Difference waves (Unrelated minus Related) are overlapped for MUS (red) and NM (black) separately for Block 1 and Block 2 at Fz and Pz. Topographic voltage distribution maps of the Unrelated minus Related differences are illustrated for N200 and N400 components separately and for MUS and NM in Block 1 and Block 2. Voltage values are scaled from −2.5 to +1.0 μV.

Condition: MUS: Related: 1.88 μV, *SD* = 2.67 and Unrelated: 0.59 μV, *SD* = 3.03; *F*(1, 14) = 4.87, *p* = .05 and NM: Related: 0.30 μV, *SD* = 3.63 and Unrelated: −0.02 μV, *SD* = 2.23; *F* < 1; Group × Condition × Laterality × Anterior/Posterior interaction: *F*(4, 112) = 2.19, *p* = .08), and this effect was larger over central and parietal sites (Condition × Anterior/Posterior interaction: *F*(2, 28) = 5.05, *p* = .01; Tukey, both *p*s < .001).

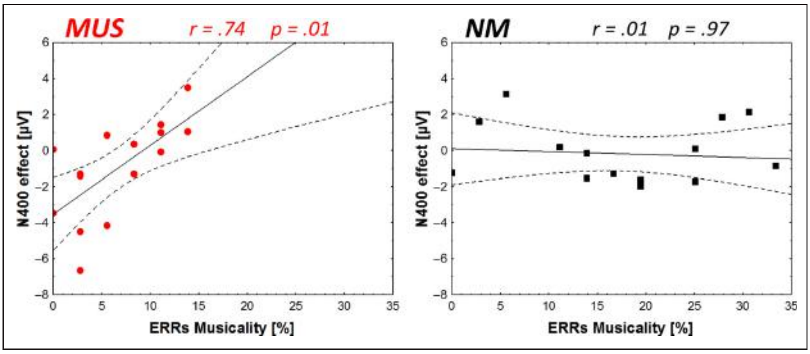


Figure 7. Correlation analyses. Correlations between the percentages of error in the musicality test (ERRs Musicality) and the sizes of the N400 effect (Unrelated minus Related) in the semantic task are illustrated for musicians (MUS: red) and for nonmusician controls (NM: black). Dotted lines represent the 95% confidence interval of the correlation line.

Long-term Memory: Matching Task (Behavior)

Results of three-way ANOVAs (i.e., 2 Groups × 3 Blocks × 2 Conditions [match vs. mismatch]) showed that MUS (20.3%, *SD* = 2.8) made fewer errors compared with NM (28.6%, *SD* = 2.9; main effect of Group: $F(1, 19) = 4.19$, $p = .05$; Figure 2C). Moreover, all participants improved their level of performance from Block 1 (31.4%, *SD* = 2.6) to Blocks 2 and 3 (22.2%, *SD* = 3.4 and 19.8%, *SD* = 3.6, respectively; main effect of Block: $F(2, 38) = 26.40, p < .001$). No significant between-group differences were found on RTs but in both groups, RTs were faster in Block 3 (965 msec, *SD* = 67) than in Block 2 (1049 msec, *SD* = 72) and slowest in Block 1 (1168 msec, *SD* = 82; main effect of Block: $F(2, 38) = 36.48, p < .001$; Figure 2D).

Long-term Memory: Semantic Task (Behavior)

Results of three-way ANOVAs (i.e., 2 Groups × 2 Blocks × 2 Conditions [related vs. unrelated]) showed that between-group differences were not significant on ERRs nor on RTs. However, all participants made fewer errors for unrelated (21.5%, *SD* = 6.0) than for related words (29.7%, *SD* = 6.3; main effect of Condition; $F(1, 19) = 6.60, p = .02$) and were faster for related (1055 msec, *SD* = 106) than for unrelated words (1151 msec, *SD* = 119; main effect of Condition; $F(1, 19) = 14.54, p = .001$; Figure 2E and F).

Relationships between Musical Aptitude, Behavioral Data, and Brain Activity

Error rates in the musicality test were correlated with the size of the N400 effects in the semantic task for MUS ($r = .74$, $p = .01$) but not for NM ($r = .01$, $p = .97$; Figure 7A and B). Moreover, the slopes of the correlations were significantly different for the two groups (main effect of Group: $F(1, 26) = 7.36, p = .01$) as well as the correlation intercepts (Group × Musicality interaction: $F(2, 26) = 6.52, p = .005$).

DISCUSSION

Summary of Results

By using an ecologic valid experimental design we tested the general hypothesis that professional musicians would learn the meaning of novel words more efficiently than control participants. Overall, both behavioral and electrophysiological data support this hypothesis. Behaviorally, musicians performed better than controls in the musicality and phonological categorization tasks. In addition, although all participants performed similarly in the matching task, musicians made significantly fewer errors in the semantic task. Finally, after 5 months, musicians remembered more words than controls as reflected by lower error rates in the matching task.

The electrophysiological markers of word learning also clearly differed between the two groups. Although control participants showed similar N100 amplitudes in all phonological categorization tasks, musicians showed an increase in N100 amplitude when categorizing the most difficult aspiration contrast. Most importantly and in line with the word learning literature, both groups showed enhanced N400 amplitudes over frontal scalp sites when learning the meaning of novel words. However, only musicians were additionally characterized by larger left-lateralized N400s in the second block compared with the first block of the word learning phase. Finally, only musicians showed the typical centroparietal distribution of the N400 effect in the matching and semantic tasks. By contrast, nonmusicians showed reversed N400 effects in both tasks over frontal sites. These findings are discussed in detail below. It is, however, important to note that cross-sectional studies, as the one reported here, are necessary to first demonstrate differences between musically trained and untrained participants before designing longitudinal experiments to test for the causality of the reported effects.

Spatiotemporal Dynamics in the Learning Phase

Results showed clear evidence for fast brain plasticity, as reflected by the rapid development of the N400 in both groups of participants after only 3 min of novel word learning (Block 1), that is after 10 repetitions of each picture–word association (see Figure 4). This finding extends previous results on word learning showing N400 enhancements when learning a second language (McLaughlin et al., 2004), the meaning of rare words (Perfetti et al., 2005), and when learning the meaning of novel words or pseudowords from highly constrained sentence contexts (Borovsky et al., 2010, 2012; Batterink & Neville, 2011; Mestres-Missé et al., 2007). Importantly, and in line with previous work in adults (Borovsky et al., 2012; Mestres-Missé et al., 2007) and in children (François et al., 2013; Friedrich & Friederici, 2008), in both musicians and controls the N400 component to novel words was larger frontocentrally than parietally. These results are compatible with previous findings, suggesting that prefrontal and temporal brain regions are associated with the maintenance of novel information in working memory (Hagoort, 2014) and the acquisition of word meaning (Rodriguez-Fornells et al., 2009).

Importantly, however, the N400 increase from the first to the second block of novel word learning was only significant in musicians (see Figure 4A). In line with our hypothesis, this is taken to suggest faster encoding of novel word meaning in musicians than in controls. In addition, word learning in musicians was accompanied by a rapid shift of the N400 component distribution from frontal to centroparietal sites. This shift in distribution is in line with the hypothesis that musicians have already integrated novel words representations in semantic memory (Batterink &

Neville, 2011) and that the N400 centroparietal distribution in musicians reflects access to semantic memory (Kutas & Federmeier, 2011). By contrast, in control participants, the N400 component remained larger frontally throughout the entire learning phase, suggesting that nonmusicians had not yet integrated the novel words' meaning into established semantic networks. This interpretation can be directly tested in future experiments by increasing the number of repetitions of picture–word associations. Under such conditions, a frontal to centroparietal shift in N400 distribution should also be found in nonmusicians.

Evidence for Rapidly Established Representations of Novel Words in the Matching and Semantic Tasks

As expected, all participants, whether musicians or controls, were able to learn the nine picture–word associations, as reflected by error rates below chance level (i.e., 50% errors) in both the matching (mean = 17%) and semantic tasks (mean = 27%; see Figure 2C and E). To follow the dynamics of word learning within each task, results were compared between the first and the second block of trials. Interestingly, the level of performance increased with repetition in both tasks, indicating a still ongoing learning effect. Moreover, both musicians and controls showed clear matching effects with lower error rates and faster RTs for match compared with mismatch words (Boddy & Weinberg, 1981). However, and in contrast to typical semantic priming effects (Meyer & Schvaneveldt, 1971), both groups of participants made more errors for semantically related words than for semantically unrelated words. Although unexpected, this result may reflect a response bias towards rejection (i.e., considering the word as unrelated to the picture) as a consequence of task difficulty generating higher uncertainty (Gigerenzer, 1991). In other words, when participants were not certain whether the pictures and the words were semantically related (e.g., "honey" and "bear"), they tended to respond that they were unrelated. By contrast, results for RTs conform to the literature (i.e., faster RTs for semantically related than for unrelated words; Meyer & Schvaneveldt, 1971). Although the presence of a speed–accuracy trade-off limits the interpretation of this priming effect, faster RTs for semantically related words are indicative that new pictures that had not been seen before in the experiment did activate the representations of semantically related newly learned words.

Turning to the influence of music training, musicians and nonmusician controls performed similarly in the matching task but musicians outperformed controls in the semantic task. This contrastive pattern of results suggests that the two tasks tap into different memory systems. To decide whether the newly learned word matched the picture in the matching task, participants had to retrieve the specific picture–word associations that were stored in episodic memory during the word learning phase. By contrast, in the semantic task participants had to retrieve general information from semantic memory because the novel pictures that were presented before the newly learned words had not been seen before in the experiment. In line with the centroparietal shift in N400 distribution observed at the end of the word learning phase, the finding that musicians outperformed nonmusicians in the semantic task is taken as evidence that musicians had already integrated the novel words' meanings into semantic memory so that priming effects generalized to new pictures.

ERPs in the matching and semantic tasks also add support to this interpretation. In musicians and for both tasks, the N400 over centroparietal sites was larger for unexpected (mismatch/unrelated) than for expected words (match/related; Figures 5 and 6). This sensitivity to word characteristics and this scalp distribution correspond to the N400 component, typically considered as the electrophysiological marker of the integration of novel words' meanings into semantic memory (Borovsky et al., 2012; Batterink & Neville, 2011; Mestres-Missé et al., 2007) and "as reflecting the activity in a multimodal long-term memory system that is induced by a given input stimulus during a delimited time window as meaning is dynamically constructed" (Kutas & Federmeier, 2011, p. 22; see also Steinbeis & Koelsch, 2008, for N400s elicited by single chords incongruous with the preceding context). By contrast, in nonmusician controls, the N400 effect was reversed over frontocentral sites in both tasks, with larger N400 for expected (match/related) than for unexpected words (mismatch/unrelated). This finding was surprising based on previous results showing that the N400 is larger for unrelated than for related words in both lexical decision tasks (Borovsky et al., 2012) and semantic priming experiments (Mestres-Missé et al., 2007). Because the amount of music training was not controlled in these experiments, a possible explanation is that some of the participants had musical skills, hence influencing the results. However, stimulus and task differences are more likely to account for this discrepancy. Most experiments on ERPs and word learning in adults used similar designs with novel words embedded in sentence contexts. Here and as in experiments conducted with children (Friedrich & Friederici, 2008; Torkildsen et al., 2008) and with adults (Dobel et al., 2009), participants learned novel words through picture–word associations. It is thus possible that this learning mode was more demanding than learning the meaning of novel words from sentence contexts. Along these lines, we interpret the reversed N400 effect in nonmusicians in the matching and semantic tasks as showing that nonmusicians had not yet fully learned the picture–word associations (as reflected by the frontal distribution of the N400 in these tasks that is similar to the frontal N400 found during word learning) and that they were still building up new episodic memory traces based on the correct information provided by the matching words (as reflected by larger N400 increase from the first to the second block for match than for mismatch words; see Figure 5A). As mentioned above, these interpretations can be tested by increasing the number of trials in the learning phase as well as in the matching and semantic tasks and by comparing different word learning designs (e.g., novel words in sentence contexts vs. picture–word associations).

Finally, although the main focus of this experiment was on the modulations of N400 amplitude with novel word learning, results also revealed interesting effects on the N200 component. In both groups, the N200 components in the matching task were larger for mismatching than for matching words over parietal sites (see Figure 5). Insofar as the N200 has been associated with categorization processes (Friedrich & Friederici, 2008), mismatching words possibly required more effortful categorization processes than matching words. However, there is also evidence that the N200 component reflects early contextual influences (van den Brink, Brown, & Hagoort, 2001) and phonological processing (Connolly & Phillips, 1994). Thus, mismatching words possibly elicited larger N200 than matching words because they were unexpected based on the picture context and/or at the phonological level. In the semantic task, the increase in N200 amplitude for semantically unrelated compared with related words was only found in musicians, again suggesting that word categorization was possibly more efficient in musicians and/or that musicians were more sensitive to the context (i.e., the picture) or to the phonological characteristics of novel words than nonmusicians.

Evidence for Long-lasting Representations of Novel Words

To our knowledge, this is the first experiment comparing long-term memory (after 5 months) for novel words in musicians and controls. Clearly, results need to be considered with caution because of the high dropout rate. Nevertheless, they point to improved long-term memory in musicians compared with controls in the matching task (lower error rates) that was always performed first. The memory traces of the words that have been learned 5 months before therefore seem stronger in musicians than in nonmusicians. By contrast, no between-group differences were found in the semantic task possibly because both groups of participants similarly benefited from the reactivation of memory traces during the matching task. Taken together, these results point to long-lasting effects of rapidly established word representations during word learning, and they open new perspectives to further test for the influence of music training on long-term memory.

Evidence for Transfer Effects from Music Training to Word Learning

In summary, behavioral and electrophysiological data showed that music training improved several aspects of word learning. How can we account for the influence of music training on novel word learning?

Cascading Effects from Perception to Word Learning

The first interpretation, in terms of cascading effects, is that enhanced auditory perception and auditory attention (Strait et al., 2015) in musicians drive the facilitation observed in word learning through different subsequent steps (i.e., building up new phonological representations and attaching meaning to them, storing this new information in short- and long-term memory). In support of this interpretation, the error rate in the musicality test was correlated with the size of the N400 effect in the semantic task in musicians but not in controls (Figure 7), thereby clearly pointing to a relationship between auditory perception/attention and word learning.

Moreover, in line with the hypothesis of improved speech discrimination in musicians when the task is most difficult (Diamond, 2013; Schön et al., 2004), musicians outperformed control participants in the phonological categorization of tonal and aspiration contrasts, but both groups performed equally well for simple voicing contrasts (/Ba/ vs. /Pa/). Thus, music training was most useful to discriminate phonemes that are contrastive in Thai but that do not belong to the French phonemic repertoire (Figure 2A). It should be noted in this respect that, although vowel length is not contrastive in French, controls nevertheless performed as well as musicians in this task, possibly because the natural difference in vowel duration was large enough (270 msec) to be easily perceived by nonmusician controls. Taken together, these findings add support to the hypothesis of transfer effects between music and speech (Asaridou & McQueen, 2013; Besson et al., 2011), with near transfer for acoustic cues common to music and speech such as frequency and with far transfer to unfamiliar linguistic cues such as aspiration.

Note also that the N100 amplitude in musicians was enhanced for the nonnative aspiration contrast (see Figure 3). Because this task was more difficult for French native speakers than the other tasks, as revealed by behavioral data, the increased N100 may reflect increased focused attention (Strait et al., 2015) and mobilization of neuronal resources in musicians. An alternative but complementary interpretation is that this increase in N100 amplitude reflected increased neural synchronicity and structural connectivity in musicians, who are typically more sensitive to the acoustic-phonetic properties of speech sounds than nonmusicians who showed similar N100 components in all tasks (Bidelman et al., 2014; Elmer et al., 2013; Chobert et al., 2012; Musacchia et al., 2007; Wong & Perrachione, 2007).

Finally, and in line with the cascade hypothesis, it may be that the very nature of the stimuli (monosyllabic Thai words) is at least partly responsible of the musician's advantage in the semantic task. Because musicians are more sensitive than nonmusicians to variations in pitch, duration, and VOT, this possibly helped them build precise phonological representations that were then more easily associated to a novel word meaning. One way to test for this interpretation would be to use a similar design but with written instead of spoken words thereby potentially reducing the musicians' acoustic-phonetic advantage.

Multidimensional Facilitation Effects

The second interpretation, that is by no means contradictory but rather complementary to the cascading interpretation, is that the multidimensional nature of music training independently improved the several functions that are necessary for word learning. There is already some evidence in the literature for enhanced working and verbal memory (George & Coch, 2011; Ho et al., 2003) and executive functions (Zuk et al., 2014; Moreno et al., 2011; Rogalsky et al., 2011; Pallesen et al., 2010) with musicianship. In line with these results, musicians showed enhanced auditory short-term memory (digit span; George & Coch, 2011). Moreover, as reported in previous experiments, we also found significant between-group differences in the ERP component related to semantic integration and memory (N400; Mestres-Missé et al., 2007; Perfetti et al., 2005; McLaughlin et al., 2004). However, in contrast to previous results, we found no evidence for increased working memory (reversed digit span; George & Coch, 2011) or for improved nonverbal intelligence (PM47) in adult musicians. Note that the assessment of general cognitive abilities was quite limited in the present experiment because of time constraints. It is thus possible that general cognitive abilities other than the ones tested here could have influenced the observed between-group differences in word learning. Future experiments will aim at including more tests targeting at elucidating different cognitive functions (selective and sustained attention, working, short-term, and long-term memory, executive functions). Finally, it could also be argued that musicians performed better than nonmusicians because they were generally more motivated (Corrigall, Schellenberg, & Misura, 2013). Although this is difficult to control for, it is important to note that participants were not informed about the aim of the experiment until the end of the session. Only that we were interested in language, music, and the brain. Thus, as they discovered task by task what the experimental session was about, it is unlikely that general motivation accounted for the present behavioral and electrophysiological effects.

Conclusions

In summary, by recording ERPs during both learning and test phases (matching and semantic tasks) of novel word learning, results revealed fast changes in brain activity after only a few minutes of exposure to picture–word associations. Moreover, these changes were stronger in musicians than in controls. Specifically, the frontoparietal shift of the N400 in the word learning phase (i.e., without motor responses) only developed in musicians, which we interpret as an electrophysiological signature of "fast mapping" (Carey, 1978). To our knowledge, this is the first report showing that music training influences semantic processing. As a future step, we plan to use a longitudinal approach with nonmusician controls trained with music so as to test for the causal influence of music training on word learning. Finally, these results also open new perspectives to further investigate the influence of music training on long-term memory for applications of music training in the domain of native and second language learning (Moreno, Lee, Janus, & Bialystok, 2015; Chobert & Besson, 2013) and for using this type of experimental design in clinical research to specify the stages of word learning that are most deficient.

Acknowledgments

We would like to thank all the participants, Chotiga Pattamadilok for registering the auditory stimuli, and Benjamin Furnari for his help with data acquisition and analyses. The work of E. D. M. Ba., M. D. I., and M. B., carried out within the Labex BLRI (ANR-11-LABX-0036), has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the program "Investissements d'Avenir" (ANR-11-IDEX-0001-02). E. D. was supported by a doctoral fellowship from the BLRI and M. Ba. by a doctoral fellowship from the French Ministry of Research and Education.

Reprint requests should be sent to Eva Dittinger, Laboratoire de Neuroscience Cognitives, Université Aix Marseille, Centre Saint Charles, 3 Place Victor Hugo, Marseille, France, 13331, or via e-mail: eva.dittinger@blri.fr.

Notes

1. Following phonetic transcription in Thai, 1 refers to low-tone, 0 to midtone, p^h to aspirated voicing, and the colon to long vowel duration.
2. Voice Onset Time (VOT) is defined as the interval between the noise burst produced at consonant release and the waveform periodicity associated with vocal cord vibrations (Lisker & Abramson, 1967) .
3. Pictures were chosen from the Snodgrass and Vanderwart (1980) pictures' set but were found on the Internet to ensure sufficient resolution and quality.

REFERENCES

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.

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience, 23,* 3181–3196.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology, 2,* 94.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience, 40,* 2662–2673.

Boddy, J., & Weinberg, H. (1981). Brain potentials, perceptual mechanism and semantic categorization. *Biological Psychology, 12,* 43–61.

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer [computer program]. Retrieved from www.praat.org.

Borovsky, A., Elman, J., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development, 8,* 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition, 116,* 289–296.

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences, 3,* 923–940.

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, 24,* 956–967.

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience, 23,* 3874–3887.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal words of spoken sentences. *Journal of Cognitive Neuroscience, 6,* 256–266.

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Frontiers in Psychology, 4,* 222.

Diamond, A. (2013). Executive functions. *Annual Review Psychology, 64,* 135–168.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: Evidence from the N400m. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 364,* 3697–3709.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35–39.

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex, 49,* 2812–2821.

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex, 22,* 650–658.

Fitzroy, A. B., & Sanders, L. D. (2013). Musical expertise modulates early processing of syntactic violations in language. *Frontiers in Psychology, 3,* 603.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex, 23,* 2038–2043.

Friederici, A. D., & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences, 19,* 329–338.

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *NeuroReport, 19,* 1757–1761.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., et al. (2002). A crosslinguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience, 14,* 1076–1087.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and nonmusicians. *Journal of Neuroscience, 23,* 9240–9245.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia, 49,* 1083–1094.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology, 2,* 83–115.

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

Ho, Y. C., Cheung, M. C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology, 17,* 439–450.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *Journal of Neuroscience, 29,* 3019–3025.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports, 1,* 1–6.

Jasper, H. H. (1958). The ten–twenty system of the International Federation. *Clinical Neurophysiology, 10,* 371–375.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage, 47,* 735–744.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7,* 302–307.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

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

Kühnis, J., Elmer, S., & Jäncke, L. (2014). Auditory evoked responses in musicians during passive vowel listening are modulated by functional connectivity between bilateral auditory-related brain regions. *Journal of Cognitive Neuroscience, 26,* 2750–2761.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203–205.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalogry and Clinical Neurophysiology, 69,* 218–233.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language ans Speech, 10,* 1–28.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). "Musical syntax" is processed in the area of Broca: An MEG-study. *Nature Neuroscience, 4,* 540–545.

Marie, C., Delogu, F., Lampis, G., Olivetti Belardinelli, M., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23,* 2701–2715.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioural and electrophysiological evidence. *Journal of Cognitive Neuroscience, 19,* 1453–1463.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience, 7,* 703–704.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex, 17,* 1858–1866.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 20,* 227–234.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22,* 1425–1433.

Moreno, S., Lee, Y., Janus, M., & Bialystok, E. (2015). Short-term second language and music training induces lasting functional brain changes in early childhood. *Child Development, 86,* 394–406.

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, 19,* 712–723.

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

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, U.S.A., 104,* 15894–15898.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., et al. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS One, 5,* 11120.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 23,* 811–814.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences, 999,* 58–75.

Peretz, I., Vuvan, D., Lagrois, M.-É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society, Series B, Biological Sciences, 370,* 20140090.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1281–1292.

Raven, J. C., Corporation, P., & Lewis, H. K. (1962). *Coloured progressive matrices: Sets A, AB, B*. London, UK: Oxford Psychologist Press.

Rodriguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & De Diego-Balaguer, R. (2009). Neuropsychological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 364,* 3711–3735.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *Journal of Neuroscience, 31,* 3843–3852.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15,* 511–514.

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,* 688–694.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41,* 341–349.

Seppänen, M., Hämäläinen, J., Pesonen, A. K., & Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Frontiers in Human Neuroscience, 6,* 43.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory, 6,* 174–215.

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS One, 3,* 2226.

Strait, D. L., Parbery-Clark, A., O'Connell, S., & Kraus, N. (2013). Biological impact of preschool music classes on processing speech in noise. *Developmental Cognitive Neuroscience, 6,* 51–60.

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience, 12,* 94–104.

Torkildsen, J. V. K., Svangstu, J. M., Friis Hansen, H., Smith, L., Simonsen, H. G., Moen, I., et al. (2008). Productive vocabulary size predicts event-related potential correlates of fast mapping in 20-month-olds. *Journal of Cognitive Neuroscience, 20,* 1266–1282.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13,* 967–985.

Wagner, A. W., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science, 281,* 1188–1190.

Wang, X., Ossher, L., & Reuter-Lorenz, P. A. (2015). Examining the relationship between skilled music training and attention. *Consciousness and Cognition, 36,* 169–179.

Wechsler, D. (2003). *Wechsler Intelligence Scale for children* (4th ed.). San Antonio, TX: The Psychological Corporation.

Woldorff, M., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology, 79,* 170–191.

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

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One, 9,* 99868.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 12 May 2017
doi: 10.3389/fnhum.2017.00233

Fast Brain Plasticity during Word Learning in Musically-Trained Children

Eva Dittinger[1,2*†], *Julie Chobert*[1†], *Johannes C. Ziegler*[3] *and Mireille Besson*[1]

[1]Laboratoire de Neurosciences Cognitives (LNC, UMR 7291), CNRS, Aix-Marseille University, Marseille, France, [2]Laboratoire Parole et Langage (LPL, UMR 7309), CNRS, Aix-Marseille University, Aix-en-Provence, France, [3]Laboratoire de Psychologie Cognitive (LPC, UMR 7290), CNRS, Aix-Marseille University, Marseille, France

Children learn new words every day and this ability requires auditory perception, phoneme discrimination, attention, associative learning and semantic memory. Based on previous results showing that some of these functions are enhanced by music training, we investigated learning of novel words through picture-word associations in musically-trained and control children (8–12 year-old) to determine whether music training would positively influence word learning. Results showed that musically-trained children outperformed controls in a learning paradigm that included picture-sound matching and semantic associations. Moreover, the differences between unexpected and expected learned words, as reflected by the N200 and N400 effects, were larger in children with music training compared to controls after only 3 min of learning the meaning of novel words. In line with previous results in adults, these findings clearly demonstrate a correlation between music training and better word learning. It is argued that these benefits reflect both bottom-up and top-down influences. The present learning paradigm might provide a useful dynamic diagnostic tool to determine which perceptive and cognitive functions are impaired in children with learning difficulties.

Keywords: word learning, musical expertise, picture-word associations, semantic memory, N400

Edited by:
Juliana Yordanova,
Institute of Neurobiology (BAS),
Bulgaria

Reviewed by:
Markus J. Hofmann,
University of Wuppertal, Germany
Viktor Müller,
Max Planck Institute for Human
Development, Germany

Correspondence:
Eva Dittinger
eva.dittinger@blri.fr

†These authors have contributed equally to this work.

Received: *21 November 2016*
Accepted: *21 April 2017*
Published: *12 May 2017*

Citation:
Dittinger E, Chobert J, Ziegler JC and Besson M (2017) Fast Brain Plasticity during Word Learning in Musically-Trained Children. Front. Hum. Neurosci. 11:233. doi: 10.3389/fnhum.2017.00233

INTRODUCTION

Learning new words is a specifically human faculty that mobilizes several perceptual and cognitive abilities: sound perception and discrimination, attention, associative learning, and semantic memory (Perfetti et al., 2005; Davis and Gaskell, 2009). Here, we investigated the temporal dynamics of word learning in a novel word learning paradigm using both behavioral measures and Event-Related Potentials (ERPs). The main question was whether musically-trained children would outperform controls in terms of novel word learning and semantic association skills and whether this would be reflected in dynamic brain plasticity measures (ERPs).

It is well known that music training improves auditory perception and attention (Kraus and Chandrasekaran, 2010). Moreover, it has been shown that music training also enhances several aspects of language processing, including phoneme and syllable perception (Musacchia et al., 2007; Chobert et al., 2011, 2014; Marie et al., 2011; Elmer et al., 2012, 2014; Parbery-Clark et al., 2012; Kühnis et al., 2013; Bidelman et al., 2014), the processing of pitch and prosody (Schön et al., 2004; Thompson et al., 2004; Delogu et al., 2006; Magne et al., 2006; Marques et al., 2007; Wong et al., 2007; Lima and Castro, 2011; Bidelman et al., 2013), phonological processing and reading (Anvari et al., 2002; Moreno et al., 2009; Corrigall and Trainor, 2011; Huss et al., 2011), speech segmentation

(François et al., 2013), and syntactic processing (Jentschke and Koelsch, 2009; Gordon et al., 2015). However, it remains an open question whether music training improves associative learning and semantic memory, two of the processes that are at the heart of word learning but that are not necessarily language-specific (Markson and Bloom, 1997).

The ERP methodology is well-suited to examine the temporal dynamics of word learning and brain plasticity as reflected by modulations of ERP components. Previous research on word learning has shown that the N400, a negative-going component that typically develops between 300 ms and 600 ms post-stimulus onset (Kutas and Hillyard, 1980), increased in amplitude when meaningless items acquire meaning and then decreased with further repetitions. This effect has been demonstrated in infants (from 9–24 months old; Friedrich and Friederici, 2008; Torkildsen et al., 2009; Junge et al., 2012; Borgström et al., 2015; Friedrich et al., 2015) and in adults (see below). However, to our knowledge, it has not yet been investigated in children and the present study was intended to fill this gap by testing children between 8 and 12 years old.

The increase in N400 amplitude with word learning is typically very fast. This effect has been observed in native adult English-speakers after 14 h of learning the meaning of novel French words (McLaughlin et al., 2004), after less than 1 or 2 h of learning novel word-picture associations (Dobel et al., 2009, 2010), after 45 min of learning the meaning of rare words (e.g., "clowder"; Perfetti et al., 2005) and even after a single exposure if a novel word or pseudoword was presented in a strongly constrained and meaningful context (Mestres-Missé et al., 2007; Borovsky et al., 2010, 2012; Batterink and Neville, 2011). Moreover, this fast increase in N400 amplitude is typically largest over fronto-central brain regions (FN400), as demonstrated in language-learning tasks and it possibly reflects speech segmentation and the building-up of novel word meaning (De Diego Balaguer et al., 2007; Mestres-Missé et al., 2007), two processes that may develop in parallel (François et al., 2017). While the scalp distribution of ERP components does not necessarily reflect the activation of directly underlying brain structures, it is nevertheless interesting that fronto-central brain regions are also found to be involved in the maintenance of novel information in working- or short-term memory, in the formation of new associations (Hagoort, 2014) and/or the construction of word representations in episodic memory (Wagner et al., 1998; Rodríguez-Fornells et al., 2009). Further exposures then allow for the integration of these novel items into existing lexical networks (Dumay and Gaskell, 2007), with recent results emphasizing the role of sleep in these processes (Tamminen et al., 2010, 2013; Friedrich et al., 2015). Once these "novel" representations are stabilized, the N400 is largest over centro-parietal regions as typically found for the N400 to already known words (Kutas et al., 1988).

Recently, we conducted an experiment in young adults, professional musicians and non-musicians (Dittinger et al., 2016) to test the hypothesis that music training would positively influence novel word learning. This hypothesis builds up on the results reviewed above and on recent behavioral and brain imaging results suggesting that music and language not only involve common sensory-perceptual processes (both at sub-cortical and cortical levels, Kraus and Chandrasekaran, 2010; Asaridou and McQueen, 2013) but also attentional (Patel, 2008; Tervaniemi et al., 2009; Strait et al., 2010, 2015; Perruchet and Poulin-Charronnat, 2013) and short-term memory resources (Ho et al., 2003; George and Coch, 2011) as well as executive functions (Pallesen et al., 2010; Moreno et al., 2011; Rogalsky et al., 2011; Zuk et al., 2014) that are involved in novel word learning. Results showed that adult musicians outperformed non-musicians in the most difficult task that required participants to map newly acquired words onto semantic associates. Moreover, the shift in N400 scalp distribution from frontal sites (FN400) during meaning acquisition, to parietal sites when word meaning is stabilized (N400) was faster in musicians than in non-musicians. Interestingly, results also showed that the amplitude of both the N200 and N400 components in the semantic task was larger to unrelated than to related words in musicians but not in non-musicians. The aim of the present experiment was to determine whether these results would replicate when comparing children with and without music training. The general procedure, inspired from Wong and Perrachione (2007), and the specific hypotheses are described below (see **Figure 1**).

Children were first asked to categorize unfamiliar Consonant-Vowel syllables with respect to voicing, vowel length and aspiration. For this purpose, we used Thai monosyllabic words. Thai is a tonal and a quantitative language in which both tonal (i.e., 5 tones) and vowel length contrasts are linguistically relevant for understanding word meaning (e.g., /pa1/ low tone with a short vowel means "to find" and /pa:1/ low tone with a long vowel means "forest"; Gandour et al., 2002). We hypothesized that if music training reinforces auditory perception and attention, children with music training should make fewer errors in the phonological categorization tasks than control children especially when the task is the most difficult, that is, for the phonological contrasts that do not belong to the French phonemic repertory (e.g., aspiration; Dobel et al., 2009).

After the phonological categorization task, the same children were then asked to learn the meaning of novel words through picture-word associations. Based on the previous results reviewed above in adults and in infants, we hypothesized that both an N200 and an N400 over frontal regions (FN400) would develop during word learning in all children but that this effect would develop faster in children with music training as was recently shown in adult musicians compared to non-musicians.

Following picture-word learning, children were then tested for the efficiency of learning using two tasks: a matching task in which they decided whether the picture-word associations matched or mismatched those seen in the word learning phase and a semantic task in which new pictures were presented that were semantically related or unrelated to the newly-learned words. Based on previous results (Meyer and Schvaneveldt, 1971), we expected mismatching and semantically unrelated words to be associated with higher error rates (ERRs) and/or slower Reaction Times (RTs) than matching

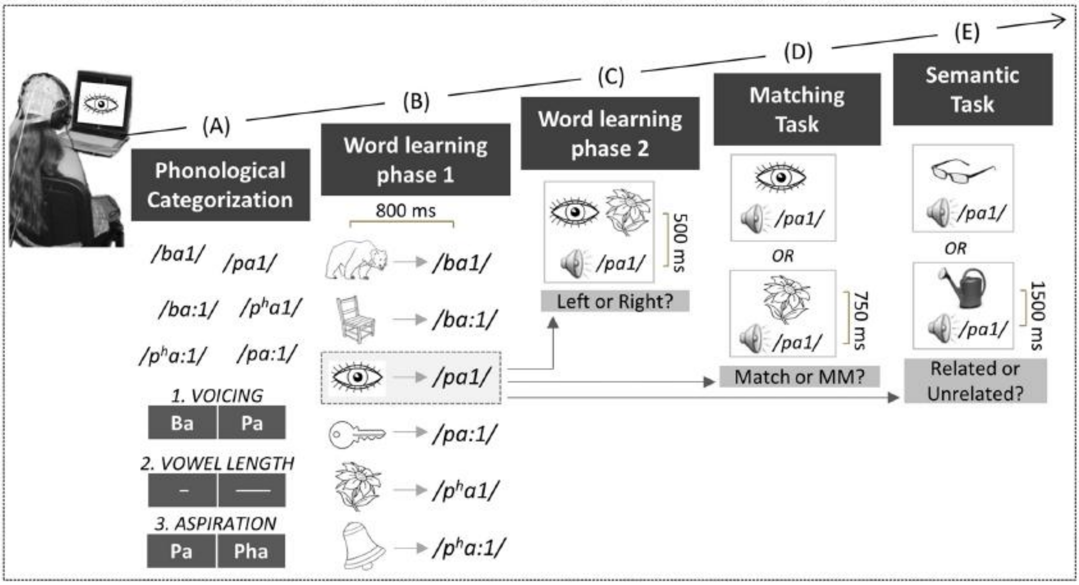


FIGURE 1 | Experimental design. Children performed a series of tasks: First, in the phonological categorization tasks **(A)**, six natural Thai mono-syllabic words had to be categorized based on voicing (Task 1), vowel length (Task 2), or aspiration contrasts (Task 3). Second, in the word learning phase 1 **(B)**, each word was paired with its respective picture. Third, in the word learning phase 2 **(C)**, two of the pictures were presented simultaneously on the screen, together with an auditory word that matched one of the two pictures. Fourth, in the matching task **(D)**, the auditory words were presented with one of the pictures, either matching or mismatching the previously learned associations. Fifth, in the semantic task **(E)**, the auditory words were presented with novel pictures that were either semantically related or unrelated to the words.

and related words. Such typical semantic priming effects were predicted in all children, showing that they all learned the picture-word associations and that learning generalized to new pictures. Moreover, we expected these behavioral effects to be accompanied by larger N400s in all children to mismatching and semantically unrelated words than to matching/related words over parietal brain regions, as typically reported for the N400 to already known words in children (Holcomb et al., 1992; Juottonen et al., 1996; Hahne et al., 2004) and in adults (Kutas and Federmeier, 2011, for review). Finally, based on our previous results with adults (Dittinger et al., 2016), we predicted lower error rates and larger N200 and N400 effects (the difference between mismatching/semantically unrelated words and matching/related words) in children with music training than in controls.

MATERIALS AND METHODS

Participants

A total of 32 children, native speakers of French without known hearing or neurological deficits participated in this experiment with 16 children that were involved in extra-scholar music training (MUS) and 16 children not involved in music training (NM), except for obligatory music lessons at school. However, participants in the NM group participated at least in one extra-scholar activity that was not related to music (e.g., sports, painting, dance), which suggests that both groups of children benefitted from stimulating extra-scholar environments.

Three children (2 with and 1 without music training) were excluded based on their level of performance in the matching and semantic tasks (i.e., percentage of error ± 2 standard deviations away from the mean) and six children (2 with and 4 without music training) because of too many artifacts in the electrophysiological data, leading to an attrition rate of 28% which is not uncommon in ERP studies with children (De Boer et al., 2005). The final group of musically-trained children (MUS) comprised six boys and six girls with three left-handers (mean age = 134.0 months, SD = 13.5) and the group of control children (NM) six boys and five girls with one left-hander (mean age = 124.5 months, SD = 20.0; $F_{(1,21)} = 1.84$, $p = 0.19$).

In the music group, children practiced music for an average of 4.9 years (4–7 years; 5 children played the piano, 2 the trumpet, 2 the trombone, 2 the violin, and 1 the saxophone). None of the children was bilingual and all children had similar socio-economic background ranging from middle to low social class as determined from the parents' profession and according to the criteria of the National Institute of Statistics and Economic Studies (MUS: 4.4 and NM: 4.3). The protocol was approved by the local Ethical Review Committee of Aix-Marseille University, and the study was conducted in accordance with local norms and guidelines for the protection of human subjects. All children agreed to participate in the experiment once the procedure had been explained to them. Children were also told that they could stop the experiment at anytime if they felt uncomfortable (none did). Finally, at least one parent accompanied each child to the laboratory and signed an informed consent form in accordance with the Declaration of Helsinki before the experiment. Children were given presents at the end of the session to thank them for their participation. The experiment lasted for 2.5 h, which included the pose of the electrocap.

Screening Measures

Cognitive and Language Abilities

Standardized psychometric tests were used to examine language-related abilities, phonological awareness (phoneme fusion), reading abilities (regular words, irregular words, pseudo-words; ODEDYS, Jacquier-Roux et al., 2005) as well as short-term memory (forward Digit Span, WISC-IV, Wechsler, 2003), visual attention (NEPSY, Korkman et al., 1998), and non-verbal intelligence (progressive matrices, PM47, Raven, 1976).

Musical Aptitude

Children performed two musicality tests (adapted from the MBEA battery; Peretz et al., 2003) consisting in judging whether pairs of piano melodies were same or different, based either on melodic or on rhythmic information.

Experimental Stimuli

Auditory Stimuli

Six natural Thai monosyllabic words were selected for the experiment: /ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/, /p^{h}a:1/. These words varied in vowel length, with short (/ba1/, /pa1/ and /p^{h}a1/; 261 ms on average) and long vowels (/ba:1/, /pa:1/ and /p^{h}a:1/; 567 ms on average), in Voice Onset Time (VOT; /ba1/ and /ba:1/, VOT = −140 ms vs. /pa1/ and /pa:1/, VOT = 5 ms), and in aspiration (/pa1/ and /pa:1/, VOT = 5 ms vs. /p^{h}a1/ and /p^{h}a:1/, VOT = 77 ms). Stimuli were recorded by a female Thai-French bilingual, ensuring that all words were produced naturally. For each word, five versions were digitally recorded in order to reproduce natural speech variability. Fundamental frequency was similar for each word (F0: 175 Hz) and sound pressure level was normalized across all words to a mean level of 70 dB using *Praat* software (Boersma and Weenink, 2011).

Visual Stimuli

For the learning phase, six pictures representing familiar objects (i.e., bear, flower, key, chair, bell, eye) were selected based on the standardized set of 260 pictures (that are matched for name and image agreement, familiarity and visual complexity) built by Snodgrass and Vanderwart (1980). The same pictures as in the learning phase were presented in the matching task. For the semantic task, 36 new pictures that the children had not seen before in the experiment and that were semantically related or unrelated to the meaning of the newly-learned words were chosen from the internet by two of the authors (JC and MB). Semantic relatedness between new and old pictures (that is, those previously presented during the word learning phase and those presented in the semantic task) was confirmed by results of pre-tests with pilot children.

Experimental Tasks

Children were tested individually in a quiet experimental room while they sat in a comfortable chair at about 1 m from a computer screen. Auditory stimuli were presented through HiFi headphones (Sennheiser, HD590). Visual and auditory stimuli presentation, as well as the collection of behavioral data, were controlled by the “Presentation” software (NeuroBehavioral Systems, Version 11.0). Children performed six concatenated tasks (see **Figure 1**).

Phonological Categorization Task

Children performed three different phonological categorization tasks that lasted for 2.5 min each (see **Figure 1A**). All six Thai monosyllabic words were presented in each task using a 2500 ms Stimulus-Onset-Asynchrony (SOA). Children were asked to categorize them based upon different features in each task: (1) voicing contrast (e.g., /ba1/ vs. /pa1/); (2) vowel length (e.g., short: /ba1/ vs. long /ba:1/); and (3) aspiration contrast (e.g., /pa1/ vs. /p^{h}a1/). For each task, the contrast was visually represented on the screen (see **Figure 1A**). Response side and task order were counterbalanced across children. Children were asked to press as quickly and as accurately as possible the left or right hand response button according to whether the auditory words matched the visual representation on the left or right side of the screen. Each monosyllabic word was presented 10 times in a pseudo-randomized order with the constraints of no immediate repetition of the same word, and no more than four successive same responses.

Word Learning Phase 1

Children were asked to learn the meaning of each word previously presented in the phonological categorization task through picture-word associations. No behavioral response was required, but children were asked to remember the words for subsequent tests. The picture was presented first, and then followed after 800 ms by one of the six words. For instance, a drawing of an eye was followed by the auditory presentation of the word /pa1/ and thus /pa1/ was the word for eye in our “foreign” language (see **Figure 1B**). Two different lists were built so that across children different pictures were associated with different words. Total trial duration was 2800 ms. Each of the six picture-word pairs was presented 20 times, resulting in 120 trials that were pseudo-randomly presented (i.e., no immediate repetition of the same association) in two blocks of 3 min each. To closely follow the brain dynamics involved in word learning, ERPs in each block were further divided into two sub-blocks for a total of four sub-blocks (i.e., Block 1: trials 1–30; Block 2: trials 31–60; Block 3: trials 61–90 and Block 4: trials 91–120).

Word Learning Phase 2

To consolidate learning, children performed a task in which two different pictures were simultaneously presented on the left and right sides of the screen, followed after 500 ms by one of the six words (see **Figure 1C**). Children were asked to press the left response key if the word matched the picture on the left side of the screen or the right key if the word matched the right-side picture (half of the stimuli in each condition). Visual feedback regarding response correctness was given, followed by the presentation of the correct picture-word pair to strengthen the association. Total trial duration was 6000 ms. Each of the six picture-word pairs was presented 10 times, resulting in 60 trials that were pseudo-randomly presented (i.e., no immediate repetition of the same association and no more than four successive same responses), within two blocks of 3 min each. Behavioral data were analyzed but not ERPs because the procedure was complex and comprised too many events.

Matching Task

One of the six pictures was presented, followed after a 750 ms delay by an auditory word that matched or mismatched the associations previously learned. For instance, while the drawing of an eye followed by /pa1/ was a match, the drawing of a flower followed by /pa1/ was a mismatch (see **Figure 1D**). Children gave their responses by pressing one out of two response keys as quickly and as accurately as possible. Some examples were given before starting the task. Response hand was counter-balanced across children. At the end of the trial, a row of X's appeared on the screen. Children were asked to blink during this time period (1500 ms; total trial duration: 4750 ms) in order to minimize eye movement artifacts during word presentation. Each word was presented 20 times, half in the match and half in the mismatch conditions. The total of 120 trials was pseudo-randomly presented (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 5 min each.

Semantic Task

One of the new pictures was presented, followed after 1500 ms by a semantically related or unrelated word. For instance, while the picture of glasses was semantically related to the previously learned word /pa1/ (i.e., "eye"), the picture of a watering can was semantically unrelated to /pa1/ (see **Figure 1E**). Children were asked to decide as quickly and as accurately as possible if the auditory word was semantically related to the new picture. Responses were given by pressing one of two response keys. Response hand was counter-balanced across participants and some examples were given before starting the task. At the end of the trial a row of X's appeared on the screen, and children were asked to blink during this time period (1500 ms; total trial duration: 7000 ms). Each word was presented 12 times but none of the new pictures were repeated, so that on each trial the words were always associated with a different related or unrelated picture. Half of the picture-word pairs were semantically related and half were semantically unrelated. A total of 72 trials was presented pseudo-randomly (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 4.2 min each.

EEG Data Acquisition

The Electroencephalogram (EEG) was continuously recorded at a sampling rate of 512 Hz with a band-pass filter of 0–102.4 Hz by using a Biosemi amplifier system (Amsterdam, BioSemi Active 2) with 32 active Ag-Cl electrodes (Biosemi Pintype) located at standard positions according to the International 10/20 System (Jasper, 1958). EEG recordings were referenced on-line to a common electrode (CMS) included in the headcap (next to Cz). Two additional electrodes were placed on the left and right mastoids and data were re-referenced off-line to the average activity of the left and right mastoids, filtered with a bandpass filter from 0.1–40 Hz (slope of 12 dB/oct). The electro-oculogram (EOG) was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Electrode impedance was kept below 5 kΩ. EEG data were analyzed using the Brain Vision Analyzer software (Version 1.05.0005; Brain Products, Gmbh). Independent component analysis (ICA) and inverse ICA were used to identify and remove components associated with vertical and horizontal ocular movements. Finally, baseline correction, DC-detrend and removal of artifacts above a gradient criterion of 10 μV/ms or a max-min criteria of 100 μV over the entire epoch were applied resulting in an average of 12% of rejected trials. For each child, ERPs were time-locked to word onset, and segmented (including a 200 ms baseline) into 1200 ms epochs in the phonological categorization tasks and into 1700 ms epochs in the other tasks (i.e., word learning phase 1, matching and semantic tasks). To increase the signal to noise ratio, all responses were considered to compute the individual averages.

Statistical Analysis

Analysis of Variance (ANOVAs) were computed using the Statistica software (Version 12.0, StatSoft Inc., Tulsa). For errors (ERRs) and RTs in each task, ANOVAs included Group (MUS vs. NM) as a between-subject factor. Additional within-subject factors were Task (voicing vs. vowel length vs. aspiration) in the phonological categorization task and Condition (match vs. mismatch or related vs. unrelated, respectively) in the matching and semantic tasks.

ERPs in the phonological categorization task were analyzed by computing N100 maximum amplitudes in the 90–160 ms latency band. During the word learning phase 1, as well as in the matching and semantic tasks, ERPs were analyzed by computing the mean amplitudes of the N200 and N400 components in specific latency bands defined from visual inspection of the traces and from previous results in the literature ANOVAs included Group (MUS vs. NM) as a between-subject factor, Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; right: F4, C4, P4), and Anterior/Posterior (frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4) as within-subject factors. Additional within-subjects factors were Task (voicing vs. vowel length vs. aspiration) in the phonological categorization task, Block (1 vs. 2 vs. 3 vs. 4) in the word learning phase 1, and Condition (match vs. mismatch or related vs. unrelated, respectively) in the matching and semantic tasks. *Post hoc* Tukey tests (reducing the probability of Type I errors) were used to determine the origin of significant main effects and interactions. Finally, to examine the relationship between musical aptitude (i.e., ERRs in the musicality tasks) and word learning (i.e., ERRs in the semantic task), a linear regression model was fitted, with level of word learning as the dependent variable and level of musical aptitude as the predictor.

RESULTS

Screening Measures

Cognitive and Language Abilities

Results of univariate ANOVAs showed no significant group differences in auditory short-term memory ($F_{(1,21)}$ = 0.03, p = 0.88, η^2 = 0.001), visual attention ($F_{(1,21)}$ = 0.22, p = 0.65, η^2 = 0.01), word reading ($F_{(1,21)}$ = 2.42, p = 0.13, η^2 = 0.10), phoneme fusion ($F_{(1,21)}$ = 0.65, p = 0.43, η^2 = 0.03), or nonverbal intelligence (PM47, $F_{(1,21)}$ = 3.71, p = 0.07, η^2 = 0.15).

Musical Aptitude

Results of two-way ANOVAs (i.e., 2 Groups × 2 Tasks) showed that MUS made fewer errors (19.3%) than NM (30.2%; main effect of Group: $F_{(1,21)}$ = 5.76, p = 0.03). In addition, all children performed better on the rhythmic (18.6%) than the melodic task (31.0%; main effect of Task: $F_{(1,21)}$ = 15.44, p < 0.001) with no Group by Task interaction ($F_{(1,21)}$ = 0.01, p = 0.93).

Experimental Tasks

Phonological Categorization

Behavioral data

Results of two-way ANOVAs [i.e., 2 Groups × 3 Tasks (voicing vs. vowel length vs. aspiration)] showed that MUS (9.7%) made overall fewer errors than NM (17.2%; main effect of Group: $F_{(1,21)}$ = 7.17, p = 0.01; see **Figure 2A**). Furthermore, across children, the error rate was lower in the voicing and vowel length tasks (9.3% and 10.3%, respectively) than in the aspiration task (20.7%; Tukey, both ps < 0.001; main effect of Task: $F_{(2,42)}$ = 11.54, p < 0.001; no Group by Task interaction).

Electrophysiological data

The N100 component (90–160 ms) was evaluated by means of a 2 × 3 × 3 × 3 ANOVA (i.e., 2 Groups × 3 Tasks × 3 Laterality positions (left vs. midline vs. right) × 3 Anterior/Posterior positions [frontal vs. central vs. parietal]). As can be seen on **Figure 3**, the N100 amplitude was larger in the aspiration task (−6.01 μV, SD = 1.96) compared to the other tasks (−4.42 μV, SD = 1.54; Tukey, p = 0.03; main effect of Task: $F_{(2,42)}$ = 4.01, p = 0.03). No significant main effect of Group ($F_{(1,21)}$ = 1.02, p = 0.32) or interaction including the Group factor was obtained (Group × Task × Anterior/Posterior interaction: $F_{(4,84)}$ = 1.87, p = 0.12).

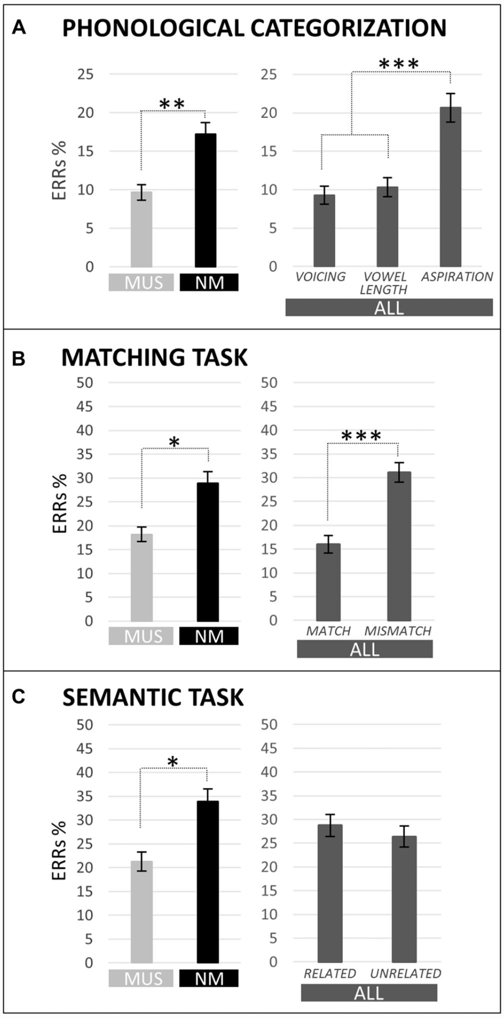


FIGURE 2 | Percentages of error (ERRs) are compared for children with music training (MUS) in light gray and for children without music training (NM) in black in the different tasks (main effect of group). Moreover, ERRs in the different tasks are also shown averaged across children (ALL; main effect of condition) in dark gray. For the phonological categorization tasks **(A)**, results are illustrated for voicing, vowel length and aspiration. For the matching task **(B)**, results are illustrated for Match and Mismatch words. For the semantic task **(C)**, results are illustrated for semantically Related and Unrelated words. Level of significance is indicated by stars with *p < 0.05, **p < 0.01 and ***p < 0.001.

Word Learning Phase 1

Electrophysiological data

The N400 (330–860 ms) as well as the N200 (200–330 ms) were evaluated by means of 2 × 4 × 3 × 3 ANOVAs (i.e., 2 Groups × 4 Blocks [1 vs. 2 vs. 3 vs. 4] × 3 Laterality × 3

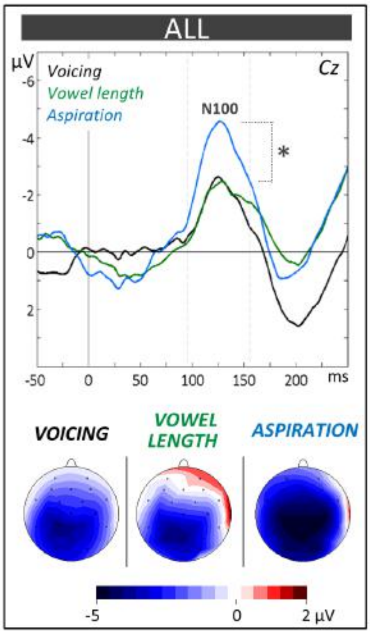


FIGURE 3 | Phonological categorization tasks. N100 components at the Central (Cz) electrode and topographic voltage distribution maps are compared between tasks averaged across children (ALL). Voltage values of topographic maps are scaled from –5 μV to +2 μV. Mean number of trials contributing to the averages (*n* = 1228). In this and subsequent ERPs figures, time in milliseconds is in abscissa and the amplitude of the effects in microvolt is in ordinate. Time zero corresponds to word onset and negativity is plotted upwards. Latency windows for statistical analyses are indicated with gray dotted lines and the level of significance is indicated by stars with $^*p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$.

Anterior/Posterior positions). For all children and in line with previous results, the N400 component was larger over frontal (−4.89 μV, SD = 1.52) and central (−4.62 μV, SD = 1.23) sites compared to parietal sites (−1.87 μV, SD = 1.32; Tukey, both *p*s < 0.001; main effect of Anterior/Posterior: $F_{(2,42)} = 33.59$, $p < 0.001$). In addition, while the Group × Block interaction was only marginally significant ($F_{(3,63)} = 2.17$, $p = 0.10$), both the increase in N400 from Block 1 to Block 2 and the decrease from Block 2 to Block 3 were significant for MUS (Block 1: −2.79 μV, SD = 1.66; Block 2: −5.46 μV, SD = 2.10; Block 3: −2.23 μV, SD = 2.04; Tukey, 1 vs. 2: $p = 0.02$, and 2 vs. 3: $p = 0.005$; main effect of Block: $F_{(3,33)} = 5.34$, $p = 0.004$) but not for NM (main effect of Block: $F_{(3,30)} = 1.13$, $p = 0.35$; see **Figures 4A,B** for topographic distributions of effects). Similarly, for the N200, the increase from Block 1 to Block 2 was significant in MUS but not in NM (MUS: Block 1: −1.64 μV, SD = 2.35; Block 2: −4.68 μV, SD = 1.98; Tukey, 1 vs. 2: $p = 0.01$; main effect of Block: $F_{(3,33)} = 3.82$, $p = 0.02$; and NM: main effect of Block: $F_{(3,30)} = 0.19$, $p = 0.90$).

Word Learning Phase 2

Behavioral data

No significant group differences were found on ERRs (MUS: 16.0% and NM: 20.2%; main effect of Group: $F_{(1,21)} = 1.12$, $p = 0.30$).

Matching Task

Behavioral data

Results of two-way ANOVAs [i.e., 2 Groups × 2 Conditions (match vs. mismatch)] showed that MUS (18.2%) made significantly fewer errors than NM (28.9%; main effect of Group: $F_{(1,21)} = 5.54$, $p = 0.03$; see **Figure 2B**) and that all children made fewer errors for match (16.0%) than for mismatch words (31.1%; main effect of Condition: $F_{(1,21)} = 22.66$, $p < 0.001$). The Group by Condition interaction was not significant ($F_{(1,21)} = 1.82$, $p = 0.19$). No significant between-group differences and no Group by Condition interaction were found on RTs ($F_{(1,21)} = 0.13$, $p = 0.72$, and $F_{(1,21)} = 0.32$, $p = 0.58$, respectively) but, in line with error rates, RTs were faster for match (1284 ms) than for mismatch words (1461 ms; main effect of Condition: $F_{(1,21)} = 35.13$, $p < 0.001$).

Electrophysiological data

The N400 (300–550 ms) and the N200 (200–300 ms) components were evaluated by means of 2 × 2 × 3 × 3 ANOVAs (2 Groups × 2 Conditions × 3 Laterality × 3 Anterior/Posterior positions). Analysis of the N400 revealed a significant Group × Laterality × Anterior/Posterior interaction ($F_{(4,84)} = 2.44$, $p = 0.05$). Results of separate group analyses showed larger N400 amplitude in MUS for mismatch (−4.15 μV, SD = 1.95) compared to match words across all scalp sites (−2.75 μV, SD = 2.19; main effect of Condition: $F_{(1,11)} = 5.08$, $p = 0.05$; see **Figures 5A,B**). In NM, the N400 effect was localized over central electrodes (mismatch: −5.81 μV, SD = 1.15 and match words: −4.62 μV, SD = 0.89) with slightly larger differences over the right than left hemisphere (Condition × Laterality × Anterior/Posterior interaction: $F_{(4,40)} = 2.79$, $p = 0.04$; see **Figures 5A,B**). Finally, analyses of the N200 showed larger amplitude in MUS compared to NM over the left hemisphere (MUS: −3.73 μV, SD = 1.50 and NM: −1.57 μV, SD = 1.57; Group × Laterality: $F_{(2,42)} = 3.20$, $p = 0.05$) but no interaction involving the factor Condition.

Semantic Task

Behavioral data

Results of two-way ANOVAs [i.e., 2 Groups × 2 Conditions (related vs. unrelated)] showed that MUS (23.1%) made significantly fewer errors than NM (33.9%; main effect of Group: $F_{(1,21)} = 4.50$, $p = 0.05$; see **Figure 2C**) and that all children made as many errors to related (29.7%) than to unrelated words (27.2%; main effect of Condition: $F_{(1,21)} = 0.39$, $p = 0.54$). The Group by Condition interaction was not significant on ERRs ($F_{(1,21)} = 0.20$, $p = 0.66$). RTs were faster to semantically related (1996 ms) than to unrelated words (2296 ms; main effect of Condition: $F_{(1,21)} = 58.42$, $p < 0.001$) with no significant between-group differences and no Group by Condition interaction ($F_{(1,21)} = 0.29$, $p = 0.60$, and $F_{(1,21)} = 0.32$, $p = 0.58$, respectively).

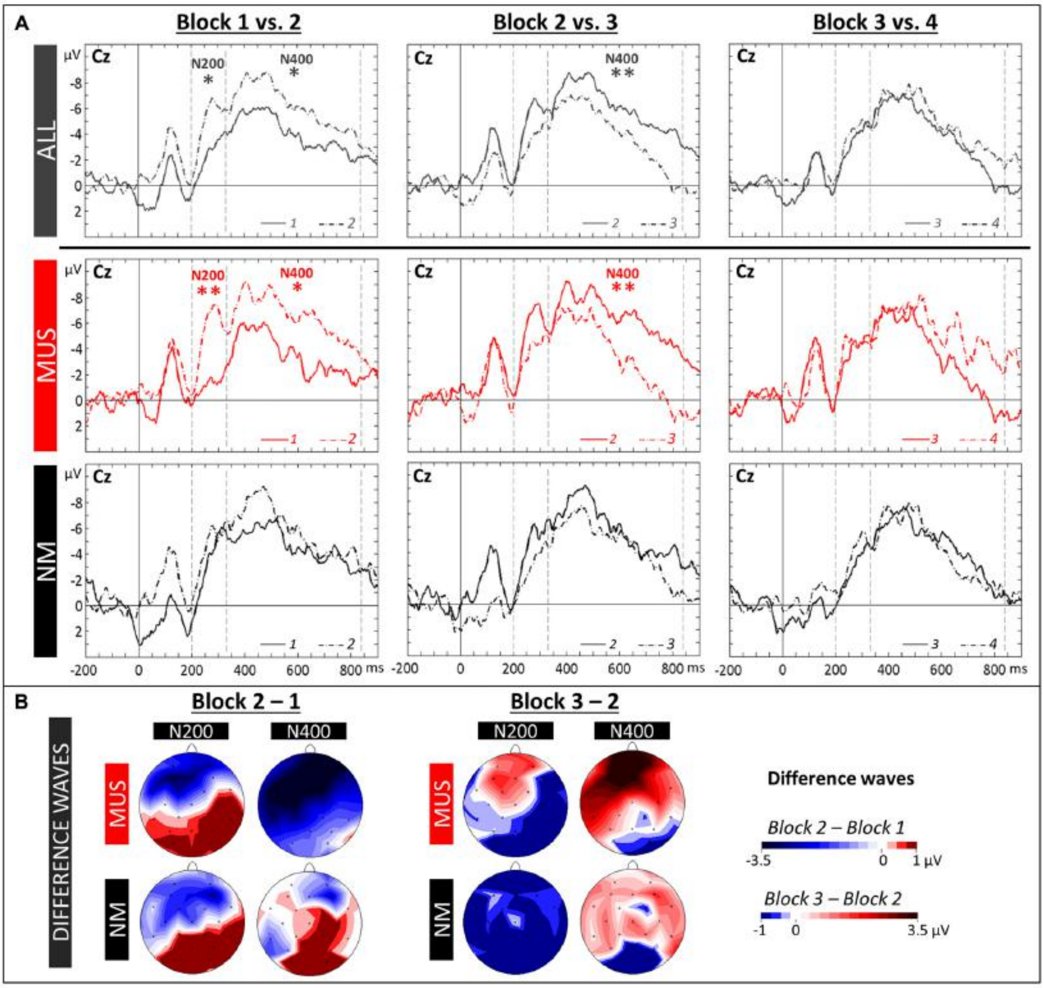


FIGURE 4 | Word learning phase 1. (A) ERPs recorded at the Central (Cz) electrode in the four Blocks are overlapped separately for Block 1 and 2 (left column), Block 2 and 3 (central column) and Block 3 and 4 (right column). Recordings are presented averaged across all children (ALL, dark gray lines) and for musician (MUS: red lines) and non-musician children (NM: black lines). Mean number of trials contributing to the averages (ALL = 607, MUS = 317, NM = 290). **(B)** Topographic voltage distribution maps of the differences between two blocks (2–1: Block 2 minus Block 1, and 3–2: Block 3 minus Block 2) are illustrated for the N200 and N400, separately for MUS and for NM. Voltage values are scaled from –3.5 μV to +1.0 μV and –1.0 μV to +3.5 μV, respectively.

Electrophysiological data

The N400 (300–640 ms) and the N200 (200–300 ms) components were evaluated by means of 2 × 2 × 3 × 3 ANOVAs (2 Groups × 2 Conditions × 3 Laterality × 3 Anterior/Posterior positions). Analysis of the N400 revealed a significant Group × Condition × Anterior/Posterior interaction ($F_{(2,42)} = 2.64$, $p = 0.05$). Results of separate analyses (see **Figure 6A** for MUS and NM, and **Figure 6B** for topographic distributions of effects) showed that for MUS, the N400 over parietal regions was larger to unrelated (−2.72 μV, SD = 1.32) than to related words (−1.35 μV, SD = 1.09; Tukey, $p = 0.03$; Condition × Anterior/Posterior interaction: $F_{(2,22)} = 3.79$, $p = 0.04$). This effect was not significant for NM (Condition × Anterior/Posterior interaction: $F_{(2,20)} = 1.07$, $p = 0.36$). However, as previously found in NM adults, the N400 effect was reversed with larger N400 for related (−5.03 μV, SD = 0.81) than for unrelated words at left central sites (−3.38 μV, SD = 0.78; Tukey, $p = 0.04$; Condition × Laterality × Anterior/Posterior interaction: $F_{(4,40)} = 2.20$, $p = 0.09$).

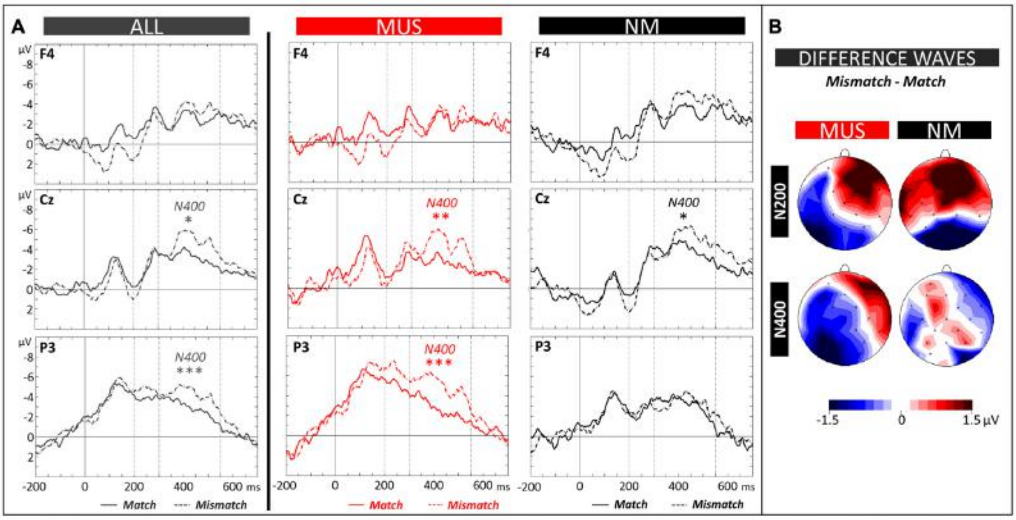


FIGURE 5 | Matching task. (A) ERPs recorded at representative electrodes (right frontal (F4), central (Cz) and left parietal (P3)) are overlapped for Match (solid lines) and Mismatch words (dotted lines), averaged across all children (ALL, dark gray lines) and for musician (MUS, red lines) and non-musician children (NM, black lines). Mean number of trials contributing to the averages (ALL = 1215, MUS = 634, NM = 581). **(B)** Topographic voltage distribution maps of the differences between conditions (Mismatch minus Match) are illustrated for the N200 and N400 and for MUS and NM. Voltage values are scaled from –1.5 μV to +1.5 μV.

Similarly, analysis of the N200 revealed significant Group × Condition and Group × Condition × Anterior/Posterior interactions ($F_{(2,21)} = 5.21$, $p = 0.03$ and $F_{(2,42)} = 5.39$, $p = 0.008$). Results of separate analysis showed that for MUS the N200 over central and parietal regions was larger to unrelated (−4.48 μV, SD = 1.37 and −4.90 μV, SD = 1.10, respectively) than to related words (−3.00 μV, SD = 1.11 and −2.57 μV, SD = 0.69; Tukey, $p = 0.05$ and < 0.001, respectively; Condition × Anterior/Posterior interaction: $F_{(2,22)} = 3.74$, $p = 0.04$) with no significant differences for NM (main effect of Condition: $F_{(1,10)} = 1.46$, $p = 0.25$, and Condition × Anterior/Posterior interaction: $F_{(2,20)} = 0.40$, $p = 0.67$).

Relationship between Musical Aptitude and Word Learning

A highly significant correlation was found between musical aptitude and word learning ($R^2 = 0.31$, $F_{(1,21)} = 10.94$, $p = 0.003$), which reflected the fact that children with fewer errors in the musical aptitude task (i.e., musically-trained children) achieved higher levels of word learning (i.e., fewer errors in the semantic task; see **Figure 7**). As screening measures showed a trend towards group differences on nonverbal intelligence (PM47, $p = 0.07$), and as age differences although not significant ($p = 0.19$) may influence word learning performance, two separate partial correlations were computed controlling for these variables. In both cases, the partial correlation between musical aptitudes and word learning remained highly significant when controlling for PM47 ($r = 0.51$, $p = 0.01$) or age ($r = 0.51$, $p = 0.02$).

DISCUSSION

This series of experiments revealed three main findings. First, ERPs recorded in the word learning phase showed that the temporal dynamics of novel word learning, as reflected by significant modulations of N200 and FN400 amplitudes after only a few minutes of picture-word associative learning, was faster in children with music training than in control children. Second, while all children were able to learn the meaning of new words, music training was associated with more efficient learning of picture-word associations as reflected by both behavioral and electrophysiological data in the matching and semantic tasks. Finally, a fronto-parietal network was involved in word learning with a shift of the distribution of the N200 and N400 components from frontal regions in the learning phase to parietal regions during the test phase (matching and semantic tasks). These findings are discussed below.

Fast Brain Plasticity in the Word Learning Phase

Recording ERPs during the four blocks of the learning phase allowed us to precisely follow the temporal dynamics of the learning process. ERPs averaged across all children clearly showed that large changes in brain activity occurred very rapidly during the acquisition of word meaning (see **Figure 4A**). As hypothesized based on previous results in adults (Mestres-Missé et al., 2007; Borovsky et al., 2012; Dittinger et al., 2016; François et al., 2017) and in infants

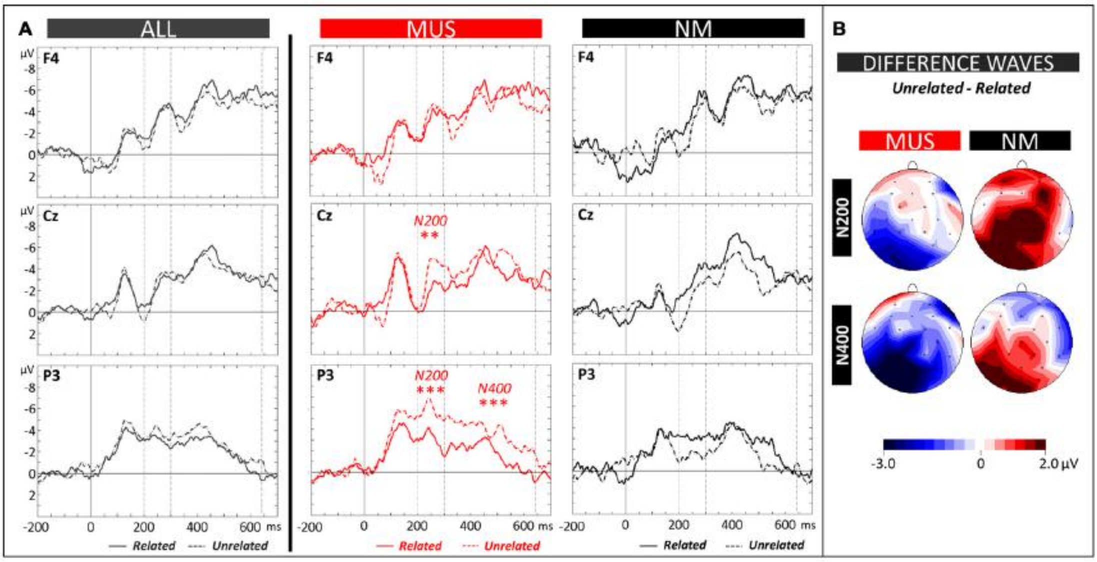


FIGURE 6 | Semantic task. (A) ERPs recorded at representative electrodes (right frontal (F4), central (Cz) and left parietal (P3)) are overlapped for Related (solid lines) and Unrelated words (dotted lines), averaged across all children (ALL, dark gray lines) and for musician (MUS, red lines) and non-musician children (NM, black lines). Mean number of trials contributing to the averages (ALL = 730, MUS = 380, NM = 350). **(B)** Topographic voltage distribution maps of the differences between conditions (Unrelated minus Related) are illustrated for the N200 and N400 and for MUS and NM. Voltage values are scaled from –3.0 μV to +2.0 μV.

(Torkildsen et al., 2006; Friedrich and Friederici, 2008), results showed an increased long-lasting negativity from Block 1 to Block 2 over fronto-central sites, comprising both an N200 and an FN400 components, taken to reflect learning of novel picture-word associations (see **Figures 4A,B**, topographic maps). It is notable that the N200 component is more clearly visible than in previous experiments, possibly due to auditory rather than visual word presentation. Moreover, the overall amplitude of the N200 and FN400 components is much larger, and the FN400 component is longer-lasting, in children than in adults (Dittinger et al., 2016; see **Figure 4**). The differences between Block 1 and Block 2 were localized over fronto-central regions. This scalp distribution is very similar to previous results in word segmentation experiments (Cunillera et al., 2009; François et al., 2014) and is compatible with previous findings suggesting that prefrontal and temporal brain regions are associated with the maintenance of novel information in working memory (Hagoort, 2014) and with the acquisition of word meaning (Rodríguez-Fornells et al., 2009). What is most remarkable is that these amplitude modulations were observed after only 3 min of learning novel word meanings (that is after only 10 repetitions of each picture-word association), thereby showing clear evidence for fast mapping (Carey, 1978) as reflected by fast changes in brain activity. Importantly, and as previously found in adults (Dittinger et al., 2016), these effects were significant in musically-trained children but not in children without music training. Thus, in line with our hypothesis, these results showed evidence for faster encoding of novel word meaning in musically-trained children. Interestingly, and strikingly similar to previous results in word segmentation experiments (Cunillera et al., 2009; François et al., 2014), FN400 amplitude was already decreased from Block 2 to Block 3 (i.e., after 3–4 min), possibly due to repetition effects (Rugg, 1985) that contribute to learning. Cunillera et al. (2009) interpret their similar findings in light of the time-dependent hypothesis (Raichle et al., 1994; Poldrack et al., 1999) following which increased activation (as reflected by the FN400) is only found during the initial learning period and quickly decreases when words have been identified, or, as in our experiment, when meaning has been attached to the auditory word-form. Finally, no differences were found between Block 3 and Block 4 possibly because all children had reached a learning threshold.

Turning to the N200 component, and in contrast to what was previously found in adult non-musicians (Dittinger et al., 2016), the differences between Block 1 and Block 2 were not significant in children with no specific music training. Insofar as the N200 reflects categorization processes (Friedrich and Friederici, 2008), it may be that these children had not yet learned to categorize the correct word with the correct picture. Alternatively, and based on recent results by Du et al. (2014) showing enhanced N200 amplitude when Chinese compound words are repeated in priming experiments, it may be that adult non-musicians were more sensitive to the repetition of

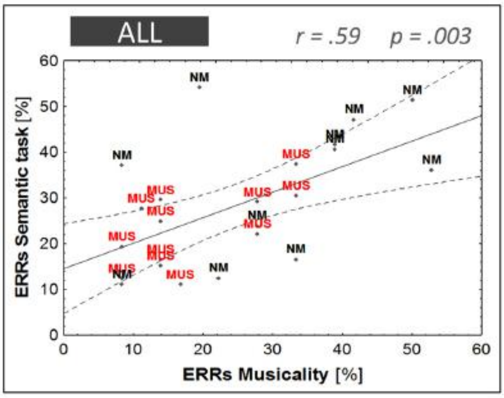


FIGURE 7 | Linear regression model. Musical aptitude (ERRs in the musicality task) against word learning performance (ERRs in the semantic task). Children with music training are illustrated in red (MUS) and children without music training in black (NM).

words in the learning phase than children with no music training.

Testing Novel Word Learning in the Matching and Semantic Tasks

All children were able to learn the six picture-word associations within a short learning phase (around 12 min total time for both word learning phases 1 and 2) as shown by the low percentage of errors in the active learning phase (<21% in both groups) and in the matching and semantic tasks (between 13% and 38% across groups). Importantly, the level of performance in both tasks and in both groups was above chance level (50%) and far from ceiling or floor effects thereby showing that the level of task difficulty was not too easy nor too difficult. In line with previous findings in the literature and with findings in adults using a similar design (Dittinger et al., 2016), results in the matching task showed clear matching effects with lower error rates and faster RTs to matching than to mismatching words thereby showing that all children had learned the picture-word associations presented in the word learning phase. Moreover, this learning effect generalized to new pictures in the semantic task, as revealed by faster RTs to auditory words semantically related to new pictures than to unrelated words (Meyer and Schvaneveldt, 1971; Dittinger et al., 2016). However, the semantic priming effect was not significant on error rates. While surprising, this finding possibly reflects a response bias towards rejection: when children were not certain whether pictures and words were semantically related (e.g., “honey” and “bear”), they tended to respond that they were unrelated. This interpretation is in line with the adult results showing that participants made significantly fewer errors for unrelated than for related words.

Finally, and perhaps most importantly, while the matching (on both errors and RTs) and semantic priming effects (on RTs) were significant in both groups (no Group by Condition interaction), musically-trained children made significantly fewer errors than controls in both the matching and the semantic tasks (main effect of Group), suggesting that they had learned the meaning of novel words more efficiently than controls. Importantly, the effect of musicianship in children was very similar to what was found in adults (Dittinger et al., 2016). In the semantic task, adult musicians outperformed adult non-musicians. Moreover, the level of performance was similar in both groups (musician children [23.1%] and adults [23.6%]; non-musician children [33.9%] and adults [30.5%] thereby showing that the level of task difficulty was similar for children [learning 6 novel words] and adults [learning 9 novel words]). In the matching task, adult musicians made fewer errors than adult non-musicians but, in contrast to children, this difference did not reach significance possibly because the matching task was too easy to reveal a between-group difference in adults.

Comparison of the electrophysiological data in the matching task between children and adults also revealed interesting differences. While children without music training showed a typical N400 effect over central electrodes (N400 larger to mismatch than match words), adults without music training showed a reversed N400 effect over frontal electrodes (see Figure 5 of Dittinger et al., 2016) that we interpreted as showing that they had not yet fully integrated the meaning of novel words into pre-existing semantic networks. Following this interpretation, non-musician children, by showing typical N400 effects, were faster in integrating the meaning of novel words than non-musician adults. However, this speculation needs further support to be convincing since adult non-musicians performed as well as adult musicians in the matching task but children with music training outperformed control children. Finally, as in the word learning phase, the N200 effect was significant in adults but not in children with no music training, again possibly because adults were more sensitive to word repetition (Du et al., 2014) than children. Maybe more interestingly, children without music training showed an N400 effect without an N200 effect in the matching task, thereby supporting the hypothesis that both components reflect independent processes (e.g., Du et al., 2014; see Hofmann and Jacobs, 2014; for a detailed discussion of this issue).

Turning to the semantic task and in line with the behavioral results showing that children with music training learned the meaning of novel words more efficiently than control children, N200 and N400 amplitudes were significantly larger for unrelated than for related words over parietal regions in children with music training but not in non-musician children (see **Figure 6**). Again, these results are very similar to previous results in adults (Dittinger et al., 2016) showing significant N200 and N400 semantic priming effects over parietal regions in adult musicians but not in non-musicians. Only the N200 effect was significant in adult non-musicians, again pointing to the independance of these two components (Du et al., 2014). By contrast, it is striking that, similar to adult non-musician results in the matching task, reversed

N400 effects (larger N400 to related than to unrelated words) were found in the semantic task, both in children without music training, over left central sites (see **Figure 6B**) and in non-musician adults, over frontal sites (see Figure 6 of Dittinger et al., 2016). Below we propose an interpretation of these surprising results that showed up in two independent samples.

Results in the word learning literature have shown that the N400 is larger for semantically unrelated than for related words in both lexical decision tasks (Borovsky et al., 2012) and semantic priming experiments (Mestres-Missé et al., 2007). This is taken as evidence that novel words are processed differently based on previously learned associations and that, with training, the meaning of novel words is rapidly integrated into semantic memory networks (Mestres-Missé et al., 2007; Batterink and Neville, 2011; Borovsky et al., 2012). Based on this interpretation, the different N400 effects for children with and without music training in the semantic task suggest that while musically-trained children had already integrated the meaning of the novel words into semantic memory, as reflected by typical N400 effects, this was not yet the case for control children (reversed N400 effects). In other words, while all children were able to retrieve the specific picture-word associations that were stored in episodic memory during the word learning phase, as reflected by typical N400 effects in the matching task, generalization of learning as seen through priming effects from new pictures semantically related to the novel words could possibly take longer for control children than for musically-trained children. In sum, differences between musically-trained and untrained participants (both children and adults, Dittinger et al., 2016) were larger when the task required retrieving general information from semantic memory in the semantic task than retrieving specific picture-word associations in the matching task.

Finally, in contrast to the frontally-distributed N400 component during the early stages of learning discussed above, the N400 effect in the test phase was clearly centro-parietally distributed. Thus, when the meaning of words was already learned, as in the matching and semantic tasks (see **Figures 5B**, **6B**), and as in typical N400 experiments with known words (Kutas et al., 1988), the N400 showed a more parietal scalp distribution that possibly reflects access to the meaning of words already stored in semantic memory or the integration of novel words meaning in existing semantic networks (Batterink and Neville, 2011). In sum, by recording ERPs both in the word learning phase and in the matching and semantic tasks from the same participants, we found a clear fronto-parietal shift in N400 scalp distribution with learning (compare **Figures 4**, **5**, **6**). Importantly, this shift in N400 distribution from the acquisition to the consolidation of novel word meaning was also found in adults (Batterink and Neville, 2011; Dittinger et al., 2016).

The Cascade and Multi-Dimensional Interpretations

We previously proposed two complementary bottom-up and top-down interpretations to account for the advantage of musician compared to non-musician adults in novel word learning (Dittinger et al., 2016). Following the "cascade" interpretation (bottom-up), increased auditory sensitivity is the driving force behind enhanced word learning in musicians. According to this view, enhanced auditory perception and attention in musicians (Kraus and Chandrasekaran, 2010; Besson et al., 2011; Strait et al., 2015) allow one to build clear and stable phonological representations (Anvari et al., 2002; Corrigall and Trainor, 2011) that are more easily discriminable and consequently easier to associate with specific meanings and to store in semantic memory. Previous reports provided clear evidence that music training improves sensitivity of auditory-related brain regions (Schneider et al., 2002; Elmer et al., 2013; Kühnis et al., 2014) and fosters the ability to focus and maintain attention on auditory stimuli (Magne et al., 2006; Moreno et al., 2009; Tervaniemi et al., 2009; Strait et al., 2010, 2015; Corrigall and Trainor, 2011).

In line with these results, the level of performance in the three phonological categorization tasks (voicing, vowel length and aspiration) was significantly higher in musically-trained children than in controls (see **Figure 2A**). This supports the hypothesis that music training is associated with clearer and more stable phonological representations. This, in turn, may facilitate the learning of new picture-word associations in the word learning phase. However, independently of music training, the N100 amplitude was largest to the unfamiliar, non-native aspiration contrast (see **Figure 3**). This result differs from previous ones in adults showing larger N100s to the aspiration contrast only in professional musicians (Dittinger et al., 2016). It may be that the differential sensitivity to familiar and unfamiliar phonetic contrasts decreases from childhood to adulthood and that music training helps to maintain this sensitivity. This interpretation needs to be further tested in future experiments.

Following the multi-dimensional interpretation, music training not only improves auditory sensitivity but also other functions that are relevant for novel word learning. For instance, there is evidence that music training enhances short-term memory (Ho et al., 2003; George and Coch, 2011) and executive functions (Pallesen et al., 2010; Moreno et al., 2011; Rogalsky et al., 2011; Zuk et al., 2014). In line with this interpretation, the present results showed that music training influenced associative learning and semantic integration as reflected by larger modulations of the N200 and N400 components in the matching and semantic tasks (McLaughlin et al., 2004; Perfetti et al., 2005; Mestres-Missé et al., 2007). Moreover, there is also evidence from at least one longitudinal intervention study that 1 year of music training is enough to enhance verbal and performance IQ as compared to drama lessons (Schellenberg, 2004). Consistent with these findings, musically-trained children in our study showed a trend for higher nonverbal IQs (as measured with the PM47, $p = 0.07$) than controls. It is thus possible that children with music training performed better in the matching and semantic tasks not because music training enhanced auditory perception or different aspects of language processing but because, in general, increased cognitive abilities improved word learning (Banai and Ahissar, 2013; Zatorre, 2013). This is a difficult issue. One could indeed try to match children's level of performance on several cognitive abilities (e.g., working and short-term memory, general intelligence). However, this might result in a selection bias against musically-trained children if superior cognitive abilities were a direct consequence of music training. Coming back to our results, it is notable that children performing higher in the musicality tests also performed higher in the most difficult semantic task (see **Figure 7**), and that this correlation remained highly significant also when controlling for the influence of nonverbal general intelligence. Thus, while music training is likely to influence high level cognitive functions that could facilitate word learning, the results found in the present experiment do not seem to be mediated by nonverbal intelligence. In sum, facilitated word learning in children with music training probably results from the strong interplay between improved auditory perception and higher cognitive functions so that the cascade and multidimensional interpretations are best considered as complementary.

CONCLUSION

Our results showed that all children were able to learn the meaning of novel words and that, similar to previous results found in adults (Batterink and Neville, 2011; Dittinger et al., 2016) word learning was associated with a fronto-parietal shift of the topographical distribution of the N400 and N200 components that developed with learning. Importantly, a few years of music training (4.5 years on average) was found to positively correlate with word learning: musically-trained children performed higher than controls in both the matching and the semantic tasks and the electrophysiological markers of word leaning (the N200 and N400 effects) were larger in children with than without music training. To our knowledge, this is the first report of a relationship between music training and the semantic aspects of language processing in children. Importantly, these results extend previous findings showing that music training enhanced phonological awareness, reading development, word comprehension and syntactic processing (Anvari et al., 2002; Jentschke and Koelsch, 2009; Corrigall and Trainor, 2011). These results also support the hypothesis that second language learning is facilitated by musical training (Slevc, 2012; Chobert and Besson, 2013; Moreno et al., 2015) and taken together they provide strong evidence for the importance of music classes in primary school.

Limitations and Perspectives

The first limitation of the present study is the small number of children in each group. Although we tested a relatively large group of 32 children with the aim of having 16 participants in each group, several children had to be discarded for technical reasons. Nevertheless, two main arguments support the robustness of our findings. First, even with a small sample size, the effects of main interest were significant (and therefore statistically valid, Friston, 2012) in musically-trained children and not significant in control children, thereby showing clear between-group differences. Second, as discussed above, the main effects found for children in the different experiments described here are remarkably similar to those previously found using a very similar paradigm with musician and non-musician adults (Dittinger et al., 2016). Thus, the correlation between music training and better novel word learning (both in behavior and ERPs) was replicated in two independent samples of participants.

The second limitation is that, while we would like to attribute the reported differences between musically-trained and untrained children to music training, the present experiment does not allow to rule out that differences other than music training accounted for the observed between-group differences. The only way to demonstrate the causal role of music training is to conduct a longitudinal study with non-musician children trained with music and to compare results with another group of non-musician children trained with an equally interesting non-musical activity. However, before conducting such longitudinal studies to ascertain the origins of the differences, it is first of primary importance to demonstrate differences between musically-trained and control children in novel word learning, and this was the aim of the present study.

Finally, the series of experiments used in this paradigm allowed us to test for auditory perception of linguistic and non-linguistic sounds (musicality tests), for auditory attention and for associative and semantic memory. Thus, an interesting perspective would also be to use this paradigm as a diagnostic tool to determine which specific computations and cognitive functions are impaired in children with learning difficulties or in patients with degenerative disorders.

AUTHOR CONTRIBUTIONS

MB, JCZ and JC designed and supervised the research; JC collected EEG data and ED analyzed the EEG data; MB and ED wrote the manuscript, and JCZ and JC contributed to the manuscript.

ACKNOWLEDGMENTS

We would like to thank the children who participated in the experiment and their parents, Eric Castet, Agustin Lage, Pedro A. Rojas-Lopez and Pedro Valdès-Sosa for very useful statistical advices, and Stefan Elmer for helpful suggestions for data analyses. JC was supported by a post-doctoral fellowship from the "Fondation de France" and ED by a doctoral fellowship from the "Brain and Language Research Institute" (BLRI). The BLRI partly supported this research (ANR-11-LABX-0036 (BLRI) and ANR-11-IDEX-0001-02 (A*MIDEX)).

REFERENCES

Anvari, S. H., Trainor, L. J., Woodside, J., and Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130. doi: 10.1016/S0022-0965(02)00124-8

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

Banai, K., and Ahissar, M. (2013). Musical experience, auditory perception and reading-related skills in children. *PLoS One* 8:e75876. doi: 10.1371/journal.pone.0075876

Batterink, L., and Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *J. Cogn. Neurosci.* 23, 3181–3196. doi: 10.1162/jocn_a_00013

Besson, M., Chobert, J., and Marie, C. (2011). Transfer of training between music and speech: common processing, attention and memory. *Front. Psychol.* 2:94. doi: 10.3389/fpsyg.2011.00094

Bidelman, G. M., Hutka, S., and 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:e60676. doi: 10.1371/journal.pone.0060676

Bidelman, G. M., Weiss, M. W., Moreno, S., and Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* 40, 2662–2673. doi: 10.1111/ejn.12627

Boersma, P., and Weenink, D. (2011). "*Praat: Doing Phonetics by Computer (computer program)*". Available online at: www.praat.org

Borgström, K., Torkildsen, J. V. K., and Lindgren, M. (2015). Substantial gains in word learning ability between 20 and 24 months: a longitudinal ERP study. *Brain Lang.* 149, 33–45. doi: 10.1016/j.bandl.2015.07.002

Borovsky, A., Elman, J. L., and Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Lang. Learn. Dev.* 8, 278–302. doi: 10.1080/15475441.2011.614893

Borovsky, A., Kutas, M., and Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition* 116, 289–296. doi: 10.1016/j.cognition.2010.05.004

Carey, S. (1978). "The child as word learner," in *Linguistic Theory and Psychological Reality*, eds M. Halle, J. Bresnan and G. A. Miller, (Cambridge, MA: The MIT Press), 264–293.

Chobert, J., and Besson, M. (2013). Musical expertise and second language learning. *Brain Sci.* 3, 923–940. doi: 10.3390/brainsci3020923

Chobert, J., François, C., Velay, J.-L., and Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* 24, 956–967. doi: 10.1093/cercor/bhs377

Chobert, J., Marie, C., François, C., Schön, D., and Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *J. Cogn. Neurosci.* 23, 3874–3887. doi: 10.1162/jocn_a_00088

Corrigall, K. A., and Trainor, L. J. (2011). Associations between length of music training and reading skills in children. *Music Percept.* 29, 147–155. doi: 10.1525/mp.2011.29.2.147

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., et al. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *Neuroimage* 48, 541–553. doi: 10.1016/j.neuroimage.2009.06.069

Davis, M. H., and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioral evidence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3773–3800. doi: 10.1098/rstb.2009.0111

De Boer, T., Scott, L. S., and Nelson, C. A. (2005). "ERPs in developmental populations," in *Event-Related Potentials: A Methods Handbook*, ed. T. Handy (Cambridge, MA: The MIT Press), 263–297.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., and Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PLoS One* 2:e1175. doi: 10.1371/journal.pone.0001175

Delogu, F., Lampis, G., and Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? *Cogn. Process.* 7, 203–207. doi: 10.1007/s10339-006-0146-7

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., and Besson, M. (2016). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *J. Cogn. Neurosci.* 28, 1584–1602. doi: 10.1162/jocn_a_00997

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., et al. (2010). New names for known things: on the association of novel word forms with existing semantic information. *J. Cogn. Neurosci.* 22, 1251–1261. doi: 10.1162/jocn.2009.21297

Dobel, C., Lagemann, L., and Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3697–3709. doi: 10.1098/rstb.2009.0158

Du, Y., Zhang, Q., and Zhang, J. X. (2014). Does N200 reflect semantic processing?—An ERP study on Chinese visual word recognition. *PLoS One* 9:e90794. doi: 10.1371/journal.pone.0090794

Dumay, N., and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychol. Sci.* 18, 35–39. doi: 10.1111/j.1467-9280.2007.01845.x

Elmer, S., Hänggi, J., Meyer, M., and Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex* 49, 2812–2821. doi: 10.1016/j.cortex.2013.03.007

Elmer, S., Klein, C., Kühnis, J., Liem, F., Meyer, M., and Jäncke, L. (2014). Music and language expertise influence the categorization of speech and musical sounds: behavioral and electrophysiological measurements. *J. Cogn. Neurosci.* 26, 2356–2369. doi: 10.1162/jocn_a_00632

Elmer, S., Meyer, M., and Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* 22, 650–658. doi: 10.1093/cercor/bhr142

François, C., Chobert, J., Besson, M., and Schon, D. (2013). Music training for the development of speech segmentation. *Cereb. Cortex* 23, 2038–2043. doi: 10.1093/cercor/bhs180

François, C., Cunillera, T., Garcia, E., Laine, M., and Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia* 98, 56–67. doi: 10.1016/j.neuropsychologia.2016.10.006

François, C., Jaillet, F., Takerkart, S., and Schön, D. (2014). Faster sound stream segmentation in musicians than in nonmusicians. *PLoS One* 9:e101340. doi: 10.1371/journal.pone.0101340

Friedrich, M., and Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants. *Neuroreport* 19, 1757–1761. doi: 10.1097/WNR.0b013e328318f014

Friedrich, M., Wilhelm, I., Born, J., and Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nat. Commun.* 6:6004. doi: 10.1038/ncomms7004

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *Neuroimage* 61, 1300–1310. doi: 10.1016/j.neuroimage.2012.04.018

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y., et al. (2002). A cross-linguistic FMRI study of spectral and temporal cues underlying phonological processing. *J. Cogn. Neurosci.* 14, 1076–1087. doi: 10.1162/089892902320474526

George, E. M., and Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* 49, 1083–1094. doi: 10.1016/j.neuropsychologia.2011.02.001

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., and Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev. Sci.* 18, 635–644. doi: 10.1111/desc.12230

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

Hahne, A., Eckstein, K., and Friederici, A. D. (2004). Brain signatures of syntactic and semantic processes during children's language development. *J. Cogn. Neurosci.* 16, 1302–1318. doi: 10.1162/0898929041920504

Ho, Y.-C., Cheung, M.-C., and Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology* 17, 439–450. doi: 10.1037/0894-4105.17.3.439

Hofmann, M. J., and Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: from behavioral to brain data. *Neurosci. Biobehav. Rev.* 46, 85–104. doi: 10.1016/j.neubiorev.2014.06.011

Holcomb, P. J., Coffey, S. A., and Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Dev. Neuropsychol.* 8, 203–241. doi: 10.1080/87565649209540525

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex* 47, 674–689. doi: 10.1016/j.cortex.2010.07.010

Jacquier-Roux, M., Valdois, S., and Zorman, M. (2005). *Outil de Dépistage des Dyslexies*. Grenoble: Cogni-Sciences.

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.

Jentschke, S., and Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage* 47, 735–744. doi: 10.1016/j.neuroimage.2009.04.090

Junge, C., Cutler, A., and Hagoort, P. (2012). Electrophysiological evidence of early word learning. *Neuropsychologia* 50, 3702–3712. doi: 10.1016/j.neuropsychologia.2012.10.012

Juottonen, K., Revonsuo, A., and Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cogn. Brain Res.* 4, 99–107. doi: 10.1016/0926-6410(96)00022-5

Korkman, M., Kirk, U., and Kemp, S. (1998). *NEPSY: A Developmental Neuropsychological Assessment*. San Antonio, TX: The Psychological Corporation.

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* 11, 599–605. doi: 10.1038/nrn2882

Kühnis, J., Elmer, S., and Jäncke, L. (2014). Auditory evoked responses in musicians during passive vowel listening are modulated by functional connectivity between bilateral auditory-related brain regions. *J. Cogn. Neurosci.* 26, 2750–2761. doi: 10.1162/jocn_a_00674

Kühnis, J., Elmer, S., Meyer, M., and Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an EEG study. *Neuropsychologia* 51, 1608–1618. doi: 10.1016/j.neuropsychologia.2013.04.007

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Kutas, M., and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205. doi: 10.1126/science.7350657

Kutas, M., Van Petten, C., and Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalogr. Clin. Neurophysiol.* 69, 218–233. doi: 10.1016/0013-4694(88)90131-9

Lima, C. F., and Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion* 11, 1021–1031. doi: 10.1037/a0024521

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

Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305. doi: 10.1162/jocn.2010.21413

Markson, L., and Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature* 385, 813–815. doi: 10.1038/385813a0

Marques, C., Moreno, S., Castro, S. L., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. *J. Cogn. Neurosci.* 19, 1453–1463. doi: 10.1162/jocn.2007.19.9.1453

McLaughlin, J., Osterhout, L., and Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* 7, 703–704. doi: 10.1038/nn1264

Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* 17, 1858–1866. doi: 10.1093/cercor/bhl094

Meyer, D. E., and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependance between retrieval operations. *J. Exp. Psychol.* 90, 227–234. doi: 10.1037/h0031564

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., and Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* 22, 1425–1433. doi: 10.1177/0956797611416999

Moreno, S., Lee, Y., Janus, M., and Bialystok, E. (2015). Short-term second language and music training induces lasting functional brain changes in early childhood. *Child Dev.* 86, 394–406. doi: 10.1111/cdev.12297

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., and Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* 19, 712–723. doi: 10.1093/cercor/bhn120

Musacchia, G., Sams, M., Skoe, E., and Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U S A* 104, 15894–15898. doi: 10.1073/pnas.0701498104

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., et al. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS One* 5:e11120. doi: 10.1371/journal.pone.0011120

Parbery-Clark, A., Tierney, A., Strait, D. L., and Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience* 219, 111–119. doi: 10.1016/j.neuroscience.2012.05.042

Patel, A. D. (2008). *Music, Language and The Brain*. Oxford: Oxford University Press.

Peretz, I., Champod, A. S., and Hyde, K. (2003). Varieties of musical disorders. the montreal battery of evaluation of amusia. *Ann. N Y Acad. Sci.* 999, 58–75. doi: 10.1196/annals.1284.006

Perfetti, C. A., Wlotko, E. W., and Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1281–1292. doi: 10.1037/0278-7393.31.6.1281

Perruchet, P., and Poulin-Charronnat, B. (2013). Challenging prior evidence for a shared syntactic processor for language and music. *Psychon. Bull. Rev.* 20, 310–317. doi: 10.3758/s13423-012-0344-5

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 10, 15–35. doi: 10.1006/nimg.1999.0441

Raichle, M. E., Fiez, J. A., Videen, T. O., MacLeod, A. M., Pardo, J. V., Fox, P. T., et al. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cereb. Cortex* 4, 8–26. doi: 10.1093/cercor/4.1.8

Raven, J. C. (1976). *Standard Progressive Matrices: Sets A, B, C, D & E*. Oxford, UK: Oxford Psychologists Press.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., and de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3711–3735. doi: 10.1098/rstb.2009.0130

Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology* 22, 642–647. doi: 10.1111/j.1469-8986.1985.tb01661.x

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* 15, 511–514. doi: 10.1111/j.0956-7976.2004.00711.x

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., and Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* 5, 688–694. doi: 10.1038/nn871

Schön, D., Magne, C., and Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349. doi: 10.1111/1469-8986.00172.x

Slevc, L. R. (2012). Language and music: sound, structure and meaning. *Wiley Interdiscip. Rev. Cogn. Sci.* 3, 483–492. doi: 10.1002/wcs.1186

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174

Strait, D. L., Kraus, N., Parbery-Clark, A., and Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29. doi: 10.1016/j.heares.2009.12.021

Strait, D. L., Slater, J., O'Connell, S., and Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Dev. Cogn. Neurosci.* 12, 94–104. doi: 10.1016/j.dcn.2015.01.001

Tamminen, J., Lambon Ralph, M. A., and Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *J. Neurosci.* 33, 15376–15381. doi: 10.1523/JNEUROSCI.5093-12.2013

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., and Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *J. Neurosci.* 30, 14356–14360. doi: 10.1523/JNEUROSCI.3028-10.2010

Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., and Friederici, A. D. (2009). Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. *Eur. J. Neurosci.* 30, 1636–1642. doi: 10.1111/j.1460-9568.2009.06955.x

Thompson, W. F., Schellenberg, E. G., and Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion* 4, 46–64. doi: 10.1037/1528-3542.4.1.46

Torkildsen, J. V. K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., et al. (2009). Brain dynamics of word familiarization in 20-month-olds: effects of productive vocabulary size. *Brain Lang.* 108, 73–88. doi: 10.1016/j.bandl.2008.09.005

Torkildsen, J. V. K., Sannerud, T., Syversen, G., Thormodsen, R., Simonsen, H. G., Moen, I., et al. (2006). Semantic organization of basic-level words in 20-month-olds: an ERP study. *J. Neurolinguistics* 19, 431–454. doi: 10.1016/j.jneuroling.2006.01.002

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281, 1188–1191. doi: 10.1126/science.281.5380.1188

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th Edn.). San Antonio, TX: The Psychological Corporation (WISC-IV).

Wong, P. C. M., and Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28, 565–585. doi: 10.1017/s0142716407070312

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science* 342, 585–589. doi: 10.1126/science.1238414

Zuk, J., Benjamin, C., Kenyon, A., and Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One* 9:e99868. doi: 10.1371/journal.pone.0099868

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Testing Relationships between Music Training and Word Learning across the Lifespan by using Cross-Sectional and Comparative Electrophysiological Approaches in Children, young Adults, and older Adults

Eva Dittinger[1,2,3], Johanna Scherer[4], Lutz Jäncke[4,5], Mireille Besson[1†] & Stefan Elmer[4†]

[1]CNRS & Aix-Marseille University, Laboratoire de Neurosciences Cognitives (LNC, UMR 7291), Marseille, France
[2]CNRS & Aix-Marseille University, Laboratoire Parole et Langage (LPL, UMR 7309), Aix-en-Provence, France
[3]Brain and Language Research Institute (BLRI), Aix-en-Provence, France
[4]Division Neuropsychology (Auditory Research Group Zurich, ARGZ), Institute of Psychology, University of Zurich, Switzerland
[5]University Research Priority Program (URRP) "Dynamic of Healthy Aging", Zurich, Switzerland

† Shared last authorship

Abbreviated title: The developmental trajectory of word learning

Declarations of interest: none

E-mail addresses:

ED: eva.dittinger@blri.fr

JS: johanna.scherer@uzh.ch

LJ: lutz.jancke@uzh.ch

MB: mireille.besson@univ-amu.fr

SE: s.elmer@psychologie.uzh.ch

Correspondence to:

Eva Dittinger, Laboratory of Cognitive Neuroscience, Centre Saint-Charles, Pole 3C, Case C, 3 place Victor Hugo, 13331, Marseille cedex 3, France.
Email: eva.dittinger@blri.fr

Stefan Elmer, Institute of Psychology, Division, Neuropsychology, University of Zurich, Binzmühlestrasse 14/25, 8050 Zurich, Switzerland.
Email: s.elmer@psychologie.uzh.ch

Date: June 27, 2018

ABSTRACT

Learning the meaning of novel words is a multifaceted cognitive task that is omnipresent in everyday life. However, until now, it is still unclear how the maturation of cognitive functions, neural ripening and degeneration influence this ability. Moreover, how cognitive stimulating activities like music influence novel word learning throughout the lifespan is still an open issue. In the present study, we used electroencephalography (EEG) in association with a cross-sectional approach in children, young adults, and older adults, and evaluated behavioral and electrophysiological markers (i.e., N200 and N400 components) of word learning in order to describe qualitative and quantitative trajectories across lifespan. Furthermore, we tested whether music training may have a beneficial influence on cognitive functioning during word learning, and whether this intervention can potentially be used for counteracting aging processes by pandering neural reserve capacity. Our results provide evidence for a facilitation of word learning from childhood to adulthood that is mirrored by lower error rates, faster reaction times, and a more efficient access to lexical-semantic representations. Importantly, the faculty of word learning is preserved in older adults who, however, seem to show some difficulties in "cognitive multitasking" situations. Across all age cohorts, music training seems to benefit word learning, and this positive influence is manifested in terms of behavioral and neural facilitation. However, the musician's advantage in older adults only is less clear-cut, compared to children and young adults. These results have important implications for education and healthy aging.

Keywords: EEG, word learning, lexical-semantic access, cognitive functions, plasticity, music training

INTRODUCTION

The aging human brain is characterized by grey (Crivello, Tzourio-Mazoyer, Tzourio, & Mazoyer, 2014; Jäncke, Mérillat, Liem, & Hänggi, 2015) and white (Bender, Völkle, & Raz, 2016; Hirsiger et al., 2016) matter changes in cortical (Crivello et al., 2014) and subcortical (Ziegler et al., 2012) brain regions, that are accompanied by an inter-individual variable decline of several cognitive functions (Shafto & Tyler, 2014), including attention (Erel & Levy, 2016; Fama & Sullivan, 2015), short-term- and working memory (Rhodes & Katz, 2017; Rieckmann, Pudas, & Nyberg, 2017), as well as episodic memory (Fjell et al., 2016). However, due to compensatory functional mechanisms that are reflected by the recruitment of auxiliary brain regions or increased functional connectivity, there is no simple relationship between aging, brain changes, and the preservation of cognitive functions (Shafto & Tyler, 2014). Furthermore, although older adults often demonstrate auditory dysfunctions (Giroud et al., 2018) that render speech processing in noisy environments particularly difficult (Coffey, Mogilever, & Zatorre, 2017), core language functions like lexical-semantic access and syntactic processing are relatively unaffected by aging (Shafto & Tyler, 2014). Nevertheless, since increased longevity is generally associated with a higher prevalence of cognitive impairments that lead to social and healthcare burden (Rechel et al., 2013), there is increased interest in the scientific community to test for the influence of different leisure activities on cognitive functioning that might counteract the detrimental effect of aging on the central nervous system (Lazarou et al., 2017; Olsen et al., 2015; Rogenmoser, Kernbach, Schlaug, & Gaser, 2018).

In the last decades, the investigation of professional musicians (MUS; Bermudez, Lerch, Evans, & Zatorre, 2009; Jäncke, 2009; Schön, Magne, & Besson, 2004) and children undergoing music training (Magne, Schön, & Besson, 2006; Zuk, Benjamin, Kenyon, & Gaab, 2014) has attracted increasing attention in the field of cognitive neuroscience. A main reason for the fast development of this specific branch of research is that professional MUS are particularly suited to be used as in vivo model for evaluating functional and structural brain changes that are mediated by a high amount of deliberate practice (Ericsson, Krampe, & Heizmann, 2007; Krampe & Ericsson, 1996). Currently, there is evidence from both cross-sectional studies in adults (Bangert & Schlaug, 2006; Elmer, Hänggi, Meyer, & Jäncke, 2013; Schön et al., 2004) and longitudinal studies in children and adolescents (Chobert, Francois, Velay, & Besson, 2014; Francois, Chobert, Besson, & Schön, 2013; Habibi et al., 2017; Moreno et al., 2009; Nan et al., 2018; Tierney, Krizman, & Kraus, 2015) indicating that music training promotes functional and structural changes in the primary and associative auditory cortex (Elmer, Meyer, & Jancke, 2012; Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005), motor- and somatosensory cortex (Bangert & Schlaug, 2006; Münte, Altenmüller, & Jäncke, 2002), ventral and dorsal part of the prefrontal cortex (Bermudez et al., 2009; Sluming, Brooks, Howard, Downes, & Roberts, 2007), as well as in inferior-parietal brain regions (Bermudez et al., 2009) and in the corpus callosum (Elmer, Hänggi, & Jäncke, 2016; Schlaug, Jäncke, Huang, & Steinmetz, 1995). Such functional and structural differences between musically trained and untrained individuals have been shown to be related to superior auditory perception (Kraus & Chandrasekaran, 2010; Seppänen, Hämäläinen, Pesonen, & Tervaniemi, 2013; Vuust, Brattico, Seppänen, Näätänen, & Tervaniemi, 2012), several aspects of speech processing (Besson, Chobert, & Marie, 2011) as well as to an optimization of cognitive functions (Dittinger, Valizadeh, Jäncke, Besson, & Elmer, 2017; Schulze & Koelsch, 2012; Zuk et al., 2014). In particular, MUS are often characterized by behavioral and functional advantages in processing acoustic cues varying in pitch, voice-onset time, duration, and prosody (Chobert et al., 2014; Elmer et al., 2013; Francois & Schön, 2011; Kühnis, Elmer, Meyer, & Jäncke, 2013; Magne et al., 2006; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011), and frequently demonstrate increased short-term memory-, working memory-, and attention functions compared to non-musicians (NM; Schulze & Koelsch, 2012; Strait, Slater, O'Connell, & Kraus, 2015; Zuk et al., 2014). Finally, it is noteworthy to mention that music training has also been proposed to decelerate the effects of brain aging (Rogenmoser et al., 2018), to postpone age-related temporal speech processing deficits (Parbery-Clark, Anderson, Hittner, & Kraus, 2012), and to have a beneficial effect on speech-in-noise perception and working memory functions (Parbery-Clark, Skoe, Lam, & Kraus, 2009; Zendel & Alain, 2012).

Recently, Dittinger and colleagues (Dittinger et al., 2016; Dittinger, Chobert, Ziegler, & Besson, 2017; Dittinger, Valizadeh, et al., 2017) proposed a novel experimental approach that enables to evaluate the influence of music training on speech-related cognitive functions in a holistic manner by using concatenated word learning tasks. Thereby, the authors compared electrophysiological markers reflecting phonological categorization (i.e., N200 component; Connolly & Phillips, 1994; Friedrich & Friederici, 2008), attentional functions (i.e., N200 component; Patel & Azzam, 2005), cognitive control (i.e., N200 component; Kropotov, Ponomarev, Pronina, & Jäncke, 2017; Kropotov, Ponomarev, Tereshchenko, Müller, & Jäncke, 2016), as well as semantic processing and verbal memory functions (i.e., N400 component;

Kutas & Federmeier, 2011) between young adult professional MUS and NM (Dittinger et al., 2016; Dittinger, Valizadeh, et al., 2017) as well as between children with and without music training (Dittinger, Chobert, et al., 2017) while the participants learned novel words. In particular, young adults and children (1) learned the meaning of novel words through picture-word associations (i.e., learning phase), (2) judged whether combinations of pictures and words matched or mismatched those previously learned (i.e., matching task), and (3) generalized the meaning of the learned words to semantically affine pictures (i.e., semantic task; see Figure 1). The results of these previous studies consistently revealed a behavioral advantage of musically-trained compared to untrained individuals in novel word learning that was accompanied by a faster development of the N200 and N400 components during learning, possibly reflecting auditory categorization, attentional functions, and verbal memory functions (Dittinger et al., 2016). Furthermore, during the matching and semantic tasks only MUS demonstrated a posterior N400 effect (i.e., difference between mismatching and matching, or semantically unrelated and related words), that is taken to reflect the successful integration of the newly learned words into lexical-semantic memory (Kutas & Federmeier, 2011). In this context, it is noteworthy to emphasize that the investigation of the neural correlates of word learning is not only advantageous for examining the intertwining of bottom-up and top-down processes that are mediated by music training, but also for testing the development of speech processing and cognitive functions across the lifespan in an ecologically valid manner. In fact, learning the meaning of novel words through picture-word associations is a multifaceted cognitive task that relies on several parallel processes (i.e., “cognitive multitasking”), including an accurate discrimination of spectral and temporal speech attributes, phonological categorization, attentional functions, associative-, working-, episodic-, as well as lexical-semantic memory (Dittinger et al., 2016).

Notwithstanding that a better understanding of the neural underpinnings underlying perceptual, linguistic, and cognitive aspects of word learning in infants and adults is a newsworthy topic (Havas, Laine, & Rodríguez Fornells, 2017; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009), currently it is still under debate whether (1) this ability changes throughout lifespan, (2) whether the preservation of lexical-semantic functions may compensate the deterioration of perceptual and cognitive functions in older adults, and (3) whether children, young adults, and older adults differ in the recruitment of phonological categorization mechanisms, working memory functions, or associative memory, episodic memory and lexical-semantic memory. This topic is rendered even more interesting by the fact that at least three factors, that are known to have an influence on word learning, namely speech perception, vocabulary knowledge, and working memory functions, are differentially affected by aging (Daneman & Green, 1986). In fact, while older adults normally have a larger vocabulary than children and adults (Ferrand, 1998) enabling them to anchor the meaning of novel words more efficiently on established lexical-semantic representations, at the same time they are characterized by hearing impairments (Giroud et al., 2018) and reduced working memory capacity (Cabeza, Anderson, Locantore, & McIntosh, 2002; Long, Shaw, Lisa, Long, & Shaw, 2000).

In the present EEG study, we evaluated (1) the influence of lifelong music training on novel word learning in a population of older adults by applying the same set of concatenated picture-word association tasks previously used by Dittinger and colleagues for examining word learning in children (Dittinger, Chobert, et al., 2017) and young adults (Dittinger et al., 2016) with and without music training. Furthermore, (2) we compared the data of older adult participants with those of the children (Dittinger, Chobert, et al., 2017) and young adults (Dittinger et al., 2016) in order to estimate the developmental trajectory of word learning across the lifespan. Finally, (3) we also evaluated putative interactions between aging, music training, and word learning by testing the influence of music training on the preservation of speech-related cognitive functions (Rogenmoser et al., 2018). According to the previous results of Dittinger and colleagues, we expected that (1) older adult musicians would outperform age-matched non-musicians and that this superiority would be reflected by lower error rates as well as by a faster development of the N200 and N400 components. Furthermore, (2) by taking into account the developmental trajectory, we predicted that N200 and N400 amplitudes would be reduced as a function of aging (Shafto & Tyler, 2014). Finally, (3) we predicted that music training would have a beneficial effect on word learning across all three cohorts.

MATERIALS & METHODS

Participants

In the present EEG study, we measured 14 older adult professional MUS and 14 NM who did not differ in age (MUS: mean age = 60.6 years, SD = 5.7; NM: mean age = 58.5 years, SD = 6.1; $F_{(1,24)}$ = .81, p = .38) or socioeconomic status (criteria of the National Institute of Statistics and Economic Studies). Due to EEG artefacts (i.e., movement and muscle artefacts), two participants had to be excluded from further analyses, resulting in 13 MUS (5 female; five pianists, two saxophonists, two clarinetists, one flautist, one violinist, one cembalo and one organ player) and 13 NM (6 female) who entered data analysis. None of the participants reported present or past neurological or audiological deficits, and all individuals had an unremarkable audiological status (i.e., all tested frequencies could be heard below a threshold of 30 dB, frequency-range of 250-8000 Hz) as revealed by pure tone audiometry (MAICO ST 20, MAICO Diagnostic, GmbH, Berlin). All older adults were right-handed (Annett, 1970) and native German speakers. Furthermore, although bilinguals were excluded, the participants spoke on average three foreign languages fluently (MUS: 3.1, NM: 3.3), but none of them had experience with Thai or another tonal language. All MUS started music training before the age of ten years (mean age of commencement = 6.6 years, SD = 1.0 years), and at the time of EEG measurements they still practiced their musical instrument for at least two hours a day (mean hours/day during the last year = 3.1 h, SD = 0.9 h), resulting in an average of 54 years of musical practice (range = 44-64 years, SD = 3.3 years). NM never regularly played a musical instrument (i.e., no formal music lessons). The local ethic committee (Kantonale Ethikkommission of the Canton of Zurich) approved the study (in accordance with the declaration of Helsinki) and written consent was obtained from all participants who were paid for participation.

A second target of the present work was to address relationships between music training and word learning from a developmental perspective. With this purpose in mind, the data of the older adult participants were additionally compared with those previously collected by Dittinger et al., in children (i.e., musically trained and untrained children; Dittinger, Chobert, et al., 2017) and young adults (i.e., professional musicians and non-musicians; Dittinger et al., 2016). The demographic data of the children, young and older adults are summarized in Table 1, whereas a more comprehensive description of the samples of children and young adults is reported elsewhere (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017).

Psychometric and autobiographical data of the older adults sample

Cognitive ability: Standardized psychometric tests showed that the older adult MUS and NM were comparable in crystalline intelligence ($F_{(1,26)} < 1$; Mehrfachwahl-Wortschatz-Intelligenztest, “MWT-B”, Lehrl, 1977), fluid intelligence ($F_{(1,26)} = 1.72$, $p = .20$; Kurztest für allgemeine Intelligenz, “KAI”; Lehrl & Fischer, 1992), verbal memory as well as verbal learning abilities ($F_{(1,26)} < 1$; Verbaler Lern- und Merkfähigkeitstest, VLMT; Helmstaedter, Lendt, & Lux, 2001).

Table 1. Overview of the demographic data of children, young and older adult participants. MUS = musicians, NM = non-musicians, SD = standard deviation.

	Children		**Young Adults**		**Older Adults**	
	12 MUS	**11 NM**	**15 MUS**	**15 NM**	**13 MUS**	**13 NM**
Mean age (years)	11.1	10.4	25.1	25.7	60.6	58.5
SD	*1.1*	*1.7*	*3.9*	*4.8*	*5.7*	*6.1*
% Females	50	45	53	53	39	46
Mother tongue	French	French	French	French	German	German
Years music training	4.9	0	17	0	54	0
SD	*1.1*	--	*4.1*	--	*3.3*	--
Age of onset music training	5.9	--	8.1	--	6.6	--
SD	*1.0*		*2.4*		*1.0*	
Musical aptitude (ERRs musicality task)	19.3	30.2	6.7	17.6	9.5	16.2

History of music training and musical aptitudes: History of music training was assessed by an in-house questionnaire (Elmer et al., 2012) that was specifically designed to collect data about the age of onset of music training, the instruments played, the number of years of music training, and the estimated number of training hours per day/week during every three-year-period of life. Furthermore, musical aptitudes were tested by asking participants to judge whether pairs of piano melodies were same or different in terms of melody or rhythm. These two musicality tests (i.e., adapted from the MBEA battery, Peretz, Champod, & Hyde, 2003) confirmed better performance in MUS than NM, and this was especially the case in the melodic condition (percentages of error: MUS: melody = 10.3 %, rhythm = 8.6 %; NM: melody = 25.6 %, rhythm = 6.8 %; Tukey, melody = .03, rhythm = .98; Group x Task interaction: $F_{(1,24)} = 5.85$, $p = .02$). Please consider that for reasons of simplification of the manuscript, the psychometric and autobiographic data of the children and young adults are not described here but can be found in the two previous studies of Dittinger et al. (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017).

Experimental stimuli for all three cohorts

Auditory stimuli: Nine natural Thai monosyllabic words were selected for the experiment (the same as those previously used by Dittinger et al., 2016). These words varied in vowel duration, with short (/ba1/, /pa1/ and /pha1/; 261 ms on average) and long vowels (/ba:1/, /pa:1/, /pha:1/, /ba:0/, /pa:0/ and /pha:0/; 531 ms on average), fundamental frequency, with low-tone (/ba1/, /pa1/, /pha1/, /ba:1/, /pa:1/ and /pha:1/; F0 = 175 Hz on average) and mid-tone vowels (/ba:0/, /pa:0/ and /pha:0/; F0 = 218 Hz on average) as well as in voicing contrasts (/ba1/, /ba:1/ and /ba:0/, Voice Onset Time (VOT) = -144 ms *versus* /pa1/, /pa:1/ and /pa:0/, VOT = 3 ms) and aspiration contrasts (/pa1/, /pa:1/ and /pa:0/, VOT = 3 ms *versus* /pha1/, /pha:1/ and /pha:0/, VOT = 77 ms). Stimuli were recorded by a female Thai-French bilingual. For each word 5 versions were recorded and presented during the experiment in order to reproduce natural speech variability. Sound pressure level was normalized across all words to a mean level of 70 dB by using the Praat software (http://www.fon.hum.uva.nl/praat/).

Visual stimuli: For the learning phase, nine pictures representing familiar objects (i.e., bear, flower, key, chair, bell, eye, strawberry, train, glass) were selected based on the standardized set of 260 pictures built by Snodgrass & Vanderwart (Snodgrass & Vanderwart, 1980), matched for name and image agreement, familiarity, and visual complexity. The same pictures as in the learning phase were also presented in the matching task. For the semantic task, 54 new pictures that the participants had not seen before in the experiment and that were semantically related or unrelated to the meaning of the newly-learned words were chosen from the internet by two of the authors (ED and MB). In a pilot experiment, a sample of university students (n = 60; age range = 19-25 years) judged the semantic relatedness of the new pictures and those presented in the learning phase. The students' ratings confirmed that half of the presented pairs were semantically related and the other half were semantically unrelated.

Experimental tasks

Participants were tested individually in a quiet experimental room (i.e., Faraday cage) while they sat in a comfortable chair at about 1 meter from a computer screen. Auditory stimuli were presented through HiFi headphones (Sennheiser, HD590) at 70dB sound pressure level. Visual and auditory stimuli presentation, as well as the collection of behavioral data, was controlled by the "Presentation" software (Neurobehavioral Systems, Albany, California, version 11.0).

Learning phase: Participants were instructed to learn the meaning of the nine words (i.e., in children we only used 6 words) by means of picture-word associations. For instance, a drawing of an eye was followed by the word /p^ha1/ and thus, /p^ha1/ was the word for eye in our "foreign" language (see Figure 1A). Each of the nine picture-word pairs was presented 20 times, resulting in 180 trials that were pseudo-randomly presented (i.e., no immediate repetition of the same association) in two blocks of 3 min each. The picture was presented first followed by one out of the nine auditory words (i.e., stimulus-onset asynchrony, SOA = 750 ms), and total trial duration was 2000 ms. In the learning phase, we used two different lists, so that across participants different pictures were associated with different words. During the learning phase, no behavioral response was required, but the participants were informed that the learning process would be evaluated in the successive tasks.

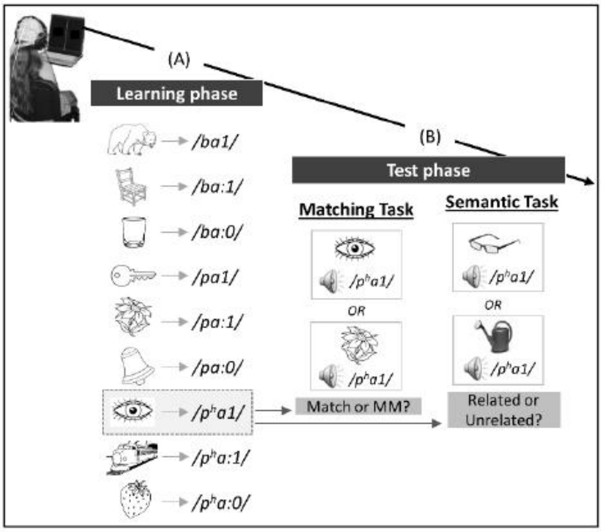


Figure 1. Experimental design. (A) Participants learned novel words through picture-word associations. (B) Afterwards, participants performed two tasks consisting of deciding whether pictures and words presented matched or mismatched the previously learned associations (matching task), and whether novel pictures were semantically related or unrelated to the learned words (semantic task).

Matching task: In the matching task, participants were exposed to one of the nine pictures that was followed, after 750 ms (SOA), by one of the words that matched or mismatched the previously learned associations. For instance, while the drawing of an eye followed by /p^{h}a1/ (i.e., eye) was a match, the drawing of a flower followed by /p^{h}a1/ was a mismatch (see Figure 1B). Participants were asked to press one out of two response keys as quickly and accurately as possible (i.e., response hand was counter-balanced across participants). Furthermore, at the end of each trial a row of "X" appeared on the screen, and participants were asked to blink during this time period (1000 ms; total trial duration: 3750 ms) in order to minimize eye movement artifacts during word presentation. Each word was presented 20 times, 10 times in match- and mismatch conditions. The total of 180 trials was pseudo-randomly presented (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 5.6 min each.

Semantic task: During the semantic task, new pictures that were semantically related or unrelated to those previously learned were used. The presentation of a new picture was followed (SOA = 1500 ms) by one of the words that was semantically related or unrelated. For instance, while the picture of glasses was semantically related to the previously learned word /p^{h}a1/ (i.e., eye), the picture of a watering can was semantically unrelated to /p^{h}a1/ (see Figure 1B). Participants were asked to press one out of two response keys as quickly and accurately as possible. A familiarization task including four trials was administrated before starting the task, and response hand was counter-balanced across participants. At the end of the trial a row of “X” appeared on the screen, and participants were asked to blink during this time period (1000 ms, total trial duration = 4500 ms). Each word was presented 12 times but none of the new pictures were repeated, so that on each trial the word was associated with a different related or unrelated picture. Half of the picture-word pairs were semantically related and half were semantically unrelated. A total of 108 trials was presented pseudo-randomly (i.e., no immediate repetition of the same association and no more than four successive same responses) within two blocks of 4 min each.

EEG data acquisition in the older adult sample

The EEG was continuously recorded at a sampling rate of 1000 Hz with a high-pass filter of 0.1 Hz by using an EEG amplifier system (Brainproducts, Munich, Germany) with 32 active Ag/Cl electrodes located at standard positions according to the international 10/20 system (Jasper, 1958). The reference electrode was placed on the nose, and electrode impedance was kept below 10 kΩ. EEG data were analyzed using the Brain Vision Analyzer software (Version 1.05.0005 & Version 2.1.0; Brain Products, GmbH). All data were filtered offline with a bandpass filter of 0.1-30 Hz (slope of 24 dB/oct), and independent component analysis (ICA) and inverse ICA were used to identify and correct vertical and horizontal ocular movements. Finally, DC-detrend (length 100 ms) and removal of artifacts above a maximum-minimum criterion of 75 µV over the entire epoch were applied automatically. For each participant, event-related potentials (ERPs) were computed time-locked to the onset of the words, segmented into 1700 ms epochs (i.e., including a 200 ms pre-stimulus baseline), and averaged for each condition. Individual averages were then used to compute the grand average across MUS, NM as well as all participants. Only trials with correct responses entered statistical analyses. Please consider that the pre-processing procedures applied in the cohorts of children and young adults are reported in the two previous studies of Dittinger et al. (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017).

Statistical analyses

In the present work, we addressed two different topics with a focus on (1) putative relationships between music training and the preservation of speech-related cognitive functions in older adult participants, and (2) on the intertwining of word learning, music training, and aging (i.e., integration of the data of the two previous Dittinger et al. studies). In order to address these two main purposes, separate statistical analyses were performed for the older adults and the lifespan perspective by using analyses of variance (ANOVAs, Statistica software version 12.0, StatSoft Inc., Tulsa). The evaluation of error rates (ERRs) and reaction times (RTs) always included Group (MUS vs. NM) as between-subject factor and Condition (i.e., match vs. mismatch in the matching task; related vs. unrelated in the semantic task) as within-subject factor. Otherwise, for the comparative analyses across cohorts (i.e., lifespan perspective), the factor Age (older adults vs. young adults vs. children) was treated as an additional between-subject factor.

The EEG analyses focused on ERPs that are sensitive to phonological categorization, attention, and cognitive control (N200) as well as to verbal memory functions and lexical-semantic access (N400; Dittinger et al., 2016; Dittinger, D'Imperio, & Besson, 2018; Kutas & Federmeier, 2011). In particular, we evaluated mean amplitudes (i.e., based on previous literature and visual inspection of the data) of the N200 and N400 components elicited during the learning phase (N200: 250-350 ms, N400: 350-550 ms), and the matching and semantic tasks (N200: 250-350 ms, N400: 350-450 ms). The ANOVAs always included Group (MUS vs. NM) as a between-subject factor and Anterior/Posterior [frontal (F3, Fz, F4) vs. central (C3, Cz, C4) vs. parietal (P3, Pz, P4)] as within-subject factor[1]. In addition, the ANOVAs included specific factors for each task. To follow the learning dynamics during the learning phase, the ANOVA included the factor Block [Block 1 (trials 1-45) vs. 2 (trials 46-90) vs. 3 (trials 91-135) vs. 4 (trials 136-180)]. By contrast, during the test phase (i.e., matching and semantic tasks), the factor Condition (match vs. mismatch or related vs. unrelated) was included in order to evaluate

---

[1] Based on previous results (Dittinger et al., 2016; Dittinger, Chobert, Ziegler, & Besson, 2017) showing no strong lateralization of the N200 and N400 effects, and in order to simplify the comparative approach, no laterality factor was considered in the EEG analyses.

whether words have been learned. Furthermore, in a similar way as for the evaluation of the behavioral data, the comparative analyses (i.e., lifespan perspective) included the factor Age (older adults vs. young adults vs. children) as an additional between-subject factor. Post-hoc Tukey tests (i.e., reducing the probability of Type I errors) were used to determine the origin of significant main effects or interactions.

RESULTS

OLDER ADULTS

Behavioral data

The evaluation of ERRs in the matching and semantic tasks showed that the main effects of Group or the interactions involving the Group factor did not reach significance (matching task: main effect of Group: $F_{(1,24)} = 2.63$, $p = .12$; Group x Condition interaction: $F_{(1,24)} = 2.85$, $p = .10$; semantic task: main effect of Group: $F_{(1,24)} = 1.77$, $p = .20$; Group x Condition interaction: $F < 1$). However, all participants made fewer errors to match (13.2 %, SD = 1.8) than to mismatch words (27.5 %, SD = 1.9; main effect of Condition: $F_{(1,24)} = 46.02$, $p < .001$), with no significant difference between related (31.3 %, SD = 2.7) and unrelated words in the semantic task (35.5 %, SD = 2.1; main effect of Condition: $F_{(1,24)} = 1.28$, $p = .27$).

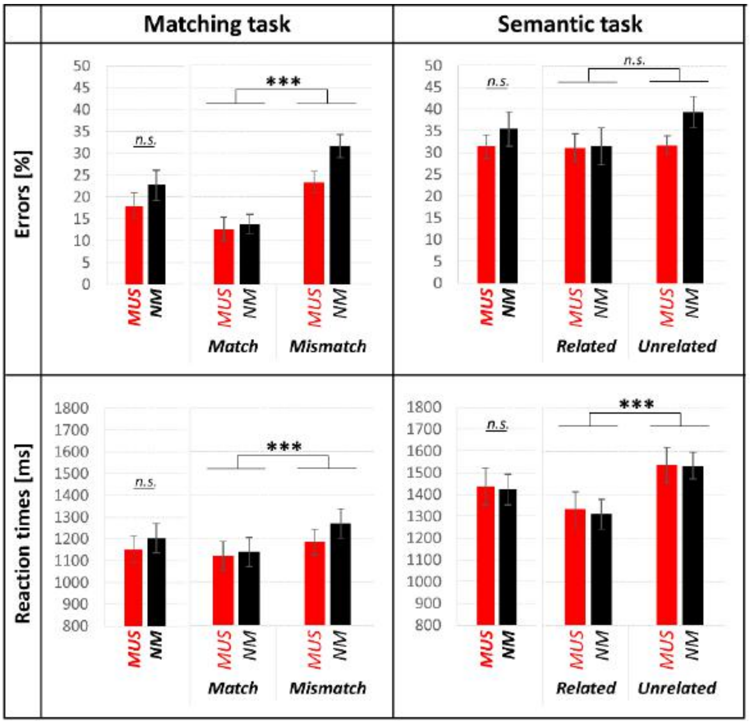


Figure 2. Older adults: behavioral data. Percentages of errors (first row) and reaction times (second row) in the matching (left column) and semantic (right column) tasks are shown for MUS (red) and NM (black).

In line with ERRs, no significant between-group differences were found on RTs in the matching and semantic tasks (matching task: main effect of Group: F < 1; Group x Condition interaction: $F_{(1,24)}$ = 1.68, p = .21; semantic task: main effect of Group and Group x Condition interaction: both F < 1) but RTs were faster to match (1129 ms, SD = 47) or to related words (1323 ms, SD = 53) compared to mismatch (1226 ms, SD = 45) or unrelated words (1536 ms, SD = 51; matching task: main effect of Condition: $F_{(1,24)}$ = 13.97, p = .001; semantic task: main effect of Condition: $F_{(1,24)}$ = 52.60, p = .001).

Electrophysiological data

In the learning phase, and in line with previous results, the N400 component was larger over frontal (-0.25 µV, SD = 1.01) and central (0.22 µV, SD = 1.04) sites compared to parietal sites (1.08 µV, SD = 1.13; Tukey, both p < .001; main effect of Anterior/Posterior: $F_{(2,48)}$ = 21.99, p < .001). In addition, both the N400 and N200 components were larger in Block 4 (N400: -0.03 µV, SD = 0.99 and N200: -0.45 µV, SD = 1.17) than in Block 1 (N400: 1.28 µV, SD = 1.05; Tukey, p = .05, main effect of Block: $F_{(3,72)}$ = 3.04, p = .03 and N200: 0.95 µV, SD = 1.28, Tukey, p = .04; main effect of Block: $F_{(3,72)}$ = 3.95, p = .01, see Figure 3A). However, the N400 increase was only significant for MUS, with significantly larger N400 in Block 2 (-0.21 µV, SD = 1.91) compared to Block 1 (2.09 µV, SD = 1.80; Tukey, p = .03; main effects of Block: MUS: $F_{(3,36)}$ = 3.06, p = .04 and NM: $F_{(3,36)}$ = 1.14, p = .35; Group x Block x Anterior/Posterior interaction: $F_{(6,144)}$ = 2.32, p = .04). No significant between-group differences were revealed for the N200 component.

In the matching task, the main effect of Group or the interactions involving the Group factor were not significant neither for the N400 (main effect of Group: F < 1; Group x Condition interaction: $F_{(1,24)}$ = 1.71, p = .20; Group x Anterior/Posterior interaction: F < 1) nor for the N200 (main effect of Group: F < 1; Group x Condition interaction: $F_{(1,24)}$ = 1.76, p = .20; Group x Anterior/Posterior interaction: $F_{(2,48)}$ = 1.46, p = .24). However, participants showed larger N400 and N200 to mismatch (N400: -3.01 µV, SD = 0.63 and N200: -3.78 µV, SD = 0.53) than to match words (N400: -1.95 µV, SD = 0.81 and N200: -3.06 µV, SD = 0.61) over parietal sites (Condition x Anterior/Posterior interactions: N400: $F_{(2,48)}$ = 10.65, p < .001 and N200: $F_{(2,48)}$ = 16.95, p < .001, Tukey, both p < .001, see Figure 3B).

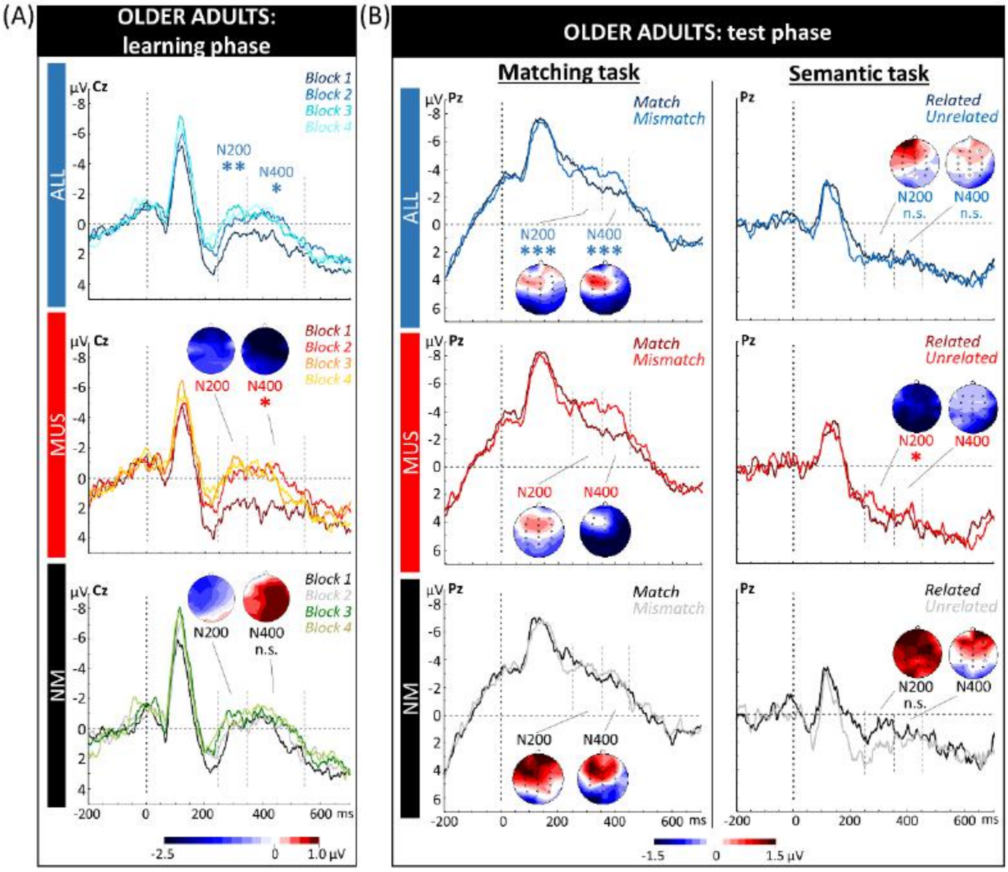


Figure 3. Older adults: electrophysiological data. ERPs recorded (A) at the central electrode (Cz) in the learning phase, and (B) at the parietal electrode (Pz) in the matching and semantic tasks, are overlapped for (A) blocks and (B) conditions, separately for all participants (ALL), musicians (MUS), and non-musicians (NM). Voltage values are scaled from (A) −2.5 to +1.0 µV and (B) −1.5 to +1.5 µV. In this and subsequent figures, time in milliseconds is reported in the abscissa and the amplitude of the effects in microvolt is depicted in the ordinate. Time zero corresponds to word onset and negativity is plotted upwards. The level of significance is represented by stars with * p < .05, ** p < .01, ***p < .001, and n.s. = not significant. Topographic voltage distribution maps of the differences between (A) the first two blocks (Block 2 minus 1) and (B) the two conditions (mismatch minus match, and unrelated minus related, respectively) are illustrated for the N200 and N400 components.

Finally, in the semantic task the N400 was smaller in MUS (2.69 µV, SD = 0.78) compared to NM (0.50 µV, SD = 0.78; main effect of Group: $F_{(1,24)} = 4.01$, p = .05, see Figure 3B), but the N400 effect (i.e., the difference between unrelated and related words) was not significant (main effect of Condition: F < 1, Condition x Anterior/Posterior interaction: $F_{(2,48)} = 1.06$, p = .36). Nevertheless, in MUS, the N200 was larger in response to unrelated (2.47 µV, SD = 1.56) than to related words (3.44 µV, SD = 1.67; main effect of Condition: $F_{(1,12)} = 6.72$, p = .02) with no such difference in NM (main effect of Condition: $F_{(1,12)} = 3.76$, p = .08; Group x Condition interaction: $F_{(1,24)} = 7.92$, p = .01).

LIFESPAN PERSPECTIVE

Behavioral data

In the matching task, MUS made overall fewer errors than NM, and all participants made fewer errors to match than to mismatch words (main effects of Group and Condition: $F_{(1,69)}$ = 10.88, p = .002 and $F_{(1,69)}$ = 69.59, p < .001, respectively; see Table 2 and Figure 4). Specifically, MUS made fewer errors than NM in response to mismatch words (Tukey, p < .001), with no between-group differences for matching words (Tukey, p = .66; Group x Condition interaction ($F_{(1,69)}$ = 6.74, p = .01; see Table 2). Young adults also made fewer errors than children (Tukey, p = .03; main effect of Age: $F_{(2,69)}$ = 3.81, p = .03; see Table 2). More specifically, young adults made fewer errors than older adults and children to mismatch words (Tukey, p = .06 and p = .002, respectively) but not to match words (Tukey, both p = .99; Age x Condition interaction: $F_{(2,69)}$ = 5.27, p = .007; see Table 2).

According to RTs, the main effect of Group was not significant (F < 1), but RTs were faster to match compared to mismatch words (main effect of Condition: $F_{(1,69)}$ = 54.84, p < .001; see Figure 4). RTs were also faster in older and young adults compared to children (Tukey, p = .02 and p < .001, respectively; main effect of Age: $F_{(2,69)}$ = 10.67, p < .001). Finally, older adults responded faster than children to mismatch words (Tukey, p = .02) but young adults responded faster than children to both match and mismatch words (Tukey, p = .01 and p < .001, respectively; Age x Condition: $F_{(2,69)}$ = 8.20, p < .001; see Table 2).

Table 2. Error rates (ERRs) and reaction times (RTs) are shown for all participants (ALL), musicians (MUS) and non-musicians (NM), averaged across conditions (Overall), separately for match and mismatch trials in the matching task, and for related and unrelated trials in the semantic task. Values in bold relate to significant effects.

		ERRs [%]			**RTs [ms]**		
		Overall	*Match/ Related*	*Mismatch/ Unrelated*	*Overall*	*Match/ Related*	*Mismatch/ Unrelated*
Matching task	ALL	20.3	**14.5**	**26.0**	1203	**1151**	**1255**
	MUS	**17.0**	13.0	**20.9**	1185	1143	1228
	NM	**23.5**	15.8	**30.7**	1221	1160	1282
	Older adults	20.3	13.2	**27.5**	**1177**	1129	**1226**
	Young Adults	**16.8**	14.2	**19.4**	**1060**	**1042**	**1078**
	Children	**23.6**	15.9	**30.8**	**1372**	**1283**	**1459**
Semantic task	ALL	29.7	30.9	28.4	1616	**1509**	**1722**
	MUS	**26.1**	28.3	23.8	1613	1513	1712
	NM	**33.2**	33.5	32.9	1618	1505	1732
	Older adults	**33.4**	31.3	35.5	**1429**	1323	1536
	Young Adults	**27.1**	**31.8**	**22.4**	**1270**	1209	1333
	Children	28.5	29.7	27.2	**2146**	1996	2296

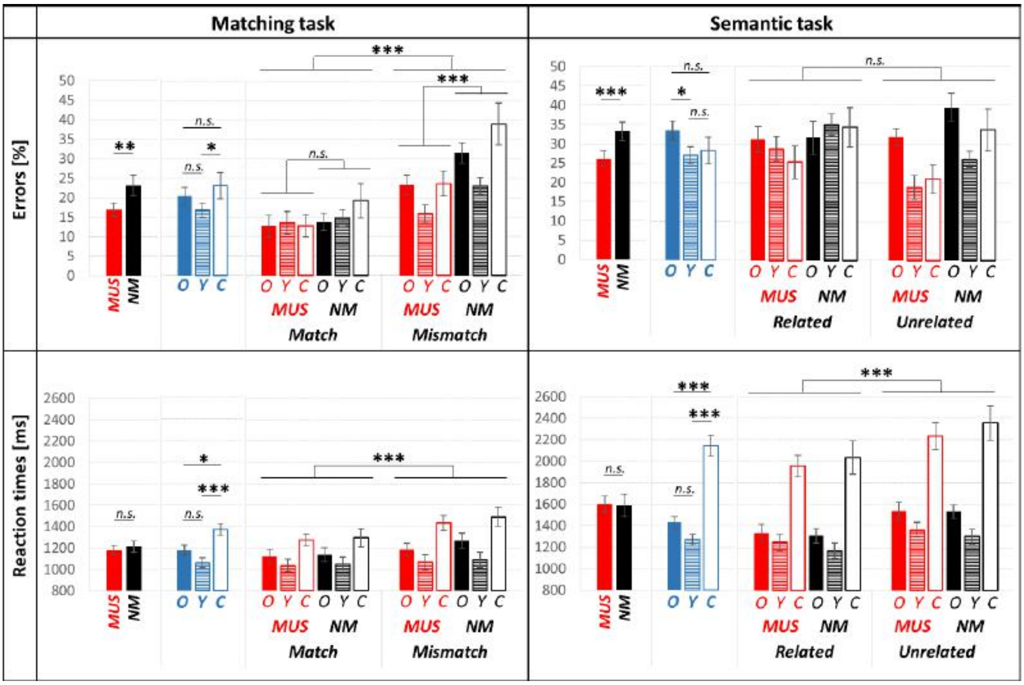


Figure 4. Lifespan perspective: behavioral data. Percentages of errors (first row) and reaction times (second row) in the matching (left column) and semantic (right column) tasks are shown for all participants (blue), MUS (red), NM (black), as well as for older adults (O), young adults (Y) and children (C).

In the semantic task, MUS made overall fewer errors than NM (main effect of Group: $F_{(1,69)} = 11.32$, $p = .001$), and older adults committed more errors than young adults (Tukey, $p = .04$; main effect of Age: $F_{(2,69)} = 3.35$, $p = .04$). Furthermore, while young adults made more errors to related compared to unrelated words (main effect of Condition: $F_{(1,24)} = 12.48$, $p = .002$), this difference was not significant either in older adults (main effect of Condition: $F_{(1,24)} = 1.28$, $p = .27$) or in children (main effect of Condition: $F < 1$; Age x Condition: ($F_{(2,69)} = 3.98$, $p = .02$; see Figure 4).

The evaluation of RTs in the semantic task did not reveal a significant main effect of Group ($F < 1$), but all participants responded faster to related than to unrelated words (main effect of Condition: $F_{(1,69)} = 148.99$, $p < .001$; see Figure 4 and Table 2), and older and young adults responded faster than children (Tukey, both $p < .001$; main effect of Age: $F_{(2,69)} = 49.68$, $p < .001$).

Electrophysiological data

In the learning phase, we only revealed a significant increase in N400 and N200 amplitudes from Block 1 to Block 2 in MUS (main effects of Block: MUS: N400: $F_{(1,37)} = 26.81$, p

< .001; N200: $F_{(1,37)} = 21.61$, $p < .001$; and NM: N400: $F_{(1,36)} = 2.51$, $p = .12$; N200: $F_{(1,36)} = 1.48$, $p = .23$; Group x Block interactions: N400: $F_{(1,73)} = 6.49$, $p = .01$; N200: $F_{(1,73)} = 6.62$, $p = .01$, see Table 3 and Figure 5). Moreover, the increase in N200 amplitudes was significant in young adults and children (main effects of Block: $F_{(1,29)} = 14.02$, $p < .001$ and $F_{(1,22)} = 8.34$, $p = .009$, respectively), but not in older adults (main effect of Block: $F_{(1,25)} = 2.37$, $p = .14$; Age x Block x Anterior/Posterior: $F_{(4,146)} = 3.23$, $p = .01$). Additionally, in children, this increase was largest over frontal and central sites (Tukey, $p < .001$ and $p = .004$, respectively; Block x Anterior/Posterior interaction: $F_{(2,44)} = 4.10$, $p = .02$). Finally, N400 and N200 components were largest in children, intermediate in young adults, and smallest in older adults (Tukey, all $p < .05$; main effects of Age: N400: $F_{(2,73)} = 29.97$, $p < .001$ and N200: $F_{(2,73)} = 12.49$, $p < .001$; see Table 3 and Figure 5).

Table 3. Mean amplitudes (µV) and standard deviations (SD) are shown for all participants (ALL), musicians (MUS) and non-musicians (NM), averaged across conditions (Overall), separately for Blocks during the training phase (Block 1, Block 2), for match and mismatch trials in the matching task, and for related and unrelated trials in the semantic task. Values in bold relate to significant effects.

			N400						**N200**					
			Overall		***Block 1/ Match/ Related***		***Block 2/ Mismatch/ Unrelated***		***Overall***		***Block 1/ Match/ Related***		***Block 2/ Mismatch/ Unrelated***	
			µV	***SD***	***µV***	***SD***	***µV***	***SD***	***µV***	***SD***	***µV***	***SD***	***µV***	***SD***
Learning		ALL	-1.87	0.48	**-1.28**	**0.45**	**-2.46**	**0.52**	-0.39	0.55	**-0.76**	**0.57**	**-1.96**	**0.54**
		MUS	-1.60	0.35	**-0.70**	**0.71**	**-2.50**	**0.78**	0.49	0.50	**-0.15**	**0.94**	**-2.07**	**0.88**
		NM	-2.15	0.36	-1.80	0.71	-2.33	0.87	-1.45	0.46	-1.38	0.61	-1.85	0.61
		Older adults	**0.68**	**0.44**	1.28	0.78	0.07	0.89	**-3.12**	**0.53**	0.95	1.29	0.02	1.29
		Young adults	**-2.10**	**0.41**	-1.85	0.73	-2.35	0.83	**-0.76**	**0.57**	**-1.10**	**0.71**	**-1.81**	**0.58**
		Children	**-4.20**	**0.46**	-3.29	0.83	-5.11	0.95	**-1.96**	**0.54**	**-2.12**	**0.90**	**-4.12**	**0.81**
Matching task		ALL (parietal)	-2.02	0.34	**-1.61**	**0.36**	**-2.42**	**0.32**	-3.42	0.30	**-2.80**	**0.31**	**-4.04**	**0.30**
		MUS	-1.81	0.45	-1.42	0.86	-2.21	0.81	-2.24	0.39	-1.85	0.76	-2.63	0.73
		NM	-1.97	0.46	-1.96	0.88	-1.98	0.82	-2.27	0.40	-2.22	0.77	-2.31	0.74
		Older adults	**-2.02**	**0.55**	-1.92	1.07	-2.13	1.00	**-2.54**	**0.49**	-2.45	0.93	-2.64	0.90
		Young adults	**-0.18**	**0.52**	-0.19	0.99	-0.16	0.93	**-1.14**	**0.46**	-0.71	0.87	-1.58	0.83
		Children	**-3.47**	**0.59**	-2.96	1.14	-3.99	1.06	**-3.08**	**0.52**	-2.96	0.99	-3.20	0.95
	MUS	Older adults	-2.57	0.78	-2.25	1.51	-2.89	1.41	**-3.03**	**0.64**	-2.69	1.32	-3.38	1.27
		Young adults	0.58	0.73	0.75	1.41	0.42	1.31	**-0.15**	**0.64**	0.46	1.23	-0.75	1.18
		Children	-3.45	0.82	-2.75	1.57	-4.15	1.47	**-3.55**	**0.67**	-3.33	1.37	-3.77	1.32
	NM	Older adults	-1.47	0.78	-1.58	1.51	-1.37	1.40	-2.06	0.74	-2.21	1.33	-1.90	1.26
		Young adults	-0.94	0.73	-1.13	1.40	-0.75	1.32	-2.14	0.69	-1.87	1.23	-2.41	1.17
		Children	-3.50	0.85	-3.16	1.64	-3.83	1.54	-2.61	0.80	-2.58	1.44	-2.63	1.38
Semantic task		ALL (parietal)	1.06	0.32	**1.29**	**0.34**	**0.83**	**0.29**	0.18	0.34	**0.52**	**0.38**	**-0.17**	**0.30**
		MUS	-0.14	0.36	0.08	0.76	-0.36	0.62	0.04	0.39	**0.68**	**0.86**	**-0.59**	**0.72**
		(parietal)	1.57	0.47	**2.05**	**0.49**	**1.08**	**0.46**						
		NM	-1.09	0.37	-1.29	0.77	-0.89	0.63	-0.44	0.40	-0.84	0.88	-0.04	0.74
		(parietal)	0.55	0.44	0.52	0.47	0.58	0.34						
		Older adults	**1.74**	**0.45**	1.69	0.94	1.78	0.76	**2.24**	**0.49**	2.07	1.06	2.41	0.89
		Young adults	**0.15**	**0.42**	0.22	0.87	0.08	0.71	**0.68**	**0.45**	1.11	0.99	0.25	0.83
		Children	**-3.73**	**0.48**	-3.72	1.00	-3.73	0.81	**-3.51**	**0.52**	-3.41	1.13	-3.60	0.95
	MUS	Older adults	**2.85**	**0.64**	2.96	1.33	2.74	1.08	2.95	0.69	3.44	1.51	2.47	1.26
		Young adults	**0.50**	**0.59**	0.75	1.24	0.25	1.00	1.19	0.64	1.86	1.40	0.52	1.18
		Children	-3.78	0.66	-3.48	1.38	-4.08	1.12	-4.01	0.72	-3.26	1.57	-4.76	1.32
	NM	Older adults	0.62	0.63	0.41	1.34	0.82	1.08	1.52	0.68	0.70	1.51	2.35	1.26
		Young adults	-0.20	0.59	**-0.31**	**1.24**	**-0.10**	**1.01**	0.17	0.65	0.36	1.40	-0.02	1.17
		Children	-3.68	0.69	-3.97	1.44	-3.39	1.17	-3.01	0.75	-3.57	1.64	-2.45	1.37

In the matching task, N400 and N200 components were larger to mismatch than to match words over parietal sites in all participants (Tukey, both p < .001; Condition x Anterior/Posterior interactions: N400: $F_{(2,146)}$ = 9.23, p < .001 and N200: $F_{(2,146)}$ = 28.03, p < .001). Furthermore, the N400 was larger in older adults and children than in young adults (Tukey, p = .05 and p < .001, respectively; main effect of Age: $F_{(2,73)}$ = 9.03, p < .001), but only in MUS the N200 was larger in older adults and children compared to young adults (MUS: Tukey, both p = .006 and p = .002, respectively; main effect of Age: $F_{(2,37)}$ = 8.73, p < .001; NM: main effect of Age: F < 1; Group x Age: $F_{(2,73)}$ = 3.28, p = .04; see Table 3 and Figure 6).

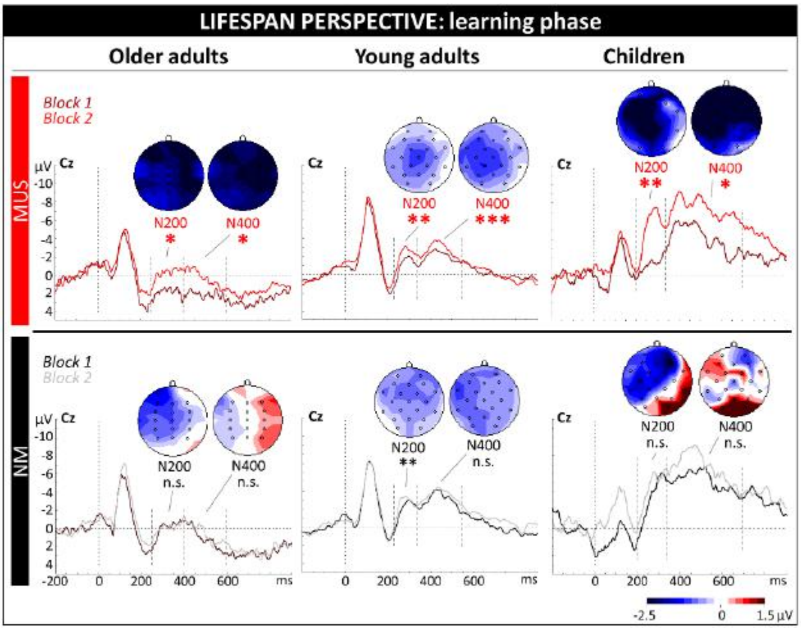


Figure 5. Lifespan perspective: electrophysiological data learning phase. ERPs recorded at the central electrode (Cz) in the learning phase are overlapped for blocks, separately for MUS and NM, and for older adults (left), young adults (central) and children (right). Topographic voltage distribution maps of the differences between the two blocks (Block 2 minus 1) are illustrated for the N200 and N400 components. Voltage values are scaled from −2.5 to +1.5 µV.

In the semantic task, the N400 was larger to unrelated compared to related words over parietal electrodes for MUS (Tukey, p < .001; Condition x Anterior/Posterior interaction: $F_{(2,74)}$ = 8.61, p < .001), but not for NM (Condition x Anterior/Posterior interaction: $F_{(2,72)}$ = 2.74, p = .08; Group x Age x Condition x Anterior/Posterior interaction: $F_{(4,146)}$ = 2.42, p = .05; see Table 3 and Figure 6). Moreover, in young adult NM this effect was reversed, with larger N400 in response to related than unrelated words over frontal sites (Tukey, p = .004; Condition x

Anterior/Posterior interaction: $F_{(2,28)} = 10.19$, $p < .001$; Age x Condition x Anterior/Posterior interaction, $F_{(4,74)} = 2.57$, $p = .04$). Similarly, for the N200, the semantic priming effect was significant only in MUS (main effect of Condition: $F_{(1,39)} = 10.63$, $p = .002$) but not in NM (main effect of Condition: $F_{(1,38)} = 1.74$, $p = .20$; Group x Condition: $F_{(1,73)} = 9.43$, $p = .003$, see Table 3 and Figure 6). Both the N400 and N200 components were largest in children, intermediate in young adults, and smallest in older adults (Tukey, all $p < .05$; main effects of Age: N400: $F_{(2,73)} = 36.04$, $p < .001$ and N200: $F_{(2,73)} = 34.55$, $p < .001$), with largest differences over frontal scalp positions (N400: Tukey, $p = .01$; Age x Anterior/Posterior interaction: $F_{(4,146)} = 5.70$, $p < .001$; N200: Tukey, $p = .03$; Age x Anterior/Posterior interaction: $F_{(4,146)} = 6.60$, $p < .001$). However, the smaller N400 in older adults compared to young adults was significant for MUS only (MUS: Tukey, $p = .05$; main effect of Age: $F_{(2,37)} = 21.56$, $p < .001$; and NM: Tukey, $p = .53$; main effect of Age: $F_{(2,36)} = 15.10$, $p < .001$; Group x Age x Condition x Anterior/Posterior interaction: $F_{(4,146)} = 2.42$, $p = .05$).

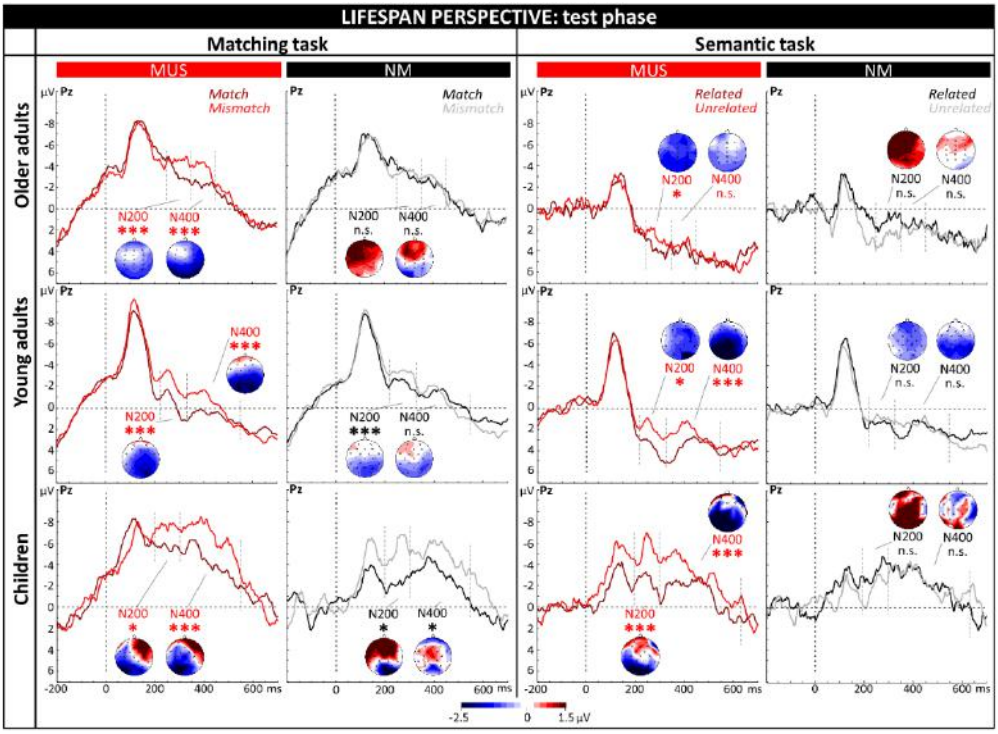


Figure 6. Lifespan perspective: electrophysiological data test phase. ERPs recorded at the parietal electrode (Pz) in the matching (left column) and semantic (right column) tasks, are overlapped for conditions, separately for MUS and NM, and for older adults (first row), young adults (second row) and children (third row). Topographic voltage distribution maps of the differences between the two conditions (mismatch minus match, and unrelated minus related, respectively) are illustrated for the N200 and N400 components. Voltage values are scaled from −2.5 to +1.5 µV.

DISCUSSION

In the present EEG study, we focused on three topics, namely on (1) whether older adult MUS show a similar beneficial effect of music training on word learning, (2) the developmental trajectory of word learning across the lifespan, and (3) on putative interactions between music training, development, and word learning. In turn, we will separately discuss the results related to these purposes by integrating behavioral and electrophysiological data.

Music training and novel word learning in older adults

During the learning phase, older adults demonstrated fast brain plasticity effects with significantly increased N200 and N400 amplitudes from Block 1 to Block 4 that developed after only five minutes of learning picture-word associations (see Figure 3A, "ALL"). Moreover, in line with previous literature on word learning (Batterink & Neville, 2011; Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2010; Dittinger et al., 2016; McLaughlin, Osterhout, & Kim, 2004; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007; Perfetti, Wlotko, & Hart, 2005), N400s to novel words were clearly larger at fronto-central scalp sites than at parietal electrodes. Such frontal N400s are in line with previous results showing the contribution of frontal brain regions to the acquisition of word meaning (Rodríguez-Fornells et al., 2009) and to the maintenance of novel information in working memory (Hagoort, 2014). Notably, N200 and N400 amplitudes also showed a close similarity to the temporal dynamics of these two components previously described in children and young adults (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017), possibly indicating a common neural principle underlying word learning across the lifespan.

The fact that older adults were able to learn novel words was also evidenced by behavioral data of the matching task, indicating lower ERRs and faster RTs to match compared to mismatch words (see Figure 2). During the same task, older adults also demonstrated significantly increased N200 and N400 amplitudes in response to mismatch compared to match words (i.e., N200 and N400 effects), and both components were most pronounced at parietal scalp sites (see Figure 3B, "ALL"). Such a posterior distribution of the N200 and N400 effects is typically observed during linguistic tasks requiring access to lexical-semantic knowledge (Kutas & Federmeier, 2011). However, it is also possible that parietal N200s and N400s reflect the engagement of processes needed to retrieve information from episodic memory. The topographical manifestation of these processes clearly differs from the frontal N400 described in the learning phase. Taken together, behavioral and electrophysiological results converge on the view that older adults were able to rapidly learn new words through picture-word associations as well as to retrieve these new associations from episodic memory.

Results in the semantic task were not clear-cut, and we only found weak evidence for the integration of the meaning of novel words into semantic memory that was reflected by faster RTs in response to related compared to unrelated words (see Figure 2). In contrast, ERRs and N400 components (no significant differences between related and unrelated words, see Figure 3B, "ALL") did not reveal evidence for such an integration into semantic memory. Therefore, results possibly indicate that older adults were able to recall newly-formed associations between pictures and words from episodic memory, but the meaning of the novel words was not sufficiently integrated into semantic networks to generalize to new pictures that have not been seen before in the experiment. This interpretation will be discussed in more details in the lifespan section, where we will also integrate previous results of young adults and children.

The evaluation of the influence of music training on word learning revealed that only older adult MUS were characterized by a rapid increase of the N400 from Block 1 to Block 2 in the learning phase. This result is in line with previous studies (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017), that was taken to reflect the faster encoding of novel words in MUS compared to NM. However, according to a visual inspection of the ERPs (see Figure 3A), it seems that in Block 1 the N400 was larger in older adult NM compared to MUS, whereas in Block 4 the N400 was similar between the two groups. Accordingly, an alternative line of argumentation is that older adult NM possibly already engaged the full spectrum of processing resources for word learning in Block 1, with no further modulatory effects (i.e., N400 increase or decrease) in the subsequent blocks. By contrast, older adult MUS possibly recruited their processing resources more gradually, as reflected by increased N400 amplitudes across the blocks. Such between-group differences in the developmental dynamics of the N400 in the learning phase, as well as its realignment at the end of the learning phase, could contribute to explain why both older adult MUS and NM performed in a comparable manner in both the matching and semantic tasks (see Figure 2). Thereby, it is important to mention that the lack of between-group differences in the behavioral data of the matching task is in line with the results previously reported in young adults (Dittinger et al., 2016), whereas the similar performance of the two groups in the semantic task was somehow unexpected. It is possible that the cognitive advantages, that have previously been shown to be mediated by music training (George & Coch, 2011; Moreno et al., 2011; Schellenberg, 2004; Strait et al., 2015; Zuk et al., 2014), decline with aging, and that putative between-group differences disappear especially in tasks that constitute a situation of "cognitive multitasking". In fact, in the semantic task, new pictures have to be maintained in working memory while at the same time comparing them with those previously learned in order to judge the degree of semantic affinity and initiate a decision. In this context, previous work has suggested that older adults have more difficulties than younger adults in multitasking situations, and that they generally exhibit inefficient activation patterns in brain regions supporting executive functions and cognitive control mechanisms (Allen, Lien, Ruthruff, & Voss, 2014; Nashiro, Qin, O'Connell, & Basak, 2018).

In the semantic task, the evaluation of the electrophysiological data revealed two interesting effects of musical expertise. First, only older adult MUS demonstrated significantly larger N200 amplitudes in response to unrelated compared to related words. As mentioned in the introduction, N200 amplitude is modulated by different factors including phonological processing (Connolly & Phillips, 1994; Friedrich & Friederici, 2008), attention (Patel & Azzam, 2005), and cognitive control (Kropotov et al., 2016, 2017), that are possibly better preserved in older MUS than NM. Moreover, the N200 effect could also indicate a sort of phonological mismatch negativity (Connolly & Phillips, 1994) that originated from a pre-activation of the words and their respective phonological properties in short-term memory while seeing the pictures (always presented first), and from the comparison of this mnemonic representations with the auditory presented words. Importantly, since this effect was only uncovered in MUS, results suggest that musical expertise influenced novel word processing and categorization in the semantic task (Dittinger et al., 2016). Finally, since N200 amplitudes have also been proposed to be influenced by contextual effects during word recognition (van den Brink, Brown, & Hagoort, 2001), the N200 effect in MUS may also be interpreted as an early semantic priming effect that developed faster in older adult MUS compared to NM. However, the N200 did not evolve into clear N400 effects in musicians or non-musicians, questioning whether words have been integrated into semantic networks.

The significant main effect of musical expertise we revealed in the N400 latency window constitutes a second important result of the semantic task. However, this effect started at around 350 ms, and originated from a larger positive shift in older adult MUS compared to NM (see Figure 3B). By taking into account the higher degree of N400-related adaptations in MUS from Block 1 to Block 2 during the learning phase as well as the compensation of the same component in NM during the matching task (i.e., similar N400 amplitudes and effects in both groups), it is conceivable to assume that NM more strongly recruited additional top-down functions (i.e., attention, short-term memory, and working memory) for managing the semantic task. Otherwise, it is also possible that the anticipatory strategy in MUS (as reflected by the N200 effect, described above) might be at the origin of the increased N400-related positivity observed in MUS compared to NM. In particular, we might suggest that due to pre-activation of the words, MUS were less dependent on the subsequent recruitment of episodic memory functions, resulting in larger positive deflections (Polich, 2007).

In summary, our results suggest that older adults were generally able to rapidly encode picture-word associations and to learn novel words of a “foreign” language. However, the lack of semantic priming effects in terms of ERRs and N400 effects during the semantic task leads us to conclude that older adults were probably not able to integrate the meaning of the novel words into semantic memory. Therefore, future studies should use more word repetitions in order to facilitate consolidation processes. Furthermore, even though we found some evidence for an influence of musical expertise on novel word learning in older adult participants, the beneficial effect was rather weak compared to children (Dittinger, Chobert, et al., 2017) and young adults (Dittinger et al., 2016). One possible reason for these rather mild effects is that the higher amount of deliberate practice of older adult MUS (i.e., compared to children undergoing music training and young musicians) might be counteracted by brain aging. Specifically, the decrease of daily music training in older compared to younger adults may not suffice to counteract cognitive and neural decline. Therefore, future studies should try to disentangle the effects of long-life music training from daily music practice in older adults in order to gain additional insights into this speculative argumentation.

Novel word learning across the lifespan: the effect of age

Speech and language processing undergo several qualitative and quantitative changes in the developmental trajectory (Shafto & Tyler, 2014). For example, aging is accompanied by an expansion of vocabulary, a differential recruitment of lexical-semantic networks, as well as a shift from controlled to automatic processing (Shafto & Tyler, 2014). Such a developmental perspective is reflected by the behavioral data (i.e., ERRs and RTs) of the matching and semantic tasks. In fact, during the matching task adults demonstrated lower ERRs compared to children, and both older and young adults responded faster than children. Additionally, young adults made fewer errors than older adults. Interestingly, behavioral between-group differences mainly originated from mismatch trials, indicating that children and older adults were more prone than adults to consider a word that mismatched with the picture as a matching word. This perspective is compatible with the fact that mismatch trials were more difficult than match trials, suggesting that age-related differences may be more pronounced for difficult compared to simple task conditions. Finally, in the semantic task older adults committed more errors than young adults, and both older and young adults responded faster than children. These results might be related to different trajectories of cognitive functions and neural ripening across the lifespan (Giorgio et al., 2010; Gogtay et al., 2004).

During the learning phase, children and young adults were characterized by an increase of N200 amplitude from Block 1 to Block 2, and children demonstrated the largest difference between the two blocks at anterior and central electrodes. Furthermore, both N200 and N400 amplitudes were smallest in older adults, intermediate in young adults, and largest in children. The fast development of the N200 component in children and young adults leads to suggest a higher degree of neural plasticity in these two cohorts compared to older adults, whereas the decrease of N200 and N400 amplitudes across the lifespan may indicate that development and aging are accompanied by less neural effort in the recruitment of associative memory functions. In fact, a similar decrease of N200 amplitudes as we found in the present study across lifespan has previously been documented by Enoki and colleagues (Enoki, Sanada, Yoshinaga, Oka, & Ohtahara, 1993) in a sample of 164 participants in the age range between 4-77 years. Thereby, the authors argued that increased N200 amplitudes in children may correlate with a wider range of attentional focus and reflect immaturity of information processing (Friedman, Brown, Vaughan, Cornblatt, & Erlenmeyer-Kimling, 1984). In line with this previous work, the fast N200 development we observed over frontal scalp sites in children, together with increased N400 amplitudes in the same cohort compared to older and young adults, may point to an underdevelopment of prefrontal functions (Gogtay et al., 2004). Accordingly, unripe cognitive functions may lead to a larger mobilization of cognitive resources and attention functions that are needed for learning novel words. This line of argumentation is compatible with the behavioral data of the matching task, indicating that adults showed lower ERRs than children and both older and young adults performed faster than children.

During the matching task, the N400 component was differentially influenced by age. In fact, both children and older adults were characterized by increased amplitudes compared to young adults. These results possibly suggest that children and older adults invested more neural resources than adults for accomplishing the task. However, in children such a “working harder” processing mode was not that efficient as in older adults. In fact, children demonstrated longer RTs compared to older adults and committed more errors than young adults. However, it is important to mention that older adults were closer to children in terms of ERRs and more similar to young adults regarding RTs. Furthermore, children had to learn fewer words compared to adults (i.e., six vs. nine words). Finally, during the semantic task, both the N200 and N400 components were largest in children, intermediate in young adults, and smallest in older adults. Such a decrease of N200 and N400 amplitudes as a function of aging might reflect neural and cognitive facilitation for accessing phonological and lexical-semantic information. However, this perspective is, at least partially, in contrast with the behavioral and physiological data showing that older adults committed more errors than young adults and were not characterized by a clear N400 effect. In order to better elucidate the complex relationships between ERPs and cognitive functions, future studies should use more sophisticated psychological screening procedures (i.e., psychometric tests targeting at better capturing attention, short-term-, working-, episodic- and semantic memory) and try to correlate these measure with behavioral and electrophysiological data.

Novel word learning across the lifespan: the effect of musicianship

An additional target of this work, was to determine whether the previously described advantage of young adult professional MUS and children undergoing music training in word learning (Dittinger et al., 2016; Dittinger, Chobert, et al., 2017) would also be detectable in older adult MUS. Such a purpose is particularly fruitful in that it enables to estimate the positive influence of music training on word learning by taking into account interactions between this specific activity, development, and aging in a large group of participants ranging from childhood to late adulthood. Furthermore, until now, it is still unclear whether lifelong music training might have a beneficial influence on the aging brain (Rogenmoser et al., 2018; Strait & Kraus, 2014). The analyses of the N200 and N400 components during the learning phase revealed that only MUS (i.e., averaged across age) demonstrated increased amplitudes in the second- compared to the first block of the learning phase. Such short-term brain changes in MUS have previously been described by other groups (Seppänen et al., 2013), and posit a high degree of neural efficiency for learning speech sounds as a function of music training (Dittinger et al., 2016; Elmer, Albrecht, Valizadeh, François, & Rodríguez-Fornells, 2018; Elmer, Greber, Pushparaj, Kühnis, & Jäncke, 2017). Notably, such an increased neural efficiency also translated to a behavioral advantage of MUS (across the three-age groups) in both the matching and semantic tasks. In fact, MUS were overall characterized by lower ERRs compared to NM, indicating a positive influence of musical expertise on the formation of novel associations in episodic memory as well as on the integration of novel associations into semantic memory. Furthermore, these results emphasize a relationship between music training and an optimization of cognitive functions.

During the matching task, between-group differences were larger on mismatch compared to match trials, suggesting that NM were more prone than MUS to consider a word that mismatched with the picture as a matching word (i.e., as mentioned earlier in older adults and children compared to young adults). These results also extend previous literature showing that difficult tasks are typically more sensitive for detecting expertise-related influences (Schön et al., 2004). Finally, it is important to accentuate that although the effects in both tasks were significant across the three age cohorts, within the sample of older adults the effects were rather small (i.e., see previous section), and the overall better performance of MUS compared to NM mainly originated from children and young adults.

Notably, during the matching task only within the MUS group children and older adults demonstrated increased N200 amplitudes compared to young adults. Since this effect was not observed within the group of NM, and children and older adults are generally characterized by more susceptible cognitive functioning compared to young adults (Dempster, 1992; Luna et al., 2001), results are interpreted as indicating a compensatory influence of music training. In this vein, previous work reported a positive influence of music training on attention functions (Kraus & Chandrasekaran, 2010), phonetic processing (Chobert et al., 2014; Elmer et al., 2012; Kühnis et al., 2013) as well as inhibitory functions (Zuk et al., 2014) that was mirrored by modulations of the N200 component. Previous studies also carved out that children and older adults often demonstrate dysfunctional phonological processing (Carroll, Snowling, Hulme, & Stevenson, 2003; Froehlich et al., 2018) and executive functions (Brocki & Bohlin, 2004; Kennedy & Raz, 2009). Consequently, children and older adults seem to have benefited from music training that possibly acts as a vehicle for compensating suboptimal auditory and cognitive processing requirements.

During the semantic task, MUS performed better than NM, and this behavioral advantage was accompanied by a multifaceted constellation of neural responses. First of all, only MUS showed N200 and N400 effects that originated from increased brain responses to incongruent- compared to congruent trials. Since N200 and N400 effects with a characteristic distribution over posterior electrodes are classically considered to be sensitive markers for the integration of information into lexical-semantic memory (Kutas & Federmeier, 2011; Patel & Azzam, 2005), results lead to suggest that music training has a beneficial influence on mapping the meaning of lexically affine items. However, as previously mentioned, it is also possible that these effects were mediated by phonological priming mechanisms or recollection from episodic memory. Importantly, it should be reminded that for older adults only no significant N400 effect was obtained. Therefore, although significant effects could be detected in MUS across the three age groups, our data lead to suggest that the number of trials was not sufficient for older adults to integrate novel words into semantic networks or that the advantage of MUS in semantic processing is less pronounced in older adults.

Finally, our analyses also uncovered smaller N400 amplitudes in older adult MUS compared to young adult MUS over frontal electrodes. This result possibly reflects a more automated processing of the novel words. Certainly, longitudinal approaches in combination with multimodal analyses (i.e., multimodal imaging and psychometric/behavioral data) are strictly requested to infer causal relationships between music training, cognitive functioning, and word learning across the lifespan.

LIMITATIONS

A limitation of this study is that two cohorts (i.e., children and young adults) were measured in Marseille, whereas the older adult participants were measured in Zurich. In this context, it is important to mention that the aspirated phoneme of the Thai word /p^{h}a/ is part of the German phonetic repertoire but constitutes a non-native phoneme for French speakers. Consequently, we cannot exclude that this specific stimulus might have influenced the results.

Such a contingency is, for example, supported by a recent study of Dittinger and colleagues (Dittinger et al., 2018) who demonstrated that French MUS are better able to distinguish between the native (/p/) and the non-native phonemes (/p^h/), as reflected by behavioral data and larger N200 and P300 amplitudes. A second limitation of this study is that the stimuli we used were recorded by a female Thai-French bilingual speaker, this possibly making the stimuli easier to perceive for French than Swiss German participants.

CONCLUSION

In the present work, we used a cross-sectional and comparative approach in three cohorts of participants in order to infer whether (1) music training might have a neuroprotective influence on word learning in older adults, (2) the ability to learn new words through picture-word associations changes throughout lifespan, and (3) to how music training interacts with word learning and aging. Our results corroborate that (1) all participants were able to learn novel words very fast, (2) music training facilitates the formation of new associations in episodic memory (i.e., lower ERRs in the matching task), and that (3) music practice facilitates the integration of novel words into semantic networks in children and young adults. By contrast, in older adults influences of music training on semantic processing are less clear-cut. These results confirm qualitative and quantitative differences in the ability to learn new words across lifespan that vary as a function of neural ripening, maturation of cognitive functions, age-related decline, and music training.

ACKNOWLEDGMENTS

We would like to thank all the participants for having contributed to the experiment. ED’s work was supported by the Labex BLRI (ANR-11-LABX-0036), managed by the French National Agency for Research (ANR) under the program "Investissements d'Avenir" (ANR-11-IDEX-0001-02). SE and LJ were supported by the Swiss National Science Foundation (grant SNF320030B_138668 to LJ). The authors declare no competing financial interests.

REFERENCES

Allen, P. A., Lien, M.-C., Ruthruff, E., & Voss, A. (2014). Multitasking and Aging: Do Older Adults Benefit from Performing a Highly Practiced Task? *Experimental Aging Research*, *40*(3), 280–307.

Annett, M. (1970). A Classification of Hand Preference by Association Analysis. *British Journal of Psychology*, *61*(3), 303–321.

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *The European Journal of Neuroscience*, *24*(6), 1832–1834.

Batterink, L., & Neville, H. (2011). Implicit and Explicit Mechanisms of Word Learning in a Narrative Context: An Event-related Potential Study. *Journal of Cognitive Neuroscience*, *23*(11), 3181–3196.

Bender, A. R., Völkle, M. C., & Raz, N. (2016). Differential aging of cerebral white matter in middle-aged and older adults: A seven-year follow-up. *NeuroImage*, *125*, 74–83.

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*, *19*(7), 1583–1596.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is Enough: N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. *Language Learning and Development*, *8*(3), 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–296.

Brocki, K. C., & Bohlin, G. (2004). Executive Functions in Children Aged 6 to 13: A Dimensional and Developmental Study. *Developmental Neuropsychology*, *26*(2), 571–593.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults. *NeuroImage*, *17*(3), 1394–1402.

Carroll, J., Snowling, M., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Dev Psychol*, *39*, 913–923.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). 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*, *24*(4), 956–967.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49–69.

Connolly, J. F., & Phillips, N. A. (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Crivello, F., Tzourio-Mazoyer, N., Tzourio, C., & Mazoyer, B. (2014). Longitudinal Assessment of Global and Regional Rate of Grey Matter Atrophy in 1,172 Healthy Older Adults: Modulation by Sex and Age. *PLOS ONE*, *9*(12), e114478.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review*, *12*(1), 45–75.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast Brain Plasticity during Word Learning in Musically-Trained Children. *Frontiers in Human Neuroscience*, *11*.

Dittinger, E., D'Imperio, M., & Besson, M. (2018). Enhanced neural and behavioural processing of a nonnative phonemic contrast in professional musicians. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13939

Dittinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2017). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Human Brain Mapping*, *39*(2), 722–734.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex*, *22*(3), 650–658.

Elmer, S., Albrecht, J., Valizadeh, S. A., François, C., & Rodríguez-Fornells, A. (2018). Theta Coherence Asymmetry in the Dorsal Stream of Musicians Facilitates Word Learning. *Scientific Reports*, *8*(1).

Elmer, S., Greber, M., Pushparaj, A., Kühnis, J., & Jäncke, L. (2017). Faster native vowel discrimination learning in musicians is mediated by an optimization of mnemonic functions. *Neuropsychologia*, *104*, 64–75.

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, *221*(1), 331–344.

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, *49*(10), 2812–2821.

Enoki, H., Sanada, S., Yoshinaga, H., Oka, E., & Ohtahara, S. (1993). The effects of age on the N200 component of the auditory event-related potentials. *Cognitive Brain Research*, *1*(3), 161–167.

Erel, H., & Levy, D. A. (2016). Orienting of visual attention in aging. *Neuroscience & Biobehavioral Reviews*, *69*, 357–380.

Ericsson, K. A., Krampe, R. T., & Heizmann, S. (2007). Can We Create Gifted People? In *Ciba Foundation Symposium 178 - The Origins and Development of High Ability* (pp. 222–249). Wiley-Blackwell.

Fama, R., & Sullivan, E. V. (2015). Thalamic structures and associated cognitive functions: Relations with age and aging. *Neuroscience and Biobehavioral Reviews, Neuroscience and Biobehavioral Reviews*, *54*, *54*, 29, 29–37.

Ferrand, L. (1998). Age differences in word and language processing, *Annual Review of Psychology, 98*, 371–371.

Fjell, A. M., Sneve, M. H., Storsve, A. B., Grydeland, H., Yendiki, A., & Walhovd, K. B. (2016). Brain Events Underlying Episodic Memory Changes in Aging: A Longitudinal Investigation of Structural and Functional Connectivity. *Cerebral Cortex*, *26*(3), 1272–1286.

Francois, C., Chobert, J., Besson, M., & Schon, D. (2013). Music Training for the Development of Speech Segmentation. *Cerebral Cortex*, *23*(9), 2038–2043.

Francois, C., & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cerebral Cortex (New York, N.Y.: 1991)*, *21*(10), 2357–2365.

Friedman, D., Brown, C., Vaughan, H. G., Cornblatt, B., & Erlenmeyer-Kimling, L. (1984). Cognitive brain potential components in adolescents. *Psychophysiology*, *21*(1), 83–96.

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants: *NeuroReport*, *19*(18), 1757–1761.

Froehlich, E., Liebig, J., Morawetz, C., Ziegler, J. C., Braun, M., Heekeren, H. R., & Jacobs, A. M. (2018). Same Same But Different: Processing Words in the Aging Brain. *Neuroscience*, *371*, 75–95.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.

Giorgio, A., Watkins, K., Chadwick, M., James, S., Winmill, L., Douaud, G., … James, A. (2010). Longitudinal changes in grey and white matter during adolescence. *Neuroimage*, *49*, 94–103.

Giroud, N., Hirsiger, S., Muri, R., Kegel, A., Dillier, N., & Meyer, M. (2018). Neuroanatomical and resting state EEG power correlates of central hearing loss in older adults. *Brain Structure and Function*, *223*(1), 145–163.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179.

Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A. A., Leahy, R. M., … Damasio, H. (2017). Childhood Music Training Induces Change in Micro and Macroscopic Brain Structure: Results from a Longitudinal Study. *Cerebral Cortex*, 1–12.

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

Havas, V., Laine, M., & Rodríguez Fornells, A. (2017). Brain signatures of early lexical and morphological learning of a new language. *Neuropsychologia*, *101*, 47–56.

Helmstaedter, C., Lendt, M., & Lux, S. (2001). *Verbaler Lern- und Merkfähigkeitstest*. Hogrefe, Göttingen.

Hirsiger, S., Koppelmans, V., Mérillat, S., Liem, F., Erdeniz, B., Seidler, R. D., & Jäncke, L. (2016). Structural and functional connectivity in healthy aging: Associations for cognition and motor behavior. *Human Brain Mapping*, *37*(3), 855–867.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biol. Rep*, *1*(78), 10–3410.

Jäncke, L., Mérillat, S., Liem, F., & Hänggi, J. (2015). Brain size, sex, and the aging brain. *Human Brain Mapping*, *36*(1), 150–169.

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Kennedy, K. M., & Raz, N. (2009). Aging White Matter and Cognition: Differential Effects of Regional Variations in Diffusion Properties on Memory, Executive Functions, and Speed. *Neuropsychologia*, *47*(3), 916–927.

Krampe, R. T., & Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *Journal of Experimental Psychology. General*, *125*(4), 331–359.

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

Kropotov, J., Ponomarev, V. A., Pronina, M., & Jäncke, L. (2017). Functional indexes of reactive cognitive control: ERPs in cued go/no-go tasks. *Psychophysiology, 54,* 1899-1915.

Kropotov, J., Ponomarev, V., Tereshchenko, E. P., Müller, A., & Jäncke, L. (2016). Effect of Aging on ERP Components of Cognitive Control., Effect of Aging on ERP Components of Cognitive Control. *Frontiers in Aging Neuroscience, Frontiers in Aging Neuroscience*, *8*, *8*, 69–69.

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, *51*(8), 1608–1618.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Lazarou, I., Parastatidis, T., Tsolaki, A., Gkioka, M., Karakostas, A., Douka, S., & Tsolaki, M. (2017). International Ballroom Dancing Against Neurodegeneration: A Randomized Controlled Trial in Greek Community-Dwelling Elders With Mild Cognitive impairment. *American Journal of Alzheimer's Disease & Other Dementias®*, *32*(8), 489–499.

Lehrl, S. (1977). *Mehrfachwahl-wortschatz intelligenz test (MWT-B)*. Perimed, Erlangen.

Lehrl, S., & Fischer, B. (1992). Kurztest für allgemeine Basisgrössen der Informationsverarbeitung (KAI) (Aufl. Vless, Ebersberg, Vol. 3).

Long, L., Shaw, R. J., Lisa, Long, L. L., & Shaw, R. J. (2000). ADULT AGE DIFFERENCES IN VOCABULARY ACQUISITION. *Educational Gerontology*, *26*(7), 651–664.

Luna, B., Thulborn, K. R., Munoz, D. P., Merriam, E. P., Garver, K. E., Minshew, N. J., … Sweeney, J. A. (2001). Maturation of widely distributed brain function subserves cognitive development. *NeuroImage*, *13*(5), 786–793.

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.

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), 2701–2715.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-Term Music Training Enhances Verbal Intelligence and Executive Function. *Psychological Science*, *22*(11), 1425–1433.

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), 712–723.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews. Neuroscience*, *3*(6), 473–478.

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J., & Desimone R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences of the United States of America, 115* (28), E6630-E6639.

Nashiro, K., Qin, S., O'Connell, M. A., & Basak, C. (2018). Age-related differences in BOLD modulation to cognitive control costs in a multitasking paradigm: Global switch, local switch, and compatibility-switch costs. *NeuroImage*, *172*, 146–161.

Olsen, R., Pangelinan, M., Bogulski, C., Chakravarty, M., Luk, G., Grady, C., & Bialystok, E. (2015). The effect of lifelong bilingualism on regional grey and white matter volume. *Brain Res*, *1612*.

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience*, *4*.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International Journal of Medical Sciences*, *2*(4), 147.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1281–1292.

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148.

Rechel, B., Roberts, B., Richardson, E., Shishkin, S., Shkolnikov, V. M., Leon, D. A., … McKee, M. (2013). Health and health systems in the Commonwealth of Independent States. *The Lancet*, *381*(9872), 1145–1155.

Rhodes, R. E., & Katz, B. (2017). Working memory plasticity and aging. *Psychology and Aging*, *32*(1), 51–59.

Rieckmann, A., Pudas, S., & Nyberg, L. (2017). Longitudinal Changes in Component Processes of Working Memory. *eNeuro*, *4*(2), ENEURO.0052-17.2017.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3711–3735.

Rogenmoser, L., Kernbach, J., Schlaug, G., & Gaser, C. (2018). Keeping brains young with making music. *Brain Structure and Function*, *223*(1), 297–305.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–514.

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005). Structural, Functional, and Perceptual Differences in Heschl's Gyrus and Musical Instrument Preference. *Annals of the New York Academy of Sciences*, *1060*(1), 387–394.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music: Schulze & Koelsch. *Annals of the New York Academy of Sciences*, *1252*(1), 229–236.

Seppänen, M., Hämäläinen, J., Pesonen, A.-K., & Tervaniemi, M. (2013). Passive sound exposure induces rapid perceptual learning in musicians: Event-related potential evidence. *Biological Psychology*, *94*(2), 341–353.

Shafto, M. A., & Tyler, L. K. (2014). Language in the aging brain: The network dynamics of cognitive decline and preservation. *Science*, *346*(6209), 583–587.

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(14), 3799–3806.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(2), 174–215.

Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: Musicians as a model of auditory learning. *Hearing Research*, *308*, 109–121.

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104.

Tierney, A. T., Krizman, J., & Kraus, N. (2015). Music training alters the course of adolescent auditory development. *Proceedings of the National Academy of Sciences*, *112*(32), 10062–10067.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., & Tervaniemi, M. (2012). Practiced musical style shapes auditory skills. *Annals of the New York Academy of Sciences*, *1252*(1), 139–146.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*(2), 410–417.

Ziegler, G., Dahnke, R., Jäncke, L., Yotter, R. A., May, A., & Gaser, C. (2012). Brain structural trajectories over the adult lifespan. *Human Brain Mapping*, *33*(10), 2377–2389.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLOS ONE*, *9*(6), e99868.

Chapter 5

FUNCTIONAL CONNECTIVITY DURING NOVEL WORD LEARNING

Dittinger, E., Valizadeh, S.A., Jäncke, L., Besson, M., & Elmer, S. (2017). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Human Brain Mapping.* doi: 10.1002/hbm.23877. [Epub ahead of print]

Received: 31 May 2017 | Revised: 13 October 2017 | Accepted: 23 October 2017
DOI: 10.1002/hbm.23877

WILEY

RESEARCH ARTICLE

Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians

Eva Dittinger[1,2,3] | Seyed Abolfazl Valizadeh[4,5] | Lutz Jäncke[4,6] | Mireille Besson[1] | Stefan Elmer[4]

[1]CNRS & Aix-Marseille Univ, Laboratoire de Neurosciences Cognitives (LNC, UMR 7291), Marseille, France
[2]CNRS & Aix-Marseille Univ, Laboratoire Parole et Langage (LPL, UMR 7309), Aix-en-Provence, France
[3]Brain and Language Research Institute (BLRI), Aix-en-Provence, France
[4]Auditory Research Group Zurich (ARGZ), Division Neuropsychology, Institute of Psychology, University of Zurich, Zurich, Switzerland
[5]Sensory-Motor System Lab, Institute of Robotics and Intelligence Systems, Swiss Federal Institute of Technology, Zurich, Switzerland
[6]University Research Priority Program (URRP) "Dynamic of Healthy Aging", Zurich, Switzerland

Correspondence
Eva Dittinger, Laboratory of Cognitive Neuroscience, Centre Saint-Charles, Pole 3C, Case C, 3 place Victor Hugo, 13331 Marseille cedex 3, France.
Email: eva.dittinger@blri.fr and Stefan Elmer, Institute of Psychology, Division Neuropsychology, University of Zurich, Binzmühlestrasse 14/25, 8050 Zurich, Switzerland.
Email: s.elmer@psychologie.uzh.ch

Funding information
Labex BLRI, Grant/Award Number: ANR-11-LABX-0036; Institut Convergence ILCB, Grant/Award Number: ANR-16-CONV-0002; Swiss National Science Foundation, Grant/Award Number: SNF 320030_163149

Abstract

Current models of speech and language processing postulate the involvement of two parallel processing streams (the dual stream model): a ventral stream involved in mapping sensory and phonological representations onto lexical and conceptual representations and a dorsal stream contributing to sound-to-motor mapping, articulation, and to how verbal information is encoded and manipulated in memory. Based on previous evidence showing that music training has an influence on language processing, cognitive functions, and word learning, we examined EEG-based intracranial functional connectivity in the ventral and dorsal streams while musicians and nonmusicians learned the meaning of novel words through picture–word associations. In accordance with the dual stream model, word learning was generally associated with increased beta functional connectivity in the ventral stream compared to the dorsal stream. In addition, in the linguistically most demanding "semantic task," musicians outperformed nonmusicians, and this behavioral advantage was accompanied by increased left-hemispheric theta connectivity in both streams. Moreover, theta coherence in the left dorsal pathway was positively correlated with the number of years of music training. These results provide evidence for a complex interplay within a network of brain regions involved in semantic processing and verbal memory functions, and suggest that intensive music training can modify its functional architecture leading to advantages in novel word learning.

KEYWORDS
dorsal and ventral streams, functional connectivity, music training, plasticity, word learning

1 | INTRODUCTION

Biologically grounded models of speech and language processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2009, 2012; Hagoort, 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) postulate that a bilaterally organized ventral stream, projecting from the posterior superior temporal gyrus (i.e., pSTG) and the inferior parietal lobule (i.e., IPL) toward the temporal pole, contributes to mapping phonological representations onto lexical-semantic representations. The left pSTG has been proposed to support phoneme recognition based on low-level acoustic features, with a higher selectivity for

---

Mireille Besson and Stefan Elmer shared last authorship.

consonant-vowel (CV) syllables compared to other natural sounds (Jäncke, Wüstenberg, Scheich, & Heinze, 2002; Leaver & Rauschecker, 2010). In addition, the middle-posterior part of the middle temporal gyrus [i.e., MTG, Brodmann area (BA) 21] has been identified as an important region for word recognition (Binder et al., 1997; Friederici, 2012; Hickok & Poeppel, 2004; Saur et al., 2008), whereas the anterior temporal lobe accommodates lexical-semantic processing at the sentence-level and grammatical or compositional aspects of speech processing (DeWitt & Rauschecker, 2012; Friederici, 2012; Hickok & Poeppel, 2007).

The second main pathway involved in language processing is strongly lateralized to the left hemisphere and mediates impulse propagation between the pSTG and the IPL and Broca's area (Hickok & Poeppel, 2007; Rilling et al., 2008). This dorsal stream facilitates articulatory processes as well as the coupling of the speech signal with articulatory representations. Such sensorimotor integration mechanisms in the left hemisphere have previously been shown to support complex verbal memory functions (Rauschecker, 1998; Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009; Schulze, Vargha-Khadem, & Mishkin, 2012), for example, when it comes to imagine syllables (Tian, Zarate, & Poeppel, 2016) or to perform rhyming tasks (Pillay, Stengel, Humphries, Book, & Binder, 2014).

The IPL [comprising the angular and supramarginal gyri (AG/SMG)] constitutes a posterior point of intersection of the two processing streams. This brain region has previously been shown to house a variety of supramodal conceptual functions that are important for language, including pre-lexical-, lexical-, and verbal mnemonic functions, phonological categorization (Joanisse, Zevin, & McCandliss, 2007; Turkeltaub & Coslett, 2010), semantic processing (Ferreira, Göbel, Hymers, & Ellis, 2015; Obleser & Kotz, 2010; Price, 2010), short-term memory (Buchsbaum et al., 2011; Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003) and working memory (Buchsbaum et al., 2011). By contrast, Broca's area represents the anterior point of convergence of the two processing streams and has been proposed to accommodate both domain-general and language-selective functions (Fedorenko, Duncan, & Kanwisher, 2012). Regarding language processing, this territory has been associated with the planning and execution of articulation (Eickhoff, Heim, Zilles, & Amunts, 2009), linguistic- and phonological processes (Friederici, 2012; Rilling et al., 2008), short-term memory (Hickok et al., 2014), working-memory (Clos, Amunts, Laird, Fox, & Eickhoff, 2013; Fedorenko et al., 2012; Hagoort, 2014) and episodic- and semantic memory (Johnson, Saykin, Flashman, McAllister, & Sparling, 2001).

Nowadays, it is generally acknowledged that professional musicians can serve as a reliable model for better understanding how experiential factors interact with perceptual and higher cognitive functions (Münte, Altenmüller, & Jäncke, 2002). This field of research brought to light training-related functional (Zuk, Benjamin, Kenyon, & Gaab, 2014) and structural (Bermudez, Lerch, Evans, & Zatorre, 2009) changes in several brain regions situated along the two processing streams: in the pSTG (Bermudez et al., 2009; Elmer, Meyer, & Jancke, 2012) in the IPL (Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011), in the somatosensory and motor cortices (Bangert & Schlaug, 2006; Hyde et al., 2009), in the ventral (Sluming, Brooks, Howard, Downes, & Roberts, 2007) and dorsal part of the frontal lobe (Bermudez et al., 2009), and in different subregions of the temporal lobe (Gaser & Schlaug, 2003; Oechslin, Meyer, & Jäncke, 2010). In this context, it is noteworthy to mention that such plastic changes are not restricted to local brain regions but are rather part of complex neural networks. In fact, previous work showed that musicians are characterized by increased functional and structural connectivity between the left pSTG/IPL and Broca's area (Halwani, Loui, Rüber, & Schlaug, 2011; Klein, Liem, Hänggi, Elmer, & Jäncke, 2016; Oechslin, Imfeld, Loenneker, Meyer, & Jäncke, 2010), between the bilateral auditory-related cortices (Elmer, Hänggi, & Jäncke, 2016; Kühnis, Elmer, & Jäncke, 2014), between the auditory-related cortex and the somatosensory cortex (Klein et al., 2016), and within temporal brain regions (Jäncke, Langer, & Hänggi, 2012).

Previous studies have highlighted several brain regions that are responsive to both speech and music stimuli, including the primary and associative auditory cortex (Farbood, Heeger, Marcus, Hasson, & Lerner, 2015; Rogalsky, Rong, Saberi, & Hickok, 2011), the inferior parietal lobe and the ventral part of the prefrontal cortex (Farbood et al., 2015) as well as brain regions situated in the middle-anterior part of the temporal lobe (Farbood et al., 2015; Koelsch, 2011). Interestingly, several brain regions previously shown to be altered as a function of music training are distributed along the ventral and dorsal processing streams (Bermudez et al., 2009). Moreover, the influence of music training on the structural organization of the brain has also been shown to be accompanied by behavioral advantages in several domains of speech processing (see Besson, Chobert, & Marie, 2011 for review), including voice-onset time (VOT; Chobert, Francois, Velay, & Besson, 2014; Elmer et al., 2012), duration (Kühnis, Elmer, Meyer, & Jäncke, 2013; Tervaniemi et al., 2009), pitch (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011), timbre (Bidelman, Weiss, Moreno, & Alain, 2014), and linguistic and emotional prosody (Lima & Castro, 2011; Marques, Moreno, Castro, & Besson, 2007). Notably, recent results also revealed that both short- and long-term music training have a positive influence on diverse cognitive functions, including auditory- (Strait, Slater, O'Connell, & Kraus, 2015) and visual attention (Wang, Ossher, & Reuter-Lorenz, 2015), short-term-, working- and verbal memory (George & Coch, 2011; Schulze & Koelsch, 2012), and on general executive functions (Zuk et al., 2014). Based on such a multifaceted influence of music training on speech processing and cognition, Dittinger and collaborators recently examined the influence of music training on novel word learning in adults (Dittinger et al., 2016) and in 8–12-year-old children (Dittinger, Chobert, Ziegler, & Besson, 2017). Thereby, the authors focused on the development of the N400 component and on the N400 effect (unexpected minus expected words) as an index of novel word learning (frontal N400 or FN400 in a learning phase) and retrieval from episodic and semantic memory (in matching and semantic tasks). As a main result, the authors revealed a behavioral advantage of musicians and musically-trained children during the cognitively most demanding "semantic task" that was associated with larger N400 effects over centro-posterior scalp sites. These findings are particularly interesting in that they lead to suggest a relationship between music training and semantic facilitation during word learning.

In this EEG study, we re-evaluated the data of the professional musicians and non-musicians of the Dittinger et al. (2016) study by focusing on intracranial functional connectivity in the theta, alpha, and beta frequency bands to examine the dynamic interplay between key regions of the ventral- (i.e., between the IPL and middle part of the middle temporal lobe) and dorsal streams (i.e., between the IPL and Broca's region). Although the functional role of these different frequency bands is not yet well understood, low frequency oscillations in the theta range have been proposed to support neural communication between distinct brain regions over long-range circuits (Polanía, Nitsche, Korman, Batsikadze, & Paulus, 2012; Ward, 2003), information integration (Ward, 2003), and mnemonic processes (Albouy, Weiss, Baillet, & Zatorre, 2017). Theta oscillations may also contribute to "packing" the multitime speech (Giraud & Poeppel, 2012) and possibly music signal (Doelling & Poeppel, 2015) into units of the appropriate temporal granularity and to the processing of syntactic and semantic violations (Bastiaansen, Magyari, & Hagoort, 2010; Carrus, Koelsch, & Bhattacharya, 2011). Otherwise, beta oscillations have been shown to be involved in a wide range of cognitive functions such as working memory (Siegel, Warden, & Miller, 2009) or decision making (Wong, Fabiszak, Novikov, Daw, & Pesaran, 2016), and to language processing (Carrus et al., 2011). They also seem to play an important role in music processing and are increased with music training (Doelling & Poeppel, 2015). Finally, alpha oscillations may support short-term and working memory functions (Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012) and anticipation and expectation processes (Weisz, Hartmann, Müller, Lorenz, & Obleser, 2011).

According to the results of Dittinger et al. (2016), and based on the literature reviewed above, we expected increased connectivity in musicians in the bilateral ventral stream which contributes to access lexical-semantic information. Furthermore, we predicted that neural oscillations in the left dorsal stream, which has been shown to be functionally (Klein et al., 2016) and structurally (Halwani et al., 2011) influenced by music training, will be more strongly synchronized for facilitating the preactivation of the phonological word forms, sensory-to-motor coupling mechanisms (Cunillera et al., 2009; López-Barroso et al., 2013), and possibly the manipulation of information in working memory (Albouy et al., 2017).

2 | MATERIALS AND METHODS

2.1 | Participants

According to previous work (Elmer et al., 2012; Klein et al., 2016), we applied strict criteria and only evaluated the data of musicians who started music training before the age of 10 years. Consequently, one subject was excluded from the analyses (i.e., age of commencement = 15 years). The professional musicians (MUS, totally 14, 8 women; 3 pianists, 2 accordionists, 4 violinists, 1 cellist, 1 guitarist, 1 hornist, 1 tubist, and 1 flautist; mean number of practice years = 17.4, range = 11–24, $SD = 4.1$) and nonmusicians (NM, totally 15, 8 women) who participated in the present study did not differ in age (MUS: mean age = 25.0 years, age range = 19–30, $SD = 4.0$; NM: mean age = 25.7 years, age range = 19–33, $SD = 4.8$; $F_{(1,27)} = .20$, $p = .66$) and were comparable in education ($F_{(1,27)} = 0.48$, $p = .50$) and in socio-economic background (criteria of the National Institute of Statistics and Economic Studies; MUS = 4.4; NM = 4.9; $F_{(1,27)} = 1.05$, $p = .31$). All participants were native French speakers and none of them was bilingual but all had some knowledge of English as a second language. Most importantly, none of the participants had previous experience with the Thai language. None of the subjects reported past or current audiological or neurological deficits. The study was conducted in accordance with the declaration of Helsinki, and all participants gave their informed consent before enrolling in the experiment.

2.2 | Cognitive capability and musical aptitudes

Standardized psychometric tests were used to assess between-group differences in short-term and working memory [forward and reverse Digit Span, WISC-IV (Wechsler, 2003)], visual attention [NEPSY (Korkman, Kirk, & Kemp, 1998)], and fluid intelligence [progressive matrices, PM47 (Raven, Corporation, & Lewis, 1962)]. Musical aptitudes were quantified by means of two subtests [adapted from the MBEA battery (Peretz, Vuvan, Lagrois, & Armony, 2015)] consisting of listening to short pairs of piano sequences and deciding whether the sequences were rhythmically or tonally same or different.

2.3 | Stimulus material

2.3.1 | Auditory stimuli

Auditory stimuli consisted of nine natural monosyllabic Thai words, namely/ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/, /p^{h}a:1/, /ba:0/, /pa:0/, /p^{h}a:0/. The words varied in vowel duration, with short- (/ba1/, /pa1/ and /p^{h}a1/; 261 ms on average) and long vowels (/ba:1/, /pa:1/, /p^{h}a:1/, /ba:0/, /pa:0/ and /p^{h}a:0/; 531 ms on average). They also varied in fundamental frequency, with low tone- (/ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/ and /p^{h}a:1/; F0 = 175 Hz on average) and mid tone vowels (/ba:0/, /pa:0/ and /p^{h}a:0/; F0 = 218 Hz on average). Furthermore, words contained voicing contrasts (/ba1/, /ba:1/ and /ba:0/, VOT = −144 ms *vs.*/pa1/, /pa:1/ and /pa:0/, VOT = 3 ms) as well as aspirated contrasts (/pa1/, /pa:1/ and /pa:0/, VOT = 3 ms *vs.*/p^{h}a1/, /p^{h}a:1/ and /p^{h}a:0/, VOT = 77 ms). The stimuli were registered by a female Thai-French bilingual speaker, digitally recorded, and normalized to a mean level of 70 dB by using the Praat software (http://www.fon.hum.uva.nl/praat/).

2.3.2 | Visual stimuli

Nine pictures representing familiar objects (i.e., bear, flower, key, chair, bell, eye, strawberry, train, and glass) were selected from the set of 260 pictures built by Snodgrass and Vanderwart (1980). These pictures are black-and-white line drawings standardized on name agreement, image agreement, familiarity, and visual complexity. For the "semantic task" (i.e., see the experimental procedure), new pictures that the participants had not seen before in the experiment and that were semantically related or unrelated to the original pictures (and, consequently, to the novel words meaning) were chosen from the internet by two of the authors (ED and

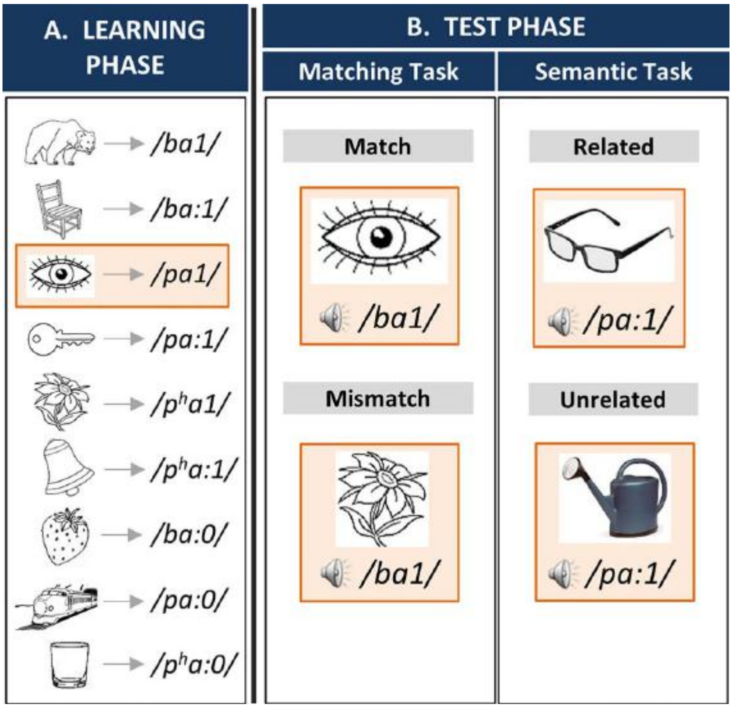


FIGURE 1 Schematic representation of the experimental design [Color figure can be viewed at wileyonlinelibrary.com]

MB). Sixty students from our university (age range = 19–25 years) were asked to rate the semantic relatedness between new and old pictures on a scale from 1 (= unrelated) to 7 (= related). Half of the presented pairs were semantically related and the other half were semantically unrelated. Overall, this was confirmed by the students' ratings. All pictures were presented with an angular size of 24.9°.

2.4 | Experimental procedure

Encoding and retrieval of newly learned words was assessed by means of three consecutive tasks consisting of an "associative learning phase," a "matching task," and a "semantic task" (Figure 1). During the "associative learning phase," participants had to learn the meaning of auditory words through picture-word associations. Each of the nine picture-word pairs was presented 20 times, resulting in 180 trials that were pseudorandomly presented within two blocks of 3 min each. Each picture was presented on the screen for 2000 ms (total trial duration), and one out of the nine words was binaurally presented with a stimulus onset asynchrony (SOA) of 750 ms. Picture–word associations were different across participants to avoid spurious effects due to specific associations and serial order effects. Note that the picture–word associations were arbitrary in that the learned meaning did not correspond to the real meaning of the Thai words. In the "associative learning phase," no behavioral response was required from the participants. During the "matching task", pictures were presented on the screen followed by a word (SOA = 750 ms, total trial duration = 3750 ms) that matched or mismatched the associations previously learned, and participants were asked to press one out of two response keys, accordingly. Response hand was counter-balanced across participants. A row of "x" was presented 2750 ms after picture onset, and participants were asked to blink during this time period (1000 ms) to minimize artefacts during word presentation and maximize power for independent component analysis (i.e., ICA-based artefacts correction). Each word was pseudorandomly presented 20 times, half in match- and half in mismatch conditions. A total of 180 trials were presented in two blocks of 5.6 min each. Finally, in the "semantic task," we tested for semantic integration and retrieval of the newly learned words by presenting new pictures that the participants had not seen before and that were semantically related or unrelated to the meaning of the words previously learned. In each trial, a new picture was followed by a word (SOA = 1500 ms, total trial duration = 4500 ms), and participants were asked to press one out of two response keys depending upon whether the word was semantically related or unrelated to the new picture. Response hand was counterbalanced across participants, and a familiarization task including four trials was administrated before starting the task. A row of "x" was presented 3500 ms after picture onset, and participants were instructed to blink during this period (1000 ms). Each word was presented 12 times but none of the new pictures was repeated so that on each trial the word was associated with different related (i.e., 50%) or unrelated (i.e., 50%) pictures. A total of 108 trials were presented pseudorandomly within two blocks of 4 min each. In all three tasks, auditory stimuli were presented through HiFi headphones

(HD590, Sennheiser Electronic GmbH, Wedemark, Germany) at 70dB sound pressure level and stimulus presentation, and the collection of behavioral responses, were controlled by the "Presentation" software (NeuroBehavioral Systems, Version 11.0).

2.5 | EEG data acquisition

EEG was continuously recorded with a sampling rate of 512 Hz and a band-pass filter of 0–102.4 Hz by using a Biosemi amplifier system (Amsterdam, BioSemi Active 2). Thirty-two active Ag/Cl electrodes (Biosemi Pintype) were located at standard positions according to the international 10/20 System. EEG recordings were referenced on-line to a common electrode (CMS) included in the headcap (next to Cz). Two additional electrodes were placed on the left and right mastoids. The electrooculogram (EOG) was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi, and from an electrode beneath the right eye. Electrode impedance was kept below 5 kΩ. EEG preprocessing was done by using the Brain Vision Analyzer software (Version 1.05.0005 & Version 2.1.0; Brain Products, GmbH). Thereby, all data were rereferenced off-line to the averaged left and right mastoids, filtered with a bandpass filter of 1–30 Hz (slope of 24 dB/oct), and independent component analysis (ICA) and inverse ICA were used to identify and remove components associated with vertical and horizontal ocular movements. Finally, DC detrends and artifacts above a gradient criterion of 10 μV/ms or a max–min criterion of 100 μV were automatically removed. For each participant, EEG responses to the pictures were segmented into single sweeps of 950 ms, including a pre-stimulus baseline of 200 ms. After baseline correction, the pre-stimulus period was removed, and the single sweeps were subjected to functional connectivity analyses by using the sLORETA toolbox (http://www.uzh.ch/keyinst/loreta.htm). In the present work, we explicitly abstained from evaluating functional connectivity during the period of acoustic stimulation (i.e., the words) because such a procedure would reflect the alignment of neural oscillations due to the often-reported increased auditory acuity of professional musicians rather than capture word learning per se.

2.6 | Functional connectivity analyses

Functional connectivity (sLORETA software package; http://www.uzh.ch/keyinst/loreta.htm) was evaluated using lagged coherence values as a measure of variability between two oscillatory time series in specific frequency bands (Lehmann, Faber, Gianotti, Kochi, & Pascual-Marqui, 2006; Thatcher, 2012). Lagged coherence reflects the coherence between two EEG time series independent from volume conduction since the instantaneous zero-lag contribution has been excluded mathematically (Nolte et al., 2004; Stam & van Straaten, 2012). Such a correction is desired when using estimated intracranial signals (EEG tomography), because zero-lag connectivity in a given frequency band is often due to nonphysiological effects or intrinsic physical artifacts (e.g., volume conduction and low spatial resolution) that usually affect connectivity indices. Thus, this measure of functional connectivity is thought to contain only physiological information.

In the current implementation of sLORETA (Pascual-Marqui et al., 2011), EEG data are rereferenced to an average reference, and computations are made within a realistic head model (Fuchs, Kastner, Wagner, Hawes, & Ebersole, 2002) relying on the Montreal Neurological Institute (MNI) 152 template (Mazziotta et al., 2001). The 3D solution space is restricted to cortical gray matter, as determined by the probabilistic Talairach atlas (Lancaster et al., 2000). The intracranial volume is partitioned in 6,239 voxels at 5 mm spatial resolution. sLORETA images represent the electric activity at each voxel in the neuroanatomic MNI space as the exact magnitude of the estimated current density. Anatomical labels and BAs are reported using MNI space, with correction to Talairach space (Brett, Johnsrude, & Owen, 2002).

Based on the specific hypotheses described in the introduction, we selected three regions of interest (ROI) in each hemisphere. These three ROIs (Figure 2a–c) consisted of BA 39/40 (ROI 1, AG/SMG), BA 44/45 (ROI 2, Broca's area), and BA 21 (ROI 3, middle part of the middle temporal gyrus, MTG). For functional connectivity analyses in the ventral and dorsal streams, a method using a single voxel at the centroid of the ROIs was chosen. Details on sLORETA connectivity algorithms can be found in previous reports by Pascual-Marqui et al. (2011). For each group and for each task, functional connectivity during the first 750 ms after picture presentation was computed in the theta (4–7 Hz), alpha (8–12 Hz), and beta frequency range (13–20 Hz).

2.7 | Statistical analyses

All statistical analyses were performed by using IBM SPSS statistics (version 22). Psychometric data were evaluated by means of univariate ANOVAs, and musical aptitudes were assessed by a 2 × 2 ANOVA (i.e., two groups and two subtests). Error rates (ERRs) and reaction times (RTs) were evaluated for both the "matching task" and the "semantic task" by means of separate univariate ANOVAs [corrected for multiple comparisons by using the Bonferroni procedure, Bonferroni-corrected p value for 2 tests (i.e., 2 tasks) = 0.03]. Functional connectivity was assessed in all three tasks (i.e., "associative learning phase," "matching task," and "semantic task") by means of separate 2 × 2 × 2 ANOVAs (i.e., two groups, two streams, and two hemispheres) for each frequency band [corrected for multiple comparisons, Bonferroni-corrected p value for 9 tests (i.e., 3 tasks × 3 frequency bands) = 0.005]. Finally, correlation analyses between the error rate in the semantic task and the cumulative number of training years (normally distributed, see Figure 5a), as well as left-hemispheric theta connectivity, were computed according to Spearman's rho (one-tailed) separately for the ventral and dorsal streams.

3 | RESULTS

3.1 | Cognitive ability and musical aptitudes

Musicians and nonmusicians did not significantly differ in terms of fluid intelligence ($F_{(1,\ 27)} = .93$, $p = .34$), verbal working memory (i.e., reverse digit span; $F_{(1,\ 27)} = 2.25$, $p = .15$), or visual attention ($F_{(1,\ 27)} = 3.89$, $p = .06$). However, musicians scored better on the short-term memory

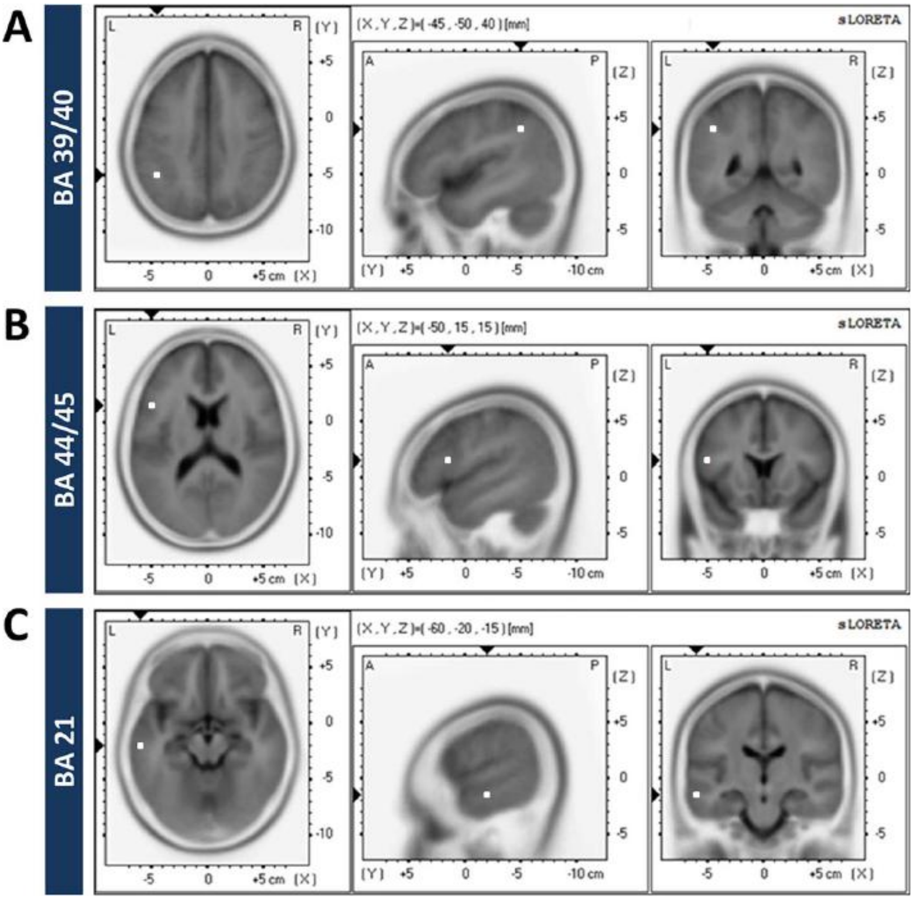


FIGURE 2 Left-hemispheric ROIs positions within the three-dimensional MNI space. (a) Centroid voxel corresponding to BAs 39/40; (b) centroid voxel corresponding to BAs 44/45; (c) centroid voxel corresponding to BA 21 [Color figure can be viewed at wileyonlinelibrary.com]

test (i.e., forward digit span test; $F_{(1,\ 27)} = 4.84$, $p = .04$), and outperformed nonmusicians on both tonal and rhythmical subtests of musical aptitudes (main effect of Group: $F_{(1,27)} = 14.01$, $p < .001$).

3.2 | Behavioral data

The evaluation of ERRs and RTs during the "matching task" did not reveal significant group differences (ERRs: $F_{(1,27)} = .28$, $p = .60$, and RTs: $F_{(1,27)} = .15$, $p = .70$; see Figure 3a). By contrast, in the "semantic task" musicians made significantly fewer errors than nonmusicians ($F_{(1,27)} = 5.02$, $p = .03$, Bonferroni-corrected; see Figure 3b), whereas the two groups did not differ in RTs ($F_{(1,27)} = .44$, $p = .51$, see Figure 3b).

3.3 | Functional connectivity

Functional connectivity did not differ between the two groups in the "associative learning phase." By contrast, in both the "matching task" and in the "semantic task," results revealed a main effect of stream in the beta frequency range that originated from increased connectivity in the bilateral ventral stream compared to the dorsal stream ("matching task": $F_{(1,27)} = 19.97$, $p < .001$ and "semantic task": $F_{(1,27)} = 14.26$, $p < .001$, Bonferroni-corrected; see Figure 4a). Moreover, in the "semantic task" we also found a significant group × hemisphere interaction effect in the theta frequency range ($F_{(1,27)} = 9.37$, $p = .005$, Bonferroni-corrected). Post-hoc univariate ANOVAs for each hemisphere yielded a significant group difference in the left ($F_{(1,27)} = 5.66$, $p = .02$) but not in the right ($F_{(1,27)} = .21$, $p = .65$) hemisphere that originated from increased theta connectivity in musicians compared to nonmusicians (Figure 4b). Finally, to further explore hemispherical contributions within the two groups, we tested the asymmetry index [i.e., asymmetry index = (connectivity left − connectivity right)/(connectivity left + connectivity right) × 0.5] against zero (i.e., no asymmetry). Results showed symmetrical hemispheric theta connectivity within the musicians group (one-sample t test against zero, $t_{(13)} = .31$, $p = .76$, two-tailed), whereas nonmusicians were characterized by a right-sided asymmetry (one-sample t test against zero, $t_{(14)} = -4.53$, $p < .001$, two-tailed).

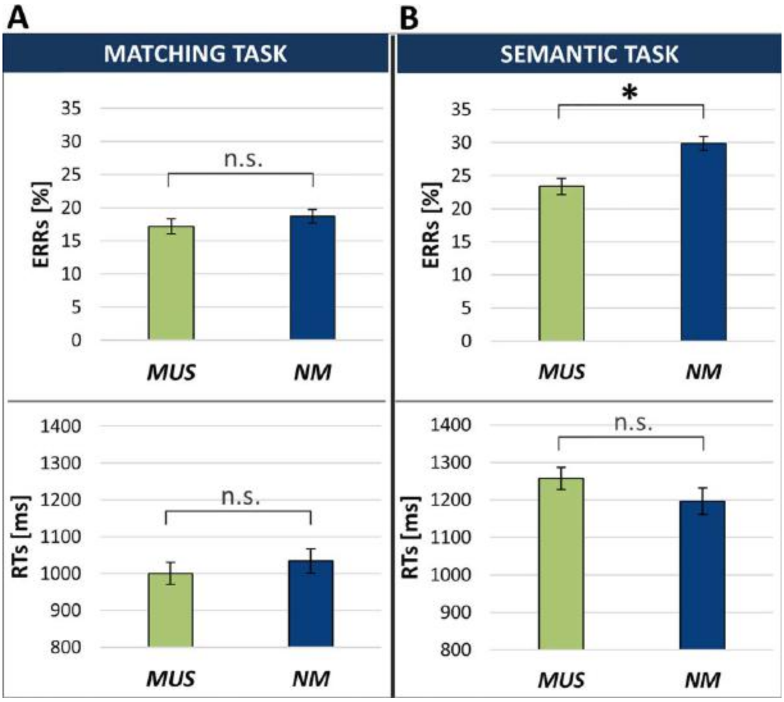


FIGURE 3 Behavioral results (ERRs: upper part and RTs: lower part) are shown separately for musicians (MUS) and nonmusicians (NM). (a) Matching task; (b) Semantic task. The bars depict SE of mean. *$p < .05$, n.s. = not significant [Color figure can be viewed at wileyonlinelibrary.com]

3.4 | Relationship between functional connectivity and music training

To verify that the increased left-hemispheric theta connectivity we revealed in musicians was related to training, we performed additional correlative analyses. In particular, within the musicians group we correlated the number of training years with left-sided theta connectivity in the "semantic task," separately for the ventral and the dorsal streams (i.e., according to Spearman's rho for non-Gaussian distributed data, one-tailed, Figure 5a). Results revealed a significant positive relationship

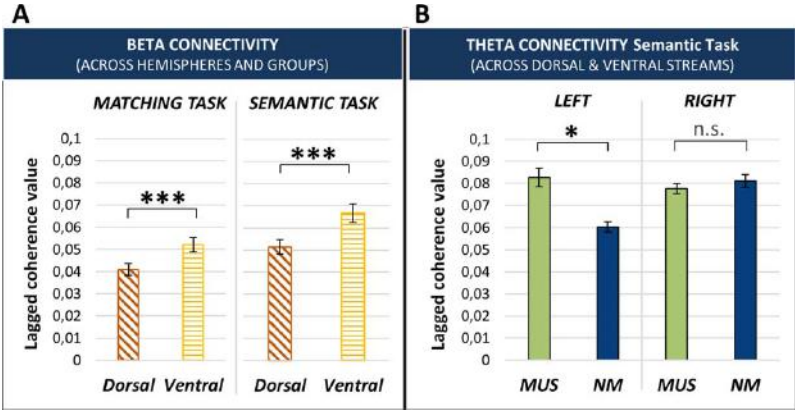


FIGURE 4 Mean functional connectivity. (a) Beta connectivity averaged across hemispheres and groups is shown separately for the tasks ("matching task" and "semantic task") and processing streams (dorsal; ventral). (b) Theta connectivity during the "semantic task" averaged across processing streams is shown separately for the hemispheres ("left" and "right"), and groups (MUS; NM). Mean values with SE bars. *$p < .05$; ***$p < .001$; n.s. = not significant [Color figure can be viewed at wileyonlinelibrary.com]

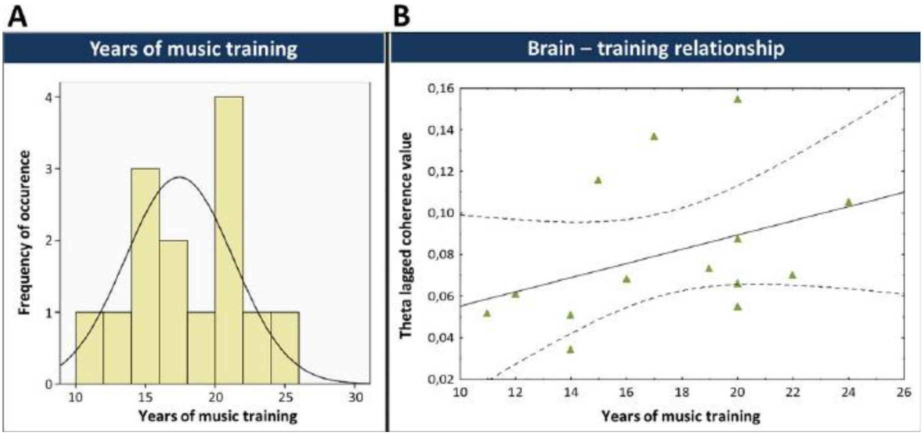


FIGURE 5 (a) Bimodal distribution of the variable "cumulative number of training years" with a superimposed Gaussian function. (b) Significant positive correlation between the left-hemispheric theta connectivity in the dorsal stream during the "semantic task" and the cumulative number of training years (one-tailed) [Color figure can be viewed at wileyonlinelibrary.com]

for the dorsal stream ($r_{(14)} = .52$, $p = .03$, Figure 5b) but not for the ventral stream ($r_{(14)} = .08$, $p = .39$). By contrast, no significant correlations were obtained between error rates in the semantic task and number of training years ($r_{(14)} = .24$, $p = .21$, one-tailed) or left-sided theta connectivity (dorsal stream: $r_{(28)} = -.12$, $p = .27$; ventral stream: $r_{(28)} = -.12$, $p = .28$, one-tailed).

4 | DISCUSSION

In this study, we compared novel word learning in musicians and nonmusicians by using three concatenated tasks that are dependent upon a variety of perceptual and cognitive functions typically involved in both music and language processing (Dittinger et al., 2016). Based on previous evidence pointing to an optimization of perceptual discrimination (Chobert et al., 2014; Marie et al., 2011), short-term and working memory (George & Coch, 2011; Schulze & Koelsch, 2012), executive functions (Zuk et al., 2014), and audio–visual integration (Bidelman, 2016; Paraskevopoulos, Kraneburg, Herholz, Bamidis, & Pantev, 2015) in musicians, we predicted that the musicians' advantage in word learning would translate into increased EEG-based functional connectivity along the ventral- and dorsal processing streams. Results consistently revealed increased beta coherence in the ventral compared to the dorsal stream during both the "matching-" and "semantic tasks." Moreover, musicians outperformed nonmusicians during the cognitively most demanding "semantic task" and exhibited increased theta coherence over the left hemisphere compared to nonmusicians. Finally, left-dorsal functional connectivity in the theta frequency range was positively correlated with the number of years of music training, leading to suggest training-related adaptations. In turn, we will discuss these findings in more details by integrating the results in neuro-linguistic and musical frameworks.

4.1 | Neural dynamics underlying word learning and semantic generalization

As expected and in line with current models of speech and language processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2012; Hagoort, 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), results generally revealed increased beta functional connectivity in the ventral stream compared to the dorsal stream during both the "matching- "and "semantic tasks." This connectivity pattern is taken to reflect the neural dynamics underlying novel word learning and, specifically the mapping of phonological information with lexical-semantic representations. Several arguments support this interpretation. First, increased activity in the IPL was observed during the learning of new phonological associations (i.e., using pseudowords; Breitenstein et al., 2005) and during the retrieval of previously learned novel words (Cornelissen et al., 2004; Hultén, Laaksonen, Vihla, Laine, & Salmelin, 2010). In addition, activity in the IPL has been shown to be modulated by the semantic richness of words (Ferreira et al., 2015), leading to suggest that this region also contributes to the retrieval of semantic information (Obleser & Kotz, 2010; Price, 2010). Second, the middle-posterior part of the MTG is involved in mediating lexical-semantic access at the word level (Friederici, 2012; Hickok & Poeppel, 2007), and the retrieval of semantic information in word learning tasks (Mestres-Missé, Càmara, Rodriguez-Fornells, Rotte, & Münte, 2008). For example, activity in the MTG is increased when participants retrieve meaningful word-forms compared to word-forms without meaning (Takashima, Bakker, van Hell, Janzen, & McQueen, 2014). Third, increased beta coherence during word encoding has previously been shown to be predictive of successful retrieval (Weiss & Rappelsberger, 2000). In addition, it has been proposed that beta oscillations contribute to the maintenance of a cognitive state (Engel & Fries, 2010), as for example, in the context of repetition priming paradigms (Ghuman, Bar, Dobbins, & Schnyer, 2008). Taken together, these results emphasize a crucial role of beta oscillations in mediating learning and retrieval of newly learned words from semantic memory.

4.2 | Novel word learning and musical expertise

In line with our main hypothesis, musicians outperformed nonmusicians in the most difficult "semantic task," in which pictures that had not been presented before in the experiment were paired with the newly learned words to test for semantic generalization. This finding is not surprising since difficult tasks are typically more sensitive for detecting expertise-related influences than simple ones (Schön, Magne, & Besson, 2004). The positive influence of music training on semantic processing extends previous findings showing a facilitation of speech perception and syntactic processing in musicians compared to nonmusicians (Besson et al., 2011; Gordon, Fehd, & McCandliss, 2015; Jentschke & Koelsch, 2009). While the specific origin of the behavioral advantage we revealed in musicians is still unclear, Dittinger et al. (2016) proposed two complementary interpretations, namely "cascading effects" and "multidimensional facilitation effects." According to "cascading effects," the behavioral advantage of musicians may be driven by enhanced auditory perception that facilitates the building-up of new phonological representations and lexical-semantic access (bottom–up effects). Otherwise, the multidimensional nature of music training may also improve several cognitive functions such as picture–word integration, short- and long-term memory that are necessary for word learning (top–down effects).

4.3 | Musicians exhibit increased theta coherence in the left hemisphere

In the "semantic task," results revealed increased theta coherence in musicians compared to nonmusicians in the left but not in the right hemisphere, irrespective of processing stream. However, this finding does not necessarily mean that the left hemisphere was more strongly involved than the right one, and especially it does not preclude a right-hemispheric contribution to language learning (Price, 2010; Vigneau et al., 2011). In fact, the additional statistical analyses performed on asymmetry index data clearly showed symmetrical theta connectivity in musicians and a right-sided asymmetry in nonmusicians. The symmetrical recruitment of the two processing streams in musicians is, at least in part, compatible with results showing that individuals with a more symmetrical distribution of the arcuate fasciculus performed better at remembering previously learned words compared to those characterized by a strong left-hemispheric asymmetry (Catani et al., 2007). Moreover, the right-sided theta asymmetry found in nonmusicians is in line with the results of Takashima et al. (2014) showing higher right- than left-hemispheric connectivity between the STG and the MTG for successfully recognized words associated with complex and colored pictures. Even though the different laterality effects we observed as a function of music training are relevant for better comprehending experiential influences, the underlying mechanisms still need to be worked out.

4.4 | Theta coherence in the left dorsal stream is increased in musicians compared to nonmusicians

In line with previous literature indicating an influence of music training on the functional (Klein et al., 2016) and structural (Halwani et al., 2011; Oechslin et al., 2010) architecture of the left dorsal stream, theta coherence was increased between the left IPL and Broca's area in musicians compared to nonmusicians. In addition, functional connectivity in the left dorsal stream was correlated with the number of years of music training, thereby pointing to an influence of music training on the neural dynamics involved in novel word learning. As increased functional connectivity in the left dorsal stream was paralleled by a behavioral advantage in the "semantic task," results are interpreted as indicating a contribution of this connectivity pattern to task-related mnemonic functions. In this context, there are several alternative explanations that may account for the mnemonic advantage of musicians. A first possibility is that increased theta synchronization between the left IPL and Broca's area contributed to build more robust multimodal memory traces by linking novel phonological representations to the respective articulatory codes (López-Barroso et al., 2013). Such sensory-to-motor coupling mechanisms are required for retrieving the phonological codes of new words (i.e., Thai) that are not part of the phonetic repertoire of the mother tongue (i.e., French). A second possibility is that musicians possibly preactivated the target words corresponding to the pictures by using articulatory strategies. A similar mechanisms has previously been described by Tian and collaborators (Tian et al., 2016) who found increased activity in a left frontal-parietal sensorimotor network when participants were instructed to imaging consonant-vowel syllables by using an articulation-based imagery strategy compared to a hearing-based one. Finally, based on recent results, it may be that increased theta oscillations in the left dorsal stream of musicians are related to the manipulation of acoustic information in working memory (Albouy et al., 2017). In fact, working memory is necessarily required for retrieving from memory the old learned pictures semantically related to the new ones while at the same time selecting the corresponding target word. Certainly, all these lines of argumentation are not conclusive, and can only be proofed by applying word learning tasks specifically targeting at increasing computational demands on sensory-to-motor coupling mechanisms, articulatory-based imagery strategies, and working memory load in a parametric manner.

5 | LIMITATIONS

A limitation of this study is that we used a source estimation approach for evaluating functional connectivity between a-priori defined brain regions without taking into account the underlying grey- and white-matter architecture. Consequently, we cannot exclude that the alignment of neural oscillations between the ROIs was indirectly mediated by other brain regions. Therefore, future studies should try to combine functional and structural connectivity with morphometric analyses in order to better comprehend the differential contribution of the dorsal and ventral streams to word learning in individuals with and without music training. In addition, the results of the present study were interpreted according to sound-to-meaning and sound-to-articulation mapping mechanisms. However, we are aware that other distinctions have been proposed by several authors (Bornkessel-Schlesewsky & Schlesewsky, 2013; see Cloutman, 2013 for review) by taking into account time-variant (i.e., dorsal stream) and time-invariant (i.e., ventral stream) neural computation in the two processing streams. According to this framework, the coupling of speech sounds with articulation and working memory functions may rely on time-dependent processes in the dorsal stream, whereas the building-up of word representations through sound-to-meaning mapping mechanisms may engage the time-independent ventral stream. Finally, increased theta functional connectivity in both the ventral and dorsal processing streams in professional musicians compared to nonmusicians does not necessarily imply that music training was the cause of these findings. For example, while the correlation between left theta connectivity in the dorsal stream and the number of years of music training was significant, no such correlation was found with the level of performance in the semantic task. Therefore, longitudinal experiments with nonmusicians are needed to demonstrate a causal link between music training and increased functional connectivity.

6 | CONCLUSIONS

In line with previous models of speech processing (Friederici, 2012; Hickok & Poeppel, 2004, 2007), we found that the bilateral ventral stream generally contributed to novel word learning, irrespective of musical expertise. In addition, we provided evidence for a relationship between the behavioral advantages of musicians in the "semantic task" and the alignment of theta oscillations in the left-sided ventral- and dorsal streams. This specific connectivity pattern was interpreted as reflecting facilitated access to lexical-semantic information though the engagement of articulation strategies and manipulation of information in working memory. Our results not only constitute a first step toward a better understanding of the influence of music training on dynamic neural networks during word learning, but also complement previous knowledge about the contribution of the ventral and dorsal streams to word learning.

ACKNOWLEDGMENTS

The authors would like to thank all the participants for having contributed to the experiment, and Chotiga Pattamadilok for registering the auditory stimuli. Furthermore, they are indebted to Mylène Barbaroux and Benjamin Furnari for their help in data acquisition. ED's and MB's work, carried out within the CNRS, the Labex BLRI (ANR-11-LABX-0036) and the Institut Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government, managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). SE and LJ were supported by the Swiss National Science Foundation (grant SNF 320030_163149 to LJ).

CONFLICT OF INTEREST

The authors declare no competing financial interests.

ORCID

Eva Dittinger http://orcid.org/0000-0002-5121-096X
Stefan Elmer http://orcid.org/0000-0003-1721-450X

REFERENCES

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*, *94*(1), 193–206.e5.

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *The European Journal of Neuroscience*, *24*(6), 1832–1834.

Bastiaansen, M., Magyari, L., & Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, *22*(7), 1333–1347.

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*, *19*(7), 1583–1596.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, *2*.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: Experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, *234*(10), 3037–3047.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662–2673.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(1), 353–362.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: A new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*(3), 958–968.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249.

Buchsbaum, B. R., Baldo, J., Okada, K., Berman, K. F., Dronkers, N., D'esposito, M., & Hickok, G. (2011). Conduction aphasia, sensory-motor integration, and phonological short-term memory - an aggregate analysis of lesion and fMRI data. *Brain and Language*, *119*(3), 119–128.

Carrus, E., Koelsch, S., & Bhattacharya, J. (2011). Shadows of music–language interaction on low frequency brain oscillatory patterns. *Brain and Language*, *119*(1), 50–57.

Catani, M., Allin, M. P. G., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., & Jones, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17163–17168.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). 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*, *24*(4), 956–967.

Clos, M., Amunts, K., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44. *NeuroImage*, *83*, 174–188.

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, *127*(2), 251–263.

Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language*, *89*(3), 617–622.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., . . . Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*(3), 541–553.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, *109*, E505–E514.

Dittinger, E., Barbaroux, M., D'imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience*, *11*.

Doelling, K. B., & Poeppel, D. (2015). Cortical entrainment to music and its modulation by expertise. *Proceedings of the National Academy of Sciences*, *112*(45), E6233–E6242.

Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *367*(1896), 2399–2421.

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, *221*(1), 331–344.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, *22*(3), 650–658.

Engel, A. K., & Fries, P. (2010). Beta-band oscillations–signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165.

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience*, *9*,

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology: CB*, *22*(21), 2059–2062.

Ferreira, R. A., Göbel, S. M., Hymers, M., & Ellis, A. W. (2015). The neural correlates of semantic richness: Evidence from an fMRI study of word learning. *Brain and Language*, *143*, 69–80.

Friederici, A. D. (2009). Pathways to language: Fiber tracts in the human brain. *Trends in Cognitive Sciences*, *13*(4), 175–181.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16* (5), 262–268.

Fuchs, M., Kastner, J., Wagner, M., Hawes, S., & Ebersole, J. S. (2002). A standardized boundary element method volume conductor model. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(5), 702–712.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory–an fMRI study with sparse temporal sampling. *NeuroImage*, *19*(4), 1417–1426.

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), 9240–9245.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.

Ghuman, A. S., Bar, M., Dobbins, I. G., & Schnyer, D. M. (2008). The effects of priming on frontal-temporal communication. *Proceedings of the National Academy of Sciences of the United States of America*, *105* (24), 8405–8409.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in Psychology*, *6*.

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

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Frontiers in Psychology*, *2*, 156.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

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

Hickok, G., Rogalsky, C., Chen, R., Herskovits, E. H., Townsley, S., & Hillis, A. E. (2014). Partially overlapping sensorimotor networks underlie speech praxis and verbal short-term memory: Evidence from apraxia of speech following acute stroke. *Frontiers in Human Neuroscience*, *8*,

Hultén, A., Laaksonen, H., Vihla, M., Laine, M., & Salmelin, R. (2010). Modulation of brain activity after learning predicts long-term memory for words. *Journal of Neuroscience*, *30*(45), 15160–15164.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical training shapes structural brain development. *Journal of Neuroscience*, *29*(10), 3019–3025.

Jäncke, L., Langer, N., & Hänggi, J. (2012). Diminished whole-brain but enhanced peri-sylvian connectivity in absolute pitch musicians. *Journal of Cognitive Neuroscience*, *24*(6), 1447–1461.

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H.-J. (2002). Phonetic perception and the temporal cortex. *NeuroImage*, *15*(4), 733–746.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, *47*(2), 735–744.

Joanisse, M. F., Zevin, J. D., & McCandliss, B. D. (2007). Brain mechanisms implicated in the preattentive categorization of speech sounds revealed using FMRI and a short-interval habituation trial paradigm. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(9), 2084–2093.

Johnson, S. C., Saykin, A. J., Flashman, L. A., McAllister, T. W., & Sparling, M. B. (2001). Brain activation on fMRI and verbal memory ability: Functional neuroanatomic correlates of CVLT performance. *Journal of the International Neuropsychological Society*, *7*(1), 55–62.

Klein, C., Liem, F., Hänggi, J., Elmer, S., & Jäncke, L. (2016). The "silent" imprint of musical training. *Human Brain Mapping*, *37*(2), 536–546.

Koelsch, S. (2011). Toward a neural basis of music perception – A review and updated model. *Frontiers in Psychology*, *2*.

Korkman, M., Kirk, U., & Kemp, S. (1998). *NEPSY: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Kühnis, J., Elmer, S., & Jäncke, L. (2014). Auditory evoked responses in musicians during passive vowel listening are modulated by functional connectivity between bilateral auditory-related brain regions. *Journal of Cognitive Neuroscience*, *26*(12), 2750–2761.

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, *51*(8), 1608–1618.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *30*(22), 7604–7612.

Lehmann, D., Faber, P. L., Gianotti, L. R. R., Kochi, K., & Pascual-Marqui, R. D. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *Journal of Physiology, Paris*, *99*(1), 29–36.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion (Washington, D.C.)*, *11*(5), 1021–1031.

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*, *110*(32), 13168–13173.

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), 2701–2715.

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), 1453–1463.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., … Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1412), 1293–1322.

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, *20*(12), 2153–2166.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews. Neuroscience*, *3*(6), 473–478.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *115*(10), 2292–2307.

Obleser, J., & Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cerebral Cortex (New York, N.Y.: 1991)*, *20*(3), 633–640.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience*, *32*(36), 12376–12383.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jäncke, L. (2010). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: A diffusion tensor imaging study. *Frontiers in Human Neuroscience*, *3*.

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch—Functional evidence of speech-relevant auditory acuity. *Cerebral Cortex (New York, NY)*, *20*(2), 447–455.

Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proceedings of the National Academy of Sciences*, *112*(40), 12522–12527.

Pascual-Marqui, R. D., Lehmann, D., Koukkou, M., Kochi, K., Anderer, P., Saletu, B., … Kinoshita, T. (2011). Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *369*(1952), 3768–3784.

Peretz, I., Vuvan, D., Lagrois, M.-E., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090–20140090.

Pillay, S. B., Stengel, B. C., Humphries, C., Book, D. S., & Binder, J. R. (2014). Cerebral localization of impaired phonological retrieval during rhyme judgment. *Annals of Neurology*, *76*(5), 738–746.

Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., & Paulus, W. (2012). The importance of timing in segregated theta phase-coupling for cognitive performance. *Current Biology: CB*, *22*(14), 1314–1318.

Price, C. J. (2010). The anatomy of language: A review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*, 62–88.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology*, *8*(4), 516–521.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.

Raven, J. C., Corporation, P., Lewis, H. K. (1962). *Coloured progressive matrices: Sets A, AB, B*. London, UK: Oxford Psychologist Press.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. J. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, *11*(4), 426–428.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3711–3735.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience*, *31*(10), 3843–3852.

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), 18035–18040.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Schulze, K., & Koelsch, S. (2012). Working memory for speech and music. *Annals of the New York Academy of Sciences*, *1252*, 229–236.

Schulze, K., Vargha-Khadem, F., & Mishkin, M. (2012). Test of a motor theory of long-term auditory memory. *Proceedings of the National*

Academy of Sciences of the United States of America, *109*(18), 7121–7125.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*(5), 771–783.

Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, *106*(50), 21341–21346.

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(14), 3799–3806.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(2), 174–215.

Stam, C. J., & van Straaten, E. C. W. (2012). Go with the flow: Use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. *NeuroImage*, *62*(3), 1415–1428.

Strait, D. L., Slater, J., O'connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278.

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. *European Journal of Neuroscience*, *30*(8), 1636–1642.

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Developmental Neuropsychology*, *37* (6), 476–496.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *77*, 1–12.

Turkeltaub, P. E., & Coslett, H. B. (2010). Localization of sublexical speech perception components. *Brain and Language*, *114*(1), 1–15.

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*(1), 577–593.

Wang, X., Ossher, L., & Reuter-Lorenz, P. A. (2015). Examining the relationship between skilled music training and attention. *Consciousness and Cognition*, *36*, 169–179.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559.

Wechsler, D. (2003). *Wechsler Intelligence Scale for children (fourth edition)*. San Antonio, TX: The Psychological Corporation (WISC-IV).

Weiss, S., & Rappelsberger, P. (2000). Long-range EEG synchronization during word encoding correlates with successful memory performance. *Brain Research. Cognitive Brain Research*, *9*(3), 299–312.

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., & Obleser, J. (2011). Alpha rhythms in audition: Cognitive and clinical perspectives. *Frontiers in Psychology*, *2*, 73.

Wong, Y. T., Fabiszak, M. M., Novikov, Y., Daw, N. D., & Pesaran, B. (2016). Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. *Nature Neuroscience*, *19*(2), 327–334.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One*, *9*(6), e99868.

How to cite this article: Dittinger E, Valizadeh SA, Jäncke L, Besson M, Elmer S. Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Hum Brain Mapp*. 2017;00:1–13. https://doi.org/10.1002/hbm.23877

Chapter 6

MUSIC TRAINING, NOVEL WORD LEARNING AND MEMORY

Dittinger, E., Korka, B., & Besson, M. (in preparation) Behavioral and electrophysiological evidence for better long-term memory in professional musicians and its contribution to novel word learning.

Behavioral and electrophysiological evidence for better long-term memory in professional musicians and its contribution to novel word learning

Eva Dittinger[1,2,3], Betina Korka[4], & Mireille Besson[1]

[1]CNRS & Aix-Marseille University, Laboratoire de Neurosciences Cognitives (LNC, UMR 7291), Marseille, France

[2]CNRS & Aix-Marseille University, Laboratoire Parole et Langage (LPL, UMR 7309), Aix-en-Provence, France

[3]Brain and Language Research Institute (BLRI), Aix-en-Provence, France

[4]Cognitive and Biological Psychology, Institute of Psychology, University of Leipzig, Germany

Abbreviated title: Musicianship, long-term memory and novel word learning

Declarations of interest: none

E-mail addresses:

ED: eva.dittinger@blri.fr

BK: betina-christiana.korka@uni-leipzig.de

MB: mireille.besson@univ-amu.fr

Correspondence to:

Eva Dittinger, Laboratory of Cognitive Neuroscience, Centre Saint-Charles, Pole 3C, Case C, 3 place Victor Hugo, 13331, Marseille cedex 3, France.

Email: eva.dittinger@blri.fr

Date: October 1, 2018

ABSTRACT

Previous studies evidenced transfer effects from professional music training to novel word learning, but whether such an advantage is driven by cascading effects from better auditory perception and attention to semantic processing, or by the interplay of perceptive and cognitive functions, has not yet been investigated. Here, we used an ecologically valid word learning design with four different sets of novel words that allowed to neutralize potential perceptive and associative learning advantages in musicians during word learning. We asked whether under such conditions musicians would still integrate novel learned words more efficiently into semantic memory, how associative learning mechanisms contribute to novel word learning, and whether musicians would remember the newly learned words better than non-musicians one month later. Behavioral and electrophysiological data failed to evidence an advantage for musicians on the day of learning (Day 1), with musicians and non-musicians integrating the novel words equally well into semantic networks. By contrast, musicians showed significantly lower error increases from Day 1 to Day 30 compared to non-musicians for all novel words. In addition, for the set of words that were perceptively difficult to discriminate, only musicians showed typical N400 effects over parietal sites on Day 30. These results show that music training strengthens long-term memory functions. Moreover, they suggest that transfer effects from music training to word learning (i.e., semantic levels of speech processing) are in large part driven by increased auditory sensitivity, but profit additionally from reinforced (long-term) memory functions. These findings highlight the positive impact of music training for the acquisition of foreign (mainly tonal and quantitative) languages.

Keywords: word learning, musical expertise, long-term memory, N400

INTRODUCTION

Playing a musical instrument at a professional level is a multidimensional task that requires acute auditory perception, focused attention, the ability to maintain auditory information in short- and long-term memory, as well as specific motor abilities. Typically, professional musicians start playing their instrument at a very young age, and train for years, many hours per week. Nowadays, there is accumulating evidence showing that such an intensive training strongly promotes brain plasticity (Münte, Altenmüller, & Jäncke, 2002). Most importantly for the present study, music training-related advantages have been shown to transfer to other domains such as language (for review, see Asaridou & McQueen, 2013; Besson, Chobert, & Marie, 2011; Kraus & Chandrasekaran, 2010), at the segmental (Bidelman, Weiss, Moreno, & Alain, 2014; Chobert, Marie, François, Schön, & Besson, 2011; Elmer, Meyer, & Jancke, 2012), supra-segmental (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Wong & Perrachione, 2007), and even syntactic level of speech processing (Fitzroy & Sanders, 2012). There is also some evidence that long-term music training positively improves auditory attention (Strait, Slater, O'Connell, & Kraus, 2015), visual attention (Wang, Ossher, & Reuter-Lorenz, 2015), working and verbal memory (George & Coch, 2011), executive functions (Zuk, Benjamin, Kenyon, & Gaab, 2014), and general intelligence (Schellenberg, 2004, 2011).

Recently, Dittinger and colleagues (Dittinger et al., 2016; Dittinger, Chobert, Ziegler, & Besson, 2017; Dittinger, Scherer, Jäncke, Besson, & Elmer, under review) showed that music training also influences the semantic level of speech processing by comparing participants with and without music training from three independent samples: children, young adults and older adults (to a smaller extent). Specifically, the authors focused on novel word learning (a multidimensional task based on both perceptive and cognitive functions) and conducted a series of experiments that comprised phonological categorization tasks, a picture-word association learning phase, as well as a test phase that included a matching and a semantic task (i.e., does the word matches or not with the previously learned picture-word association and is the word semantically related or not to a novel picture?). Both, behavioral measures and Event-Related Potentials (ERPs) were analyzed. In line with the hypotheses, results of behavioral and electrophysiological data showed that musicians outperformed non-musicians during the phonological categorization tasks based on aspiration and pitch, two non-native phonetic contrasts for French-speaking participants (Dittinger et al., 2016; Dittinger, D'Imperio, & Besson, 2018). Moreover, only musicians showed significant N400 increase in amplitude from the first to the second half of the learning phase over fronto-central regions (Batterink & Neville, 2011; Francois, Cunillera, Garcia, Laine, & Rodriguez-Fornells, 2017; McLaughlin, Osterhout, & Kim, 2004). Finally, in both tasks of the test phase, only musicians were characterized by significantly larger N400 amplitude to words that were unexpected compared to words that were expected based on previous picture-word association learning. The N400 effects (i.e., the difference between unexpected minus expected words) was largest over centro-parietal sites, a scalp distribution that is comparable to the N400 effect for known words (Kutas & Federmeier, 2011). These results at the electrophysiological level were accompanied by higher level of performance in the semantic task. Importantly, results described above for young adults (Dittinger et al., 2016) were similar in children (Dittinger et al., 2017) but less clear-cut in older adults (Dittinger et al., under review). Overall, they showed that the effects of music training transfer to semantic levels of speech processing.

Two main interpretations have been proposed to explain transfer effects from music training to novel word learning (Dittinger et al., 2016). First, according to the cascading interpretation, enhanced auditory perception and attention possibly drive transfer effects in musicians by facilitating the cascade of processing stages involved in novel word learning (speech perception, building phonological, lexical and semantic representations; Besson, Barbaroux, & Dittinger, 2017). Second, according to the multi-dimensional interpretation, that refers to the multi-dimensionality of music training, musicians not only show enhanced speech perception, but also improved attention (Strait et al., 2015), working and/or short-term memory (George & Coch, 2011; Schulze, Dowling, & Tillmann, 2012), and executive functions (Zuk et al., 2014), functions that may all contribute to novel word learning. The aim of the present series of experiment was to better understand why musicians are at the advantage in novel word learning by trying to disentangle these two interpretations.

In Experiment 1 (E1), and to reduce the influence of auditory perception and associative learning, we used an experimental design similar to Dittinger et al. (2016), with the same stimuli (see Figure 1) but with one important modification. Rather than performing only one learning phase (to learn picture-word associations) followed by only one matching task (to test for associative learning), participants performed a variable number of short learning-matching task cycles until they all reached a level of 83 % correct responses in the matching task. In this way, all participants had learned the associations equally well when they were presented with the semantic task. Consequently, if differences in novel word learning between musicians and non-musicians are still found using this procedure, they are unlikely to reflect better auditory perception and associative learning.

In Experiment 2 (E2), the specific influence of enhanced auditory speech perception and attention in musicians (Bidelman et al., 2014; Strait, Kraus, Parbery-Clark, & Ashley, 2010; Strait et al., 2015) was further reduced by using Finnish disyllabic words that, in contrast with the monosyllabic Thai words used by Dittinger et al. (2016), were easy to discriminate for all participants. In sum, in E1 and E2, we tested the hypothesis that if musicians and non-musicians still differ in the semantic task when the impact of auditory perception, auditory attention and associative learning has been strongly reduced, these results would show that other functions, such as semantic processing (i.e., better integration of novel words into pre-existing semantic networks), executive and/or memory functions, also contribute to explain the advantage of musicians in novel word learning.

Finally, based on previous results showing more efficient cross-modal integration in musicians than in non-musicians (Bidelman, 2016; Paraskevopoulos, Kuchenbuch, Herholz, & Pantev, 2012), Experiments 3 and 4 (E3 and E4) aimed at further investigating the processes involved in cross-modal integration of novel visual picture and auditory word associations. To this end, we manipulated the amount of semantic information available to the participants. In E3, disyllabic Finnish words were associated with non-objects, thereby providing a control condition in which participants had to learn an association between two items that, at the start of learning, were non-meaningful and unknown (i.e., Finnish words and pictures of non-objects with no semantic content, see Figure 1). By contrast, in E4, known French words were associated with meaningful, known but non-corresponding pictures (e.g., the word “fromage” -cheese-, was presented together with the picture of a hat). This experiment provided another control condition in which a new association had to be learned between two known items (i.e., known word and known, but unrelated, meaningful picture, see Figure 1). If professional musicians perform better than non-musicians in these two tasks, this will suggest that their advantage in novel word learning relies on their ability to build novel picture-word associations more efficiently than non-musicians, possibly through facilitation in cross-modal integration.

In sum, these four experiments (see Figure 1 below) were designed to test the following three issues: first, do musicians still outperform non-musicians in the semantic task when between-group differences in word perception and associative learning have been neutralized? In other words, is semantic integration faster and/or better in musicians even without the advantage of better auditory perception and attention (E1 and E2)? Second, to further investigate the contribution of associative learning and cross-modal integration to novel word learning, we manipulated the amount of semantic information available to the participants (E3 and E4). Final but not least, the third aim was to investigate what remained of novel word learning after a one-month period, and whether musicians would remember newly-learned words better than non-musicians. A positive finding would extend to long-term memory previous results showing enhanced working- and short-term memory in musicians (George & Coch, 2011; Schulze et al., 2012).

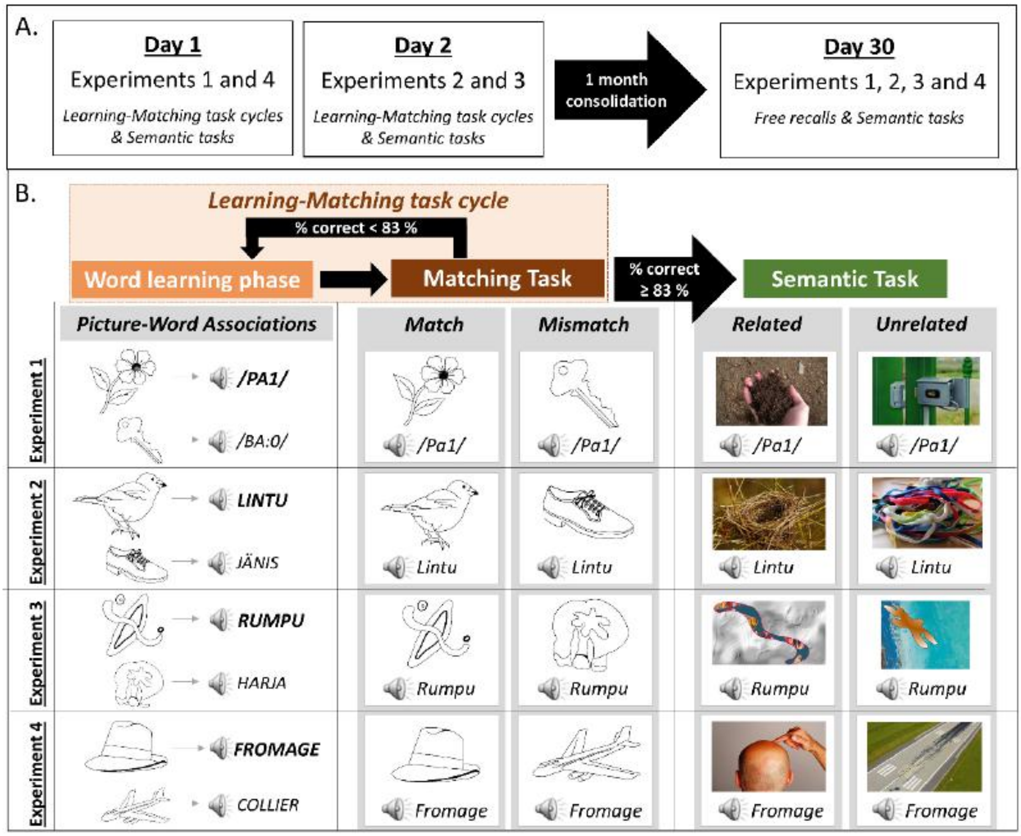


Figure 1. Experimental design. (A) Participants performed four independent novel word learning experiments on two days (Day 1 and Day 2). Levels of performance in all experiments were re-tested after 30 days. (B) Each experiment was composed of Learning-Matching task cycles and a Semantic task. The cycles included word learning phases during which each word was paired with its respective picture, and matching tasks during which words were presented with one of the pictures, either matching or mismatching the previously learned associations. Once the participant reached 83% correct responses, she/he entered the semantic task where words were presented with novel pictures that were either semantically related or unrelated to the novel words.

MATERIALS & METHODS

Participants

Thirty-three participants contributed data, out of which 17 were professional musicians (MUS, 8 women), and 16 were controls without formal music training (non-musicians, NM, 8 women), but were involved in a regular leisure activity (e.g. sports, dance, theatre). Out of these, two MUS and one NM were considered outliers based on behavioural performance (+/- 2 SD away from the mean; see Methods, Statistical Analyses), resulting in equal samples of 15 analysed datasets for each group. The experiments lasted for altogether nine hours on three different days (i.e., two sessions on consecutive days, and the third session about 30 days later). The two groups did not significantly differ in age (MUS: mean age = 25.7 years, age range = 19-36, SD = 1.5; NM: mean age = 26.0 years, age range = 20-35, SD = 1.5; $F_{(1,28)} = .02$, $p = .90$). All participants were native French speakers, had comparable education levels (university degree), and reported no past or current audiological or neurological deficits. The MUS group practiced their instruments for an average of 18.3 years (range = 11-29, SD = 5.1) and included five violinists, three clarinetists, two pianists, two flautists, one oboist, one trombonist, and one trumpeter. None of the participants was bilingual, but all spoke English as a second language and all participants had a rudimentary knowledge of a third language (mainly Spanish, Italian or German). The study was conducted in accordance with the Helsinki declaration. All participants gave their informed consent before enrolling in the experiment and received monetary compensation for participating.

Screening measures

Cognitive ability: Standardized psychometric tests were used to examine short-term- and working memory (forward and reverse Digit Span, WAIS-III; Wechsler, Coalson, & Raiford, 1997), auditory attention (Associated Responses, adapted from the NEPSY-II child battery; Korkman, Kirk, & Kemp, 2007), visual attention (D2-R; Brickenkamp, Schmidt-Atzert, & Liepmann, 2015), lexical and semantic fluency (Verbal Fluency; Cardebat, Doyon, Puel, Goulet, & Joanette, 1990), and non-verbal general intelligence (Matrices, WAIS-III; Wechsler et al., 1997).

Musical aptitude: Participants performed two musicality tests (adapted from the MBEA battery; Peretz, Champod, & Hyde, 2003) consisting in judging whether pairs of piano melodies were same or different, one test based on melodic and the other one on rhythmic information.

Experimental stimuli

Auditory stimuli: For E1, twelve natural Thai monosyllabic words were selected: /ba1/, /pa1/, /p^{h}a1/, /ba:1/, /pa:1/, /p^{h}a:1/, /ba0/, /pa0/, /p^{h}a0/, /ba:0/, /pa:0/, /p^{h}a:0/[1]. These words varied in vowel duration, with short (261 ms on average) and long vowels (531 ms on average), and in fundamental frequency, with low-tone (F0 = 175 Hz, on average) and mid-tone vowels (F0 = 218 Hz on average). Furthermore, words contained contrasts based on Voice Onset Time (VOT; /b/ = -144 ms *vs*. /p/, VOT = 3 ms *vs*. /p^{h}/ = 77 ms). Stimuli were recorded by a female Thai-French bilingual, ensuring that all words were produced naturally. For E2 and E3, 48 natural Finnish disyllabic words were selected, including non-native phonetic features for French speakers (i.e., geminate stops, e.g. Noppa, Teltta; initial and pronounced "H"s, e.g. Hanhi, Hylje; non-native vowels for French speakers, e.g. Jänis, Sähkö; for the full list see Supplementary Figure 1). Two lists of 24 words without same initial syllables were created and presented in the two Experiments, counterbalanced across participants. Finally, for E4, 24 natural French disyllabic words without same initial syllable (e.g., Fléchette, Valise, Echelle; for the full list see Supplementary Figure 1) were chosen. Words of E2, E3, and E4 were recorded by a female Finnish-French bilingual. For each word in each experiment, 4 versions were digitally recorded in order to reproduce natural speech variability. Sound pressure level was normalized across all words to a mean level of 70 dB by using the Praat software (Boersma & Weenink, 2011).

Visual stimuli: For the learning phase, black and white line drawings were selected. In E1, E2 and E4 these drawings represented familiar objects (e.g., bear, flower, banana, scissors, fork, desk; see Supplementary Figure 1) that were chosen from the Snodgrass and Vanderwart (Snodgrass & Vanderwart, 1980) pictures' set[2]. Only pictures with disyllabic French names were selected, and controlled for name agreement, image agreement, familiarity, image complexity, age of acquisition, and frequency, based on the French normative measures for the Snodgrass & Vanderwart pictures (Alario & Ferrand, 1999). For E1, the same pictures were used as in Dittinger et al. (2016), and for E2 and 4 pictures were counterbalanced across participants. In E3, line drawings corresponded to the best-rated non-objects (i.e., do not resemble a real object) created by Kroll and Potter (Kroll & Potter, 1984) by tracing parts of drawings of real objects and

---

[1] Following phonetic transcription in Thai, 1 refers to low-tone, 0 to mid-tone, p^{h} to aspirated voicing and the colon to long vowel duration

[2] Pictures were based on the standardized set of 260 pictures built by Snodgrass and Vanderwart (1980), but retraced in Office PowerPoint to ensure sufficient resolution and quality.

regularizing the resulting figures[3]. For all four experiments, the same pictures as in the learning phase were then presented in the matching task. For the semantic task of E1, E2, and E4, new pictures that the participants had not seen before in the learning-matching task cycles and that were semantically related or unrelated to the meaning of the newly-learned words, were selected from the internet. For the semantic task of E3, new visually related or unrelated pictures were created manually in Office PowerPoint (for examples, see Figure 1B). All pictures were pretested with students from our university (n = 15; age range = 18-30 years). Specifically, they were asked to rate the semantic or visual relatedness between new and old pictures on a scale from 1 to 5, with 1 being not related at all and 5 being very well related. Only picture pairs that were on average rated higher than 4 in related conditions and lower than 2 in unrelated conditions were accepted for the experiments.

Experimental tasks

Participants were tested individually in a quiet experimental room (i.e., Faraday cage), where they sat in a comfortable chair at about 1 meter from a CRT computer screen. Auditory stimuli were presented through HiFi headphones (Sennheiser, HD590) at 70dB sound pressure level. Visual and auditory stimuli presentation, as well as the collection of behavioral data, were controlled via the ‘’Presentation’’ software (NeuroBehavioral Systems, Version 11.0). Four independent Experiments were performed, two on Day 1 (always E1 and E4 with E1 first) and two on Day 2[4] (always E2 and E3 with E2 first), and each Experiment was composed of Learning-Matching task cycles, followed by a Semantic Task (see Figure 1A and 1B). Moreover, in order to test for long-term memory effects, participants performed again the semantic tasks of all four Experiments around 30 days (range: 24-41 days) after Day 1.

Learning-Matching task cycle

One learning-matching task cycle consisted of one learning block (i.e., one presentation of each picture-word association), followed by one matching task (i.e., one match and one mismatch condition for each picture-word association). The number of cycles was based on the participant’s percentage of correct responses in the matching task (and thus variable across participants): as long as the percentage was below 83 % correct responses, the participant

---

[3] As for Snodgrass and Vanderwart (1980) pictures, non-objects from Kroll and Potter (1984) were retraced in Office PowerPoint to ensure sufficient resolution and quality.

[4] To simplify, we refer to both Day 1 and Day 2 as “Day 1” (i.e., initial learning session) throughout the paper.

performed another learning-matching task cycle. Once the participant reached at least 83 % correct responses twice in a row, the semantic task was presented (described below).

Word learning phase: Participants were asked to learn the meaning of the novel words using picture-word associations. In E1, 12 picture-word associations had to be learned, based on known pictures and natural Thai monosyllabic words previously used by Dittinger et al. (2016). In E2, E3 and E4, 24 picture-word associations had to be learned, with known pictures and natural Finnish disyllabic words in E2, non-objects and natural Finnish disyllabic words in E3, and known pictures and known French disyllabic words in E4 (see Figure 1B). The number of associations to be learned in each experiment was determined based on pilot studies showing that associations were clearly more difficult to learn in E1 than in the other experiments. To give some examples, in E1, a drawing of a flower was followed by the auditory presentation of the word /pa1/ and thus, /pa1/ was the word for flower in our "foreign" language. In E2, a drawing of a bird, was followed by the word "Lintu", therefore "Lintu" was the word for bird. In E4, "fromage" (meaning cheese in French) was the word for hat. And finally, in E3, participants were asked to learn the visual appearance (given by the picture of the non-object) of the novel word "Rumpu" (see Figure 1B). In one block of learning, each picture-word association was presented once, resulting in 12 trials for E1, and 24 trials for E2, E3 and E4. The picture was presented first and then followed after 750 ms by one of the words. Total trial duration was 2000 ms. Different lists were built, so that across participants, different pictures were associated with different words (together with the counterbalancing of Finnish words and known pictures mentioned above; for one example of a learning list see Supplementary Figure 1). No behavioral response was required from the participants during this word learning phase. Duration of one learning block was about half a minute for E1, and one minute for E2, E3, and E4.

Matching task: One of the pictures was presented, followed after 750 ms by an auditory word that matched or mismatched the associations previously learned in the word learning phase. After 2750 ms, a row of "X"s appeared on the screen for 1000 ms, and participants were asked to blink during this time period in order to minimize eye movement artifacts during the next trial. Total trial duration was 3750 ms. To give some examples, in E1, the drawing of a flower followed by /pa1/ (i.e., flower) was a match and the drawing of a key followed by /pa1/ was a mismatch (see Figure 1B, for more examples in E2, E3 and E4). Participants were asked to press one out of two response keys accordingly, as quickly and accurately as possible, after having heard the word. Response hand was counter-balanced across participants. In one run of the matching task, each word was presented twice, once in a match- and once in a mismatch condition. The total of 24 trials in E1 and 48 trials in E2, E3 and E4 were pseudo-randomly presented (i.e., no more than four successive “matches” or “mismatches”). Duration of one matching block was 1.5 min in E1 and 3 min in E2, E3, and E4.

Semantic Task

One of the new pictures was presented, followed after 1500 ms by an auditory word that was semantically (i.e., E1, E2 and E4) or visually (i.e., E3) related, or unrelated. After 3500 ms, a row of “X”s appeared on the screen for 1000 ms, and participants were asked to blink during this time period. Total trial duration was 4500 ms. For instance, while the picture of soil was semantically related to the previously learned word /pa1/ (i.e., “flower”) in E1, the picture of a lock was semantically unrelated to /pa1/ (see Figure 1B for more examples in E2, E3 and E4). Participants were asked to press one out of two response keys accordingly, as quickly and accurately as possible, after having heard the word. Response hand was counter-balanced across participants. Four trials were presented first to familiarize participants with the task. Each word was presented 12 times, but none of the new pictures were repeated, so that on each trial the word was associated with a different related or unrelated picture. Half of the picture-word pairs were semantically or visually related and half were semantically or visually unrelated. A total of 144 trials in E1, and 288 trials in E2, E3 and E4 were presented pseudo-randomly (i.e., no consecutive associations referring to the same word, and no more than four successive ”related” or “unrelated” associations) within two blocks in E1, and within four blocks in E2, E3, and E4. Duration of one block was 5.4 min.

Long-term memory session

To test for long-term memory of the novel words, participants came back about 30 days after the first experimental session, on average. For each Experiment, they were first asked to freely recall the words and to write them down, before starting the semantic task. Then, they performed the semantic tasks. The order of Experiments was counter-balanced across participants for this session.

EEG data acquisition and analysis

The Electroencephalogram (EEG) was continuously recorded at a sampling rate of 512 Hz with a band-pass filter of 0-102.4 Hz by using a Biosemi amplifier system (Amsterdam, BioSemi Active 2) with 32 active Ag/Cl electrodes (Biosemi Pintype) located at standard positions according to the international 10/20 System (Jasper, 1958). The electro-oculogram (EOG) was recorded from flat-type active electrodes placed 1 cm to the left and right of the external canthi and from an electrode beneath the right eye. Two additional electrodes were placed on the left and right mastoids. Electrode impedance was kept below 5 kΩ. EEG data were analyzed using the Brain Vision Analyzer software (Version 1.05.0005; Brain Products, Gmbh). All data were re-referenced off-line to the averaged left and right mastoids, filtered with a bandpass filter from 0.1-30 Hz (slope of 24 dB/oct), and independent component analysis (ICA) and inverse ICA were used to identify and remove components associated with vertical and horizontal ocular movements. Data were segmented into 1200 ms epochs, time-locked to word onset, and including a 200 ms baseline. DC-detrend and removal of artifacts above a gradient criterion of 10 µV/ms or a max-min criterion of 100 µV over the entire epoch, were applied. Averages were computed for each participant and for each condition in each experiment and these individual averages were then averaged into the grand averages across all participants.

Statistical analyses

Analyses of Variance (ANOVAs) were computed using the Statistica software (Version 12.0, StatSoft Inc., Tulsa). ANOVAs always included Group (MUS vs. NM) as between-subject factor, and specific factors for each task. Regarding data registered on Day 1, univariate ANOVAs (only the Group factor) were computed on the number of learning cycles. For the matching and semantic tasks, ANOVAs were computed on error rates (ERRs) and on reaction times (RTs), first by considering each Experiment separately, and second by including Experiment (E1 vs. E2 vs. E3 vs. E4) as an additional within-subject factor (for this between-experiments comparison, only additional significant effects are reported). ANOVAs included Condition (Matching task: match *vs*. mismatch and Semantic task: related *vs*. unrelated) as within-subject factors. ANOVAs in the matching task also included Moment in the learning-matching cycles (Matching task: Start vs End[5]) as an additional within-subject factor.

Regarding data registered at Day 30 in the semantic task, ANOVAs were computed on the difference in ERRs and in RTs between Day 30 and Day 1 for both related and unrelated words. As for Day 1, ANOVAs were first computed separately for each Experiment and second,

---

[5] "Start" refers to the trials of the matching tasks of the first two learning-matching task cycles, and "End" refers to the trials of the matching tasks of the last two learning-matching task cycles.

by including Experiment (E1 vs. E2 vs. E3 vs. E4) as an additional within-subject factor (as noted above, for this analyses, only additional significant effects are reported).

For each experiment, the N400 component was analyzed in the semantic task by computing the mean amplitude in the 400-550 ms latency band. Only correct responses were considered in these analyses. ANOVAs always included Group (MUS *vs*. NM) as a between-subject factor, and Condition (related *vs*. unrelated) as within-subject factor, together with Laterality (left: F3, C3, P3; midline: Fz, Cz, Pz; right: F4, C4, P4) and Anterior/Posterior (frontal: F3, Fz, F4; central: C3, Cz, C4; parietal: P3, Pz, P4). As for behavior, ANOVAs were first computed for each Experiment separately, and second by including Experiment (E1 vs. E2 vs. E3 vs. E4) as an additional within-subject factor (as noted above, only additional significant effects are reported in this case). Post-hoc Tukey tests (reducing the probability of Type I errors) were used to determine the origin of significant main effects and interactions.

RESULTS

Psychometric measures

Results of univariate ANOVAs showed no significant differences between groups regarding general intelligence ($F < 1$), auditory or visual attention ($F_{(1, 28)} = 1.55$, $p = .22$ and $F_{(1, 28)} = 1.96$, $p = .17$, respectively), nor verbal or semantic fluency (both $F < 1$). By contrast, MUS showed better working and short-term memory abilities than NM ($F_{(1, 28)} = 6.90$, $p = .01$ and $F_{(1, 28)} = 12.83$, $p = .001$, respectively).

Musicality task

Results of a 2 × 2 ANOVA (i.e., 2 Groups × 2 Tasks) showed that MUS made fewer errors (8.1 %, SD = 1.6) than NM (15.2 %, SD = 1.6; main effect of Group: $F_{(1, 28)} = 9.32$, $p = .005$), and all participants performed better on the rhythmic (9.3 %, SD = 1.1) than on the melodic task (14.1 %, SD = 1.8; main effect of Task: $F_{(1, 28)} = 6.56$, $p = .02$) with no Group × Task interaction.

Learning-Matching task cycles

Number of cycles: Results of univariate ANOVAs for each experiment showed that for all four experiments, the number of cycles necessary to reach the threshold did not significantly differ for musicians and non-musicians (main effects of Group: E1, E2, E3, E4: all $F < 1$). Regarding the comparison between experiments, results of the ANOVA (2 Groups x 4 Experiments) showed that participants needed more cycles in E1 (9.0 cycles, SD = 0.5) and fewer cycles in E4 (2.4 cycles, SD = 0.1) compared to E2 (3.6 cycles, SD = 0.2; Tukey, p < .001 and p = .02, respectively) and E3 (3.8 cycles, SD = 0.2; Tukey, p < .001 and p = .005, respectively; main effect of Experiment: $F_{(3, 84)}$ = 105.08, p < .001).

Matching task: Results of separate ANOVAs (2 Groups x 2 Moments x 2 Conditions; see Table 1A for F- and p-values, and Table 2 for ERR- and RT- values) for each experiment showed that for all four experiments, ERRs and RTs did not significantly differ between musicians and non-musicians. Moreover, in all four Experiments, participants made fewer errors and responded faster in the End- compared to the Start-session. Finally, participants made fewer errors and responded faster for match than for mismatch words in E1. By contrast, participants made fewer errors for mismatch than for match words in E2, E3 and E4, with no differences on RTs. More specifically, while this pattern of results for ERRs was obtained in the Start session of E2, E3 and E4, in the End session, ERRs were not significantly different for match and mismatch words (Tukey, all p < .001; Moment x Condition interactions: E2: $F_{(1,28)}$ = 21.74, p < .001; E3: $F_{(1,28)}$ = 31.56, p < .001; E4: $F_{(1,28)}$ = 21.49, p < .001).

Table 1. Results of statistical analyses on error rates (ERRs) and reaction times (RTs), in the different tasks and experiments (E1, E2, E3 and E4). F- and p-values are shown, and significant p-values are in bold and italic.

		A. Matching task				B. Semantic task D1				C. Semantic task D30			
		ERRs		RTs		ERRs		RTs		ERRs		RTs	
		F (1,28)	*p*	*F (1,28)*	*p*	*F (1,28)*	*p*	*F (1,28)*	*p*	*F (1,28)*	*p*	*F (1,28)*	*p*
Main effects of Group	E1	.42	.52	.00	.99	.32	.58	.07	.79	.82	.37	.10	.75
	E2	.69	.41	.51	.48	1.14	.29	.27	.61	7.50	***.01***	2.90	.10
	E3	.02	.88	1.10	.31	.02	.90	.01	.94	4.93	***.04***	1.63	.21
	E4	.39	.54	1.45	.24	1.55	.22	3.47	.07	5.56	***.03***	7.16	***.01***
Main effects of Condition	E1	5.71	***.02***	4.21	***.05***	4.76	***.04***	49.46	***.001***	1.18	.29	15.93	***.001***
	E2	38.45	***.001***	.40	.53	68.34	***.001***	67.26	***.001***	5.86	***.02***	23.12	***.001***
	E3	41.87	***.001***	.18	.67	75.01	***.001***	73.89	***.001***	6.94	***.01***	21.17	***.001***
	E4	25.65	***001***	2.90	.10	66.03	***.001***	41.80	***.001***	22.57	***.001***	8.94	***.006***
Group x Condition	E1	.24	.63	.00	.98	.01	.96	.03	.86	9.16	***.005***	.43	.52
	E2	.02	.88	1.02	.32	.80	.38	.25	.62	3.59	.07	.40	.53
	E3	.01	.93	2.83	.10	.02	.88	.83	.37	2.39	.13	1.02	.32
	E4	1.30	.26	.04	.84	.00	.99	1.51	.23	.27	.60	.94	.34
Main effects of Moment	E1	365.63	***.001***	139.85	***.001***								
	E2	337.20	***.001***	159.91	***.001***								
	E3	302.00	***.001***	137.80	***.001***								
	E4	54.46	***.001***	97.81	***.001***								
Group x Moment	E1	1.81	.19	2.25	.14								
	E2	2.03	.17	.03	.87								
	E3	1.11	.30	.14	.71								
	E4	.69	.41	1.42	.24								

Regarding the comparison between experiments (2 Groups x 4 Experiments x 2 Conditions), in the Start-session participants made more errors and responded slower in E1 and made fewer errors and responded faster in E4 compared to E2 and E3 (Tukey, ERRs and RTs: all

$p < .001$). By contrast, in the End-session participants performed similarly (ERRs and RTs) in E1, E2 and E3, with fewer errors and faster RTs only in E4 (Tukey, ERRs: all $p < .01$; and RTs: all $p < .001$; Moment x Experiment interactions: ERRs: $F_{(3,84)} = 29.21$, $p < .001$; and RTs: $F_{(3,84)} = 10.87$, $p < .001$).

Table 2. Matching tasks: error rates and reaction times (standard deviations in brackets)

		E1	E2	E3	E4
% error	MUS	27.0 (1.2)	21.7 (1.5)	22.1 (1.4)	9.2 (1.7)
	NM	28.1 (1.2)	19.9 (1.5)	22.4 (1.4)	7.7 (1.6)
	Start	44.8 (2.1)	34.0 (2.4)	36.3 (2.3)	15.2 (2.9)
	End	10.3 (1.1)	7.6 (1.0)	8.3 (1.1)	1.6 (0.7)
	Match Overall	24.7 (2.1)	27.0 (2.1)	29.1 (1.9)	11.2 (2.2)
	Start	43.1 (2.4)	44.0 (2.3)	48.0 (2.3)	20.4 (2.8)
	End	6.3 (1.2)	10.0 (1.1)	10.1 (1.1)	2.0 (0.6)
	Mismatch Overall	30.4 (2.0)	14.6 (2.0)	15.5 (2.2)	5.7 (1.4)
	Start	46.5 (1.2)	24.0 (2.5)	24.6 (2.6)	10.1 (1.7)
	End	14.3 (1.6)	5.1 (0.7)	6.4 (0.9)	1.3 (0.5)
RTs [ms]	MUS	1212 (42)	1130 (42)	1130 (45)	906 (36)
	NM	1212 (42)	1172 (42)	1197 (45)	844 (36)
	Start	1423 (60)	1297 (48)	1310 (52)	981 (44)
	End	1002 (34)	1006 (43)	1017 (46)	770 (34)
	Match Overall	1189 (48)	1147 (40)	1166 (47)	862 (35)
	Start	1398 (50)	1296 (32)	1313 (40)	973 (29)
	End	980 (26)	998 (33)	1019 (32)	751 (26)
	Mismatch Overall	1235 (41)	1156 (47)	1161 (46)	889 (40)
	Start	1448 (43)	1298 (38)	1308 (35)	988 (36)
	End	1023 (28)	1014 (33)	1015 (34)	790 (25)

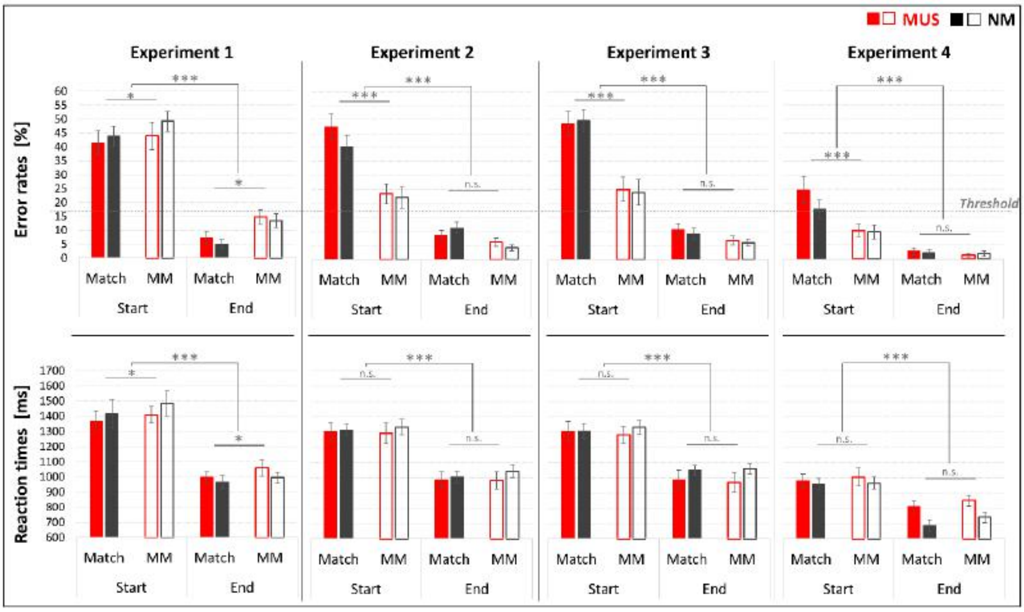


Figure 2. Error rates (upper row) and Reaction Times (RTs; lower row) in the matching tasks are shown separately for the four experiments. Results for Match (full bars) and Mismatch (MM; empty bars) words are illustrated for musicians (MUS; red) and for non-musician (NM; black), and averaged across the first two (Start) and the last two (End) learning-matching task cycles. The learning threshold (i.e., 83 % correct responses) is indicated. In this and subsequent figures, the level of significance is represented by stars with * $p < .05$, ** $p < .01$, *** $p < .001$, and n.s. $p > .05$.

Semantic Task, Day 1

Results of separate ANOVAs (i.e., 2 Groups x 2 Conditions; see Table 1B for F- and p-values, and Table 3 for ERR- and RT-values) in the semantic task of each experiment on Day 1 showed that for all four experiments, ERRs and RTs did not significantly differ between musicians and non-musicians. Moreover, participants made more errors and responded faster to related than to unrelated trials in all four Experiments (see Figure 3A).

Regarding the comparison between experiments (i.e., 2 Groups x 4 Experiments x 2 Conditions), participants made most errors in E1, intermediate errors in E3, and fewest errors in E2 and E4 (Tukey, all $p < .001$; main effect of Experiment: $F_{(3,84)} = 80.36$, $p < .001$). More specifically, for related trials error rates were higher in E1 and E3 compared to E2 and E4 (Tukey, all $p < .001$), while for unrelated trials error rates were higher in E1 compared to E2, E3 and E4 (Tukey, all $p < .001$; Experiment x Condition: $F_{(3,84)} = 11.84$, $p < .001$). For RTs, participants responded faster to unrelated trials in E3 and E4 compared to E1 and E2 (Tukey, all $p < .001$), while for related trials RTs were similar in all four Experiments (Tukey, all $p > .10$; Experiment x Condition: $F_{(3,84)} = 5.24$, $p = .002$).

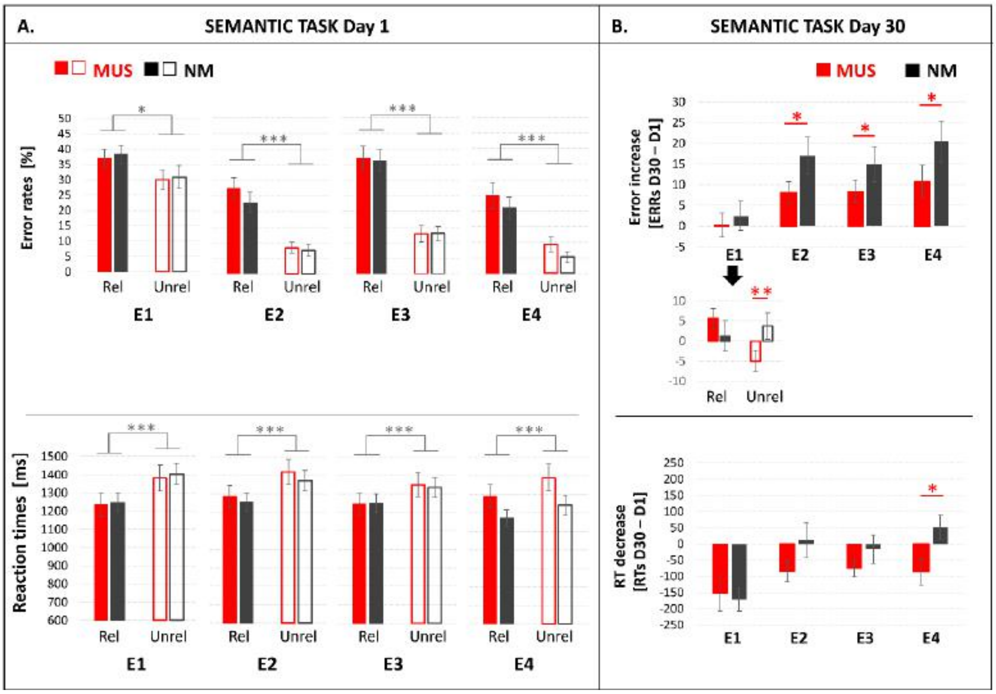


Figure 3. (A) Error rates (upper panel) and Reaction Times (RTs; lower pannel) in the semantic tasks on Day 1 are shown separately for the four Experiments. Results for Related (Rel; full bars) and Unrelated (Unrel; empty bars) words are illustrated for musicians (MUS; red) and non-musicians (NM; black). (B) Error increase (upper panel) and RT decrease (lower panel) from Day 1 to Day 30 in the semantic tasks are shown.

Table 3. Semantic tasks: error rates and reaction times (RTs; standard deviations in brackets)

		E1	E2	E3	E4
% error Day 1	MUS	33.5 (1.5)	17.9 (1.8)	24.9 (2.0)	17.2 (2.3)
	NM	34.7 (1.5)	15.1 (1.9)	24.5 (2.0)	13.1 (2.3)
	Related	37.7 (1.7)	25.2 (2.1)	36.6 (2.5)	23.0 (2.4)
	Unrelated	30.4 (2.2)	7.8 (1.0)	12.8 (1.4)	7.3 (1.1)
Error increase Day 30	MUS	0.3 (1.7)	8.1 (2.3)	8.4 (2.1)	10.7 (2.9)
	NM	2.5 (1.7)	17.0 (2.3)	14.9 (2.1)	20.4 (2.9)
	Related	3.4 (2.3)	16.9 (2.9)	16.4 (2.7)	24.2 (3.1)
	Unrelated	-0.6 (2.1)	8.2 (1.9)	7.0 (1.9)	7.0 (2.4)
RTs Day 1	MUS	1308 (45)	1352 (53)	1297 (47)	1339 (50)
	NM	1325 (46)	1314 (52)	1292 (46)	1207 (51)
	Related	1242 (31)	1271 (36)	1246 (33)	1232 (34)
	Unrelated	1392 (36)	1395 (40)	1342 (33)	1315 (39)
RT decrease Day 30	MUS	-152 (43)	-83 (40)	-75 (33)	-85 (36)
	NM	-172 (44)	12 (40)	-16 (33)	51 (36)
	Related	-121 (34)	21 (30)	-3 (28)	18 (27)
	Unrelated	-203 (30)	-93 (30)	-88 (23)	-52 (29)

Turning to the electrophysiological data (see Figure 4), results of separate ANOVAs for each experiment (i.e., 2 Groups x 2 Conditions x 3 Laterality x 3 Anterior/Posterior; see Table 4 for µV values) showed that in E1, the N400 was larger to unrelated than to related trials over parietal sites (Condition x Anterior/Posterior: $F_{(2,56)} = 3.27$, $p = .05$). Furthermore, the Group x Condition x Laterality and Group x Anterior/Posterior interactions were significant in E1 ($F_{(2, 56)} = 3.23$, $p = .05$ and $F_{(2, 56)} = 3.40$, $p = .04$, respectively). Separate ANOVAs for musicians and non-musicians revealed larger N400 to related than to unrelated trials over frontal sites in MUS (related: -2.62 µV, SD = 1.43; unrelated: -1.90 µV, SD = 1.40; Tukey, $p = .02$; Condition x Anterior/Posterior: $F_{(2, 28)} = 5.19$, $p = .01$). By contrast, in NM, N400 was larger to unrelated than to related trials over midline and right hemisphere (unrelated: -0.03 µV, SD = 1.37, and -0.78 µV, SD = 1.16, respectively; related: 0.53 µV, SD = 1.45, and -0.01 µV, SD = 1.20, respectively; Tukey, $p = .03$ and $p = .002$, respectively; Condition x Laterality: $F_{(2, 28)} = 3.32$, $p = .05$). In E2 and E3, the N400 was larger to unrelated than to related trials over parietal sites, and to related than to unrelated trials over frontal sites (Tukey, E2 and E3 all $p < .001$; Condition x Anterior/Posterior interactions: E2: $F_{(2,56)} = 44.93$, $p < .001$; and E3: $F_{(2,56)} = 53.09$, $p < .001$). Finally, in E4, the N400 was larger to related than to unrelated trials over frontal and central sites (Tukey, both $p < .001$; Condition x Anterior/Posterior: $F_{(2,56)} = 21.62$, $p < .001$).

Regarding the comparison between experiments (i.e., 2 Groups x 4 Experiments x 2 Conditions x 3 Laterality x 3 Anterior/Posterior), N400 amplitude was smaller in E1 than in E2, E3 and E4 (Tukey, all $p < .001$; main effect of Experiment: $F_{(3,84)} = 16.91$, $p < .001$).

Table 4. Semantic tasks: N400 amplitudes in µV, standard deviations in brackets

		E1	E2	E3	E4
Day 1	Related	-0.24 (1.50)	-2.17 (1.28)	-1.93 (1.08)	-2.33 (1.56)
	frontal	-1.76 (0.97)	-3.38 (0.82)	-2.66 (0.75)	-3.46 (0.97)
	central	-0.46 (0.93)	-2.58 (0.81)	-2.35 (0.66)	-2.80 (0.97)
	parietal	1.50 (0.83)	-0.55 (0.74)	-0.78 (0.60)	-0.74 (0.89)
	Unrelated	-0.32 (1.55)	-2.00 (1.16)	-2.11 (1.09)	-1.80 (1.48)
	frontal	-1.61 (0.94)	-2.37 (0.67)	-1.93 (0.70)	-2.37 (0.90)
	central	-0.49 (0.99)	-2.46 (0.73)	-2.58 (0.68)	-2.33 (0.94)
	parietal	1.15 (0.90)	-1.17 (0.71)	-1.83 (0.58)	-0.70 (0.83)
Day 30	Related	0.69 (1.50)	-0.79 (1.26)	-0.68 (1.30)	-0.62 (1.24)
	frontal	-0.52 (0.96)	-1.55 (0.83)	-1.43 (0.86)	-1.43 (0.82)
	central	0.34 (0.93)	-1.16 (0.81)	-1.10 (0.76)	-0.96 (0.78)
	parietal	2.26 (0.87)	0.34 (0.71)	0.51 (0.75)	0.52 (0.73)
	Unrelated	0.57 (1.44)	-1.03 (1.25)	-1.68 (1.09)	-1.33 (1.24)
	frontal	-0.35 (0.85)	-1.49 (0.76)	-1.82 (0.76)	-1.93 (0.87)
	central	0.33 (0.88)	-1.38 (0.79)	-2.06 (0.66)	-1.79 (0.78)
	parietal	1.73 (0.89)	-0.22 (0.73)	-1.15 (0.61)	-0.29 (0.67)

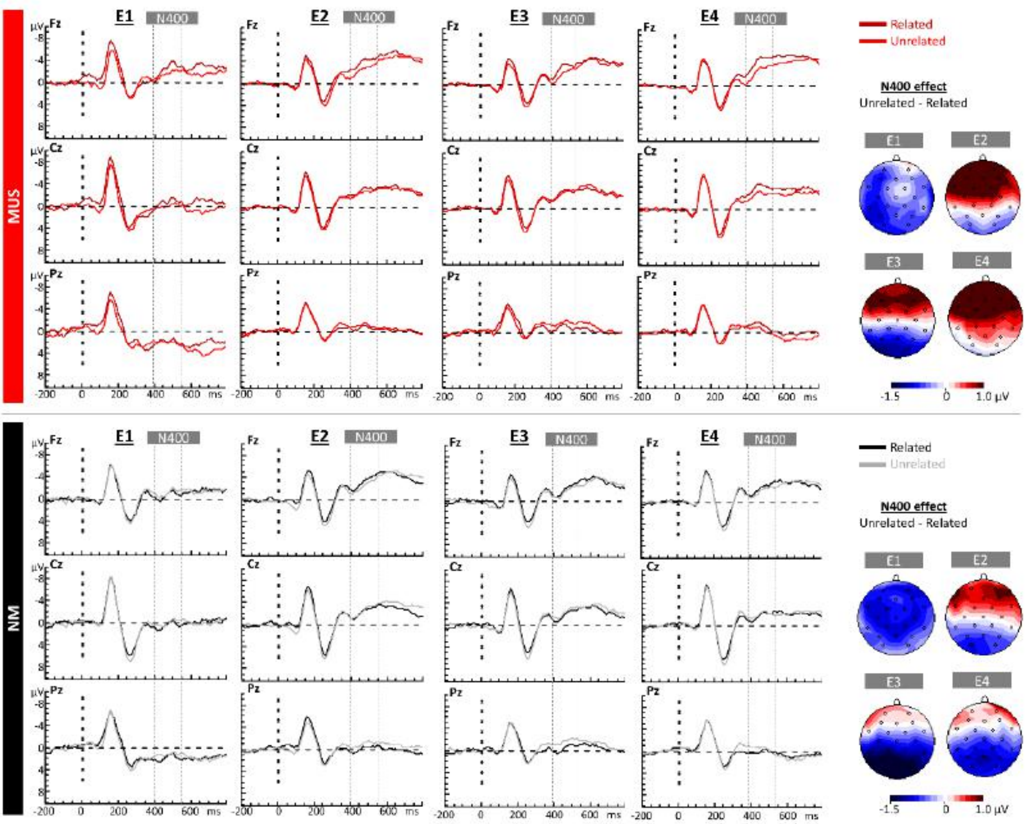


Figure 4. Semantic tasks of the four experiments (E1, E2, E3, and E4) on Day 1. ERPs recorded at frontal (Fz), central (Cz) and parietal sites (Pz) are overlapped for semantically Related (dark-red and black lines) and Unrelated (light-red and grey lines), separately for musicians (MUS; dark and light red lines) and non-musicians (NM; black and grey lines). In this and subsequent figures, time in milliseconds is in abscissa and the amplitude of the effects in microvolt is in ordinate. Time zero corresponds to word onset and negativity is plotted upwards. Latency windows for statistical analyses are indicated with gray dotted lines. Topographic voltage distribution maps of the Unrelated minus Related differences are illustrated for N400 components on Day 1 and on Day 30. Voltage values are scaled from –1.5 µV to +1.0 µV.

Semantic Task, Day 30

Results of separate ANOVAs (i.e., 2 Groups x 2 Conditions; see Table 1C for F- and p-values, and Table 3 for ERR- and RT-values) in the semantic task for each experiment on Day 30, revealed a significant Group x Condition interaction in E1 ($F_{(1,28)}$ = 9.16, p = .005). Separate ANOVAs for each Condition showed that the error increase from D1 to D30 was smaller in MUS than in NM for unrelated words only (MUS: -5.0, SD = 2.9; NM: 3.7; SD = 3.0; main effects of Group for unrelated words: $F_{(1,28)}$ = 8.98, p = .006 and for related words: $F_{(1,28)}$ = 3.84, p = .06). In the other experiments, the error increase from D1 to D30 was smaller in MUS than in NM for both types of words. RT decreases from D1 to D30 were larger in MUS than in NM only in E4. In addition, the error-increase was lower for unrelated than for related trials in E2, E3 and E4, and the RT decrease was larger for unrelated than for related words in all Experiments (see Figure 3B).

Regarding the comparison between experiments (2 Groups x 4 Experiments x 2 Conditions), error increases were smaller and RT decreases were larger in E1 compared to the other three Experiments (Tukey, all p < .001). Moreover, error increases were smaller in E3 compared to E4 (Tukey, p = .05; main effects of Experiment: ERRs: $F_{(3,84)}$ = 33.90, p < .001; and RTs: $F_{(3,84)}$ = 9.51, p < .001).

Turning to the electrophysiological data (see Figure 5), results of separate ANOVAs for each experiment (i.e., 2 Groups x 2 Conditions x 3 Laterality x 3 Anterior/Posterior; see Table 4 for µV values) showed larger N400 to unrelated than to related trials over parietal sites in E1 and E2 (Tukey, p = .009 and p = .004, respectively; Condition x Anterior/Posterior interactions: $F_{(2,56)}$ = 6.13, p = .004 and $F_{(2,56)}$ = 4.50, p = .02, respectively), and over all scalp sites in E3 and E4 (main effects of Condition: $F_{(1,28)}$ = 13.53, p < .001, and $F_{(1,28)}$ = 7.54, p = .01, respectively). The Group x Condition x Anterior/Posterior interaction was only significant for E1 ($F_{(2, 56)}$ = 3.27, p = .05). Separate ANOVAs revealed larger N400 to unrelated than to related trials over parietal sites in MUS (unrelated: 1.56 µV, SD = 1.37; related: 2.56 µV, SD = 1.45; Tukey, p < .001; Condition x Anterior/Posterior: $F_{(2, 28)}$ = 9.86, p < .001) but not in NM (main effect of Condition and Condition x Anterior/Posterior interaction: both F < 1).

Regarding the comparison between experiments (2 Groups x 4 Experiments x 2 Conditions x 3 Laterality x 3 Anterior/Posterior), the N400s were smaller in E1 than in E2, E3 and E4 (Tukey, all p < .001; main effect of Experiment: $F_{(3,84)}$ = 12.50, p < .001).

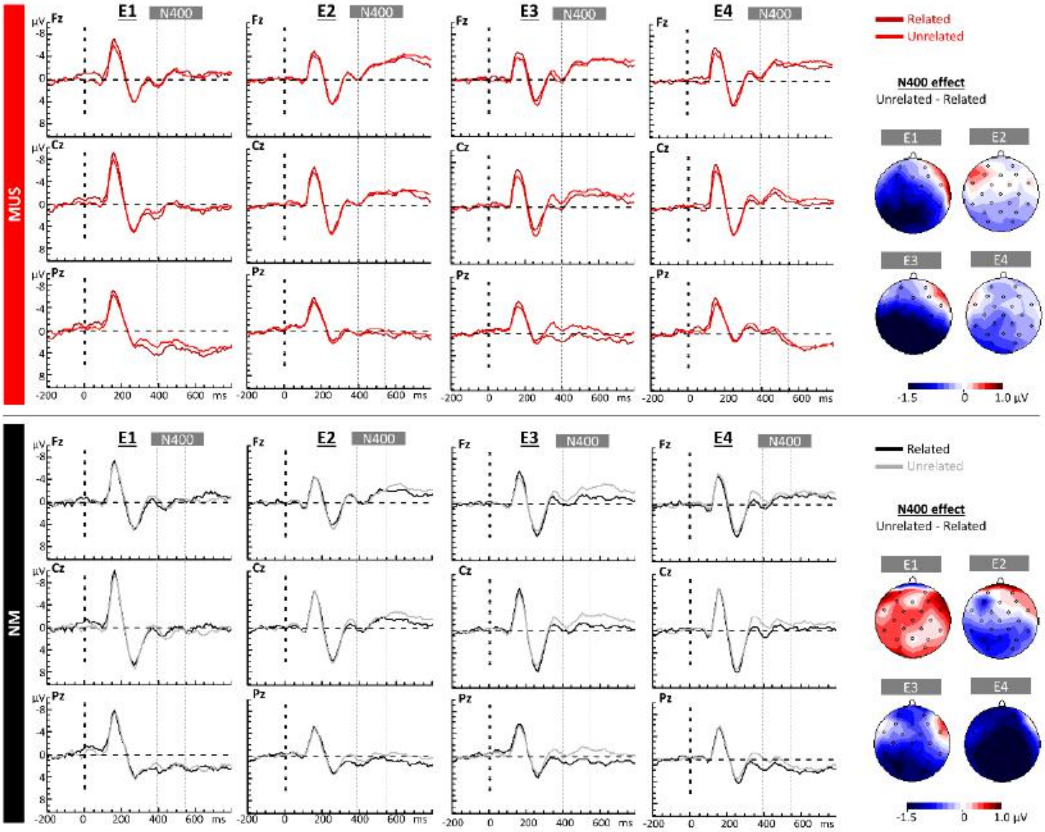


Figure 5. Semantic tasks of the four experiments (E1, E2, E3, and E4) on Day 30. ERPs recorded at frontal (Fz), central (Cz) and parietal sites (Pz) are overlapped for semantically Related (dark-red and black lines) and Unrelated (light-red and grey lines), separately for musicians (MUS; dark and light red lines) and non-musicians (NM; black and grey lines). Topographic voltage distribution maps of the Unrelated minus Related differences are illustrated for N400 components on Day 1 and on Day 30. Voltage values are scaled from –1.5 µV to +1.0 µV.

DISCUSSION

By taking into account previous results showing better novel word learning in musicians than in non-musicians (Dittinger et al., 2016), here we tried to better understand the origin of such an advantage. We adapted the experimental design of Dittinger et al. (2016), and led all participants to reach a learning threshold (i.e., 83 % correct responses) before performing a semantic task. This task modification reduced the influence of auditory perception and associative learning on semantic processing. We asked, whether musicians would still outperform non-musicians in the semantic task, how audio-visual integration contributed to the musician's advantage in novel word learning, and whether musicians would remember the newly-learned words better than non-musicians. Overall, results of both behavioral and electrophysiological data failed in evidencing a musician's advantage in semantic processing or associative learning on the day of learning, but showed better long-term memory of the novel words. These results point to the importance of auditory perception and attention during facilitated novel word learning in musicians in the Dittinger et al. study (2016). Moreover, they suggest that the interaction of better perception and long-term memory as a function of music training leads to long-term advantages in novel word learning in musicians. These findings are discussed in detail below.

Day 1: Learning-matching task cycles

In order to neutralize the influence of auditory perception and associative learning, participants performed a variable number of short learning-matching task cycles until they all reached a level of 83 % correct responses in the matching task. Successful learning in all four experiments was reflected by lower error rates and faster reaction times in the End- compared to the Start- session. Overall, participants needed five cycles to reach the learning threshold (i.e., five repetitions during the learning phases, and five repetitions during the matching tasks). Such fast learning is not surprising, in view of the word learning literature evidencing that novel word encoding can be successful only with few repetitions (Batterink & Neville, 2011; Borovsky, Elman, & Kutas, 2012; Dobel et al., 2010; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007; Perfetti, Wlotko, & Hart, 2005). Based on previous results in the Dittinger et al. study (2016), showing faster N400 increases with learning in musicians than in non-musicians, we expected between-group differences on the number of cycles. However, this was not the case, and it is possible that fast plasticity effects that were visible on electrophysiological data in Dittinger et al. (2016), were not visible at a behavioral level here. Analyzing electrophysiological data of such learning-matching task cycles may shed some more light on this question in future studies.

Interestingly, the number of cycles varied as a function of the experiment, with significantly more cycles in E1 (N = 9) and significantly fewer cycles in E4 (N = 2.4). This likely results from the different stimuli used in the different experiments. While in E4, novel words were familiar French words, in E1, novel words were unfamiliar monosyllabic Thaï words (including aspiration, tonal and duration variations) that were difficult to discriminate. Thus, participants clearly needed more cycles in E1 even if the number of words that needed to be learned was half the number of words in the other three Experiments (i.e., 12 words compared to 24 words in E2, E3 and E4)[6]. In E1, participants needed on average nine cycles to reach the threshold (i.e., 18 repetitions) which is in line with the 20 repetitions of each picture-word association and with the level of performance in the matching task in the Dittinger et al. (2016) study.

The interpretation of E1 being the most difficult and E4 being the easiest experiment is in line with results in the Start session of the matching tasks showing the highest error rates and slowest reaction times in E1, and the lowest error rates and fastest reaction times in E4. Importantly, while E4 remained the easiest task in the End session, participants reached a comparable level in E1, E2 and E3 (as shown by similar error rates and reaction times). Thus, although difficult words need more repetitions during learning, they can also be encoded with accuracy. That words were successfully learned in E1 was further reflected by fewer errors for match than for mismatch words, and this result is in line with previous literature (Boddy & Weinberg, 1981; Dittinger et al., 2016). By contrast, in E2, E3 and E4 error rates were higher for match than for mismatch words in the Start session and similar for match and mismatch words in the End session. We might interpret this result as suggesting that typical matching effects only occur with a certain number of repetitions and after deeper encoding, as was possibly the case for words in E1. Alternatively, results may also reflect different response strategies as a function of the nature of the stimuli: in E2, E3 and E4, words were easy to discriminate, and it might have been possible to recognize the mismatching words based on more shallow information, without necessarily remembering the picture-word association. Such a strategy, leading to an advantage for mismatch trials, would not have been possible in E1, where words needed to be discriminated based on fine-grained auditory perception, and deep encoding is likely to have been necessary to respond correctly to match and mismatch trials.

Finally, and most important for the aim of this study, error rates and reaction times in the matching task, were similar for musicians and non-musicians in all four experiments, thereby reaching the intended aim for this study. Importantly, these results form the basis for assuming that the participants' performance on the semantic tasks (on Day 1 and Day 30) was only minimally influenced by auditory perception and learning-related processes.

---

[6] The number of words to be learned in each experiment was decided based on results of pilot studies.

Day 1: Semantic task

For Day 1, the aim of the study was two-fold: first, we aimed at answering the question of whether musicians would still outperform non-musicians in the semantic task once between-group differences in word perception and associative learning had been neutralized (E1 and E2). And second, we tried to better understand the contribution of audio-visual integration to the musician's advantage in novel word learning (E3 and E4). Behavioral and electrophysiological results in the semantic tasks did not reveal significant between-group differences in any of the four experiments, thereby failing to evidence a musician's advantage in semantic integration or audio-visual integration on Day 1.

In regard to the Dittinger et al. study (2016), these results suggest the following conclusions: first, both musicians and non-musicians were able to rapidly generalize the novel words' meanings to novel concepts even after fewer learning trials (i.e., about 18 repetitions in E1, and 5-8 in E2, E3, and E4) than in Dittinger et al. (2016; 20 repetitions). That novel words had been integrated into semantic networks was reflected by typical semantic priming effects (Meyer & Schvaneveldt, 1971) with shorter reaction times for related than for unrelated trials in all four experiments. However, in all four experiments participants also made more errors to related than to unrelated words, thereby producing a speed-accuracy trade off. In Dittinger et al. (2016) such a speed-accuracy trade off was also present, and was interpreted as a response bias towards rejection (i.e., considering the word as unrelated to the picture), as a consequence of task difficulty generating higher uncertainty (Gigerenzer, 1991). That is, when participants were not certain whether a picture and a word were semantically related, they tended to respond that they were not related. This interpretation is likely to also account for the results obtained here.

Turning to electrophysiological data, both musicians and non-musicians had integrated the novel words into semantic networks in E1, E2 and E3, as reflected by N400 effects with larger amplitudes to unrelated than to related words over parietal sites (Dittinger et al., 2016; Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). By contrast, in E4, results showed larger N400 to related than to unrelated words (inversed N400 effects) over frontal electrodes. As discussed in Dittinger et al. (2016), such reversed N400 effects may reflect ongoing building up and consolidation of episodic memory traces even during test phases, and are in line with previous literature evidencing increased frontal N400 amplitudes during learning of meaningful novel words (Cunillera et al., 2009; De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007; Francois et al., 2017). Moreover, since only related trials provided correct information on the novel words' meanings, it is not surprising that related trials elicited larger amplitudes compared to unrelated trials. If this interpretation is correct, results in E4 possibly reflect that words had not been integrated into semantic networks at that point of time (as reflected by the lack of typical N400 effects over parietal sites), even though E4 seemed to be the easiest experiment on Day 1 (as reflected by fastest learning and lowest error rates in the matching and semantic tasks). In this respect, results seem to show that the apparition of semantic priming effects is not necessarily synchronized with behavioral performance.

Moreover, inversed frontal effects, in addition to parietal N400 effects, were also obtained in E2 and E3, and in E1 for musicians only. In fact, we did obtain between-group differences regarding the N400 component in E1: while musicians showed larger N400 to related than to unrelated words over frontal sites, non-musicians were characterized by larger N400 to unrelated than to related words over midline and right hemisphere. This result stands in contrast to results in the Dittinger et al. study (2016), in which musicians showed parietal N400 effects and non-musicians inversed frontal effects. This finding could suggest that, once the musician's advantage on auditory perception and associative learning had been neutralized, semantic integration (at the electrophysiological level) was not similar between groups, but even slower in musicians compared to non-musicians. In regard to results on Day 30 (discussed below), we might speculate whether such slower semantic integration in musicians (as reflected by inversed frontal rather than parietal N400 effects), led to better encoding and better long-term memory traces compared to non-musicians. However, since this result was unexpected in view of our hypothesis based on the previous Dittinger et al. study (2016), and since we did not obtain similar results in E2, E3 and E4, this interpretation should be taken with caution.

In view of our second aim, general associative learning mechanisms most probably did not play a facilitative role during better word learning in musicians in Dittinger et al. (2016). This interpretation rests upon results in our control conditions of associating two items (i.e., picture and word; E3 and E4), that did not show any between-group differences, independent of whether picture-word associations contained semantic information or not. These findings are surprising in regard to two previous studies evidencing better audio-visual integration in musicians: Bidelman (Bidelman, 2016) investigated audio-visual integration in musicians and non-musicians by using a double-flash illusion, and showed that musicians were both faster and more accurate at processing concurrent audio-visual cues than non-musicians, evidencing better multisensory integration and audio-visual binding in musicians. Second, Paraskevopoulos and colleagues (Paraskevopoulos et al., 2012) used magnetoencephalography to identify the neural correlates of an audio-visual incongruency response generated due to violations of a congruent rule (i.e., "the higher the pitch of the tone, the higher the position of the circle"). Musicians scored significantly better, with increased activity in frontal, temporal, and occipital areas as a response to multisensory integration. Importantly however, the discrimination of congruent from incongruent trials did not only rely on audio-visual integration, but also on working memory, since one has to keep in mind all tone-circle pairs presented and maintain attention on the comparison of the audio-visual stimuli to the congruent rule. Based on the very different task used in Bidelman (2016), as well as on the fact that Paraskevopoulos et al. (2012) used magnetoencephalography, together with the drawback of possibly interpreting their results also as a function of working memory, we might question whether results from our experiment are comparable to these two studies. In addition, it is possible that our task, possibly reflecting a more natural situation for many participants compared to the experimental designs above, tapped into higher cross-modal integration processes that would not be different for musicians and non-musicians.

Finally, E1 and E3 were more difficult than E2 and E4 (as reflected by higher error rates). For E3, such a result was expected based on previous literature evidencing worse memory performance for words that are not associated to semantic information (Heikkilä, Alho, Hyvönen, & Tiippana, 2015; Savill, Ellis, & Jefferies, 2017; Takashima, Bakker, van Hell, Janzen, & McQueen, 2017). For E1, this finding is interesting in view of the previous Dittinger et al. studies, and in extending previous studies evidencing that difficult task conditions are more sensitive to between-group differences (Schön, Magne, & Besson, 2004): in fact, between-group differences were obtained in three independent studies using the same experimental design, with clear advantages for musically-trained participants in a sample of children (Dittinger et al., 2017) and of young adults (Dittinger et al., 2016), and with smaller transfer effects in older adults (Dittinger et al., under review). Having failed to reproduce similar results without the advantage of better auditory perception and attention in musicians here, our results suggest that between-group differences in the Dittinger et al. studies (2016, 2017; under review) were most probably driven by better auditory perception and attention.

Day 30: Semantic task

Participants were re-tested on Day 30, to investigate what remained of newly-learned words after a one-month period, and whether musicians would remember the words better than non-musicians. Interestingly, musicians outperformed non-musicians in all four experiments on Day 30. This result was reflected by smaller error increases from Day 1 to Day 30 for both related and unrelated words, except in E1 where this effect was only significant for unrelated words. RT decreases from Day 1 to D30 were also larger in MUS than in NM only in E4.

At the electrophysiological level, the N400 was larger to unrelated than to related words across all scalp sites in E3 and E4 and only over parietal sites in E1 and E2. This finding suggests that ecological valid conditions in regard to novel word learning and semantic processing (i.e., E1 and E2) elicited N400 effects restricted to parietal scalp sites, as previously reported in the literature (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). By contrast, results of control conditions of associative learning (E3 and E4) showed largely distributed effects, eventually reflecting more general processes compared to more specific semantic processes for E1 and E2.

Importantly, in E1 the N400 effect over parietal sites also differed as a function of music training with statistically significant effects in musicians but not in non-musicians. Thus, in addition to better behavioral performance in all four experiments, between-group differences on semantic processing on Day 30 were also reflected by the N400 component, but only when words are difficult to discriminate (i.e., Thai words in E1). In the other three experiments (that included disyllabic Finnish or French words) both musicians and non-musicians showed similar N400 effects.

These findings suggest better long-term memory in musicians, with larger advantages for musicians in difficult task conditions (Schön et al., 2004): first, in E1 musicians outperformed non-musicians on unrelated trials only (although between-group differences on related trials very closely passed statistical significance, with p = .06). Second, only in E4 did we obtain both smaller error increases and larger RT decreases in musicians than in non-musicians. Remembering the words learned in E4 might have been more difficult compared to E1, E2, and E3, since the real meaning of the French words, anchored in deeply encoded memory traces, probably interfered with the newly-attached meanings. This interpretation is also in line with a delayed apparition of N400 effects over parietal sites in E4 on Day 30 so that semantic integration of the novel meanings possibly relied more strongly on consolidation periods to overcome such interference effects (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Dumay & Gaskell, 2007; Takashima, Bakker, van Hell, Janzen, & McQueen, 2014). Having found better long-term memory for the words in E4 in musicians, may reflect smaller interference effects in musicians, and may be interpreted along better executive functions in musicians (Rogalsky, Rong, Saberi, & Hickok, 2011; Zuk et al., 2014). Third, the significant between-group differences on the N400 effect only in E1 (that without doubt was the most difficult experiment, see discussion above), further points to the influence of task difficulty.

A final point, related to task difficulty, deserves comments. Although E1 was very difficult for all participants, it triggered the lowest error increases and the largest reaction time decreases from Day 1 to Day 30. These results may be explained by the increased number of repetitions in the learning-matching task cycles in E1 compared to the other Experiments (i.e., 9 cycles in E1 versus 2.4 in E4, 3.6 in E2 and 3.8 in E3). This double to threefold number of repetitions may have resulted in deeper encoding with stronger memory traces for long-term memory. This interpretation is also reflected to some extent by the results of E4: while E4 was very easy on Day 1, error rates did not significantly differ from those of E2 or E3 on Day 30, and error increases in E4 were even higher compared to those in E3, suggesting that the very fast learning with the lowest number of repetitions on Day 1 was not enough to create deep memory traces. This is an interesting issue and future experiments may focus on the influence of the number of repetitions onto long-term memory, by for example testing large groups of participants and testing whether participants who learn most slowly (i.e., with the highest number of learning-matching task cycles), remember best after 30 days.

Two different interpretations possibly account for better long-term memory in musicians. First, even though we did not obtain evidence for better word encoding in musicians at a behavioral level (i.e., number of learning cycles), it is possible that musicians have encoded the novel words better compared to non-musicians. In this respect it is important to note, that the acquisition of novel words is not only a learning process, but also a memory process, since remembering words is as important as learning them (Wojcik, 2013). This means that what was remembered on Day 30, had been encoded and integrated into memory on Day 1, thereby eventually putting the conclusion according to which better novel word learning in musicians is mainly driven by auditory perception and attention (presented above) into question. Further experiments may shed more light on this issue by focusing on ERPs during novel word encoding (i.e., learning phase on Day 1), and by sorting them as a function of whether or not they are subsequently remembered on Day 30 (i.e., examining differences at encoding based on subsequent memory, the so-called Dm effect; Paller, Kutas, & Mayes, 1987).

Second, musicians may have profited better from long-term consolidation. In this respect, it is possible that they remembered the newly learned words better on Day 30, although initial encoding and integration were similar for both groups. While these two interpretations are difficult to disentangle with the present data, the important point remains that musicians showed better long-term memory performances for novel words, a result that, to the best of our knowledge, has not been shown before.

Taken together, these results extend to long-term memory transfer effects previously described from music training to working and short-term memory (George & Coch, 2011; Ho, Cheung, & Chan, 2003; Kraus, Strait, & Parbery-Clark, 2012; Schulze et al., 2012; for meta-analysis see Talamini, Altoè, Carretti, & Grassi, 2017). While results of Day 1 suggested that the musician’s advantage in novel word learning was mainly driven by auditory perception and attention, considering results of Day 30 may extend this conclusion by suggesting that the interaction of auditory perception and better cognitive functions (especially long-term memory) possibly optimizes novel word learning in musicians. Importantly, such interactions of perceptive and cognitive functions, as well as transfer effects from music training to long-term memory and novel word learning, adds evidence in favor of domain-general networks in the brain underlying perception, cognition and language processing (for review on the hotly debated issue of domain-specific versus domain-general networks in the brain, see Besson, Dittinger, & Barbaroux, 2018).

Conclusion

In sum, by applying a similar experimental design compared to Dittinger et al. (2016), and by comparing four sets of novel words, results showed three main findings: first, based on the lack of between-group differences on Day 1, we conclude that the musician’s advantage in novel word learning does not result from better semantic integration and/or associative learning, but strongly relies on better auditory perception and attention. Second, better behavioral performance on Day 30 in musicians than in non-musicians suggest that music training increases long-term memory of newly-learned words. Third, only when novel words are difficult to discriminate (E1), the musician’s better performance is accompanied by larger N400 effects. Altogether, these findings show that transfer effects from music training to novel word learning (i.e., semantic levels of speech processing) are in large part driven by an enhanced auditory system, but profit additionally from reinforced (long-term) memory functions. To the best of our knowledge, these results evidence for the first time transfer effects from music training to long-term memory. Since short- and long-term memory are crucial cognitive functions central to language processing, these results highlight the potential impact of music training starting at a young age.

Acknowledgments

We thank all the participants, as well as Chotiga Pattamadilok, Aino Niklas-Salminen and Mathilde Niklas-Salminen for registering the auditory stimuli. Moreover, we thank Ken Paller and James McQueen for helpful discussions in the building of the task design. The present work was carried out within the Labex BLRI (ANR-11-LABX-0036) and has benefited from support from the French government, managed by the French National Agency for Research (ANR), under the program “Investissements d’Avenir” (ANR-11-IDEX-0001-02). ED was supported by a doctoral fellowship from the BLRI, and BK benefitted from a grant from the A*MIDEX program (ANR-11-IDEX-0001-02) funded by the French Government under the program “Investissements d’Avenir".

REFERENCES

Alario, F. X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*(3), 531–552.

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

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41.

Batterink, L., & Neville, H. (2011). Implicit and Explicit Mechanisms of Word Learning in a Narrative Context: An Event-related Potential Study. *Journal of Cognitive Neuroscience*, *23*(11), 3181–3196.

Besson, M., Dittinger, E., & Barbaroux, M. (2018). How music training influences language processing: Evidence against informational encapsulation. In *L’année psychologique/ Topics in Cognitive Psychology*, 118, 273–288.

Besson, M., Barbaroux, M., & Dittinger, E. (2017). Music in the brain: Music and language processing. In *The Routledge Companion to Music Cognition.* Eds. R. Ashley and R. Timmers. Routledge / Taylor & Francis, New York.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, *234*(10), 3037–3047.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662–2673.

Boddy, J., & Weinberg, H. (1981). Brain potentials, perceptual mechanisms and semantic categorisation. *Biological Psychology*, *12*(1), 43–61.

Boersma, P., & Weenink, D. (2011). *"Praat: Doing phonetics by computer [computer program]", http://www.praat.org*

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is Enough: N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. *Language Learning and Development*, *8*(3), 278–302.

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2015). *d2-R : test d'attention concentrée [d2-R: test of attention]. (Éditions Hogrefe France).*

Cardebat, D., Doyon, B., Puel, M., Goulet, P., & Joanette, Y. (1990). [Formal and semantic lexical evocation in normal subjects. Performance and dynamics of production as a function of sex, age and educational level]. *Acta Neurologica Belgica*, *90*(4), 207–217.

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced Passive and Active Processing of Syllables in Musician Children. *Journal of Cognitive Neuroscience*, *23*(12), 3874–3887.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., … Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*(3), 541–553.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PloS One*, *2*(11), e1175.

Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (*under review*). Testing Relationships between Music Training and Word Learning across the Lifespan by using Cross-Sectional and Comparative Electrophysiological Approaches in Children, young Adults, and older Adults.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-lasting Word Representations. *Journal of Cognitive Neuroscience*, *28*(10), 1584–1602.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast Brain Plasticity during Word Learning in Musically-Trained Children. *Frontiers in Human Neuroscience*, *11*.

Dittinger, E., D'Imperio, M., & Besson, M. (2018). Enhanced neural and behavioural processing of a nonnative phonemic contrast in professional musicians. *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13939

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: on the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261.

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35–39.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex*, *22*(3), 650–658.

Fitzroy, A. B., & Sanders, L. D. (2012). Musical expertise modulates early processing of syntactic violations in language. *Frontiers in Psychology*, *3*, 603.

Francois, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*. 98, 56-67.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.

Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases." *European Review of Social Psychology*, *2*(1), 83–115.

Heikkilä, J., Alho, K., Hyvönen, H., & Tiippana, K. (2015). Audiovisual Semantic Congruency During Encoding Enhances Memory Performance. *Experimental Psychology*, *62*(2), 123–130.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439–450.

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY–Second Edition (NEPSY-II). *San Antonio: The Psychological Corporation.*

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

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory: Kraus et al. *Annals of the New York Academy of Sciences*, *1252*(1), 100–107.

Kroll, J. F., & Potter, M. C. (1984). Recognizing Words, Pictures, and Concepts: A Comparison of Lexical, Object, and Reality Decisions. *Journal of Verbal Learning and Verbal Behavior, 23*, 39–66.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

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), 2701–2715.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews. Neuroscience*, *3*(6), 473–478.

Paller, K. A., Kutas, M., & Mayes, A. R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology*, *67*(4), 360–371.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical expertise induces audiovisual integration of abstract congruency rules. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(50), 18196–18203.

Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*, *999*, 58–75.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1281–1292.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, *31*(10), 3843–3852.

Savill, N., Ellis, A. W., & Jefferies, E. (2017). Newly-acquired words are more phonologically robust in verbal short-term memory when they have associated semantic representations. *Neuropsychologia*, *98*, 85–97.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–514.

Schellenberg, E. G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology (London, England: 1953)*, *102*(3), 283–302.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception: An Interdisciplinary Journal*, *29*(3), 255–267.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology. Human Learning and Memory*, *6*(2), 174–215.

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), 22–29.

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–278.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language*, *167*, 44–60.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, *12*(10), e0186773.

Wang, X., Ossher, L., & Reuter-Lorenz, P. A. (2015). Examining the relationship between skilled music training and attention. *Consciousness and Cognition*, *36*, 169–179.

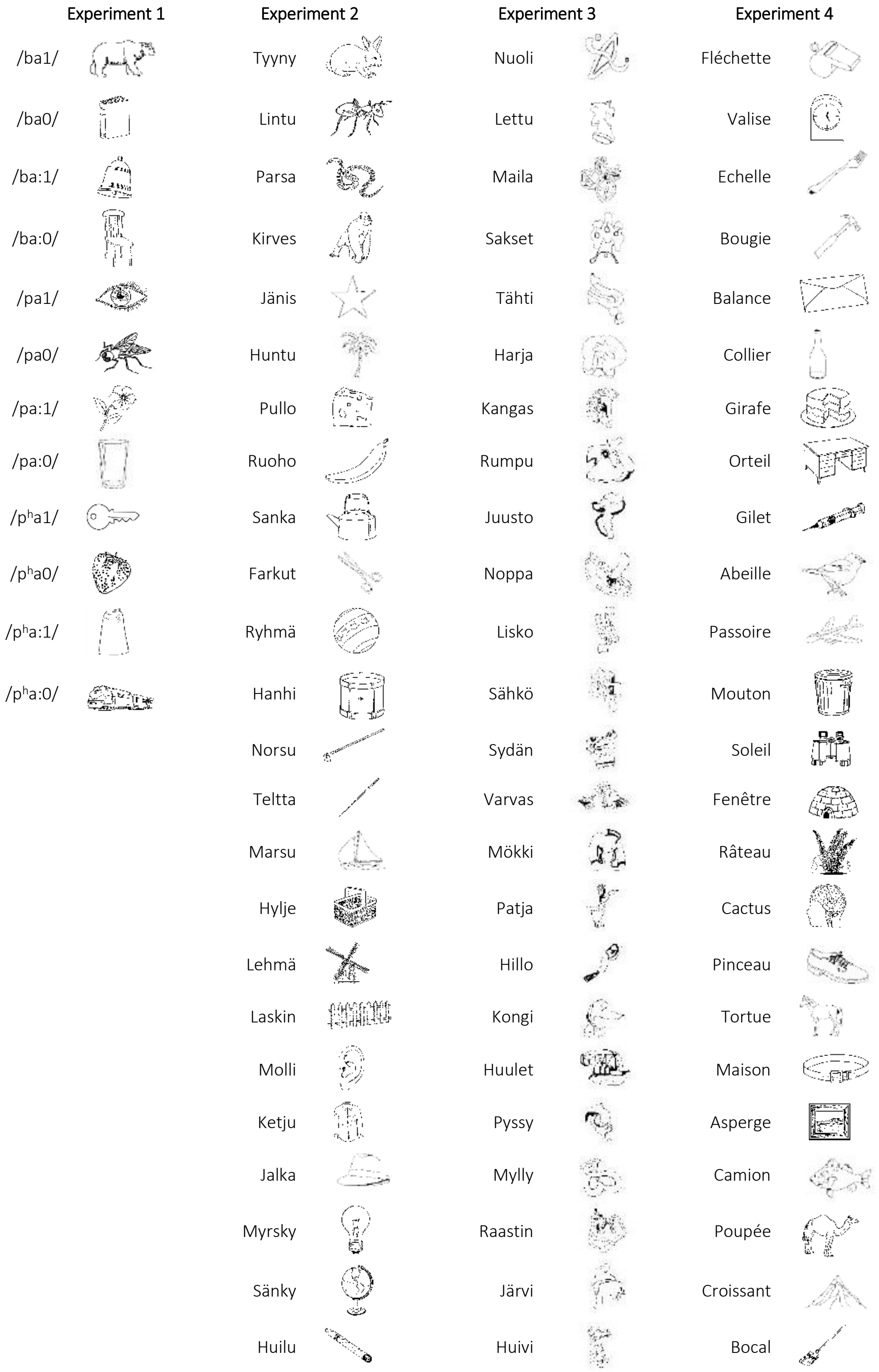
Wechsler, D., Coalson, D. L., & Raiford, S. E. (1997). *WAIS-III: Wechsler adult intelligence scale. Psychological Corporation San Antonio, TX.*

Wojcik, E. H. (2013). Remembering New Words: Integrating Early Memory Development into Word Learning. *Frontiers in Psychology*, *4*.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04).

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLOS ONE*, *9*(6), e99868.

Supplementary Material



Supplementary Figure 1. Novel words that needed to be learned are shown for the four Experiments.

Chapter 7

GENERAL DISCUSSION & CONCLUSION

Language and music are two fascinating activities of the human cognitive repertoire, encompassing basic and higher-cognitive abilities, from motor control and auditory perception to emotional and attention-related processes (Fitch, 2010; Koelsch, 2014; Kraus & Chandrasekaran, 2010; Patel, 2008; Zatorre, 2013). Regarding music processing, structural and functional brain changes as a consequence of intensive and regular music training have been frequently described (Bangert & Schlaug, 2006; Bermudez, Lerch, Evans, & Zatorre, 2009; Hyde et al., 2009; Kühnis, Elmer, Meyer, & Jäncke, 2013; Schneider, Sluming, Roberts, Bleeck, & Rupp, 2005; Sluming, Brooks, Howard, Downes, & Roberts, 2007), and have been evidenced to translate into behavioral advantages (e.g., Kühnis et al., 2013; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011; Zuk et al., 2014). Importantly, growing literature also shows that such advantages as a function of music training transfer to language processing (see Besson et al., 2011 for review), favoring a domain-general view of the underlying brain networks. However, the issue of whether music and language processing rely on shared or distinct neuronal networks in the brain, remains a matter of debate. The research presented in this thesis evaluated whether brain plasticity mechanisms induced by music training allowed advantages in novel word learning[22], and whether such potential transfer effects are best explained by cascading effects (better auditory perception drives the advantages) or by multidimensional effects (interaction of perceptive and cognitive functions). To do so, several questions related to novel word learning have been answered, and results are discussed in detail below.

First, we investigated whether professional musicians showed advantages compared to non-musicians in phoneme perception, a cornerstone for novel word learning (presented in Chapter 3). Results of both behavioral and electrophysiological data showed that musicians outperformed non-musicians in the task condition in which non-native phoneme perception was relevant. Specifically, musicians made fewer errors and showed larger differences on N200 and P300 amplitudes between native and non-native phonemes, suggesting that musicians were more sensitive to the difference between phonemes. These results are in line with several previous studies revealing transfer effects from music training to the segmental and supra-segmental levels of speech perception in foreign languages (for review see Chobert & Besson, 2013): for example, Delogu and colleagues (Delogu, Lampis, & Belardinelli, 2010) demonstrated

---

[22] Based on the auditory stimuli, including non-native attributes for French speakers, used in our experimental designs presented in Chapters 3 to 6, our results possibly refer to novel word learning of a foreign language, rather than to word acquisition in one's mother tongue. Readers may keep this information in mind when interpreting our results.

that music expertise was associated with better lexical tone processing of monosyllabic Mandarin Chinese words during a same-different task (see also Marie, Delogu, Lampis, Belardinelli, & Besson, 2011). Similarly, Sadakata and Sekiyama (Sadakata & Sekiyama, 2011) showed that musicians enjoyed an advantage in the discrimination of the timing and quality of Japanese consonants as well as the quality of Dutch vowels, suggesting better perception of acoustical features (i.e., pitch and timing) of speech in a second language. Finally, Marques and colleagues (Marques, Moreno, Castro, & Besson, 2007) showed better pitch processing in musicians, as was reflected by better behavioral performance and faster ERP latency modulations in response to prosodically congruous and incongruous sentence-final words.

By contrast, results of a recent study by Swaminathan and Schellenberg (Swaminathan & Schellenberg, 2017) showed no evidence for an association between music training and non-native phoneme discrimination. Specifically, results showed a positive correlation between rhythm perception, but not music training, and speech perception, and the authors concluded that natural abilities were a better predictor of speech perception than music training. In this context, it is interesting to make a detour to the issue of natural abilities versus training-related advantages, an issue that is highly debated. Specifically, natural abilities relate to genetic and epigenetic backgrounds that possibly predisposes each individual to experience the world in a certain way, and to seek out for activities that fulfill each individual's innate skills. One argument that has been frequently used in favor of a genetic predisposition for music is that all humans are able to perceive, produce, and enjoy music, even in the absence of formal music training. However, the genetic influences on music development and expertise have been little investigated (for review see Tan, McPherson, Peretz, Berkovic, & Wilson, 2014), compared to the growing literature on neuroplastic changes as a function of music training. As mentioned previously, one way to ascertain the influence of music training, is to conduct longitudinal studies with non-musicians trained with music or with an equally interesting non-musical activity (Chobert et al. 2014; Forgeard et al. 2008; François et al. 2013; Moreno et al. 2009). However, since longitudinal studies are time- and cost-intensive, their implementation is difficult. Thus, possibilities of interactions between neuroplastic phenomena and genetic and epigenetic predispositions remain unclear, and more developmental perspectives are needed to better understand their contribution to the "musician's brain", as recently pointed out by Zuk and Gaab (Zuk et Gaab, 2018). While this issue limits the interpretation of results of several studies (as well as of the results presented within this thesis), the recent study by Swaminathan and Schellenberg

(Swaminathan & Schellenberg, 2017) mentioned above possibly failed to find between-group differences as a function of music training due to the fact that the musicians tested in this study had on average only 4.9 years of music training (compared to about 15 years of music training for musicians tested here). Taking into consideration that certain transfer effects may only develop after extensive music training started at early ages, this negative result may not be surprising.

Coming back to our results, the reader might wonder how non-native phoneme perception relates to novel word learning. Learning new words is a complex process that consists of several steps: the perception of the novel word; the creation of a word representation in phonological short-term memory; and finally the relation of the acquired word form to pre-existing conceptual knowledge and to the encoding in long-term memory (for detailed descriptions of the different phases of word learning, see Service & Craik, 1993). In this context, it has been proposed that learners with high phonological short-term memory capacity are better at maintaining intact representations of words in short-term memory, which in turn facilitates encoding of the word representation in long-term memory (Gathercole, Willis, Braddeley, & Emslie, 1992; Gathercole & Baddeley, 1989). Importantly, a phonological mechanism directly linked to phonological short-term memory is phonological awareness, which requires the recognition and/or manipulation of syllables, phonemes, and other phonological units. In fact, upon hearing a novel word, the phonological units already stored in long-term memory are possibly activated, and reused to build the mnemonic representation of the novel word (de Jong, Seveke, & van Veen, 2000; Ramachandra, Hewitt, & Brackenbury, 2011; Service, Maury, & Luotoniemi, 2007). Since the efficiency and accuracy of this process increases with the quality of phonological representations available to the learner, better phoneme perception in musicians may therefore provide the basis for better word encoding in musicians.

At this point it is tempting to conclude that better novel word learning in musicians is a direct consequence of better auditory perception, and thus, results from a cascading effect (as described on page 35). However, I will come back to this issue further ahead in the discussion, with the conclusion that, rather than considering better auditory perception as the only driving factor, the strong interaction between auditory perception and cognitive functions optimally boosts novel word learning in musicians.

Second, the main aim of this thesis was to determine whether musicians learned the meaning of novel words better compared to non-musicians (presented in Chapter 4). Specifically, the aim was two-fold: not only were we interested in understanding whether music training facilitated novel word learning, but also whether such potential training-related advantages were present across the lifespan. Thus, a very similar experimental design (i.e., novel word learning task with picture-word associations) was used in children, younger adults as well as older adults. Musically-trained children outperformed children without music training in the matching task that tested for the efficiency of learning the picture-word associations, as well as in the semantic task that tested for semantic integration of the novel words into pre-existing semantic networks. Moreover, in both tasks, the expected N200 and N400 effects (i.e., the difference between mismatching/unrelated and matching/related trials in the matching and semantic tasks, respectively) were larger in children with music training compared to controls. Thus, results at both behavioral and electrophysiological levels showed that children with music training learned the meaning of novel words more efficiently than children without music training. Taking into account that children had followed music training for about only five years, these results have potentially strong implications for school education.

Similarly, adult professional musicians made fewer errors than non-musicians in the semantic task and in a 5-months re-test session. Additionally, these behavioral advantages were accompanied by larger N400 effects in both the semantic and matching tasks. The finding of between-group differences on electrophysiological data in both tasks, but better level of performance of musicians compared to non-musicians in the semantic task only was somewhat surprising, and may reflect the fact that the two tasks tap into different memory systems. To decide whether a word matched or mismatched a picture in the matching task, participants could refer to the specific picture-word associations stored in episodic memory during learning. By contrast, to decide whether a learned word was semantically related or unrelated to a novel picture (that the participant had not seen before during the experimental session), participants needed to retrieve more general information from semantic memory. It is thus possible that musicians built stronger memory traces in episodic memory than non-musicians, and that this helped them integrate the novel word meaning into semantic memory and possibly into long-term memory, a hypothesis that was tested later in the experiment presented in Chapter 6. This interpretation possibly explains differences at the electrophysiological level between musicians and non-musicians in both the matching and semantic tasks. Moreover, the finding that non-musicians had learned the picture-word associations similarly to musicians (as reflected by similar error rates in the matching task), but had not integrated the novel words into semantic networks (as reflected by higher error rates in non-musicians than in musicians in the semantic task), is also in line with this interpretation. However, it is also possible that musicians outperformed non-musicians only in the semantic task based on differences in task difficulty, as the semantic task was more difficult than the matching task. This interpretation is in line with previous literature demonstrating between-group differences only in difficult task conditions (Schön et al., 2004).

Finally, results with older adults revealed that the overall faculty of novel word learning was preserved with fast increase of the N400 component over fronto-central sites from the first to the last Block of the learning phase, and more efficient processing of match compared to mismatch words (i.e., lower error rates and faster reaction times) in the matching task. However, the influence of music training on novel word learning was not as clear-cut as in the other two groups of participants. Older musicians and non-musicians performed similarly (considering both error rates and reaction times) in both the matching and semantic tasks. However, in the semantic task, N200 effects were larger in musicians than in non-musicians but overall N400 amplitude was smaller in musicians. Thus, it is possible that transfer effects from music training to semantic processing in older participants were reflected by the N200 effects rather than by the N400 effect, as was found in children and young adults. The larger N200 effect in musicians could have reflected increased word activation and categorization, and the smaller N400 component could have reflected a reduced cognitive load in musicians in the semantic task so that the semantic task required less effort for musicians than for non-musicians. Nevertheless, one may be surprised that in older musicians, who showed the longest period of music training, between-group differences were not found on the N400 effect, the typical marker of semantic processing[23]. As this result was unexpected, we can only propose post-hoc interpretations.

First, previous literature has shown that, although older and younger adults show overall similar levels of performance in speech perception and semantic priming tasks, N400 amplitude can be reduced and N400 latency can be delayed with age. For example, Günter and colleagues

---

[23] Nowadays, the N400, rather than specifically reflecting semantic processing, may be more accurately described as reflecting neural activity in a multimodal and context dependent long-term memory system (Kutas & Federmeier, 2011). In this respect, smaller amplitudes may reflect a smaller amount of brain activity in response to a certain stimulus.

(Günter, Jackson, & Mulder, 1992) investigated ERP responses to congruent and incongruent final words of sentences in older academics (average age of 55 years). Results showed reduced and delayed N400 components in older compared to younger adults, suggesting that the semantic priming effect reflected quantitative rather than qualitative differences of semantic integration. Similar results were obtained by Kutas and Iragui (Kutas & Iragui, 1998), who interpreted these results as slower lexical and semantic access in older adults. More recently, Lee and Federmeier (Lee & Federmeier, 2009) also revealed age-differences on the N400 component. Following the authors' interpretation, younger adults recruit additional control mechanisms for the processing of congruent and incongruent final words of sentences, resulting in increased frontal negativities. By contrast, such top-down control mechanisms are possibly in part lost in older adults, thereby resulting in smaller N400s. In our experiment, novel word learning was very fast, and it is possible that, since N400 components in older participants are smaller and delayed, the expected N400 effect would have needed more time to install during the semantic task (although behavioral performance was overall comparable to young adults). Future studies should investigate whether slower word encoding or longer consolidation periods would increase N400 amplitudes and N400 effects in older participants, potentially bringing to light transfer effects from music training to semantic processing in older adults as we found in children and young adults. In addition, cognitive decline with age is associated with strong individual differences, and possibly relies on stronger interactions of linguistic processes, working memory resources and the reduction of control processes (Kliesch, Giroud, Pfenninger, & Meyer, 2017; Lee & Federmeier, 2009; Wlotko, Lee, & Federmeier, 2010), thereby eventually pointing to the importance of testing larger groups of participants in older adults compared to children and younger adults to obtain statistically significant effects and/or between-group differences.

Second, cognitive decline might not be counteracted by music training, and the advantages of younger musicians in novel word learning might fade away with the natural process of aging. However, some previous results revealed the importance of music training in postponing (at least in part) age-related deficits. Parbery-Clark and colleagues (Parbery-Clark, Anderson, Hittner, & Kraus, 2012) compared brainstem responses to speech syllables in young musicians and non-musicians (18-32 years), and older musicians and non-musicians (45-65 years). Results showed that professional musicians of both ages profited from better phonetic discrimination compared to their respective control group, thereby pointing to sustained transfer effects also in older adults, at least at a subcortical level. Moreover, older musicians have been shown to perform better on speech-in-noise perception, in part based on better auditory working memory compared to non-musicians (Parbery-Clark, Strait, Anderson, Hittner, & Kraus, 2011; Zendel & Alain, 2012). Thus, although music training seems to counteract parts of the aging process and seems to improve speech perception at the subcortical and cortical level, our data indicate that it might not be enough for extending to more complex tasks such as novel word learning, or to semantic processing.

Third, considering that the novel words needed to be discriminated based on acoustic features[24], it is possible that our word learning task included a strong working memory component. In this context and similar to the previous point, it is an open issue whether the transfer effects from music training to working memory in younger adults (George & Coch, 2011) persist in older adults. This is an important issue, since two factors have been shown to possibly influence novel word learning: a person's pre-existing vocabulary knowledge as well as working memory (Daneman & Green, 1986). While older adults could easily learn new vocabulary when considering their generally better vocabulary knowledge compared to younger adults (Kausler, 1991; Long & Shaw, 2000), they could also have increased difficulty with word learning because they typically show reduced working memory capacity (Craik & Jennings, 1992). Furthermore, results by Long and Shaw (Long & Shaw, 2000) showed that age differences are minimized when the task is ecologically valid, so that older adults have the ability to use their accumulated experience and knowledge to compensate for the deleterious effects of aging. Considering that our novel words included foreign phonetic contrasts, older participants probably could not profit from their extended vocabulary knowledge, therefore making the task more difficult for them, compared to children and young adults who typically learn novel words by strongly relying on their working memory capacity (Baltes, 1993; Daneman & Green, 1986). As a consequence, N400 effects possibly did not install by the end of the experimental session.

Related to the novel words used in this experimental design, it is also important to point out that the auditory stimuli were recorded by a French-Thai bilingual to guarantee natural

---

[24] To be learned novel words were natural monosyllabic Thai words, that differed on temporal and tonal variations, with three different Voice-Onset-Times (/b/ vs. /p/ vs. /pʰ/), short and long vowels (/a/ vs. /a:/) and low-tone or mid-tone pitch (/a0/ vs. /a1/). Importantly, since older adults are typically characterized by hearing impairments (Giroud et al., 2018), all participants passed a test of pure tone audiometry (MAICO ST 20, MAICO Diagnostic, GmbH, Berlin) before being accepted to participate in the study. As a consequence, all older adults in our study showed an unremarkable audiological status (i.e., all tested frequencies could be heard below a threshold of 30 dB, frequency-range of 250-8000 Hz).

pronunciation of native phonemes for French children and young adults. However, older adult participants were German speakers, and we cannot exclude that the difference in mother tongue may have influenced the experimental outcome in an unexpected way. For instance, it is possible that there was not enough time for semantic integration to install in older adults because the auditory stimuli were too difficult to perceive and to discriminate.

In sum, our comparative approach, including children with music training as well as professional musicians of younger and older age, shed some light on novel word learning across the lifespan, transfer effects from music to speech and how they develop with age. Behavioral results have indicated that participants of all ages were able to rapidly learn and integrate novel words as reflected by lower error rates and/or faster reaction times to match/ related words than to mismatch/ unrelated words. By contrast, while the typical electrophysiological marker of novel word learning (i.e., N400 effect) was clearly present in the semantic task in children and young adults, it was very small to absent in older adults. Regarding influences of music training, children with music training and young professional musicians showed clear advantages compared to participants without music training. For older adults, between-group differences were less clear-cut. While understanding the neural underpinnings of word learning in infants and young adolescents is a topic of interest, not much is known about how these learning mechanisms change throughout life, although both younger and older adults continually confront a changing world and need to adapt by acquiring new skills and information (as for example, novel words). The data presented here contribute to this field of research, and future studies are clearly needed to further explore how mechanisms implied in novel word learning change with age. Overall, based on previous literature evidencing transfer effects from music training to various perceptive and cognitive functions (for a review see Besson et al., 2017) that may also subserve language acquisition[25], it is not surprising to find transfer effects from music training to novel word learning. These findings are in line with growing research evidencing transfer effects from music training to different aspects of speech perception (Bidelman et al., 2014; Chobert, Marie, François, Schön, & Besson, 2011; Elmer et al., 2012; Kraus & Chandrasekaran, 2010; Linnavalli, Putkinen, Lipsanen, Huotilainen, & Tervaniemi, 2018; Moreno et al., 2009; Nan et al., 2018; Sadakata & Sekiyama, 2011; Tervaniemi et al., 2009; Wong & Perrachione, 2007). They point to the potential of music training in facilitating the acquisition of

---

[25] The issue of domain-specific versus domain-general perceptive and cognitive functions underlying language processing is a frequently debated hot topic that is considered further ahead in this discussion.

novel words of a foreign (mainly tonal or quantitative) language, a result that should have strong implications especially for education (Tervaniemi, Tao, & Huotilainen, 2018).

To go one step further in the analyses of electrophysiological data, we investigated the dynamics of neuronal oscillations and their synchronization across distributed brain areas during novel word learning in young adults (presented in Chapter 5). In this context, Elmer and Jäncke (Elmer & Jäncke, 2018) recently pointed to the importance of extending research from investigating local changes to the modeling of neural networks. In fact, network approaches will possibly allow for a better understanding of the neural dynamics underlying complex cognitive functions involved in word learning and to how they are modulated as a function of music training.

Results of functional connectivity analyses in the ventral and dorsal streams (according to the dual stream model of language processing; Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2012; Hagoort, 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009) showed that during the semantic task, word processing was generally associated with increased beta functional connectivity in the ventral stream compared to the dorsal stream. This result is in line with previous research showing that the ventral stream plays a role in semantic processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2012; Hagoort, 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), and that beta oscillations are indicative of a wide range of cognitive functions such as working memory (Siegel, Warden, & Miller, 2009) or decision making (Wong, Fabiszak, Novikov, Daw, & Pesaran, 2016). Moreover, beta oscillations have been shown to be involved in language processing (Carrus, Koelsch, & Bhattacharya, 2011), specifically in the extraction of global phonetic features (Bidelman, 2017), during word encoding and retrieval (Weiss & Rappelsberger, 2000), and lexical semantic memory access (Shahin, Picton, & Miller, 2009).

Turning to the impact of music training, professional musicians showed increased left-hemispheric theta connectivity in both the dorsal and ventral streams compared to adult non-musicians, suggesting that intensive music training eventually modifies the functional architecture of the brain leading to cross-domain advantages. Related to the "left-hemispheric theta connectivity", it is important to note that this finding does not mean that the left hemisphere was more strongly involved than the right one. Rather, musicians showed a more symmetrical recruitment of the two processing streams, while non-musicians were characterized by a right-sided asymmetry, as revealed by additional statistical analyses performed on asymmetry index data. This is important in highlighting the right-hemispheric contribution to language learning[26], and in view of previous results showing that individuals with a more symmetrical distribution of the arcuate fasciculus had an advantage when remembering previously learned words (Catani et al., 2007). However, this difference in lateralization between musicians and non-musicians was not predicted. It may be that the right-sided theta asymmetry found in non-musicians reflected processes similar to those at play in a recent study by Takashima et al. (2014) showing higher right-hemispheric connectivity between the STG and the MTG for successfully recognized words associated with colored pictures (as is the case in our semantic task).

Oscillations in the theta range have previously been proposed to support neural communication between distinct brain regions over long-range circuits (Polanía, Nitsche, Korman, Batsikadze, & Paulus, 2012; Ward, 2003) and mnemonic processes (Albouy, Weiss, Baillet, & Zatorre, 2017). Moreover, they possibly contribute to syllable coding at the word level (Bastiaansen, Linden, Keurs, Dijkstra, & Hagoort, 2005; Giraud & Poeppel, 2012), as well as to processing of syntactic and semantic violations (Bastiaansen, Magyari, & Hagoort, 2010). Our findings are in line with these findings, as well as with recent results by Elmer and colleagues (Elmer, Albrecht, Valizadeh, François, & Rodríguez-Fornells, 2018), in which the authors evaluated functional connectivity in the theta frequency range between the IP lobe and Broca's area (i.e., dorsal stream), as well as between the middle-posterior part of the middle temporal gyrus and Brodmann's area 21 (i.e., ventral stream) in a word learning task. Similar to our results, the authors evidenced a behavioral advantage of musicians compared to non-musicians. By contrast, musicians showed a left-hemispheric functional asymmetry in the dorsal stream in the theta frequency range (while we found a symmetrical recruitment of the two processing streams) and non-musicians showed a shift toward the right hemisphere (similar to our results). In line with previous literature showing that music training has an influence on the architecture of the left arcuate fasciculus (Oechslin, Imfeld, Loenneker, Meyer, & Jäncke, 2010) the authors concluded that music training had contributed to an increased alignment of neural oscillations

---

[26] Typically, language processing is described as strongly lateralized (Concha, Bianco, & Wilson, 2012; Hugdahl, 2000; Ocklenburg et al., 2018; Tzourio-Mazoyer, Perrone-Bertolotti, Jobard, Mazoyer, & Baciu, 2017), with a predominant role of the left hemisphere for the processing of grammar, vocabulary and meaning (Taylor & Taylor, 1990), and language production (Beaumont, 2008; Mazoyer et al., 2014).

in the dorsal, but not in the ventral stream. Interestingly, Lopez-Barroso and colleagues (López-Barroso et al., 2013) also showed that novel word learning was mediated by the dorsal but not the ventral stream. It is thus somewhat surprising that musicianship influenced both processing streams in our study. However, in both the Elmer et al (2018) and Lopez-Barroso et al (2013) studies, the task was to memorize pseudo-words that were presented in the form of auditory streams. Thus, increased functional connectivity in the dorsal stream possibly reflected neural coupling between phonetic representations stored in the inferior parietal lobe and the planning of articulation in Broca's region. Most importantly, however, no meaning was attached to these pseudo-words so that it is not surprising that functional connectivity was not increased in the ventral stream, typically more implicated in meaning processing (Bornkessel-Schlesewsky & Schlesewsky, 2013; Friederici, 2012; Hagoort, 2014; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009). In sum, our results, revealing between-group differences for both the dorsal and ventral stream, potentially point to both, better phonetic / articulatory processing and meaning processing in adult musicians compared to non-musicians.

The reader might wonder how to interpret the differences in different frequency bands (beta band for the increased ventral stream connectivity, and theta band for the between-group differences). This is a difficult issue, since the functional role of the different frequency bands is not yet well understood. Related to speech processing, authors have suggested that different units of the speech signal (e.g., envelope vs. fine structure; phoneme vs. sentential segments) possibly relate to different frequency ranges of neural oscillations, coordinating brain activity at multiple spatial and temporal scales across distant cortical regions (Giraud & Poeppel, 2012; Rufener, Oechslin, Wöstmann, Dellwo, & Meyer, 2016; Yellamsetty & Bidelman, 2018). However, since we investigated functional connectivity during the processing of monosyllabic words (i.e., very short stimuli), the different interpretations of functional significance between beta and theta oscillations (as described in the respective parts above) should be taken with caution.

Finally, a limitation of this work needs to be considered when interpreting these results: the functional connectivity analyses, based on EEG data, are estimations of brain networks (i.e., mathematical computations) at play during a certain task. In this respect, they are interesting in extending ERP analyses presented in Chapter 4. However, the accuracy with which sources of neural activity can be localized is affected by a number of factors, and future studies using methods with better spatial resolution (e.g., MRI) are needed to better define the role of the dorsal and ventral streams during novel word learning (when words have meaning), and how the two processing streams are influenced by music training. In addition, functional connectivity computes the alignment of neural oscillations in two spatially distinct, predefined regions, without information on the underlying white matter structures that eventually connect these two regions.

After having found compelling evidence that intensive music training facilitates the acquisition of novel words in children and adults, we aimed to better understand where this advantage possibly originated from. This is a complex issue, since first, novel word learning is a task including multiple functions, second, music training possibly influences multiple functions, and third, effects related to perceptive and different cognitive functions are difficult to disentangle. With this challenge in mind, we tried to adapt our experimental design in a way so that we “shifted weights” of the perceptive and cognitive functions at play. Specifically, we neutralized (i.e., reduced) the impact of auditory perception and attention, as well as learning-related processes on novel word learning to examine the impact, if any, of other cognitive functions. Using the same stimulus material as in our previous experiments (i.e., monosyllabic Thai words that are difficult to discriminate), we asked whether semantic integration would still be faster and/or better in musicians even without the advantage of better auditory perception and attention (Experiment 1). In a second experiment, we modified the stimulus material by including disyllabic Finnish words that were easy to discriminate for all participants. Moreover, two other experiments specifically aimed at better understanding the contribution of associative learning to novel word learning by using picture-word associations with (Experiment 4) or without semantic information (Experiment 3). Finally, to investigate long-term memory for novel words and to test the hypothesis that musicians would remember newly-learned words better than non-musicians, we invited participants to come back one month after the initial day of learning.

Results of behavioral and electrophysiological data in this new word learning design are presented in Chapter 6. They showed that under such conditions (i.e., neutralized perceptive and learning advantages) musicians did not outperform non-musicians anymore in the semantic task on the day of learning (i.e. Day 1). Similarly, musicians and non-musicians showed similar N400 effects on Day 1. These results were similar in all four experiments. By contrast, in all four experiments, musicians showed significantly lower error increases from Day 1 to Day 30 (i.e., retest one month later) compared to non-musicians. At the electrophysiological level, results were less consistent across the four experiments, with statistically significant between-group differences on the N400 effect on Day 30 only in response to the perceptively difficult Thai words (Experiment 1).

Several tentative (and not mutually exclusive) conclusions can be drawn from these results. First, behavioral and electrophysiological data on Day 1 failed to evidence that professional musicians are at an advantage in novel word learning compared to non-musicians when perceptual and/or learning-related facilitation is reduced. Therefore, one likely conclusion is that previously obtained between-group differences in semantic processing (presented in Chapter 4) reflected cascading effects from better auditory perception to better word learning in musicians (i.e., better semantic integration into pre-existing semantic networks). However, this conclusion should also be considered in light of the finding of better long-term memory for newly learned words in musicians compared to non-musicians on Day 30. This raises the possibility that musicians profit better from long-term consolidation. Thus, musicians may have remembered the newly learned words better on Day 30, although initial encoding and integration was similar for both groups (as shown by results on Day 1). While we typically describe the acquisition of novel words as a learning process, it is clear that it is also a memory process, since remembering words is as important as learning them (Wojcik, 2013). This means that what was remembered on Day 30, had been encoded and integrated into memory on Day 1, thereby putting the conclusion above (according to which better novel word learning in musicians is triggered by cascading effects from better auditory perception) into question. These two interpretations are difficult to disentangle with the present data, and further experiments investigating the influence of one night sleep on novel word leaning or sorting the ERPs to the novel words during the learning phase (encoding) on Day 1 as a function of whether or not they are subsequently remembered on Day 30 (i.e., examining differences at encoding based on subsequent memory, the so-called Dm effect first described by Paller, Kutas, & Mayes, 1987) may be helpful in shedding light on these issues.

In any event, the important point remains that musicians showed better long-term memory for novel words, a result that, to the best of our knowledge, has not been shown before. In addition, when novel words are difficult to discriminate (i.e., Thai words) between-group differences on semantic processing on Day 30 were also reflected by the N400 component, with statistically significant N400 effects in musicians only. By contrast, in the other three experiments (that all included disyllabic Finnish or French words) both musicians and non-musicians showed similar N400 effects. Again, these findings point to an advantage in musicians reflected at the behavioral and electrophysiological level only in difficult task conditions[27], as previously described elsewhere (Schön et al., 2004), and/or if better auditory perception and better cognitive functions can optimally interact. Taken together, these results extend to long-term memory previously described transfer effects from music training to working and short-term memory (George & Coch, 2011; Ho et al., 2003; Kraus, Strait, & Parbery-Clark, 2012; Schulze, Dowling, & Tillmann, 2012; for review see Talamini, Altoè, Carretti, & Grassi, 2017).

More generally, these results are linked to the fundamental question of whether perception and cognition are two sharply distinct systems, or whether there is cognitive penetration to perceptive processing. This question has attracted attention in the field of neuroscience, psychology and philosophy. On the one hand, results pointed to different functional areas for perception and cognition that process specific inputs, are specialized in different types of information processing, and form distinct modules (Barrett, 2005; Felleman & Van Essen, 1991; Zeki, 1978). On the other hand, other results are more in line with the idea that cognitive information influences perceptual processes, and at the same time, cognitive processes depend on perceptual information (Goldstone & Barsalou, 1998; Prinz, 2002). Nowadays, it seems that perception and cognition are considered to be interrelated (Montemayor & Haladjian, 2017; Tacca, 2011), and, based on such a framework, it is not surprising that music training positively influences both systems, that during novel word learning both systems are at play, and that the musician's advantage in novel word learning is best described by the working together of these two systems.

Our findings also contribute to the ongoing debate on domain-specific versus domain-general networks underlying language processing (for review see our recent article, appended at the end of this thesis; Besson et al., in press). Specifically, these results argue against language as an encapsulated system independent from other cognitive abilities, and support the idea of "many functions-one structure and many structures for one function" (Park & Friston, 2013). In

---

[27] That the experiment including the Thai words was the most difficult, was reflected by the fact that all participants needed clearly more repetitions to learn the Thai words at a similar level of performance compared to the Finnish and French words of the other three experiments. For more details on this issue I refer the reader to Chapter 6.

this respect, our findings are important in contrasting the traditional modular view of brain organization (Fodor, 1983), that has strongly influenced research in cognitive research for the last 35 years, and in evidencing that speech perception is processed in interaction with other cognitive abilities, such as memory. This view is also in line with the OPERA hypothesis proposed by Patel (Patel, 2008; 2014). Patel's hypothesis compares the auditory encoding demands made by musical training and by speech perception, and states that musical training benefits the neural encoding of speech based on the high precision of encoding of auditory attributes after music training (Patel, 2012). In addition, following the OPERA hypothesis (as well as our multi-dimensional interpretation), language and music are processed in interaction with other cognitive, emotional and motor abilities.

A final point deserves comments. Our results clearly point to the importance of considering the N400 effect not only over centro-parietal sites, but also over frontal scalp sites. We consistently observed inversed N400 effects (with larger amplitudes to expected than to unexpected words) over frontal scalp sites, not only in this experiment, but also in all the experiments presented in Chapter 4. Inversed frontal N400 effects possibly reflect ongoing learning processes during the test phase: the frontal distribution is in line with previous literature evidencing increased frontal N400 amplitudes during the learning of novel meaningful words (Cunillera et al., 2009; De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Lévi, 2007; François et al., 2016). Since only related trials (i.e., expected words) refer to correct picture-word associations, participants can only learn from these trials, resulting in larger N400 amplitudes than for unrelated trials (i.e., unexpected words). This interpretation is compatible with results showing that prefrontal and temporal brain regions are associated with the maintenance of novel information in working memory (Hagoort, 2014) and the acquisition of word meaning (Rodríguez-Fornells et al., 2009). Additionally, increased frontal N400s might also reflect control mechanisms recruited for the processing of expected and unexpected words (Lee & Federmeier, 2009). In sum, by taking into consideration not only the typical N400 effect over parietal sites, but also the additional inversed N400 effect over frontal scalp sites, future studies could possibly gain crucial information on the different processes involved during word encoding and integration.

Taken together, I think that this thesis provides multiple evidence that professional music training facilitates the learning of novel words of a foreign language, especially of tonal or quantitative languages in which fine-tuned auditory perception is crucial. Moreover, by investigating the underlying mechanisms, we showed that transfer effects in musicians strongly rely on auditory perception, and are additionally boosted by better long-term memory functions. The attractiveness of this research direction is possibly anchored in the worldwide growing number of multilingual speakers, as well as for patients with language impairments, considering rehabilitations via a musical route.

Future directions

Regarding possible future directions, the very same data obtained in the three experiments could be explored in different ways. Specifically, while EEG shows high temporal resolution, its spatial resolution is limited. To better understand brain networks underlying novel word learning, MRI and DTI approaches might be of interest. We must also acknowledge that the interpretation of our results is limited by the fact that we used cross-sectional designs, thereby not allowing to infer causality between music training and novel word learning. A goal for future research could be performing longitudinal studies, possibly allowing to track whether music training improves foreign novel word learning. In this respect it might be of interest starting longitudinal studies in children, and following them for several years until young adulthood. This kind of longitudinal studies, although complicated and expensive to implement, would allow to investigate in more detail the effect of age of acquisition of musical training during normal maturation, as well as the traces that early music training can leave in brain structures and functions for later language learning.

Conclusion

Research in the field of neuroscience has devoted a great effort in understanding the neural basis of language and music processing (Tillmann, 2012; Zatorre, 2013). Recent studies have revealed large overlaps in both brain structures and functions (Besson et al., 2017; Kraus & Chandrasekaran, 2010; Patel, 2008), as well as transfer effects from long-term music practice to speech perception (Bidelman & Krishnan, 2010; François & Schön, 2014). Thus, considering music to have a societal impact has become evident, as well as its implication into educational and rehabilitation programs (for reviews, see François, Grau-Sánchez, Duarte, & Rodriguez-Fornells, 2015; Tervaniemi et al., 2018). In this respect, learning novel words is a key part of language acquisition, and understanding the mechanisms behind this capacity can help us understand what makes some people better at learning foreign languages than others. Results presented in this thesis, extend the growing literature on transfer effects from music training to speech perception and cognitive functions. In addition, they are important in showing that professional music training boosts long-term memory functions, thereby possibly facilitating the memorization of novel words. Thus, our results point to memory as another piece of the fascinating puzzle of functions that subserve efficient novel word learning. They also open new perspectives related to beneficial applications of music training for pedagogical purposes, in order to improve language acquisition in children and adults, as well as for rehabilitation programs of pathological cases in which language functions are impaired. In an interconnected world, where the competence of several languages is becoming a prerequisite for social interactions and successful careers, they are of uttermost importance. I believe that the findings and suggestions provided in this work can contribute to break new ground in the field of language, as well as in investigations related to the hot topic of domain-specificity versus domain-generality of cognitive functions such as memory.

Chapter 8

REFERENCES

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex (New York, N.Y.: 1991)*, *21*(7), 1507–1518.

Agosta, F., Henry, R. G., Migliaccio, R., Neuhaus, J., Miller, B. L., Dronkers, N. F., … Gorno-Tempini, M. L. (2010). Language networks in semantic dementia. *Brain: A Journal of Neurology*, *133*(Pt 1), 286–299.

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective Entrainment of Theta Oscillations in the Dorsal Stream Causally Enhances Auditory Working Memory Performance. *Neuron*, *94*(1), 193–206.e5.

Anderson, R. C., & Freebody, P. (1981). Vocabulary Knowledge. In *Comprehension and teaching: Research reviews* (pp. 77–117). Newark, DE: International Reading Association: J.T. Guthrie.

Ashley, R., & Timmers, R. (2017). *The Routeledge Companion to Music Cognition*. Eds. R. Ashley and R. Timmers. Routledge / Taylor & Francis , New York.

Awad, M., Warren, J. E., Scott, S. K., Turkheimer, F. E., & Wise, R. J. S. (2007). A common system for the comprehension and production of narrative speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(43), 11455–11464.

Baer, L. H., Park, M. T. M., Bailey, J. A., Chakravarty, M. M., Li, K. Z. H., & Penhune, V. B. (2015). Regional cerebellar volumes are related to early musical training and finger tapping performance. *NeuroImage*, *109*, 130–139.

Bailey, J. A., & Penhune, V. B. (2010). Rhythm synchronization performance and auditory working memory in early- and late-trained musicians. *Experimental Brain Research*, *204*(1), 91–101.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41.

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*(2), 126–140.

Baltes, P. B. (1993). The aging mind: Potential and limits. *The Gerontologist*, *33*(5), 580–594.

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *The European Journal of Neuroscience*, *24*(6), 1832–1834.

Barrett, H. C. (2005). Enzymatic computation and cognitive modularity. *Mind and Language, 20(3)* 259 - 287.

Bastiaansen, M. C. M., Linden, M. van der, Keurs, M. ter, Dijkstra, T., & Hagoort, P. (2005). Theta Responses Are Involved in Lexical—Semantic Retrieval during Language Processing. *Journal of Cognitive Neuroscience*, *17*(3), 530–541.

Batterink, L., & Neville, H. (2011). Implicit and Explicit Mechanisms of Word Learning in a Narrative Context: An Event-related Potential Study. *Journal of Cognitive Neuroscience*, *23*(11), 3181–3196.

Baumgaertner, A., Weiller, C., & Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *NeuroImage*, *16*(3), 736–745.

Bayona, N. A., Bitensky, J., & Teasell, R. (2005). Plasticity and reorganization of the uninjured brain. *Topics in Stroke Rehabilitation*, *12*(3), 1–10.

Beaumont, J. G. (2008). Introduction to Neuropsychology (Second Edition). The Guilford Press.

Bengtsson, S. L., Nagy, Z., Skare, S., Forsman, L., Forssberg, H., & Ullén, F. (2005). Extensive piano practicing has regionally specific effects on white matter development. *Nature Neuroscience*, *8*(9), 1148–1150.

Bentin, S., McCarthy, G., & Wood, C.C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343-355.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*(3), 235–260.

Benz, S., Sellaro, R., Hommel, B., & Colzato, L. S. (2016). Music Makes the World Go Round: The Impact of Musical Training on Non-musical Cognitive Functions—A Review. *Frontiers in Psychology*, *6*.

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*, *19*(7), 1583–1596.

Besson, M., Dittinger, E., & Barbaroux, M. (2018). How music training influences language processing: Evidence against informational encapsulation. In *L'année psychologique/ Topics in Cognitive Psychology* (Vol. 118, pp. 273–288).

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*(1), 14–25.

Besson, M., Barbaroux, M., & Dittinger, E. (2017). Music in the brain: Music and language processing. In *The Routledge Companion to Music Cognition.* Eds. R. Ashley and R. Timmers, Routledge / Taylor & Francis , New York.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of Training between Music and Speech: Common Processing, Attention, and Memory. *Frontiers in Psychology*, *2*.

Besson, M., & Schön, D. (2003). Comparison between language and music. In *Neurosciences and music.* Eds. I. Peretz & R. Zatorre. Oxford: Oxford University Press.

Bidelman, G. M. (2016). Musicians have enhanced audiovisual multisensory binding: experience-dependent effects in the double-flash illusion. *Experimental Brain Research*, *234*(10), 3037–3047.

Bidelman, G. M. (2017). Amplified induced neural oscillatory activity predicts musicians' benefits in categorical speech perception. *Neuroscience*, *348*, 107–113.

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(3), 1240–1249.

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

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research*, *1355*, 112–125.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662–2673.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *17*(1), 353–362.

Bishop, L., & Goebl, W. (2017). Music and Movement: Musical Instruments and Performers. In *The Routledge Companion to Music Cognition*. Eds. R. Ashley and R. Timmers. Routledge / Taylor & Francis , New York.

Bokde, A. L., Tagamets, M. A., Friedman, R. B., & Horwitz, B. (2001). Functional interactions of the inferior frontal cortex during the processing of words and word-like stimuli. *Neuron*, *30*(2), 609–617.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain and Language*, *125*(1), 60–76.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is Enough: N400 Indexes Semantic Integration of Novel Word Meanings from a Single Exposure in Context. *Language Learning and Development*, *8*(3), 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*(2), 289–296.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., & Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *NeuroImage*, *25*(3), 958–968.

Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage*, *63*(3), 1334–1342.

Broca, P. (1863). L'aphasie de Broca. *Bulletin de La Société Français D'anthropologie*, *4*, 200–204.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain Research*, *1446*, 127–143.

Brown, R., & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In *Cognition and the Development of Language*. Ed. J. Hayes. Wiley.

Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY : Oxford University.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science (New York, N.Y.)*, *304*(5679), 1926–1929.

Carey, S., & Bartlett, E. (1978). Acquiring a Single New Word, *Papers and Reports on Child Language Development*, 15, 17–29.

Carrus, E., Koelsch, S., & Bhattacharya, J. (2011). Shadows of music–language interaction on low frequency brain oscillatory patterns. *Brain and Language*, *119*(1), 50–57.

Catani, M., Allin, M. P. G., Husain, M., Pugliese, L., Mesulam, M. M., Murray, R. M., & Jones, D. K. (2007). Symmetries in human brain language pathways correlate with verbal recall. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17163–17168.

Catani, M., & Mesulam, M. (2008). The arcuate fasciculus and the disconnection theme in language and aphasia: History and current state. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *44*(8), 953–961.

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

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). 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*, *24*(4), 956–967.

Chobert, J., & Besson, M. (2013). Musical Expertise and Second Language Learning. *Brain Sciences*, *3*(2), 923–940.

Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced Passive and Active Processing of Syllables in Musician Children. *Journal of Cognitive Neuroscience*, *23*(12), 3874–3887.

Concha, M. L., Bianco, I. H., & Wilson, S. W. (2012). Encoding asymmetry within neural circuits. *Nature Reviews Neuroscience*, *13*(12), 832–843.

Connolly, J. F., & Phillips, N. A. (1994). Event-Related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *Journal of Cognitive Neuroscience*, *6*(3), 256–266.

Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, *43*(1), 1–18.

Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., & Salmelin, R. (2004). Learning new names for new objects: Cortical effects as measured by magnetoencephalography. *Brain and Language*, *89*(3), 617–622.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*(2), 131–143.

Craik, F. I. M., & Jennings, J. M. (1992). Human memory. In *Handbook of aging and cognition,* Eds. F. I. M. Craik & T. A. Salthouse. Hillsdale, NJ:Erlbaum.

Cramer, S. C., Sur, M., Dobkin, B. H., O'Brien, C., Sanger, T. D., Trojanowski, J. Q., … Vinogradov, S. (2011). Harnessing neuroplasticity for clinical applications. *Brain: A Journal of Neurology*, *134*(Pt 6), 1591–1609.

Cunillera, T., Càmara, E., Toro, J. M., Marco-Pallares, J., Sebastián-Galles, N., Ortiz, H., … Rodríguez-Fornells, A. (2009). Time course and functional neuroanatomy of speech segmentation in adults. *NeuroImage*, *48*(3), 541–553.

Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effects in lexical decision and recognition memory. *Neuropsychologia*, *37*(7), 771–785.

Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, *21*(10), 1882–1892.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3773–3800.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Lévi, A.-C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PloS One*, *2*(11), e1175.

de Jong, P. F., Seveke, M.-J., & van Veen, M. (2000). Phonological Sensitivity and the Acquisition of New Words in Children. *Journal of Experimental Child Psychology*, *76*(4), 275–301.

DeFelipe, J. (2006). Brain plasticity and mental processes: Cajal again. *Nature Reviews. Neuroscience*, *7*(10), 811–817.

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.

DeWitt, I., & Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(8), E505-514.

D'Imperio, M. (2011). Prosodic representations. In *Handbook of Laboratory Phonology*. Eds. A. Cohn, C. Fougeron & M. Huffman. Oxford: Oxford University Press.

D'Imperio, M., Dittinger, E., & Besson, M. (2016). Prosodie et intonation: notions de base et données neuro-psycholinguistiques. In *Traité de neurolinguistique*. Eds S. Pinto and M. Sato. De Boeck-Supérieur.

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: on the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261.

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: evidence from the N400m. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3697–3709.

Donchin, E. (1981). Surprise!… Surprise? *Psychophysiology*, *18*(5), 493–513.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374.

Doughty, C. J., & Long, M. H. (2003). *The Handbook of Second Language Acquisition*. Blackwell Publishing Ltd, Oxford.

Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of Spoken Words. *Psychological Science*, *18*(1), 35–39.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased Cortical Representation of the Fingers of the Left Hand in String Players. *Science*, *270*(5234), 305–307.

Elmer, S., Dittinger, E., & Besson, M. (in press). One step beyond: musical expertise and word learning. In *The Oxford Handbook of Voice Perception*. Eds. S. Frühholz and P. Belin. Oxford: Oxford University Press.

Elmer, S., & Jäncke, L. (2018). Relationships between music training, speech processing, and word learning: a network perspective: Music training, speech processing, and word learning. *Annals of the New York Academy of Sciences*, *1423*(1), 10–18.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and Behavioral Correlates of Phonetic and Temporal Categorization in Musically Trained and Untrained Subjects. *Cerebral Cortex*, *22*(3), 650–658.

Elmer, S., Albrecht, J., Valizadeh, S. A., François, C., & Rodríguez-Fornells, A. (2018). Theta Coherence Asymmetry in the Dorsal Stream of Musicians Facilitates Word Learning. *Scientific Reports*, *8*, 4565.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews. Neuroscience*, *2*(10), 704–716.

Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., … Kramer, A. F. (2007). Training-induced functional activation changes in dual-task processing: an FMRI study. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(1), 192–204.

Ericsson, K. A., Krampe, R. T., & Tesch-romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 363–406.

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Frontiers in Neuroscience*, *9*.

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience*, *12*(2), 105–118.

Felleman, D. J., & Van Essen, V. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, *1*(1), 1–47.

Ferstl, E. C., & von Cramon, D. Y. (2001). The role of coherence and cohesion in text comprehension: an event-related fMRI study. *Brain Research. Cognitive Brain Research*, *11*(3), 325–340.

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP “old/new” effects: memory strength and decisional factor(s). *Neuropsychologia*, *40*(13), 2288–2304.

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitzroy, A. B., & Sanders, L. D. (2012). Musical expertise modulates early processing of syntactic violations in language. *Frontiers in Psychology*, *3*, 603.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, *45*(1), 152–170.

Ford, J. M., Roth, W. T., & Kopell, B. S. (1976). Auditory evoked potentials to unpredictable shifts in pitch. *Psychophysiology*, *13*(1), 32–39.

François, C., Cunillera, T., Garcia, E., Laine, M., & Rodriguez-Fornells, A. (2017). Neurophysiological evidence for the interplay of speech segmentation and word-referent mapping during novel word learning. *Neuropsychologia*, *98*, 56–67.

François, C., Grau-Sánchez, J., Duarte, E., & Rodriguez-Fornells, A. (2015). Musical training as an alternative and effective method for neuro-education and neuro-rehabilitation. *Frontiers in Psychology*, *6*.

François, C., Ripollés, P., Bosch, L., Garcia-Alix, A., Muchart, J., Sierpowska, J., … Rodriguez-Fornells, A. (2016). Language learning and brain reorganization in a 3.5-year-old child with left perinatal stroke revealed using structural and functional connectivity. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *77*, 95–118.

François, C., & Schön, D. (2014). Neural sensitivity to statistical regularities as a fundamental biological process that underlies auditory learning: The role of musical practice. *Hearing Research*, *308*, 122–128.

Freedman, R., Adler, L. E., Waldo, M. C., Pachtman, E., & Franks, R. D. (1983). Neurophysiological evidence for a defect in inhibitory pathways in schizophrenia: comparison of medicated and drug-free patients. *Biological Psychiatry*, *18*(5), 537–551.

Freitas, C., Farzan, F., & Pascual-Leone, A. (2013). Assessing brain plasticity across the lifespan with transcranial magnetic stimulation: why, how, and what is the ultimate goal? *Frontiers in Neuroscience*, *7*, 42.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews*, *91*(4), 1357–1392.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*(5), 262–268.

Friederici, A. D. (2015). White-matter pathways for speech and language processing. *Handbook of Clinical Neurology*, *129*, 177–186.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2458–2463.

Friedman, D., & Johnson, R. (2000). Event-related potential (ERP) studies of memory encoding and retrieval: a selective review. *Microscopy Research and Technique*, *51*(1), 6–28.

Friedrich, M., & Friederici, A. D. (2008). Neurophysiological correlates of online word learning in 14-month-old infants: *NeuroReport*, *19*(18), 1757–1761.

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications*, *6*, 6004.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cognitive Sciences*, *9*(10), 474–480.

Fuchs, M., Kastner, J., Wagner, M., Hawes, S., & Ebersole, J. S. (2002). A standardized boundary element method volume conductor model. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *113*(5), 702–712.

Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. *NeuroImage*, *19*(4), 1417–1426.

Gandour, J. T., & Krishnan, A. (2016). Processing tone languages. In *Neurobiology of language*. Eds. G.S. Hickok & S.L. Small. San Diego, CA: Academic Press.

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), 9240–9245.

Gathercole, S. E., Willis, C., Braddeley, A. D., & Emslie, H. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology*, *28*, 887–898.

Gathercole, S. E, & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*(2), 200–213.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.

Gerstein, G. L., & Perkel, D. H. (1969). Simultaneously recorded trains of action potentials: analysis and functional interpretation. *Science (New York, N.Y.)*, *164*(3881), 828–830.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Goldstein, R., & Rodman, L. B. (1967). Early Components of Averaged Evoked Responses to Rapidly Repeated Auditory Stimuli. *Journal of Speech, Language, and Hearing Research*, *10*(4), 697–705.

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*(2), 231–262.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., … Vanrumste, B. (2008). Review on solving the inverse problem in EEG source analysis. *Journal of Neuroengineering and Rehabilitation*, *5*, 25.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An Electrophysiological Study of Semantic Processing in Young and Middle-Aged Academics. *Psychophysiology*, *29*(1), 38–54.

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

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439–483.

Hagoort, P., Indefrey, P., Brown, C., Herzog, H., Steinmetz, H., & Seitz, R. J. (1999). The neural circuitry involved in the reading of German words and pseudowords: A PET study. *Journal of Cognitive Neuroscience*, *11*(4), 383–398.

Hall, J. W. (1992). *Handbook of auditory evoked responses.* Eds. Allyn & Bacon. Massachusetts.

Harvey, D. Y., Wei, T., Ellmore, T. M., Hamilton, A. C., & Schnur, T. T. (2013). Neuropsychological evidence for the functional role of the uncinate fasciculus in semantic control. *Neuropsychologia*, *51*(5), 789–801.

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage, 62*(2), 852–855.

Hensch, T. K. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews. Neuroscience*, *6*(11), 877–888.

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397.

Herholz, S. C., & Zatorre, R. J. (2012). Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, *76*(3), 486–502.

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, *4*(4), 131–138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*(1–2), 67–99.

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

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science (New York, N.Y.)*, *182*(4108), 177–180.

Hillyard, S. A., Woldorff, M., Mangun, G. R., & Hansen, J. C. (1987). Mechanisms of early selective attention in auditory and visual modalities. *Electroencephalography and Clinical Neurophysiology. Supplement*, *39*, 317–324.

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439–450.

Hoffman, J. E. (1990). Event-related potentials and automatic and controlled processes. In *Event Related Brain Potentials*. Eds. Rohrbaugh JW, Parasuraman R, Johnson R Jr. New York: Oxford University Press.

Howard, D., Patterson, K., Wise, R., Brown, W. D., Friston, K., Weiller, C., & Frackowiak, R. (1992). The cortical localization of the lexicons. Positron emission tomography evidence. *Brain: A Journal of Neurology*, *115*, 1769–1782.

Hugdahl, K. (2000). Lateralization of cognitive processes in the brain. *Acta Psychologica*, *105*(2), 211–235.

Hultén, A., Laaksonen, H., Vihla, M., Laine, M., & Salmelin, R. (2010). Modulation of Brain Activity after Learning Predicts Long-Term Memory for Words. *Journal of Neuroscience*, *30*(45), 15160–15164.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224.

Huttenlocher, P. R. (2003). Basic neuroscience research has important implications for child development. *Nature Neuroscience*, *6*(6), 541.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., & Schlaug, G. (2009). Musical Training Shapes Structural Brain Development. *Journal of Neuroscience*, *29*(10), 3019–3025.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks: The Official Journal of the International Neural Network Society*, *13*(4–5), 411–430.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144.

Janata, P., Birk, J. L., Van Horn, J. D., Leman, M., Tillmann, B., & Bharucha, J. J. (2002). The cortical topography of tonal structures underlying Western music. *Science (New York, N.Y.)*, *298*(5601), 2167–2170.

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience*, *6*(7), 682–687.

Jäncke, L. (2009). Music drives brain plasticity. *F1000 Biology Reports*, *1*(78), 10–3410.

Jäncke, L., & Langer, N. (2011). A strong parietal hub in the small-world network of coloured-hearing synaesthetes during resting state EEG. *Journal of Neuropsychology*, *5*(2), 178–202.

Jantzen, M. G., Large, E. W., & Magne, C. (2016). Editorial: Overlap of Neural Systems for Processing Language and Music. *Frontiers in Psychology*, *7*, 876.

Jasper, H. (1958). The ten twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, *47*(2), 735–744.

Jewett, D. L., Romano, M. N., & Williston, J. S. (1970). Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science (New York, N.Y.)*, *167*(3924), 1517–1518.

Jewett, D. L., & Williston, J. S. (1971). Auditory-evoked far fields averaged from the scalp of humans. *Brain: A Journal of Neurology*, *94*(4), 681–696.

Kausler, D. H. (1991). *Experimental psychology, cognition, and human aging.* New York: Springer-Verlag.

Key, A. P. F., Molfese, D. L., & Ratajczak, E. D. (2006). ERP indicators of learning in adults. *Developmental Neuropsychology*, *29*(2), 379–395.

Kliesch, M., Giroud, N., Pfenninger, S. E., & Meyer, M. (2017). Research on Second Language Acquisition in Old Adulthood: What We Have and What We Need. In *Third Age Learners of Foreign Languages*. Ed. Gabrys-Barker D. Bristol: Multilingual Matters.

Koelsch, S. (2010). Towards a neural basis of music-evoked emotions. *Trends in Cognitive Sciences*, *14*(3), 131–137.

Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nature Reviews Neuroscience*, *15*(3), 170–180.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage*, *25*(4), 1068–1076.

Koelsch, S., Gunter, T. C., v Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical “language-network” serves the processing of music. *NeuroImage*, *17*(2), 956–966.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307.

Kolb, B., & Muhammad, A. (2014). Harnessing the power of neuroplasticity for intervention. *Frontiers in Human Neuroscience*, *8*, 377.

Kraus, N. (2011). Listening in on the listening brain. https://doi.org/10.1063/1.3603917

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

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory: Kraus et al. *Annals of the New York Academy of Sciences*, *1252*(1), 100–107.

Kraus, N., & White-Schwoch, T. (2015). Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. *Trends in Cognitive Sciences*, *19*(11), 642–654.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing Research*, *166*(1–2), 192–201.

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

Kühnis, J., Elmer, S., Meyer, M., & Jäncke, L. (2013). The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: An EEG study. *Neuropsychologia*, *51*(8), 1608–1618.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. *Brain and Language*, *100*(3), 223–237.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Brain Research. Cognitive Brain Research*, *17*(1), 117–129.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science (New York, N.Y.)*, *207*(4427), 203–205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539–550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting Mental Chronometry: The P300 as a Measure of Stimulus Evaluation Time. *Science*, *197*(4305), 792–795.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, *69*(3), 218–233.

Kutas, M. & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *108*(5), 456–471.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., … Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5675–5679.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*(3), 120–131.

Lee, C., & Federmeier, K. D. (2009). Wave-ering: An ERP study of syntactic and semantic context effects on ambiguity resolution for noun/verb homographs. *Journal of Memory and Language*, *61*(4), 538–555.

Lehmann, D., Faber, P. L., Gianotti, L. R. R., Kochi, K., & Pascual-Marqui, R. D. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *Journal of Physiology, Paris*, *99*(1), 29–36.

Lieberman, P. (2016). The evolution of language and thought. *Journal of Anthropological Sciences = Rivista Di Antropologia: JASS*, *94*, 127–146.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion (Washington, D.C.)*, *11*(5), 1021–1031.

Limb, C. J., & Braun, A. R. (2008). Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. *PLOS ONE*, *3*(2), e1679.

Lindenberg, R., & Scheef, L. (2007). Supramodal language comprehension: role of the left temporal lobe for listening and reading. *Neuropsychologia*, *45*(10), 2407–2415.

Linnavalli, T., Putkinen, V., Lipsanen, J., Huotilainen, M. & Tervaniemi, M. (2018). Music playschool enhances children's linguistic skills. *Scientific Reports*, 2018; 8 (1)

Long, L. L., & Shaw, R. J. (2000). Adult age differences in vocabulary acquisition, *26*, 651–664.

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*, *110*(32), 13168–13173.

López-Barroso, D., Ripollés, P., Marco-Pallarés, J., Mohammadi, B., Münte, T. F., Bachoud-Lévi, A.-C., … de Diego-Balaguer, R. (2015). Multiple brain networks underpinning word learning from fluent speech revealed by independent component analysis. *NeuroImage*, *110*, 182–193.

Loui, P., & Przysinda, E. (2017). Music in the Brain: Areas and Networks. In *The Routledge Companion to Music Cognition*. Eds. R. Ashley and R. Timmers. Routledge / Taylor & Francis , New York.

Loveless, N. E., & Brunia, C. H. (1990). Effects of rise-time on late components of the auditory evoked potential. *Journal of Psychophysiology*, 4, 369-380.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience*, *4*(5), 540–545.

Magne, C., Besson, M., & Robert, S. (2014). Context influences the processing of verb transitivity in French sentences: more evidence for semantic−syntax interactions. *Language and Cognition*, *6*(02), 181–216.

Magne, C., Schön, D., & Besson, M. (2003). Prosodic and melodic processing in adults and children. Behavioral and electrophysiologic approaches. *Annals of the New York Academy of Sciences*, *999*, 461–476.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, *46*(1), 53–85.

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, 2701–2715.

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), 1453–1463.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: structure and processes. *Current Opinion in Neurobiology*, *11*(2), 194–201.

Mazoyer, B., Zago, L., Jobard, G., Crivello, F., Joliot, M., Perchey, G., … Tzourio-Mazoyer, N. (2014). Gaussian Mixture Modeling of Hemispheric Lateralization for Language in a Large Sample of Healthy Individuals Balanced for Handedness. *PLOS ONE*, *9*(6), e101165.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., … Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *356*(1412), 1293–1322.

McCandliss, B. D., Posner, M. I., & Givón, T. (1997). Brain Plasticity in Learning Visual Words. *Cognitive Psychology*, *33*(1), 88–110.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704.

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: neural mechanisms underlying speech parsing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *26*(29), 7629–7639.

Mestres-Missé, A., Càmara, E., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2008). Functional neuroanatomy of meaning acquisition from context. *Journal of Cognitive Neuroscience*, *20*(12), 2153–2166.

Mestres-Missé, A., Münte, T. F., & Rodriguez-Fornells, A. (2014). Mapping concrete and abstract meanings to new words using verbal contexts. *Second Language Research*, *30*(2), 191–223.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the Brain during Meaning Acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Michel, C. M., Koenig, T., Brandeis, D., Gianotti, L. R. R., & Wackermann, J. (2009). *Electrical neuroimaging*. Cambridge University Press.

Montemayor, C., & Haladjian, H. H. (2017). Perception and Cognition Are Largely Independent, but Still Affect Each Other in Systematic Ways: Arguments from Evolution and the Consciousness-Attention Dissociation. *Frontiers in Psychology*, *8*.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-Term Music Training Enhances Verbal Intelligence and Executive Function. *Psychological Science*, *22*(11), 1425–1433.

Moreno, S., & Bidelman, G. M. (2014). Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Research*, *308*, 84–97.

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*, 19, 712–723.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nature Reviews Neuroscience*, *3*(6), 473–478.

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), 15894–15898.

Näätänen, R., Gaillard, A., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, *42*(4), 313–329.

Näätänen, Risto, & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review Component Structure. *Psychophysiology*, 24 (4), 375-425.

Nagy, W. E., & Anderson, R. C. (1984). How Many Words Are There in Printed School English? *Reading Research Quarterly*, *19*(3), 304–330.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning Word Meanings From Context During Normal Reading. *American Educational Research Journal*, *24*(2), 237–270.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning Words from Context. *Reading Research Quarterly*, *20*(2), 233–253.

Nan, Y., Liu, L., Geiser, E., Shu, H., Gong, C. C., Dong, Q., Gabrieli, J., & Desimone R. (2018). Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proceedings of the National Academy of Sciences of the United States of America*, *115* (28), E6630-E6639.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian Reconstruction of Natural Images from Human Brain Activity. *Neuron*, *63*(6), 902–915.

Nessler, D., Mecklinger, A., & Penney, T. B. (2001). Event related brain potentials and illusory memories: the effects of differential encoding. *Brain Research. Cognitive Brain Research*, *10*(3), 283–301.

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., & Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, *115*(10), 2292–2307.

Norman, K.A., Polyn, S.M., Detre, G.J., & Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*(9), 424-30.

Ocklenburg, S., Friedrich, P., Fraenz, C., Schlüter, C., Beste, C., Güntürkün, O., & Genç, E. (2018). Neurite architecture of the planum temporale predicts neurophysiological processing of auditory speech. *Science Advances*, *4*(7), eaar6830.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jäncke, L. (2010). The Plasticity of the Superior Longitudinal Fasciculus as a Function of Musical Expertise: A Diffusion Tensor Imaging Study. *Frontiers in Human Neuroscience*, *3*.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806.

Paller, K. A., Kutas, M., & McIsaac, H. K. (1995). Monitoring Conscious Recollection via the Electrical Activity of the Brain. *Psychological Science*, *6*(2), 107–111.

Paller, K. A., Kutas, M. & Mayes, A. E. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67*, 360–371.

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive Control in Auditory Working Memory Is Enhanced in Musicians. *PLoS ONE*, *5*(6), e11120.

Pantev, C., & Herholz, S. C. (2011). Plasticity of the human auditory cortex related to musical training. *Neuroscience and Biobehavioral Reviews*, *35*(10), 2140–2154.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical expertise induces audiovisual integration of abstract congruency rules. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(50), 18196–18203.

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Frontiers in Aging Neuroscience*, *4*.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661.

Parbery-Clark, A., Strait, D. L., Anderson, S., Hittner, E., & Kraus, N. (2011). Musical Experience and the Aging Auditory System: Implications for Cognitive Abilities and Hearing Speech in Noise. *PLOS ONE*, *6*(5), e18082.

Park, H.-J., & Friston, K. (2013). Structural and Functional Brain Networks: From Connections to Cognition. *Science*, *342*(6158), 1238411.

Pascual-Marqui, R. D., & Lehmann, D. (1993). Comparison of topographic maps and the reference electrode: comments on two papers by Desmedt and collaborators. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *88*(6), 530–531.

Pascual-Marqui, R. D., Lehmann, D., Koukkou, M., Kochi, K., Anderer, P., Saletu, B., … Kinoshita, T. (2011). Assessing interactions in the brain with exact low-resolution electromagnetic tomography. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *369*(1952), 3768–3784.

Patel, A. D. (2008). *Music, Language, and the Brain*. Oxford: Oxford University Press.

Patel, A. D. (2012). The OPERA hypothesis: assumptions and clarifications: Patel. *Annals of the New York Academy of Sciences*, *1252*(1), 124–128.

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108.

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the event-related potential. *International Journal of Medical Sciences*, *2*(4), 147.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*(12), 976–987.

Penhune, V. B. (2011). Sensitive periods in human development: evidence from musical training. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *47*(9), 1126–1137.

Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(6), 1281–1292.

Picton, T. W., Goodman, W. S., & Bryce, D. P. (1970). Amplitude of evoked responses to tones of high intensity. *Acta Oto-Laryngologica*, *70*(2), 77–82.

Polanía, R., Nitsche, M. A., Korman, C., Batsikadze, G., & Paulus, W. (2012). The importance of timing in segregated theta phase-coupling for cognitive performance. *Current Biology: CB*, *22*(14), 1314–1318.

Polich, J. (1989). Habituation of P300 from auditory stimuli. *Psychobiology*, *17*(1), 19–28.

Polich, J. (2003). Overview of P3a and P3b. In *Detection of Change: Event-Related Potential and fMRI Findings*. Ed. J. Polich. Kluwer Academic Press, Norwell, MA.

Polich, J. (2007). Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *118*(10), 2128–2148.

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), 816–847.

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis.* Cambridge, MA: MIT Press.

Ramachandra, V., Hewitt, L. E., & Brackenbury, T. (2011). The Relationship Between Phonological Memory, Phonological Sensitivity, and Incidental Word Learning. *Journal of Psycholinguistic Research*, *40*(2), 93–109.

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*(6447), 611.

Ritter, W., Simson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, *20*(2), 168–179.

Ritter, W., Simson, R., Vaughan, H. G., & Friedman, D. (1979). A brain event related to the making of a sensory discrimination. *Science (New York, N.Y.)*, *203*(4387), 1358–1361.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3711–3735.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional Anatomy of Language and Music Perception: Temporal and Structural Factors Investigated Using Functional Magnetic Resonance Imaging. *The Journal of Neuroscience*, *31*(10), 3843–3852.

Rufener, K. S., Oechslin, M., S., Wöstmann, M., Dellwo, V., & Meyer, M. (2016). Age-Related Neural Oscillation Patterns During the Processing of Temporally Manipulated Speech. *Brain Topography, 29*, 440–458.

Rugg, M. D. (1995). Event-related potential studies of human memory. In *The cognitive neurosciences*. Ed. M. S. Gazzaniga. Cambridge, MA: The MIT Press.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257.

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science (New York, N.Y.)*, *274*(5294), 1926–1928.

Sala, G., & Gobet, F. (2017). When the music's over. Does music skill transfer to children's and young adolescents' cognitive and academic skills? A meta-analysis. *Educational Research Review*, *20*, 55–67.

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39.

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), 18035–18040.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science*, *15*(8), 511–514.

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701.

Schmithorst, V. J., & Wilke, M. (2002). Differences in white matter architecture between musicians and non-musicians: a diffusion tensor imaging study. *Neuroscience Letters*, *321*(1–2), 57–60.

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), 688–694.

Schneider, P., Sluming, V., Roberts, N., Bleeck, S., & Rupp, A. (2005). Structural, Functional, and Perceptual Differences in Heschl's Gyrus and Musical Instrument Preference. *Annals of the New York Academy of Sciences*, *1060*(1), 387–394.

Schön, D., & François, C. (2011). Musical expertise and statistical learning of musical and linguistic structures. *Frontiers in Psychology*, *2*, 167.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*(3), 341–349.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception: An Interdisciplinary Journal*, *29*(3), 255–267.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*(5), 771–783.

Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial Word Knowledge and Vocabulary Growth during Reading Comprehension. *Journal of Literacy Research*, *29*(4), 531–553.

Service, E., & Craik, F. I. M. (1993). Differences Between Young and Older Adults in Learning A Foreign Vocabulary. *Journal of Memory and Language*, *32*(5), 608–623.

Service, Elisabet, Maury, S., & Luotoniemi, E. (2007). Individual differences in phonological learning and verbal STM span. *Memory & Cognition*, *35*(5), 1122–1135.

Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, *70*(3), 259–266.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, *16*(16), 1781–1785.

Siegel, M., Warden, M. R., & Miller, E. K. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences*, *106*(50), 21341–21346.

Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., & Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage*, *17*(3), 1613–1622.

Sluming, V., Brooks, J., Howard, M., Downes, J. J., & Roberts, N. (2007). Broca's area supports enhanced visuospatial cognition in orchestral musicians. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(14), 3799–3806.

Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, *18*(1), 1–13.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*(4), 387–401.

Stam, C. J., & van Straaten, E. C. W. (2012). Go with the flow: use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. *NeuroImage*, *62*(3), 1415–1428.

Steele, C. J., Bailey, J. A., Zatorre, R. J., & Penhune, V. B. (2013). Early musical training and white-matter plasticity in the corpus callosum: evidence for a sensitive period. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(3), 1282–1290.

Sternberg, R. J., & Powell, J. S. (1983). Comprehending Verbal Comprehension. *American Psychologist*, *38*(8), 878–893.

Strait, D. L., Kraus, N., Skoe, E., & Ashley, R. (2009). Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *The European Journal of Neuroscience*, *29*(3), 661–668.

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188.

Swaab, T. Y., Brown, C., & Hagoort, P. (1998). Understanding ambiguous words in sentence contexts: electrophysiological evidence for delayed contextual selection in Broca's aphasia. *Neuropsychologia*, *36*(8), 737–761.

Swaminathan, S., & Schellenberg, E. G. (2017). Musical competence and phoneme perception in a foreign language. *Psychonomic Bulletin & Review*, *24*(6), 1929–1934.

Swanborn, M. S. L., & de Glopper, K. (2002). Impact of Reading Purpose on Incidental Word Learning From Context. *Language Learning*, *52*(1), 95–117.

Tacca, M. C. (2011). Commonalities between Perception and Cognition. *Frontiers in Psychology*, *2*.

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language*, *167*, 44–60.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS One*, *12*(10), e0186773.

Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(39), 15376–15381.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology*, *5*.

Taylor, I., & Taylor, M. M. (1990). *Psycholinguistics: Learning and using Language*. Pearson.

Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, *10*, 21.

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. *European Journal of Neuroscience*, 30, 1636–1642.

Tervaniemi, M., Tao, S., & Huotilainen, M. (2018). Promises of Music in Education? *Frontiers in Education, 3*(74).

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Developmental Neuropsychology*, *37*(6), 476–496.

Thiebaut de Schotten, M., Cohen, L., Amemiya, E., Braga, L. W., & Dehaene, S. (2014). Learning to read improves the structure of the arcuate fasciculus. *Cerebral Cortex (New York, N.Y.: 1991)*, *24*(4), 989–995.

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science*, *4*(4), 568–584.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Annals of the New York Academy of Sciences*, *999*, 209–211.

Torkildsen, J. von K., Friis Hansen, H., Svangstu, J. M., Smith, L., Simonsen, H. G., Moen, I., & Lindgren, M. (2009). Brain dynamics of word familiarization in 20-month-olds: Effects of productive vocabulary size. *Brain and Language*, *108*(2), 73–88.

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language*, *162*, 60–71.

Tzourio-Mazoyer, N., Perrone-Bertolotti, M., Jobard, G., Mazoyer, B., & Baciu, M. (2017). Multi-factorial modulation of hemispheric specialization and plasticity for language in healthy and pathological conditions: A review. *Cortex*, *86*, 314–339.

Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nature Reviews. Neuroscience*, *2*(10), 717–726.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*(7), 967–985.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the Word Repetition Effect with Event-Related Potentials. *Journal of Cognitive Neuroscience*, *3*(2), 131–150.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, *30*(4), 1414–1432.

Volgushev, M., Chistiakova, M., & Singer, W. (1998). Modification of discharge patterns of neocortical neurons by induced oscillations of the membrane potential. *Neuroscience*, *83*(1), 15–25.

Voss, J. L., & Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology*, *48*(4), 532–546.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*(2), 329–338.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., … Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*(5380), 1188–1191.

Wan, C. Y., & Schlaug, G. (2010). Music Making as a Tool for Promoting Brain Plasticity across the Life Span. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *16*(5), 566–577.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559.

Weiss, S., & Rappelsberger, P. (2000). Long-range EEG synchronization during word encoding correlates with successful memory performance. *Brain Research. Cognitive Brain Research*, *9*(3), 299–312.

Wernicke, C. (1881). *Textbook of Brain Disorders: for doctors and students*.

Wilding, E. L., & Rugg, M. D. (1997). Event-related potentials and the recognition memory exclusion task. *Neuropsychologia*, *35*(2), 119–128.

Wlotko, E. W., Lee, C.-L., & Federmeier, K. D. (2010). Language of the Aging Brain: Event-Related Potential Studies of Comprehension in Older Adults. *Language and Linguistics Compass*, *4*(8), 623–638.

Wojcik, E. H. (2013). Remembering New Words: Integrating Early Memory Development into Word Learning. *Frontiers in Psychology*, *4*.

Wong, P. C. M., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(04).

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), 420–422.

Wong, Y. T., Fabiszak, M. M., Novikov, Y., Daw, N. D., & Pesaran, B. (2016). Coherent neuronal ensembles are rapidly recruited when making a look-reach decision. *Nature Neuroscience*, *19*(2), 327–334.

Yellamsetty, A., & Bidelman, G. M. (2018). Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hearing Research*, *361*, 92–102.

Zatorre, R. J. (2013). Predispositions and Plasticity in Music and Speech Learning: Neural Correlates and Implications. *Science*, *342*(6158), 585–589.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews. Neuroscience*, *8*(7), 547–558.

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), 528–536.

Zeki, S. M. (1978). Functional specialisation in the visual cortex of the rhesus monkey. *Nature*, *274*(5670), 423–428.

Zendel, B. R., & Alain, C. (2009). Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience*, *21*(8), 1488–1498.

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*(2), 410–417.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians. *PLOS ONE*, *9*(6), e99868.

ANNEX

----

1. Experimental material: examples of the learning lists used in the novel word learning tasks (“APPSENS” – Chapter 4; “MUMEMO” – Chapter 6), including the novel words, the line drawings and semantically related pictures

2. Contributions to special volumes:

D’Imperio, M., Dittinger, E., & Besson, M. (2016). “Prosodie et intonation : notions de base et données neuro-psycholinguistiques”. In “*Traité de neurolinguistique*”, Eds S. Pinto and M. Sato, De Boeck-Supérieur, p. 133-145.

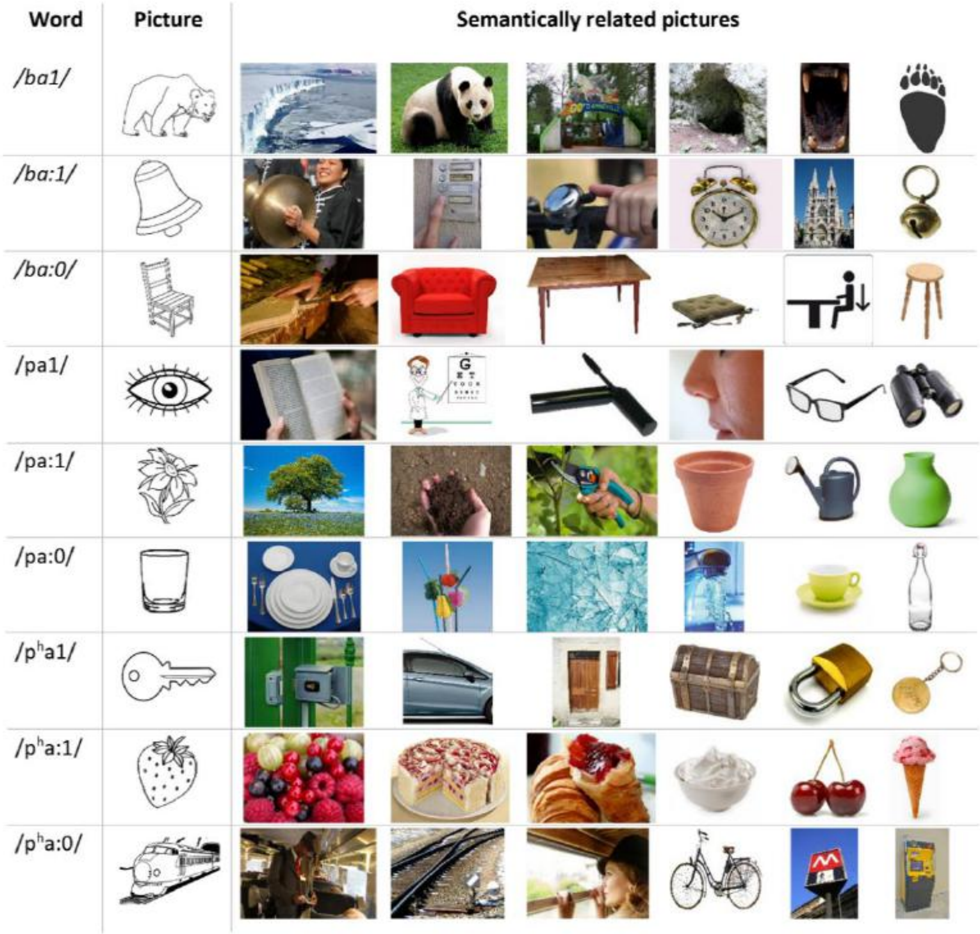
Besson, M., Barbaroux, M., & Dittinger, E. (2017). “Music in the brain: Music and language processing”. In “*The Routledge Companion to Music Cognition*”, Eds R. Ashley and R. Timmers, Routledge/ Taylor & Francis group.

Elmer, S., Dittinger, E., & Besson, M. (in press). One step beyond: musical expertise and word learning. In “*The handbook of voice perception*”, Eds S. Frühholz and P. Belin, Oxford University Press.

Besson, M., Dittinger, E. & Barbaroux, M. (in press). How music training influences language processing: evidence against informationnal encapsulation. In “*Revue année psychologique*”, Eds P. Colé and C. Pattamadilok, Humensis.

1. Experimental Material

a. APPSENS (Chapter 4)

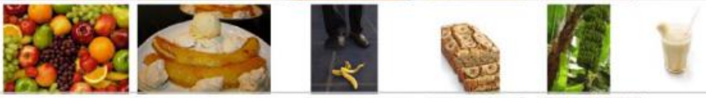


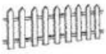
b. MUMEMO (Chapter 6)

Experiment 1:

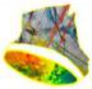
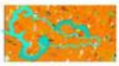
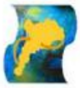
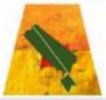
Word	Picture	Semantically related pictures
/ba1/		
/ba0/		
/ba:1/		
/ba:0/		
/pa1/		
/pa0/		
/pa:1/		
/pa:0/		
/pʰa1/		
/pʰa0/		
/pʰa:1/		
/pʰa:0/		

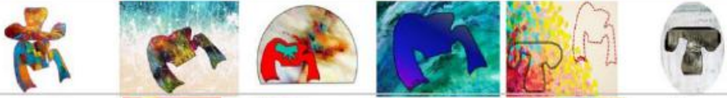
Experiment 2:

Word	Picture	Semantically related pictures
Tyyny		
Lintu		
Parsa		
Kirves		
Jänis		
Huntu		
Pullo		
Ruoho		
Sanka		
Farkut		
Ryhmä		
Hanhi		

Word	Picture	Semantically related pictures
Norsu		
Teltta		
Marsu		
Hylje		
Lehmä		
Laskin		
Molli		
Ketju		
Jalka		
Myrsky		
Sänky		
Huilu		

Experiment 3:

Word	Picture	Semantically related pictures					
Nuoli							
Lettu							
Maila							
Sakset							
Tähti							
Harja							
Kangas							
Rumpu							
Juusto							
Noppa							
Lisko							
Sähkö							

Word	Picture	Semantically related pictures
Sydän		
Varvas		
Mökki		
Patja		
Hillo		
Kongi		
Huulet		
Pyssy		
Mylly		
Raastin		
Järvi		
Huivi		

Experiment 4:

Word	**Picture**	**Semantically related pictures**
Fléchette		
Valise		
Echelle		
Bougie		
Balance		
Collier		
Girafe		
Orteil		
Gilet		
Abeille		
Passoire		
Mouton		

Word	Picture	Semantically related pictures
Soleil		
Fenêtre		
Râteau		
Cactus		
Pinceau		
Tortue		
Maison		
Asperge		
Camion		
Poupée		
Croissant		
Bocal		

2. Contributions to special volumes

CHAPITRE

10

Prosodie et intonation : notions de base et données neuro-psycholinguistiques

Mariapaola D'Imperio, Éva Dittinger, Mireille Besson

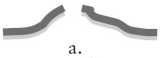
La **prosodie** est un aspect déterminant du langage parlé qui permet de véhiculer un message linguistique et paralinguistique et que nous pouvons définir comme un système grammatical complexe (Beckman, 1996) dont les sous-composantes sont l'intonation, le phrasé et l'accentuation (*intonation, phrasing, prominence*). D'un point de vue strictement phonétique, il s'agit principalement des modulations mélodiques et rythmiques produites dans un énoncé qui servent à véhiculer du sens pragmatique (comme le *focus* de phrase) ou bien à marquer la frontière des constituants syntaxiques. En effet, afin de saisir pleinement le sens ou la structure linguistique d'un énoncé, les informations lexicales, syntaxiques et sémantiques ne suffisent pas, et les modulations prosodiques doivent être prises en compte. Par exemple, la même phrase « Aujourd'hui il fait vraiment beau » peut véhiculer soit un sens littéral, si elle est prononcée avec une intonation assertive neutre, ou bien signifier son contraire si elle est prononcée avec une prosodie ironique (D'Imperio, Champagne-Lavau, Loevenbruck, et Fouillaud, 2013 ; Loevenbruck, Jannet, D'Imperio, Spini & Champagne-Lavau, 2013). La prosodie joue aussi un rôle dans la désambigüisation syntaxique, si on prend l'exemple « La vieille ferme la porte », qui peut être prononcé soit comme « La vieille | ferme la porte » ou bien « La vieille ferme | la porte » (où la rupture prosodique est marquée par la barre verticale), qui diffèrent par la fonction de l'homophone « ferme » qui revêt un rôle verbal dans la première phrase *vs* celui d'un adjectif dans la deuxième. Il s'agit aussi des premiers aspects de la parole acquis par l'enfant, et ce à partir du stade prénatal (de Boysson-Bardies, 1996), qui lui permettront d'accéder au lexique (Gerken, 1996 ; Gerken, et Aslin, 2005 ; Jusczyk, Cutler, et Redanz, 1993 ; Jusczyk, et Kemler Nelson, 1996) ainsi qu'à d'autres structures linguistiques, telles que les constituants syntaxiques de haut niveau.

Parmi les sous-systèmes prosodiques, un aspect qui a retenu l'attention de nombreux chercheurs est **l'intonation**, ou les *variations mélodiques au cours de l'énoncé qui peuvent être utilisées pour apporter une valeur pragmatique contrastive* (Beyssade, Delais-Roussarie, Marandin, et Rialland, 2004 ; Ladd, 1996/2008 ; Portes et Reyle, 2014). Dans le langage adulte, des patrons intonatifs contrastifs, tels que des accents mélodiques catégoriquement différents (qui diffèrent par exemple en alignement tonal, c'est-à-dire la position temporelle du pic mélodique à l'intérieur d'une syllabe accentuée, voir D'Imperio 2011 pour une revue), peuvent en effet déterminer à eux seuls la valeur pragmatique d'un énoncé telle que, par exemple, le contraste entre une question et une assertion, comme dans (1) ; ou signaler la portée du focus informationnel (large ou étroit) ou un contraste dans le discours, comme dans (2) ; ou bien désambigüiser un référent dans le discours (Ito et Speer, 2008), ce qui est exemplifié en (3). Plus spécifiquement, une courbe mélodique montante (voir 1b) est uti-

SOMMAIRE

lisée pour une demande de confirmation ou une question oui/non en français et aussi dans d'autres langues romanes (sauf dans des variétés d'italien méridional, comme le napolitain, voir D'Imperio et House, 1997, D'Imperio 2005a, D'Imperio 2011). Par contre, la position de l'accent nucléaire (le plus proéminent de la phrase) est souvent utilisée pour marquer une information contrastive, comme on peut voir en 2. L'accentuation peut être aussi utilisé aussi en tant que mécanisme visant à la désambigüisation de référents pronominaux, comme on peut voir en (3).

(1)



a. Jean est venu.

b. Jean est venu ?

(2)

Marie ARRIVERA demain, elle ne partira pas.

(3)

a. Anne et Robert sont allés au café hier soir.

b. Quand ils sont arrivés au comptoir, ELLE a payé, non pas Robert.

Pourtant, ces mêmes caractéristiques sont parmi les dernières à être complètement maîtrisées à la fois en production et compréhension d'une langue seconde, surtout en ce qui concerne le détail phonétique fin comme les allongements accentuels ou ceux qui marquent la présence d'une frontière prosodique (*preboundary lengthening*) ou bien l'alignement tonal (D'Imperio, Cavone, et Petrone, 2014 ; Mennen, 2004). Dans les sections suivantes, nous traiterons d'abord les aspects linguistiques de la prosodie et de l'intonation (section 1) avant de considérer les lignes de recherches les plus récentes concernant le traitement de la prosodie dans les paradigmes psycholinguistiques (section 2) et neurolinguistiques, visant à mieux comprendre les bases cérébrales de ces processus en bénéficiant des méthodes d'analyse en temps réel, telles que les mouvements oculaires, les potentiels évoqués et l'imagerie cérébrale (section 3). Dans la dernière section (section 4), nous résumerons les avancées principales dans l'étude des interactions entre prosodie et musique et nous donnerons quelques notions relatives à la prosodie émotionnelle.

1. Prosodie et intonation : notions linguistiques de base

Avant de décrire l'impact de la prosodie et de l'intonation sur le traitement et la compréhension de la parole, nous devons définir ces termes. En premier lieu, la prosodie peut être envisagée comme un système phonologique complexe qui repose sur trois sous-systèmes : l'intonation (ou modulations mélodiques de l'énoncé à caractère pragmatique), l'accentuation (proéminence au niveau du mot et de la phrase) et le phrasé (structuration en constituants de nature prosodique dont les frontières sont signalées par une rupture rythmique et/ou intonative). Nous défendons donc ici une vision structurelle de la prosodie (Beckman, 1996) dans la mesure où il s'agit d'un système qui représente plus qu'un simple ensemble de propriétés acoustiques/phonétiques (telles que la durée segmentale, l'intensité, la fréquence fondamentale [f_0] ou hauteur perçue [*pitch*], et la qualité de la voix). Tout phénomène prosodique, tel que la proéminence accentuelle ou de phrasé (*phrasing*), peut être signalé par une combinaison d'indices phonétiques différents et cette combinaison peut-être différente pour chaque langue. Par exemple, le marquage perceptif de la proéminence, qui pourrait signaler l'accent lexical (*stress*) ou mélodique *(pitch accent)*, peut se manifester aussi bien par des variations des paramètres spécifiques de durée et d'intensité que par une modulation de la f_0 qui se traduit par une variation de hauteur perçue (*pitch*). En outre, ces propriétés peuvent se combiner de manière différente dans différents contextes d'énonciation ainsi que dans différentes variétés de langues. Qui plus est, certains patrons de durée segmentale, tel l'allongement d'une voyelle en position finale de mot, pourraient également être utilisés pour signaler le phrasé (*phrasing* ou forme prosodique), c'est-à-dire le marquage d'une rupture prosodique, ou break, correspondant à la frontière d'un constituant prosodique. Ainsi, une simple correspondance biunivoque entre un élément prosodique et une seule propriété acoustique/prosodique ne peut être envisagée. En d'autres termes, l'idée qu'un accent soit forcement réalisé par une excursion de f_0 ou par un pic d'intensité (Bolinger, 1978 ; Fry, 1955, 1958) n'est pas corroborée par les données issues de langues différentes (Jun, 2005, 2014). Le concept même d'accent lexical ou de stress doit être revu, dans la mesure où la proéminence rythmique au niveau du mot peut s'exprimer à travers différentes typologies accentuelles selon la langue étudiée (voir le concept de *stress accent vs non-stress accent* de Beckman, 1986).

En même temps, concevoir les propriétés prosodiques de la parole en tant que structures abstraites ne suffirait pas non plus. Dans le modèle de la phonologie prosodique classique (Nespor et Vogel, 1986 ; Selkirk, 1984), les constituants prosodiques tels que la syllabe, le syntagme phonologique ou le syntagme intonatif, sont dérivés par des algorithmes syntaxiques qui ne prennent pas en compte la variabilité des énoncés. Or, nous savons que la même structure syntaxique peut être réalisée avec des ruptures prosodiques différentes ou avec un emplacement accentuel variable qui reflète plutôt la portée pragmatique que syntaxique d'un énoncé. Par exemple, la question « Qui a mangé le poulet aux amandes ce soir ? » peut recevoir une montée initiale (voir Welby, 2006 ; D'Imperio *et al*., à paraître) sur « aux amandes » ou bien sur « le poulet », ce qui se traduit par un contraste de nature pragmatique ou de structure de l'information perçu et interprété différemment par les auditeurs (voir German et D'Imperio, 2015 ; D'Imperio, German, et Michelas, 2012 pour la relation entre structure de l'information et focus en français ; Cutler, Dahan, et van Donselaar, 1997 pour des effets concernant l'anglais et Magne *et al*., 2005, ci-dessous, pour les bases cérébrales du focus pragmatique).

Les locuteurs peuvent mettre en relief un mot parmi d'autres à l'intérieur d'un énoncé grâce au placement de l'accent de focalisation, pour les langues telles que l'anglais ou l'italien (D'Imperio, 2002 ; Grice, D'Imperio, Savino, et Avesani, 2005 ; voir Ladd, 1996/2008 pour une revue) ou grâce à la présence d'une montée initiale en français (German et D'Imperio, 2010 ; 2015). En effet, les travaux sur le français suggèrent qu'au lieu d'une simple proéminence prosodique, les constituants focalisés ou porteurs de

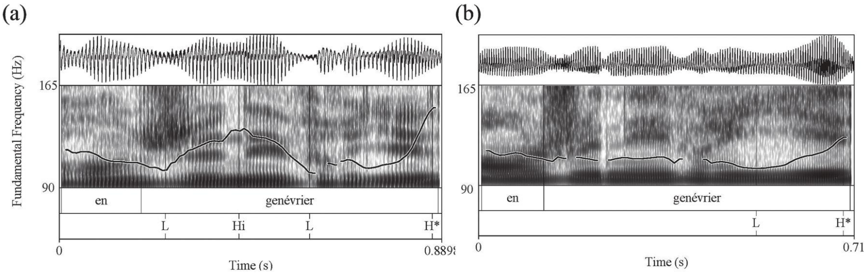


FIGURE 1. Contours de f_0 pour la phrase « en genévrier » réalisée comme un seul syntagme accentuel, avec une montée initiale (LHi) et une montée finale (LH*), en 1a, ou avec seulement une montée LH* finale (1b). Tiré de German and D'Imperio (2015).

contraste pourraient être signalés par la présence de tons de frontières spécifiques (Di Cristo, 1999 ; Féry, 2001), tels que la montée initiale LHi (voir figure 1a). Des études plus récentes montrent que la focalisation prosodique en français est un phénomène complexe. Ainsi, il existe au moins deux façons de marquer l'accent en français et la relation entre accent et prosodie ne peut être modélisée que par une approche intégrée dans laquelle différents niveaux linguistiques seraient pris en compte simultanément (D'Imperio *et al.*, 2012 ; German et D'Imperio, 2010 ; voir aussi Dohen et Lœvenbruck, 2004).

Du côté de la compréhension, il semble que les auditeurs recherchent activement les mots accentués dans le discours en raison de leur rôle sémantique central (Cutler, 1982 ; Sedivy, Tanenhaus, Spivey-Knowlton, Eberhard, et Carlson, 1995). De plus, les auditeurs sont capables d'interpréter l'information prosodique pour obtenir des indices relatifs à la structuration discursive (Hirschberg et Pierrehumbert, 1986), et à la prise de parole (turn-taking), un mécanisme qui se mettrait en place déjà à partir de la petite enfance, comme démontré par des travaux récents (Lammertink, Casillas, Benders, Post, et Fikkert, 2015). La structure intonative est également primordiale dans l'interprétation du discours : la dérivation du sens de l'intonation implique la considération simultanée des contours mélodiques (tunes), du contexte propositionnel de la phrase ainsi que du discours dans lequel ils apparaissent (Grabe, Gussenhoven, Haan, Marsi, et Post, 1998). Cependant, l'intégration entre indices prosodiques et contextuels semble être acquise plus tard que la compétence intonative simple (après 5 ans, voir Laval et Bert-Erboul, 2005), et pourrait être sévèrement altérée dans les populations pathologiques, comme dans le cas de la schizophrénie.

Cette complexité signifie aussi que les propriétés objectivement mesurables au niveau de l'énoncé, telles que le niveau de hauteur (pitch) ou la durée d'une syllabe ne donnent souvent qu'une image partielle de la prosodie de l'énoncé. Par conséquent, il est très important de décrire l'énoncé comme un système prosodique à la fois phonologique et phonétique. En d'autres termes, nous souscrivons à la position de Beckman (1996) selon laquelle le niveau prosodique du langage parlé doit être analysé comme une structure grammaticale, tout en reconnaissant la nécessité de modéliser une composante phonétique qui permet de traduire le niveau de représentation phonologique en propriétés acoustiques concrètes du signal de parole (ou représentation phonétique).

Pour ce faire, nous adoptons un modèle phonologique de l'intonation, le modèle « Autosegmental-Metrical » (théorie AM, voir Ladd, 1996/2008), issu du travail de Pierrehumbert (1980) et de ses collègues. Ce modèle est également à la base du système de transcription prosodique ToBI (Tones and Break Indices) qui a été employé pour décrire plusieurs langues du monde (Beckman et Ayers-Elam, 1993 ; voir D'Imperio, Grice, et Cangemi, 2016 pour des approches récentes à partir de ce système) et qui a aussi été récemment appliqué à des langues comme le français (Delais-Roussarie *et al.*, 2015). Selon la théorie AM, le contour tonal d'un énoncé peut être envisagé comme une série de cibles tonales (par exemple, des cibles statiques comme les niveaux hauts H (High) ou bas L (Low) qui peuvent être associées à des syllabes métriquement fortes en donnant lieu à des accents mélodiques ou pitch accents (d'où l'utilisation de la notation avec astérisque, H*, L*, etc.). Ce système codifie aussi des mouvements complexes, tels que des montées (par exemple : L+H *, voir Figure 2 ; ou L*+H) ou des chutes de f0 (par exemple, H+L*), ainsi que les tons qui délimitent les frontières prosodiques à différents niveaux (telles que H % pour une montée finale de contour questionnant en français ou un H- de syntagme intermédiaire, voir Figure 3 ; voir D'Imperio et Michelas, 2014 ; Michelas et D'Imperio, 2012 ; 2015).

Il faut remarquer que l'inventaire exact des accents mélodiques tout comme le type et le nombre de niveaux de phrasé dépendent strictement de la langue décrite. De plus, ces événements de nature tonale doivent être considérés en tant que représentations phonologiques, puisque le détail de leur réalisation phonétique (tels que l'alignement tonal exact de leur pic ou bien la synchronisation entre ton et segment ; voir D'Imperio, 2011) est également spécifique à une langue donnée. Comme souligné par les fondateurs de ToBI, ce système ne doit pas être utilisé comme un alphabet phonétique international (API) pour l'intonation. Il est donc nécessaire d'opérer au préalable une analyse phonologique des patrons intonatifs potentiels et des niveaux de rupture prosodique qui seraient contrastifs dans la langue étudiée.

Du point de vue du traitement du langage, cette analyse phonologique permet de considérer que les distinctions prosodiques perceptivement pertinentes impliquent souvent la combinaison

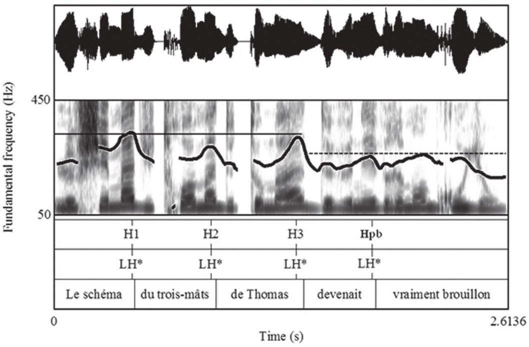


FIGURE 2. Courbe de f_0 et spectrogramme de l'énoncé « La mamie des amis de Rémy ? Elle viendra demain ». Les lignes de registre du premier et du deuxième syntagme intermédiaire sont identifiées par les deux lignes horizontales. Tiré de D'Imperio et Michelas (2014).

de plus d'un paramètre acoustique, et que ces paramètres sont influencés par le contexte segmental de réalisation (c'est-à-dire la présence et le type d'attaque consonantique ou de coda syllabique, van Santen et Hirschberg, 1994). Par exemple, un accent mélodique montant de type H* verrait son pic déplacé à la fois vers la fin ou vers le début d'une syllabe en Italien selon que la syllabe à laquelle ce ton est associé est ouverte ou fermée (D'Imperio, 2000, 2005b ; Prieto, D'Imperio, et Gili Fivela, 2005) ou bien qu'elle est placée au début ou à la fin du syntagme intonatif (Silverman et Pierrehumbert, 1990). En d'autres termes, ce modèle est basé sur l'idée que les détails prosodiques acoustiquement et perceptivement saillants ainsi que les contrastes linguistiquement pertinents auxquels ces détails contribuent (au niveau lexical, syntaxique et pragmatique) sont essentiels et doivent être décrits. Ce modèle semble aussi assez robuste en termes de pouvoir descriptif, de testabilité et par son applicabilité avérée au français (D'Imperio *et al.*, 2012 ; D'Imperio, Bertrand, Di Cristo, et Portes, 2007 ; Jun et Fougeron, 2000 ; 2002 ; Post, 2000 ; voir aussi D'Imperio, et Michelas et Portes, à paraitre).

2. Traitement de l'intonation et de la prosodie en compréhension : données psycholinguistiques

Comme nous l'avons vu ci-dessus, la prosodie est un aspect fondamental du langage parlé, pourtant longtemps considéré comme moins déterminant que la syntaxe ou la sémantique qui occupaient le devant de la scène. Le langage était alors considéré comme un système comprenant différents niveaux de traitement (phonologique, syntaxique, sémantique et éventuellement pragmatique) hiérarchiquement organisés, chaque niveau étant relativement indépendant des autres niveaux de traitement. L'enjeu majeur était de comprendre la structuration syntaxique des phrase et dans un deuxième temps, leur signification, leur portée sémantique. Or, les résultats de travaux en linguistique, psycho- et neurolinguistique ont largement remis en cause ces conceptions hiérarchiques et modulaires de l'organisation du langage et la manière dont le langage est perçu, compris et produit par les locuteurs.

Ainsi, la structuration prosodique d'un énoncé influence l'analyse linguistique à plusieurs niveaux au cours de la production et la compréhension du langage parlé (voir Wagner et Watson, 2010 pour une revue récente). Les nombreuses fonctions du marquage prosodique et intonatif concernent la segmentation lexicale (Cutler, Dahan et van Donselaar, 1997 ; Spinelli, Grimaud, Meunier et Welby, 2010), la segmentation des phrases en constituants syntaxiques, l'interprétation de la signification lin-

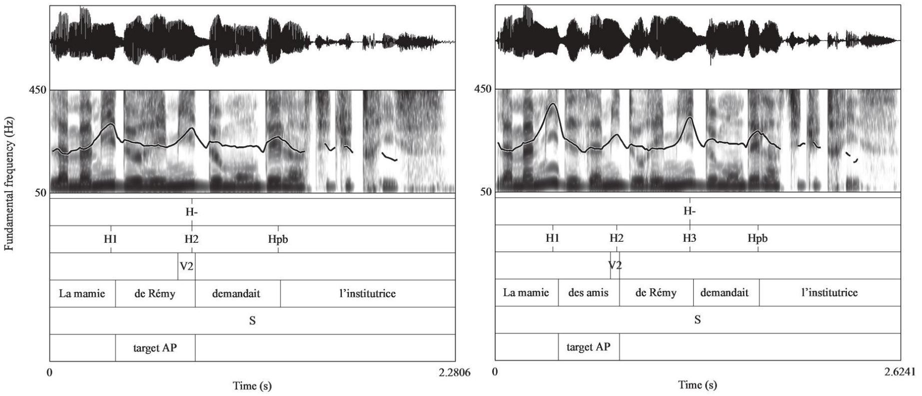


FIGURE 3. Courbes de f_0 et spectrogrammes des énoncés « La mamie de Rémy demandait l'institutrice » (gauche) et « La mamie des amis de Rémy demandait l'institutrice » (droite) produits par 2 locutrices féminines. Le ton de syntagme H- est signalé dans les deux cas, mais dans l'énoncé à droite (avec 3 AP), il est clair que ce ton de frontière provoque un rehaussement du *pitch accent* LH* de sorte qu'il revienne à la ligne de registre du syntagme intermédiaire. Tiré de D'Imperio et Michelas (2014).

guistique (à la fois sémantique et pragmatique) et extra-linguistique, ainsi que l'établissement et le maintien des fonctions discursives telles que la prise de tour de parole (voir Venditti et Hirschberg, 2003 pour une revue).

Un des rôles principaux de la prosodie et de l'intonation dans la reconnaissance des mots est d'aider à la segmentation de la chaîne continue des sons de parole en unités distinctes que sont les mots. Les résultats d'études dans de nombreuses langues (voir, par exemple, pour une revue Otake et Cutler, 1996) ont montré que les auditeurs peuvent utiliser la structure rythmique des énoncés pour déterminer le positionnement des frontières de mots. Comme la structure rythmique diffère selon les langues, les processus impliqués dans la segmentation des énoncés en mots peuvent également être spécifiques à une langue donnée. Par exemple, la segmentation se fait sur la base de la syllabe en anglais (Cutler et Norris, 1988), mais sur le positionnement de l'accent lexical (*stress*) en français (Mehler, Dommergues, Frauenfelder, et Segui, 1981) et de la more, qui est une unité prosodique plus petite que la syllabe, en japonais (Otake, Hatano, Cutler, et Mehler, 1993). Cette spécificité de la langue peut entraîner une mauvaise application des procédures de segmentation dans une langue étrangère qui auraient une structure rythmique différente (Otake et Cutler, 1996). Autre exemple : les nourrissons peuvent distinguer perceptivement des langues rythmiquement différentes, mais pas celles qui sont rythmiquement similaires (Nazzi, Bertoncini, et Mehler, 1998).

De multiples sources d'information sont traitées simultanément pour construire la structure linguistique d'un énoncé (Norris, McQueen, et Cutler, 2000 ; Spivey-Knowlton et Tanenhaus, 1998), mais nous ne savons toujours pas comment les informations pragmatiques de haut niveau sont combinées aux informations prosodiques, ou comment et à quel moment les informations acoustiques suprasegmentales (pitch, durée, qualité de la voix, intensité sonore) s'intègrent pour permettre l'identification d'un évènement phonologique de nature linguistique tel qu'un accent mélodique et/ou un ton de frontière. Ce manque de connaissances est, au moins en partie, lié au fait que la plupart des études sur la perception et la fonction de l'intonation et du phrasé prosodique en français ont été conduites en utilisant des méthodes « hors ligne » ou des jugements de pertinence pour lesquels les processus de compréhension ne sont pas indépendants des processus décisionnels (Christophe, Peperkamp, Pallier, Block, et Mehler, 2004 ; Michelas et D'Imperio 2015 ; Millotte, René, Wales, et Christophe, 2008 ; Pynte et Prieur, 1996). Cependant, des recherches plus récentes sur le placement et la catégorie d'accent mélodique (surtout de contraste ou focus), utilisant des mesures en temps réel, ont mis en évidence que le positionnement des accents contrastifs est à la base d'inférences contrastives très rapides (Dahan, Tanenhaus, et Chambers, 2002 ; Ito et Speer, 2008). Malgré ces études, il n'y a pas à l'heure actuelle de modèle de la façon dont le sens prosodique s'intègre aux autres niveaux linguistiques dans le traitement de la parole.

La relation entre le traitement du phrasé prosodique et de la structure syntaxique des énoncés est l'un des aspects les plus étudiés en psycholinguistique. Il semble que les indices prosodiques associés à la structure syntaxique soient produits à la fois en lecture et en parole spontanée. En fait, la reconnaissance de la structure syntaxique d'un énoncé serait très largement déterminée par le phrasé prosodique (Clifton, Carlson, et Frazier, 2002). Bien que la syntaxe puisse fournir des indices pour la prosodie et vice versa, la relation entre syntaxe et prosodie n'est pas univoque. Le rôle de l'unité prosodique est particulièrement complexe lorsque l'objet de l'étude est la parole conversationnelle, et non pas simplement la parole de laboratoire. Bien que la structuration prosodique soit beaucoup plus complexe, la plupart des études en traitement de la prosodie considèrent une seule unité prosodique maximale, le syntagme intonatif ou *Intonation Phrase* (IP), basé sur de petites unités hiérarchisées, comme le Syntagme Intermédiaire ou *Intermediate Phrase* (IP, voir D'Imperio et Michelas, 2014 pour le français) et le Syntagme Accentuel ou *Accentual Phrase* (AP, pour le français, voir Jun et Fougeron, 2000).

Des études récentes en français (D'Imperio et Michelas, 2014 ; Michelas et D'Imperio, 2012) ont mis en évidence les caractéristiques phonétiques et phonologiques du syntagme intermédiaire. Il aurait, entre autres, une fonction syntaxique dans la mesure où sa frontière droite semble placée entre des constituants syntaxiques majeurs (*e.g.*, entre un syntagme nominal sujet et un syntagme verbal). Cette structuration se baserait largement sur un phénomène de reset du registre tonal déterminé par la hauteur du premier pic de l'énoncé. De plus, les frontières de syntagme intermédiaire, ainsi que celles de syntagme accentuel, seraient utilisées en temps réel pour prédire le statut morphosyntaxique du mot qui suit immédiatement un fragment d'énoncé. Dans l'étude de Michelas et D'Imperio (2015), ces frontières ont été placées après le fragment « La Nana du sauna », qui pouvait être suivi, par exemple, soit par « d'Héléna » (donc interne au syntagme nominal) ou par « était.... » (donc un syntagme verbal). Nous avons modulé la hauteur tonale de la dernière syllabe du fragment, et demandé au sujet de choisir entre les deux complétions possibles. Les résultats ont montré que la hauteur tonale est utilisée efficacement dans cette tâche : comme dans le cas d'une frontière de syntagme intermédiaire, la dernière syllabe du fragment est plus haute (ou bien elle n'est pas abaissée par un phénomène de *downstep*), comme on peut voir en Figure 4b (Michelas et D'Imperio, 2015, voir aussi D'Imperio *et al.*, à paraître).

Une des méthodes qui commence à être largement utilisée dans l'étude de l'utilisation incrémentale des indices prosodiques dans le processus de compréhension du sens pragmatique (tel que le contraste) est celle des mouvements oculaires. Elle permet de mesurer les délais de fixations oculaires ainsi que leurs distributions sur une scène visuelle (*Visual World Paradigm*, Tanenhaus *et al.*, 1995). Elle permet aussi de montrer que les fixations sont plus ou moins anticipées par rapport à des objets cibles affichés sur un écran selon les modifications prosodiques des stimuli. Ainsi, cette méthode permet d'étudier les effets d'anticipation du contexte syntaxique, sémantique et prosodique. Un autre avantage de cette méthode est que les effets *en temps réel* de la prosodie peuvent être étudiés dans des milieux assez « naturel ». Par exemple, grâce à une étude récente utilisant les mouvements oculaires et exploitant le paradigme de la scène visuelle (Esteve-Gibert, Portes, Schafer, Hemforth, et D'Imperio, 2016, à paraître), nous avons montré une relation entre les différences cognitives individuelles liées aux niveaux d'empathie (*perspective taking*) et le traitement incrémental des indices prosodiques de l'énoncé dans le cas de la signalisation du contraste prosodique. En effet, les patrons de fixations oculaires des sujets les plus empathiques seraient différenciés dès lors que la partie initiale (ou *pré-nucléaire*) du syntagme intonatif est présentée. Ce résultat démontre, pour la première fois, que l'auditeur ne doit pas attendre d'avoir perçu l'accent nucléaire (et, donc, le ton de frontière qui le suit) pour opérer un processus d'inférence sémantique guidé par l'intonation.

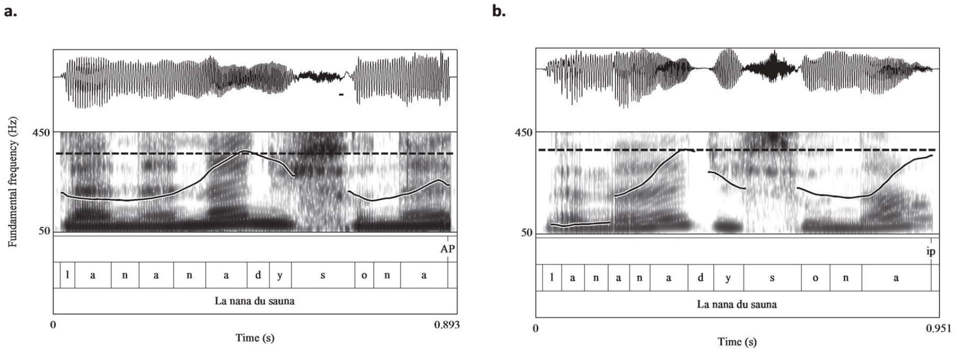


FIGURE 4. **Courbes de f_0 pour les deux fragments de phrase « La nana du sauna » dont la dernière syllabe « na » est soit associée à une frontière de type AP (a), soit à une frontière de syntagme intermédiaire de type ip (b).**

3. Traitement de l'intonation et de la prosodie en compréhension : données neurolinguistiques

Nous prendrons ici quelques exemples issus de la neurolinguistique qui soulignent les interactions étroites entre prosodie, sémantique et syntaxe et qui illustrent l'évolution vers des modèles dynamiques de la construction du sens.

La méthode des Potentiels évoqués (PEs)[1] permet de suivre, en temps réel, le déroulement des opérations mentales qui président à la compréhension du langage. Ainsi, on distingue typiquement les composantes N100 et P200, liées au traitement sensoriel et perceptif, la composante N200 souvent associée aux processus de catégorisation, la composante N400, associée au traitement sémantique, et la famille des composantes P300 qui reflètent la prise de décision et les processus liés à la détection d'un évènement pertinent et parfois surprenant, comme une violation syntaxique dans un énoncé ou le non-respect d'une frontière prosodique.

3.1. Interaction des traitements prosodiques, sémantiques et syntaxiques

L'une des premières études utilisant la méthode des PEs, réalisée par Steinhauer et collaborateurs, a permis de montrer que les indices prosodiques, en particulier les frontières prosodiques, sont utilisés en temps réel pour construire la structure syntaxique des énoncés, et sont associés à une composante positive tardive, probablement de la famille des P300 (Steinhauer, Alter et Friederici, 1999). En montrant qu'une composante des PEs était sensible au traitement prosodique, ces résultats ont ouvert la voie à d'autres études visant à examiner les relations entre la prosodie et les autres niveaux de traitement linguistique. Ainsi, nous avons comparé traitements prosodique et sémantique afin de déterminer s'ils sont indépendants ou interactifs (Astésano, Besson, et Alter, 2004). Pour ce faire, nous avons construit des énoncés prosodiquement congruents (phrases interrogatives ou déclaratives) ou incongrus (un début de phrase interrogative combiné à une fin de phrase déclarative et vice-versa) et des énoncés sémantiquement congruents (« la lumière clignotait ») ou incongrus (« la toiture conduisait »). Les participants faisaient attention à l'intonation ou à la sémantique des énoncés. Les résultats principaux montrent que la composante N400 est plus ample lorsque le dernier mot de la phrase est incongru que lorsqu'il est congruent (indépendamment de la direction de l'attention) et que l'amplitude d'une composante positive tardive, de la famille des P300, est plus ample pour les énoncés prosodiquement incongrus que congruents (lorsque les participants prêtent attention à la prosodie). Sur la base de cette première analyse, on pourrait conclure que les deux types d'incongruité sont associés à des composantes des PEs différentes et donc à des processus différents, et que, par conséquent, traitements sémantique et prosodique sont indépendants. Mais les choses ne sont pas si simples. En effet, l'effet de la valeur prosodique de l'énoncé est plus fort pour les phrases sémantiquement incongrues que pour les phrases congruentes ce qui montre que les aspects prosodiques (ici l'intonation) et sémantiques d'un énoncé sont traités en temps réel et en interaction. Ils démontrent également que la valeur sémantique de l'énoncé est comprise automatiquement (indépendamment de la direction de l'attention) alors qu'il faut prêter attention aux « monstres prosodiques » (début de phrase interrogative associé à une fin de phrase déclarative) pour qu'ils suscitent un effet au niveau électrophysiologique.

Il y a bien sûr mille manières de tester les relations entre prosodie et sémantique. Dans une autre expérience (Magne *et al.*, 2007), nous avons comparé des énoncés métriquement con-

---

[1] L'enregistrement de l'activité électrique cérébrale (EEG) au moyen d'électrodes placées sur le scalp (typiquement insérées dans un casque à électrodes) et le moyennage de cette activité en prenant comme origine des temps le début de la stimulation (la présentation d'un mot, par exemple) permet de faire émerger de l'électrogénèse corticale les variations ou composantes qui sont liées au stimulus, au mot. Les composantes des PEs sont caractérisées par leur polarité (N pour négatif, P pour positif), leur latence par rapport au début du stimulus (N100, composante négative à 100 ms), leur amplitude et leur signification fonctionnelle qui est le plus souvent encore matière à débats.

gruents, pour lesquels l'allongement de la durée syllabique qui se produit typiquement à la fin d'un mot est bien placé (« Le concours a regroupé mille candi**dats** »), et des énoncés métriquement incongrus, pour lesquels l'allongement syllabique est mal placé sur l'avant-dernière syllabe (« Le concours a regroupé mille can**di**dats »). Les énoncés, sémantiquement congruents dans les exemples ci-dessus, pouvaient également être sémantiquement incongrus (« Le concours a regroupé mille bigou**dis** » et « Le concours a regroupé mille bi**gou**dis »). Les participants faisaient attention à la métrique ou à la sémantique des énoncés. Une petite parenthèse avant de résumer les résultats obtenus. Modifier la durée d'une syllabe sans modifier sa fréquence fondamentale (f_0) ou son intensité est un problème ardu, que nous avons pu résoudre, mais qui montre clairement les interactions très fortes entre les différents paramètres acoustiques qui forment les sons. Par exemple, la perception de la durée d'un son semble dépendre non seulement de sa fréquence, mais aussi de son intensité perçue, la sonie. Retournons maintenant aux résultats qui montrent que les mots métriquement incongrus suscitent, comme nous l'avons vu pour d'autres incongruités prosodiques, des composantes positives tardives de type P300, dont la latence varie entre 300 et 800 ms post-stimulus. Or, dans ce cas, une composante N400 précède la composante positive tardive, ce qui pourrait signifier qu'un allongement syllabique mal placé sur l'avant-dernière syllabe perturbe l'accès au sens du mot. De nouveau, ces résultats illustrent les interactions qui existent entre traitements prosodique (métrique) et sémantique. Des collègues du Max Planck Institute à Leipzig, Schmidt-Kassow et Kotz (2009), ont utilisé un protocole très proche pour tester les relations entre métrique et syntaxe. Ils ont obtenu des résultats similaires ce qui les a conduits à conclure que la valeur métrique d'un énoncé influence non seulement les aspects sémantiques, mais aussi de structuration syntaxique. De manière générale, ces résultats sont en accord avec les modèles actuels de la compréhension du langage mentionnés ci-dessus, selon lesquels les différents indices disponibles (accents, intonations, phrasés, sémantiques et syntaxiques) sont utilisés en temps réel pour faciliter la compréhension des énoncés (Hagoort, 2014 ; Norris *et al*., 2000).

3.2. La fonction pragmatique de la prosodie

Dans une autre étude (Magne *et al*., 2005) nous avons testé la fonction pragmatique de la prosodie à travers le concept de focus prosodique (accent d'emphase). Évidemment, c'est un peu compliqué, mais allons-y ! Nous avons construit de courts dialogues sous forme de questions-réponses :

1) A-t-il donné une bague ou un bracelet à sa fiancée ? Il a donné une **bague** à sa fiancée (l'accent focal est là où il doit être, tout va bien)
2) A-t-il donné une bague à sa fiancée ou à sa sœur ? Il a donné une **bague** à sa fiancée (ici rien ne va plus, l'accent focal est mal placé ; peut-être que le locuteur ne souhaite pas répondre à une question dont la réponse peut prêter à conséquences…).
3) A-t-il donné une bague à sa fiancée ou à sa sœur ? Il a donné une bague à sa **fiancée** (l'accent focal est bien placé sur le dernier mot, c'est clair, tout va bien)
4) A-t-il donné une bague ou un bracelet à sa fiancée ? Il a donné une bague à sa **fiancée** (de nouveau, rien ne va plus, peut-être que le locuteur ne veut pas répondre à la question posée par peur d'incidents diplomatiques…)

Ce protocole assez complexe a néanmoins permis d'obtenir des résultats très clairs. Un accent focal mal placé suscite des variations particulières au niveau des PEs : l'accent focal est donc traité en temps réel et permet de mieux comprendre la réponse à la question posée. Ces variations sont différentes en fonction de la position de l'accent focal dans l'énoncé : au milieu de l'énoncé, il crée un effet de surprise qui se traduit par l'occurrence d'une composante de type P300 (comme vu précédemment pour les incongruités métriques et d'intonation). À la fin de l'énoncé, il crée une incompréhension (il n'y a pas d'autres informations disponibles pour éclairer le malentendu) qui est associée à une composante N400. Ces jolis résultats, qui n'ont été que très peu cités dans la littérature…, montrent que la prosodie, en soulignant la valeur pragmatique des énoncés, facilite la fonction communicative du langage.

Enfin, il est également important de noter que des travaux récents (Rufener, Oechslin, Wöstmann, Dellwo, et Meyer, 2015) ont démontré une relation entre les oscillations neurales de l'EEG dans le rang de fréquence thêta (4–8 Hz) lors de l'écoute d'énoncés qui sont modulés de manière à préserver l'enveloppe temporelle et dans le rang de fréquence gamma (30-48 Hz) lorsque la structure temporelle fine est préservée. Ces résultats vont dans le sens du modèle proposé par Poeppel (2003) qui fait l'hypothèse d'un isomorphisme entre les modulations de fréquence au niveau des rythmes de l'EEG et les oscillations acoustiques du langage parlé.

3.3. Les régions cérébrales activées par la prosodie

De nombreux travaux ont utilisé l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) pour tenter de localiser les régions cérébrales spécifiquement impliquées dans le traitement de la prosodie. Les premières études (Meyer, Alter, et Friederici, 2003 ; Meyer, Steinhauer, Alter, Friederici, et von Cramon, 2004) ont souvent comparé des énoncés linguistiquement corrects à des énoncés filtrés qui préservent les informations prosodiques d'amplitude, de fréquence et de durée, mais qui suppriment les informations morpho-syntaxiques et lexicales de telle manière que l'énoncé est incompréhensible, mais qu'il est possible de reconnaître le contour prosodique. Les résultats soulignent l'implication des régions frontales, en particulier des aires fronto-latérales de l'hémisphère droit. En outre, les énoncés dont les mots à contenu sémantique ont été remplacés par des pseudo-mots (qui sont donc syntaxiquement et prosodiquement corrects mais qui n'ont pas de sens) activent le *planum polare* et la région fronto-operculaire bilatéralement. Or, des résultats très récents (Norman-Haignere, Kanwisher, et McDermott, 2015) obtenus en utilisant des méthodes d'analyse sophistiquées montrent que le *planum polare* serait spécifiquement impliqué lors de l'écoute de sons musicaux…

Une idée, largement répandue (bien qu'elle ne soit pas nécessairement correcte), est que la prosodie est traitée dans l'hémisphère droit. En effet, les résultats des premières études ont montré des activations plutôt à droite qu'à gauche. Par exemple, Weintraub et collaborateurs (Weintraub, Mesulam, et Kramer, 1981) ont testé des participants avec lésions de l'hémisphère droit et des participants de contrôle dans une tâche de discrimination de phrases déclaratives et interrogatives ; ils ont montré que les capacités de discrimination étaient diminuées chez les patients avec lésion droite. Des résultats similaires ont été rapportés par Bryan (1989) en utilisant la même tâche, mais en ajoutant un groupe de patients avec lésion à gauche ; les patients avec lésion à droite avaient plus de difficultés à percevoir l'intonation des énoncés que les patients avec lésion à gauche et les participants de contrôle. Ainsi, en accord avec l'idée que la prosodie est la musique du langage (voir ci-dessous) et que le traitement de la musique se fait dans l'hémisphère droit, les résultats de ces deux études suggèrent que le traitement de la prosodie linguistique est aussi préférentiellement localisé dans l'hémisphère droit (pour des résultats similaires chez le participant sain voir Shipley-Brown, Dingwall, Berlin, Yeni-Komshian et Gordon-Salant, 1988).

Cependant, d'autres résultats soulignent que l'intégralité des deux hémisphères est nécessaire au traitement de la prosodie linguistique. Par exemple, Perkins *et al.* (Perkins, Baran, et Gandour, 1996) n'ont pas trouvé de différences significatives entre patients avec lésions dans les hémisphères droit ou gauche dans une tâche d'identification ou de discrimination de phrases interrogatives ou déclaratives (*i.e.*, avec intonation) et de frontières syntaxiques (*i.e.*, sans intonation). Imaizumi et collaborateurs (Imaizumi, Mori, Kiritani, Hiroshi, et Tonoike, 1998) ont obtenu des résultats similaires (activation bilatérale lors de tâches impliquant la prosodie linguistique) en utilisant la Magnétoencéphalographie (MEG). Plus récemment, Witteman et collaborateurs (Witteman, van Ijzendoorn, van de Velde, van Heuven et Schiler, 2011) ont fait une méta-analyse des études conduites chez des patients cérébro-lésés. Ces auteurs insistent sur le fait que ces études posent souvent des problèmes au niveau méthodologique, comme par exemple, le manque d'un groupe de contrôle adapté, le fait que les effets de transfert inter-hémisphérique par le corps calleux ne soient pas considérés et que l'hétérogénéité des lésions (aussi bien du type de lésions que de leur localisation) dans l'hémisphère droit (ou gauche) puisse avoir des conséquences comportementales très différentes (Wong, 2002). En fait, les résultats conduisent à penser que la spécialisation hémisphérique pour la prosodie serait plutôt relative qu'en tout ou rien. Les auteurs concluent qu'il n'y a pas de différences significatives entre les deux hémisphères pour le traitement de la prosodie linguistique, et que les lésions de l'hémisphère gauche, aussi bien que celles de l'hémisphère droit, ont un effet négatif important (voir aussi la revue de Baum et Pell, 1999 pour des conclusions similaires).

Une autre méta-analyse de la perception de la prosodie linguistique a été réalisée récemment par Belyk et Brown (2014), cette fois chez des participants sains. Ces auteurs ont considéré les données obtenues en utilisant l'IRMf ou la Tomographie par Émission de Positons (TEP) en sélectionnant les études selon des critères rigoureux (*e.g.*, utilisation de tâches expérimentales portant explicitement sur la perception de la prosodie ; une condition de contrôle incluant un traitement de bas niveau ; une analyse portant sur la différence entre tâche expérimentale et de contrôle ; des analyses sur le cerveau pris dans son ensemble et non pas centrées sur des régions d'intérêt...). De nouveau, les problèmes méthodologiques sont tels que seules 10 études résistent à cette sélection rigoureuse ! Le plus souvent, les résultats montrent des activations dans les régions typiquement impliquées dans la perception et la compréhension du langage : des activations bilatérales du gyrus frontal inférieur (pars opercularis, aire de Brodmann 44 [BA44], *i.e.*, aire de Broca et homologue de Broca dans l'hémisphère droit), du gyrus frontal moyen, du cortex auditif primaire (BA41) et de la partie postérieure du gyrus temporal supérieur, du gyrus supra-marginal (BA40) et de l'aire motrice supplémentaire plutôt à droite, et des activations du noyau caudé plutôt latéralisées à gauche.

Ainsi, les résultats de ces méta-analyses, réalisées 30 ans après les premières études sur la perception de la prosodie linguistique, soulignent de nouveau que les résultats, qu'ils soient obtenus chez patients ou des participants sains, sont très variables : ils révèlent parfois des activations unilatérales (à droite ou à gauche), par exemple des régions auditives dans le lobe temporal, et parfois des activations bilatérales, par exemple du lobe frontal (Meyer, Alter, Friederici, Lohmann, et von Cramon, 2002 ; Strelnikov, Vorobyev, Chernigovskaya, et Medvedev, 2006). La question de la spécialisation hémisphérique du traitement de la prosodie linguistique reste donc ouverte.

Pour conclure cette section, nous considérons maintenant deux hypothèses, liées aux effets du stimulus ou aux effets de la tâche, qui ont été proposées pour expliquer la variabilité des résultats concernant la localisation du traitement de la prosodie linguistique.

L'hypothèse selon laquelle les différences de latéralisation sont liées aux caractéristiques du stimulus repose sur l'idée que tous les paramètres acoustiques de la prosodie sont traités dans l'hémisphère droit et intégrés avec l'information traitée dans l'hémisphère gauche à travers le corps calleux (Klouda, Robin, Graff-Radford, et Cooper, 1988). Cependant, d'autres résultats montrent que certains paramètres acoustiques sont pris en charge par l'un ou l'autre des deux hémisphères : temporel (hémisphère gauche) *vs* spectral (hémisphère droit ; Robin, Tranel, et Damasio, 1990 ; Van Lancker et Sidtis, 1992), haute (à gauche) *vs* basse fréquence (à droite ; Ivry et Robertson, 1998), et transitions formantiques rapides (à gauche) *vs* lentes (à droite ; Schwartz et Tallal, 1980).

L'hypothèse selon laquelle les différences observées sont liées à la tâche repose plutôt sur les niveaux hiérarchiques de la structure linguistique (Behrens, 1989). Ainsi, les aspects prosodiques au niveau de la phrase (intonation, phrasé) relèveraient d'un traitement global pris en charge par l'hémisphère droit alors que les aspects prosodiques au niveau du mot (accent lexical) relèveraient d'un traitement local dans l'hémisphère gauche. Par exemple, Gandour et collègues (2003) ont montré dans une étude utilisant l'IRMf et une tâche de discrimination de la taille des fenêtres d'intégration temporelle (*i.e.*, fenêtre courte = contours de hauteur associés aux tons chinois ; fenêtre longue = contours d'intonation), que les contours de hauteur locaux (tons chinois) étaient traités dans l'hémisphère gauche alors que les contours de hauteur globaux (intonation) étaient traités bilatéralement, avec une préférence pour l'hémisphère droit. Ces résultats vont dans le sens du modèle proposé par Poeppel (2001) dans lequel les informations auditives dans une fenêtre courte (25-50 ms) sont traitées préférentiellement dans l'hémisphère gauche et dans une fenêtre longue (150-250 ms), plutôt dans l'hémisphère droit.

En conclusion, et en accord avec l'évolution des modèles linguistiques décrite dans la première section de ce chapitre, le traitement de la prosodie linguistique impliquerait une interaction dynamique entre les deux hémisphères. Cette interaction serait modulée à la fois par le traitement des paramètres acoustiques contenus dans le contour prosodique et par l'importance accordée au niveau de structuration linguistique, avec une sensibilité aux informations prosodiques suprasegmentales dans l'hémisphère droit et une dominance de l'hémisphère gauche dans le traitement des variations prosodiques au niveau segmental.

4. Prosodie et musique

On dit souvent que la prosodie est la musique du langage. Mais quelle est la réalité scientifique de cette jolie métaphore ? Comme nous l'avons vu précédemment, la prosodie repose sur quatre paramètres acoustiques, la fréquence, la durée, l'intensité et le timbre qui correspondent aux percepts de hauteur, de rythme ou de mètre, de sonie et de timbre, et sur leurs interactions. Or, la musique repose également sur ces mêmes paramètres acoustiques et sur leurs interactions ce qui a conduit les chercheurs à déterminer si le traitement des aspects prosodiques du langage et de la musique repose ou pas sur des processus communs (pour une analyse des relations entre chant et musique, voir le chapitre de Lévêque *et al.*, ce volume). Cette question qui pourrait paraître relativement simple et circonscrite a néanmoins une portée théorique importante. En effet, si langage et musique partagent des processus communs, le langage ne peut pas être considéré comme un système isolé et indépendant des autres activités cognitives, comme la musique, par exemple. Au contraire, l'acte de langage, dans sa complexité et ses différents niveaux d'organisation (phonologique, prosodique, sémantique, syntaxique, pragmatique) se construit en interaction avec les autres fonctions cognitives, telles que l'attention, la mémoire et les fonctions exécutives et motrices. On voit ainsi qu'une réponse positive remet en cause une organisation strictement hiérarchique et modulaire du langage au profit d'une organisation dynamique et interactive où le système linguistique tire profit de toutes les informations disponibles à un instant donné pour construire l'acte de langage.

4.1. Les effets de l'expertise musicale sur le traitement de la prosodie

Les musiciens ont développé une meilleure perception auditive et ils perçoivent mieux que les non-musiciens des variations de hauteur ou de durée dans des phrases musicales. Ceci n'est pas pour nous étonner compte tenu qu'un musicien professionnel âgé de 25 ans, a passé environ 30 000 heures à pratiquer son instrument de musique et donc à écouter et à produire des sons. Ce qui est plus étonnant est que cet avantage s'étend à la perception des différents aspects de la prosodie. Par exemple, dans une première série d'expériences nous avons étudié la perception de l'intonation, qui permet on l'a vu, de différencier phrases interrogatives (la hauteur monte en fin de phrase: « Tu fumes encore ? ») et phrases déclaratives (la hauteur reste stable ou descend en fin de phrase: « La petite fille lit un livre d'images »). Nous avons montré que les adultes musiciens professionnels aussi bien que les enfants, entre 9 et 12 ans, qui suivent un apprentissage musical depuis 4 à 5 ans en moyenne, sont plus sensibles aux variations d'intonation que les non-musiciens, surtout lorsque ces variations d'intonation sont difficiles à percevoir (faible variation de la f_0). Cette plus grande sensibilité se traduit par une latence plus courte et une augmentation de l'amplitude de la composante P300, associée, entre autres, à la perception consciente d'évènements pertinents (Magne, Schön, et Besson, 2006 ; Schön, Magne, et Besson, 2004). En outre, les musiciens adultes sont également plus sensibles que les non-musiciens à des variations d'intonation dans une langue étrangère qu'ils ne comprennent pas (des phrases en Portugais ; Marques, Moreno, Castro, et Besson, 2007). Enfin, nous avons utilisé le protocole expérimental décrit plus haut qui permet de manipuler l'allongement syllabique sur le dernier mot (métriquement congruent) ou avant-dernier mot (métriquement incongru ; Magne *et al.*, 2007) afin de déterminer si les musiciens sont plus sensibles à la métrique d'un énoncé (Marie, Magne, et Besson, 2010). En accord avec cette hypothèse, les résultats montrent que les musiciens font moins d'erreurs et réagissent plus fortement que les non-musiciens à la structure des mots au niveau perceptif (augmentation d'amplitude de la composante P200) ce qui semble interférer avec l'intégration du sens des mots (augmentation d'amplitude de la composante N400) dans le contexte de l'énoncé (augmentation d'amplitude d'une composante positive tardive de type P300). Ainsi, ces différents résultats suggèrent que les musiciens ont développé une plus grande sensibilité aux différents paramètres acoustiques que les non-musiciens ce qui leur permettrait de mieux percevoir des variations de hauteur ou de durée non seulement dans la musique, mais aussi dans le langage. Cependant, des facteurs autres que perceptifs, tels que l'attention auditive (très difficile à dissocier de la perception auditive) ou la mémoire auditive, peuvent également contribuer à expliquer ces résultats.

L'apprentissage musical facilite également la perception des variations de hauteur dans des langues à tons, comme le Chinois Mandarin ou le Thaï, dans lesquelles ces variations sont linguistiquement pertinentes car elles modifient le sens des mots (Bidelman, Gandour, et Krishnan, 2011). Par exemple, le mot monosyllabique /di/ prononcé sur un ton haut signifie « bas » alors que /di/ prononcé sur un ton descendant signifie « terre ». L'analyse des PEs montre que la composante N2 associée à la catégorisation et la composante P300, associée, entre autres, à la prise de décision sont respectivement de plus courte latence et de plus grande amplitude pour les musiciens que pour les non-musiciens (Marie, Delogu, Lampis, Olivetti Belardinelli, et Besson, 2011). Ces résultats confortent l'idée intuitive selon laquelle les musiciens ont des facilités pour apprendre des langues étrangères, et en particulier, peut-être, les langues à tons…

En conclusion, il est intéressant de noter un retour d'intérêt pour les similarités entre prosodie et musique. En effet, alors que de nombreux travaux au cours des 15 dernières années ont eu pour but de comparer la structure harmonique des phrases musicales et la structure syntaxique des phrases linguistiques, certains auteurs (Heffner et Slevc, 2015) argumentent de manière tout à fait convaincante en faveur d'une plus grande similarité entre structures prosodiques et musicales qu'entre structures syntaxiques et harmoniques. Ceci ouvre évidemment la voie à de nouvelles recherches…

4.2. La prosodie émotionnelle

La prosodie est aussi le véhicule privilégié pour la transmission des émotions : comment rester insensible à une voix pathétique ou refuser une demande énoncée d'une voix très douce… Les premiers travaux dans ce domaine (Schirmer, Kotz, et Friederici, 2002) ont examiné la fonction émotionnelle de la prosodie en manipulant l'intonation joyeuse ou triste d'énoncés sémantiquement neutres (« Hier, elle a passé son dernier examen). La présentation visuelle d'un mot cible qui ne correspond pas à l'intonation (*e.g.*, le mot « succès » après une intonation triste) est associée à une composante N400 plus ample que si le mot cible correspond à l'intonation (*e.g.*, le mot « succès » après une intonation joyeuse) ce qui montre que, comme la prosodie linguistique, la prosodie émotionnelle interagit avec le traitement sémantique. Des travaux suivants (Kotz *et al.*, 2003), utilisant cette fois l'IRMf, ont montré que les intonations tristes et joyeuses activent un réseau bilatéral fronto-temporal ainsi que des régions sous-corticales (les ganglions de la base). En outre, l'activation frontale est renforcée lorsque les énoncés sont filtrés de manière à ne retenir que les modulations prosodiques (le contenu linguistique n'est pas intelligible).

Qu'en est-il des effets de l'expertise musicale sur la perception de la prosodie émotionnelle ? Des travaux récents (Lima et Castro, 2011) montrent que les musiciens reconnaissent plus précisément que les non-musiciens les émotions de colère, de dégoût, de peur, de joie, de tristesse ou de surprise, portées par une phrase sémantiquement neutre (*e.g.*, « Elle a voyagé par train »). Par conséquent, si vous voulez cacher vos sentiments, mieux vaut ne pas parler à un musicien : son oreille entraînée pourrait déceler vos secrets…

5. Conclusion

Ce résumé, forcément incomplet, des travaux conduits en linguistique, psycho- et neurolinguistique sur la perception et la production de la prosodie permet néanmoins de souligner un aspect fondamentalement encourageant : l'évolution convergente des conceptions issues de ces différentes approches. En effet, les théories linguistiques ont évolué à partir d'une conception modulaire de l'organisation du langage, accordant un rôle premier à la syntaxe, vers une conception dynamique dans laquelle tous les indices linguistiques, y compris bien sûr les indices prosodiques, sont traités en temps réel et en interaction pour comprendre et produire un énoncé. Ces conclusions sont largement en accord avec celles basées sur les données acquises en psycho- et neurolinguistique et avec les modèles proposés dans ces domaines. On assiste donc à la convergence des cercles linguistique, psycholinguistique et neurolinguistique et au développement de la « neuro(psycho)linguistique » décrite par Jean-Luc Nespoulous dans l'introduction de ce livre, ce qui témoigne s'il en est encore besoin, de l'importance de l'interdisciplinarité.

Références bibliographiques

Astésano, C., Besson, M. & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18, 172-184.

Baum, S. & Pell, M. (1999). The neural bases of prosody : Insights from lesion studies and neuroimaging. *Aphasiology*, 13, 581-608.

Beckman, M.E. (1986). *Stress and non-stress accent (Netherlands Phonetics Archives, vol. VII)*. Dordrecht : Foris Publications. Pp. xiii + 239.

Beckman, M.E. (1996). The parsing of prosody. *Language and Cognitive Processes* 11, 17-67.

Beckman, M.E. & Ayers Elam, G. (1993). *Guidelines for ToBI labelling, version 3.0. Manuscript and accompanying speech materials*, Ohio State University. [Obtain by writing to tobi@ling.ohio-state.edu]

Behrens, S.J. (1989). Characterizing sentence intonation in a right hemisphere-damaged population. *Brain and Language*, 37, 181-200.

Belyk, M. & Brown, S. (2014). Perception of affective and linguistic prosody : an ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience*, 9, 1395-1403.

Beyssade, C., Delais-Roussarie, E., Marandin, J.-M. & Rialland, A. (2004). « Prosody and Information in French », in F. Corblin & H. de Swart & (eds.), *Handbook of French Semantics*, Stanford : CSLI. 477-500.

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

Bolinger, D.L. (1978). Intonation across languages. In J. Greenberg, Ed., *Universals of Human Language,* vol. 2, *Phonology.* Palo Alto, CA : Stanford University Press, pp. 471-524.

Bryan, K.L. (1989). Language prosody and the right hemisphere. *Aphasiology*, 3, 285-299.

Christophe, A., Peperkamp, S., Pallier, C., Block, E. & Mehler, J. (2004). *Journal of Memory and Language*, 51, 523-547.

Clifton, C., Carlson, K. & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, 45, 87-114.

Cutler, A. (1982). Prosody and sentence perception in English. In J. Mehler, E.C.T. Walker & M.F. Garrett, Eds., *Perspectives on Mental Representation : Experimental and Theoretical Studies of Cognitive Processes and Capacities.* Hillsdale, NJ : Erlbaum, pp. 201-216.

Cutler, A., Dahan, D. & van Donselaar, W. (1997). Prosody in the comprehension of spoken language : a literature review. *Language & Speech*, 40, 141-201.

Cutler, A. & Norris, D.G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology : Human Perception and Performance* 14, 113-121.

Dahan, D., Tanenhaus, M.K. & Chambers, C.G. (2002). Accent and reference resolution in spoken-language comprehension. Journal of *Memory and Language*, 47, 292-314.

De Boysson-Bardies, B. (1996). *Comment la parole vient aux enfants*. Odile Jacobs.

Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S.A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R. & Yoo, H. (2015) *Chapter 3 : Intonational Phonology of French : Developing a ToBI system for French, Intonation in Romance*, Oxford University Press, Editors : Sónia Frota & Pilar Prieto.

Di Cristo, A. (1999). Vers une modélisation de l'accentuation du français : Première partie. *Journal of French Language Studies*, 9, 143-180.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. Ph.D Dissertation, The Ohio State University.

D'Imperio, M. (2002). Italian intonation : An overview and some questions. *Probus*, 14, 37-69.

D'Imperio, M. (2005a). *Aspects prosodiques de l'italien de Naples : proéminence, catégories accentuelles et cibles tonales*. Mémoire HDR, Paris III (Sorbonne Nouvelle).

D'Imperio, M. (2005b). « La Phonologie de Laboratoire : finalités et quelques applications ». In Nguyen, Noël, Wauquier-Gravelines, Sophie and Durand, Jacques (Eds.) *Phonologie et phonétique : Forme et substance*. Paris : Hermès. 2005, p. 241-264.

D'Imperio, M. (2011). « Prosodic representations ». In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Handbook of Laboratory Phonology*. Oxford : Oxford University Press (section on tonal alignment).

D'Imperio, M., Bertrand, R., Di Cristo, A. & Portes, C. (2007). « Investigating phrasing levels in French : Is there a difference between nuclear and prenuclear accents ? » In J. Camacho, V. Deprez, N. Flores, L. Sanchez (Eds.), *Selected Papers from the 36th Linguistic Symposium on Romance Languages (LSRL)*. New Brunswick : John Benjamins Publishing Company, pp. 97-110.

D'Imperio, M., Cavone, R. & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, 5, 1226.

D'Imperio, M., Champagne-Lavau, M., Loevenbruck, H. & Fouillaud, L. (2013). *Perceptual evaluation of tonal and contextual cues* *to sarcasm in French. Phonetics and Phonology in Iberia* (PaPI), June 25-26 2013, Lisbon.

D'Imperio, M., German, J. & Michelas, A. (2012). A multi-level approach to focus, phrasing and intonation in French. In Elordieta, G. and Prieto, P. (eds), *Prosody and Meaning* (pp. 11-35).

D'Imperio, M., Grice, M. & Cangemi, F. (2016). *Advancing Prosodic Transcription*. Laboratory Phonology (special issue of prosodic transcription).

D'Imperio, M. & D. House (1997). « *Perception* of questions and statements in Neapolitan Italian », *Proceedings of Eurospeech* 97, 1, 251-254, Rhodes.

D'Imperio, M. & Michelas, A. (2014). Pitch scaling and the internal structuring of the Intonation Phrase in French. *Phonology*, 31, 95-122.

D'Imperio, M., Michelas, A. & Portes, C. (à paraitre). *École d'Aix et approches tonales de l'intonation du français : un aperçu des recherches depuis les années 1970*. Langue Française.

Dohen, M. & Lœvenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and post-focal deaccentuation in French. In Proceedings of ICSLP, 1313-1316.

Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B. & D'Imperio, M. (2016). *Intonation in the processing of contrast meaning in French : an eye-tracking study*. Proceedings of the Speech Prosody 2016, Boston, MA (USA).

Esteve-Gibert, N., Portes, C., Schafer, A., Hemforth, B. D'Imperio M. (à paraitre). « Intonation and empathy in the online processing of contrastive meaning in French ». *Language and Speech* (special issue on Prosody and Meaning, Eds. P. Prieto and M.V. Escandell.

Féry, C. (2001). The Phonology of Focus in French. In : A*udiatur Vox Sapientiae. A Festschrift for Arnim von Stechow*. C. Féry and W. Sternefeld (eds.). Berlin. A-V. 153-181.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguisticstress. *The Journal of the Acoustic Society of America*, 27, 765-768.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126-152.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N. & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience : an fMRI study. *Brain and Language*, 84, 318-36.

Gerken, L.A. (1996). Prosody's role in language acquisition and adult parsing. *Journal of Psycholinguistic Research*, 25, 341-352

Gerken, L.A. & Aslin, R.N. (2005). Thirty years of Reserch on Infant Speech Perception : The Legacy of Peter W. Juszyk. *Language Learning and Development*, 1, 5-21.

German, J. & D'Imperio, M. (2010). « Predicting Phrase-initial Accents in French : Length Constraints and Focus ». *Speech Prosody 5*, Mai 11-14, 2010, Chicago, USA, 100207 : 1-4.

German, J. & D'Imperio, M. (2015). The Status of the Initial Rise as a Marker of Focus in French, *Language and Speech*, 59, 165-195.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E. & Post, B. (1998). Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* 41, 63-85.

Grice, M., D'Imperio, M., Savino, M. & Avesani, C. (2005). Towards a strategy for labelling varieties of Italian. In Jun, S.-A. (ed.) *Prosodic Typology and Transcription : A Unified Approach*. Oxford : Oxford University Press. 55-83.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language : Broca's region and beyond. *Current Opinion in Neurobiology*, 28, 136-141.

Heffner, C.C. & Slevc, L.R. (2015). Prosodic Structure as a Parallel to Musical Structure. *Frontiers in Psychology*, 22, 1962.

Hirschberg, J. & Pierrehumbert, J. (1986). Intonational Structuring of Discourse. Proceedings of the 24th Meeting of the Association for Computational Linguistics, 136-144.

Imaizumi, S., Mori, K., Kiritani, S., Hiroshi, H. & Tonoike, M. (1998). Task-dependent laterality for cue decoding during spoken language processing. *Neuroreport*, 9, 899-903.

Ito, K. & Speer, S. (2008). Anticipatory effects of intonation : Eye movements during instructed visual search. *Journal of Memory and Language*. 58, 541-573.

Ivry, R. & Robertson, L. (1998). The two sides of perception. Cambridge, MA : MIT Press.

Jun, S.-A. (2005). Editor. *Prosodic Typology : The Phonology of Intonation and Phrasing*. Oxford University Press.

Jun, S.-A. (2014). Editor. *Prosodic Typology II : The Phonology of Intonation and Phrasing*. Oxford University Press.

Jun, S.-A. & Fougeron, C. (2000). A Phonological Model of French Intonation. In *Intonation : Analysis, Modeling and Technology*, ed. by Antonis Botinis. Kluwer Academic Publishers. 209-242.

Jun, S.-A. & Fougeron, C. (2002). The Realizations of the Accentual Phrase in French Intonation. *Probus*, 14,147-172.

Jusczyk, P.W., Cutler, A. & Redanz, N. (1993). Infants sensitivity to predominant word stress patterns in English. *Child Development*, 64, 675-687.

Jusczyk, P.W. & Kemler Nelson, D. (1996). Syntactic units, prosody, and psychological reality duringinfancy. In J. Morgan & K. Demuth (Eds.), *Signal to syntax : Bootstrapping from speech to grammarin early acquisition* (389-408). Cambridge, MA : MIT Press.

Klouda, G.V., Robin, D.A., Graff-Radford, N.R. & Cooper, W.E. (1988). The role of callosal connections in speech prosody. *Brain and Language*, 35, 154-171.

Kotz, S., Meyer, M., Alter, K., Besson, M., von Cramon, Y. & Friederici, A. (2003). On the lateralization of affective prosody : An event-related functional MR investigation. *Brain and Language*, 86, 366-376.

Ladd, R.D. (1996/2008). *Intonational Phonology*, Cambridge University Press.

Lammertink, I., Casillas, M., Benders, T., Post, B. & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology*, 6, 495.

Laval, V. & Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm : The role of intonation and context. *Journal of Speech Language and Hearing Research*, 48, 610-620.

Lima, C.F. & Castro, S.L. (2011). Speaking to the trained ear : musical expertise enhances the recognition of emotions in speech prosody. *Emotion Washington DC*, 11, 1021-1031.

Loevenbruck, H., Ben Jannet, M., D'Imperio, M., Spini, M. & Champagne-Lavau, M. (2013). *Prosodic cues of sarcastic speech in French : slower, higher, wider*. Proceedings of Interspeech 2013, August 25-29, Lyon, FRANCE.

Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronland-Martinet, R. & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French : Behavioural and electrophysiological evidence. *Cerebral Cortex*, 17, 2659-2668.

Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K. & Besson, M. (2005). On-line processing of « pop-out » words in spoken French dialogues. *Journal of Cognitive Neuroscience*, 17, 740-756.

Magne, C., Schön, D. & Besson, M. (2006). Musician children detect pitch violations in both music and language better than non-musician children. *Journal of Cognitive Neuroscience*, 18, 199-211.

Marie, C., Delogu, F., Lampis, G., Olivetti Belardinelli, M. & Besson, M. (2011). Influence of musical expertise on tone and phoneme processing in Mandarin Chinese, *Journal of Cognitive Neuroscience*, 23, 2701-2715.

Marie, C., Magne, C. & Besson, M. (2010). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience*, 23, 294-305.

Marques, C., Moreno, S., Castro, S.L. & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than non-musicians : behavioural and electrophysiological evidence. *Journal of Cognitive Neuroscience*, 19, 1453-1463.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior,* 20, 298-305.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543-563.

Meyer, M., Alter, K. & Friederici, A.D. (2003). Towards the cerebral substrates of sentence-level syntactic and prosodic processing. *Journal of Neurolinguistics*, 16, 277-300.

Meyer, M., Alter, K., Friederici, A.D., Lohmann, G. & von Cramon, Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, 17, 73-88.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D. & von Cramon, D.Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89, 277-289.

Michelas, A. & D'Imperio, M. (2012). When syntax meets prosody : tonal and duration variability in French Accentual Phrases. *Journal of Phonetics*, 40, 816-829.

Michelas, A. & D'Imperio, M. (2015). Prosodic boundary strength guides syntactic parsing of French utterances. *Laboratory Phonology*, 6, 119-146.

Millotte, S., René, A., Wales, R. & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 34, 874-885.

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns : Towards an understanding of the role of rhythm. *Journal of Experimental Psychology : Human Perception and Performance* 24, 756-766.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht : Foris Publications.

Norman-Haignere, S., Kanwisher, N.G. & McDermott, J.H. (2015). Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-Free Voxel Decomposition. *Neuron*, 88, 1281-1296.

Norris, D., McQueen J.M. & Cutler, A. (2000). Merging information in speech recognition : feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-325.

Otake, T. & Cutler, A. (1996). Phonological Structure and Language Processing : Cross-Linguistic Studies. *Speech Research Series*, 12, Mouton de Gruyter, Berlin, 13-44.

Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993). Mora or syllable ? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 358-378.

Perkins, J.M., Baran, J.A. & Gandour, J. (1996). Hemispheric specialization in processing intonation contours. *Aphasiology*, 10, 343-362.

Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation*. Thèse de doctorat, MIT, USA.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science*, 25, 679-693.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows : cerebral lateralization as « asymmetric sampling in time ». *Speech Communication*, 41, 245-255.

Portes, C. & Reyle, U. (2014) *The meaning of French « implication » contour in conversation*. Proceedings of the 7th International Conference on Speech Prosody, 413-417.

Post, B. (2000). Tonal and phrasal structures in French intonation. The Hague : Thesus.

Prieto, P., D'Imperio, M. & Gili Fivela, B. (2005). Pitch accent alignment in Romance : primary and secondary associations with metrical structure. *Language and Speech*, 48, 359-396.

Pynte, J. & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, 11, 165-192.

Robin, D.A., Tranel, D. & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language*, 39, 539-555.

Rufener, K.S., Oechslin, M.S., Wöstmann, M., Dellwo, V. & Meyer, M. (2015). Age-Related Neural Oscillation Patterns During the Processing of Temporally Manipulated Speech. *Brain Topography*, 29, 440-458.

Schirmer, A., Kotz, S.A., Friederici, A.D. (2002). Sex differentiates the role of emotional prosody during word processing. Brain research. *Cognitive brain research*, 14, 228-233.

Schmidt-Kassow, M. & Kotz, S. (2009). Event-related potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience*, 21, 1693-1708.

Schön, D., Magne, C. & Besson, M. (2004). The music of speech : music training facilitates pitch perception in both music and language. *Psychophysiology*, 41, 341-349.

Schwartz, J. & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science*, 207, 1380-1381.

Sedivy, J., Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K. & Carlson, G. (1995). Using intonationally marked presuppositional information in on-line language processing : Evidence from eye movements to a visual model. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ : Erlbaum, 375-380.

Selkirk, E. (1984). *Phonology and syntax. The relation between sound and structure*. MITPress, Cambridge, Massachusetts.

Shipley-Brown, F., Dingwall, W.O., Berlin, C.I., Yeni-Komshian, G. & Gordon-Salant, S. (1988). Hemispheric processing of affective and linguistic intonation contours in normal subjects. *Brain and Language*, 33, 16-26.

Silverman, K. & Pierrehumbert, J. (1990). *The Timing of Prenuclear High Accents in English, Papers in Laboratory Phonology I*, Cambridge University Press, Cambridge UK. 72-106.

Spinelli, E., Grimaud, N., Meunier, F. & Welby, P. (2010). An intonational cue to segmentation in phonemically identical sequences. *Attention, Perception & Psychophysics*, 72, 775-787.

Spivey-Knowlton, M.J. & Tanenhaus, M.K. (1998). Syntactic ambiguity resolution in context : Modeling the effects of discourse context and lexical frequency. *Journal of Experimental Psychology : Learning, Memory and Cognition*, 24, 1521-1543.

Steinhauer, K., Alter, K. & Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191-196.

Strelnikov, K.N., Vorobyev, V.A., Chernigovskaya, T.V. & Medvedev, S.V. (2006). Prosodic clues to syntactic processing : a PET and ERP study. *Neuroimage*, 29, 1127-34.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M. & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.

Van Lancker, D. & Sidtis, J. (1992). The identification of affective-prosodic stimuli by left- and right hemisphere-damaged subjects : All errors are not created equal. *Journal of Speech and Hearing Research*, 35, 963-970.

Van Santen, J.P.H. & Hirschberg, J. (1994). *Segmental Effects on Timing and Height of Pitch Contours*. Proceedings of ICSLP-94, Yokohama, September, 719-722.

Venditti, J. & Hirschberg, J. (2003). *Intonation and Discourse Processing*. Proceedings of ICPhS 2003, Barcelona.

Wagner, M. & Watson, D.G. (2010). Experimental and theoretical advances in prosody : A review. *Language and Cognitive Processes*, 25, 905-945.

Weintraub, S., Mesulam, M. & Kramer, L. (1981). Disturbances in prosody : a right-hemisphere contribution to language. *Archives of Neurology*, 38, 742-44.

Welby, P. (2006). French intonational structure : Evidence from tonal alignment. *Journal of Phonetics* 34 : 343-371.

Witteman, J., van Ijzendoorn, M., van de Velde, D., van Heuven, V.J.P. & Schiler, N.O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception : a meta-analysis of the lesion literature. *Neuropsychologia*, 49, 3722-38.

Wong, P.C.M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin*, 59, 83-95.

4

MUSIC IN THE BRAIN

Music and Language Processing

Mireille Besson, Mylène Barbaroux, and Eva Dittinger

Introduction

Music and language[1] produce phenomenologically different experiences and require different abilities (for example, playing the violin is not speaking). However, they also share interesting commonalities: both are formed of structured sequences of auditory events that unfold in time and they rely on the same acoustic parameters (frequency, duration, intensity and timbre). These similarities have opened the intriguing possibility that musical training, by enhancing sensitivity to aspects that are common to music and language, positively influences language processing (the "cascade" hypothesis). As described below, such transfer effects have indeed been demonstrated from musical training to different levels of language processing.[2] Importantly, however, recent results have shown that musical training also influences cross-modal integration (Paraskevopoulos, Kraneburg, Herholz, Bamidis, & Pantev, 2015) and higher-order cognitive functions such as attention, working memory, short and long-term memory, and executive functions that are of primary importance to processing language (the multi-dimensional hypothesis). As a consequence, a currently debated issue is whether improvements in these cognitive functions are mediating the impact of musical training on language processing. However, as we will argue below, these two interpretations are probably best considered as complementary.

Transfer Effects from Music to Language

Below we review evidence for the influence of musical training and musical abilities on different aspects of language processing and executive functions. Results were obtained using different methodologies that provide relevant information on behavior (percent correct and reaction times, RTs), on the time-course of information processing (event-related potentials, ERPs) and on the activation of brain structures (functional magnetic resonance imaging, fMRI).

At the segmental level (consonants, vowels and syllables), musical training is positively correlated to the processing of brief speech sounds at multiple stages of the auditory system, from the brainstem (Wong, Skoe, Russo, Dees, & Kraus, 2007) to cortical regions (Chobert, Francois, Velay, & Besson, 2014; Elmer, Meyer, & Jancke, 2012). Musical training also enhances the discrimination of Mandarin tones in native English-speakers (Bidelman, Gandour, & Krishnan, 2011) and of lexical tones in Italian speakers (Delogu, Lampis, & Belardinelli, 2010). By recording ERPs, Marie et al. (2011) showed that both tone discrimination and higher-order decision processes were more efficient in musicians than in non-musicians. By contrast, the automatic orientation of attention was not different in the two groups.

Categorical perception is fundamental to speech perception by allowing listeners to categorize continuous acoustic changes in the speech signal into discrete phonetic categories. Interestingly, Bidelman and Alain (2015) demonstrated an influence of musical training, in both younger and older musicians, on the categorical perception of speech sounds (vowels) by recording both brainstem and cortical evoked responses. Thus, increased auditory sensitivity may be one of the driving forces behind enhanced categorical perception and enhanced speech processing in musicians.

At the supra-segmental level (words, sentences, discourse), early research has shown that adult musicians and children with musical training are more sensitive than non-musicians to linguistic prosody (e.g., final pitch rise in sentences; see Besson, Chobert, & Marie, 2011 for review; Gordon & Magne, this volume) and to emotional prosody (Lima & Castro, 2011). Since most of these findings were obtained from cross-sectional studies, it is possible that genetic pre-dispositions for music and musical abilities were influencing the results. To determine whether improvements in linguistic pitch processing were causally linked to musical training, Moreno and colleagues (2009) compared non-musician children (8–12 year old) before and after 6 months of music or painting training. Enhanced perception of prosodic intonation was found only in the music group, together with better reading abilities of complex words, providing evidence that musical training was directly related to these improvements.

At the phonological level, there is evidence that musical abilities are predictive of phonological skills in children (Anvari, Trainor, Woodside, & Levy, 2002) and in adults (Slevc & Miyake, 2006). Moreover, musical training positively influences phonological and reading abilities in dyslexic children (Habib et al., 2016). These results from cross-sectional studies are in line with those of a longitudinal study with 6–7 year old children showing that two months of rhythm-based training produced roughly comparable enhancements on a variety of standardized tests of phonological processing than an equivalent amount of training of phonological skills (Bhide, Power, & Goswami, 2013). They are also in line with the conclusions of an interesting meta-analysis of longitudinal studies recently conducted by Gordon et al. (2015), showing that musical training significantly improves phonological awareness skills, even if the effect sizes are small. By contrast, these analyses also showed that the evidence for an impact of musical training on reading has not yet been convincingly demonstrated.

Speech segmentation is necessary for speech comprehension as is clearly exemplified when learning a foreign language that is first perceived as a continuous stream of nonsense words. François and colleagues (2013) used a longitudinal approach over two school-years with 8 year old children to examine the impact of training in music, compared to painting, on the ability to extract "words" from a continuous stream of meaningless sung syllables. Implicit recognition of meaningless words steadily increased over the two years of music—but not of painting training—and this was associated with modulations of a fronto-central negative component.

Syntactic processing: since both music and language are structured sequences of events that unfold in time and since both syntax and harmony may rely on similar processing principles, several studies have investigated the influence of musical training on syntactic processing. Jentschke & Koelsch (2009) reported that violations of harmonic structure elicited larger cortical responses in musically trained children than in controls possibly because the former made better use of the prosodic and rhythmic cues that constrain syntactic constructions (Cason & Schön, 2012). Recent results by Gordon, Fehd, & McCandliss (2015) are in line with this hypothesis by showing that children with stronger rhythmic abilities also showed higher grammatical competence, as measured by their ability to produce sentences with relevant grammatical constructions.

Executive functions: what is meant by "executive functions and cognitive control" (generally concerned with top-down processes that control behavior) is currently debated in the literature (Diamond, 2013). Following Miyake and collaborators (2000), executive functions are linked to cognitive flexibility, inhibitory control, and working memory. However, executive functions are difficult to disentangle from one another as well as from short-term and long-term memory,[3] and contrastive results have been reported in the literature. For instance, results of several experiments showed that adult musicians outperformed non-musicians in working memory tasks based on musical and verbal stimuli (George & Coch, 2011; Schulze et al., 2011). However, it is still debated whether these findings reflect music-related improvements in working memory, short-term or long-term verbal memory, the influence of enhanced selective attention and cognitive control or the use of different strategies. Inconsistent findings have also been reported in children (Degé, Kubicek, & Schwarzer, 2011; Schellenberg, 2011).

Recently, Zuk and collaborators (2014) tested both adult and children musicians and non-musicians, selected according to strict criteria, and used a large standardized battery of executive functions. Adult musicians outperformed non-musicians on tests of cognitive flexibility and working memory, but not on tests of inhibitory control and processing speed. Children with musical training also differed from those without musical training on verbal fluency and trail making but not on tests of working memory and inhibitory control. By contrast, results of longitudinal studies showed larger improvements in verbal intelligence and inhibitory control in children from kindergarten involved in an intensive and interactive computer-based musical training program for 20 days compared to children involved in visual arts training (Moreno et al., 2011). Moreover, Roden and collaborators (2012) reported improved verbal memory in children who followed 45 minutes of weekly musical training during 18 months compared to children involved in natural science training or in no training. Finally, in a longitudinal study including a wide age-range (6 to 25 years old) and a large number of participants (N=352), Bergman Nutley and collaborators (2014) showed that musical training was positively associated with spatial and verbal working memory as well as with processing speed and general reasoning.

Clearly more experiments are needed to better understand the influence of musical training on executive functions as well as short-term and long-term memory by testing groups of participants with wide range of musical abilities, by using standardized tests when they are available and by trying to control for the effects of the many different factors that can influence the results (socio-economic status, bilingualism, etc.).

Word learning: Learning the meaning of new words is a multi-dimensional task that requires both perceptive and higher-order cognitive abilities. Since there is evidence that musicians show improved auditory perception and attention together with enhanced working and verbal memory, they should be at an advantage to learn new words in a foreign language like Thai, for example, in which both tones (i.e., patterns of rising and/or falling pitch) and vowel length contrasts change the meaning of a word.[4] Support for this hypothesis was provided by Wong & Perrachione (2007) and by Cooper & Wang (2012) who showed that both tone pitch identification and musical aptitudes were significantly correlated with word learning success in adult English native speakers.

Recently, we conducted a series of experiments (see Figure 4.1A) to further examine the influence of professional musical training on word learning (Dittinger et al., in press). Participants performed a series of experiments including word categorization, learning the meaning of monosyllabic words through picture-word associations, a matching, and a semantic task to test for learning and semantic generalization. Musicians outperformed non-musicians when word categorization was most difficult because it involved phonetic contrasts that belonged to the Thai but not to the French phonemic repertoire. In the learning phase, the N400 component, taken as an index that words had acquired meaning (Mestres-Misse, Rodriguez-Fornells, & Munte, 2007), developed after only 3 minutes of training, thereby reflecting fast brain plasticity in adults. Importantly, this N400 developed faster in musicians than in non-musicians (see Figure 4.1B). Moreover, in both the matching and semantic tasks, the N400 effect (i.e., the difference between mismatching/semantically unrelated and matching/related words) was larger in musicians than in non-musicians (see Figure 4.1C). The N400 effect was larger over centro-parietal regions in musicians and more frontally distributed in non-musicians, which was taken as evidence that musicians were more efficient at integrating the meaning of novel words into semantic networks. Finally, musicians were better than non-musicians at maintaining these novel words in long-term memory (tested five months later).

How can we account for these effects and more generally for the influence of musical training at various levels of language processing?

Two main interpretations, the cascade and multi-dimensional interpretations, can account for the results reviewed above. In the cascade interpretation (bottom-up), increased sensitivity to low-level acoustic parameters such as pitch or duration, that are common to music and speech, drives the influence of musical training at different levels of language processing (e.g., phonetic, phonologic, prosodic, syntactic and semantic; Besson et al., 2011; Cooper & Wang, 2012; Wong & Perrachione, 2007). In other words, because musicians perceive speech sounds better than non-musicians, they are more sensitive to prosodic cues such as pitch and rhythm and they form more accurate phonological representations. This cascades upward, facilitating the construction of more stable lexical representations as well as higher levels of language processing (e.g., syntactic structures, word learning and semantic processing). By contrast, in the multi-dimensional interpretation (top-down), both language and music are processed in interaction with other cognitive, emotional and motor functions that are enhanced by musical training (similar to the OPERA hypothesis proposed by Patel, 2014). Indeed, playing a musical instrument proficiently is a multi-dimensional ability requiring auditory and visual perception, auditory-visuo-motor integration, selective and divided attention, motor control, memory, cognitive control, and emotion.

In sum, it may come as no surprise that extensive training of these different abilities in musicians, from auditory perception to cognitive control, facilitate various levels of language processing (Kraus, Strait, & Parbery-Clark, 2012). In this respect the cascade and multi-dimensional hypotheses are complementary with both bottom-up and top-down processes

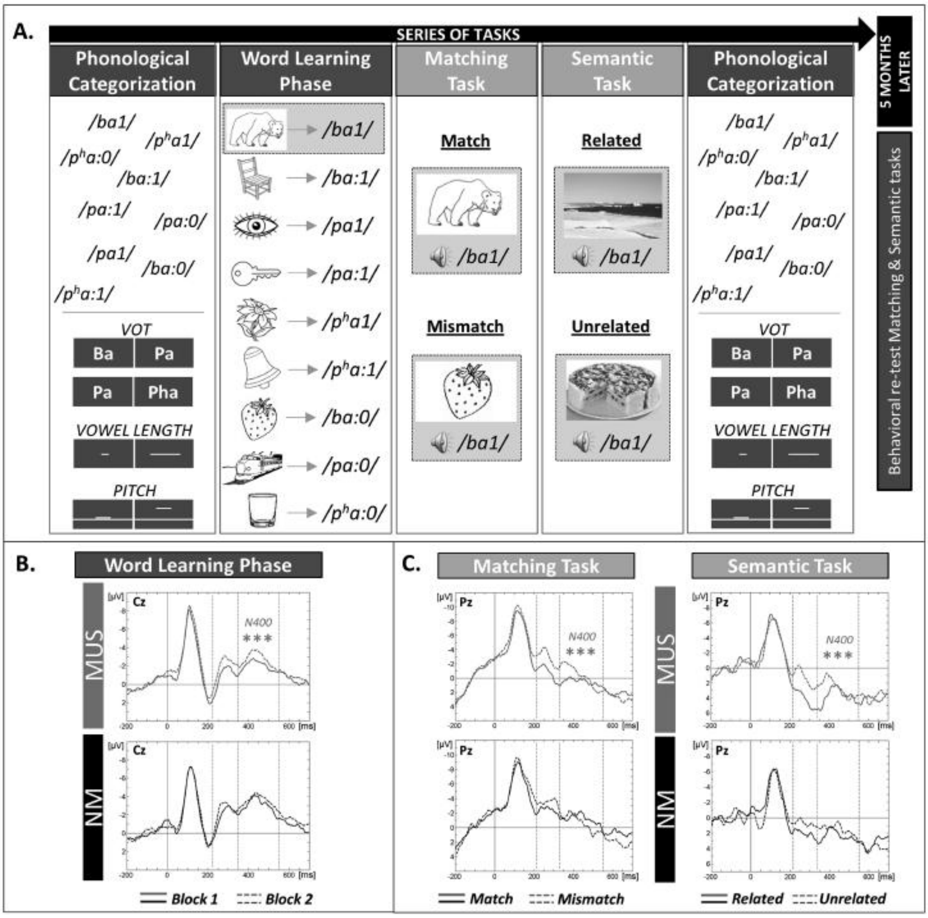


Figure 4.1 (A) Illustration of the series of tasks used in the experiment. Phonological categorization task of mono-syllabic words based on pitch, vowel length and voice-onset-time (VOT). In the word learning phase, participants learned the meaning of new words through picture-word associations. In the matching task, participants were asked to decide whether picture-word pairs matched or mismatched those from the learning phase and, in the semantic tasks, if novel pictures were semantically related or unrelated to the newly-learned words. Finally, the matching and semantic tasks were performed again five months later in a subset of participants to test for long-term memory. (B) The increase in N400 amplitude from the first to the second part of the learning phase (block 1 vs. block 2) was larger in musicians than in non-musicians. (C) The differences (N400 effect) between match/semantically related and mismatch/semantically unrelated words were also larger in musicians than in non-musicians.

probably at play to various degrees in most experimental designs. Moreover, the complementarity of the cascade and multi-dimensional interpretations is also illustrated by the results of neuro-imaging experiments investigating the neural basis of music and speech processing, a topic to which we now turn.

Shared vs. Distinct Neural Networks for Music and Speech

Whether similar levels of processing in language and music activate similar brain regions, and whether musical expertise influences brain structures considered as speech-specific, are important issues for our understanding of the anatomo-functional organization of the brain.

Broca's Area: From Speech Production and Syntax to Executive Functions

It has long been considered that the region of Broca, in the left inferior frontal cortex, was specific to speech production and to processing of syntactic structures. However, results of early experiments using fMRI showed that Broca's area was also activated when processing musical structures (Maess, Koelsch, Gunter, & Friederici, 2001). These results led to the conclusion that Broca's area was involved in the processing of syntax, defined as the rules that structure sequences of events that unfold in time, independently of whether these events form linguistic sentences or musical phrases. However, there is now evidence that Broca's area is also activated by the processing of phonological, lexical and semantic information (Sahin, Pinker, Cash, Schomer, & Halgren, 2009), thereby calling into question its syntax-specificity and the idea that syntactic constructions are processed independently from lexico-semantic information. Finally, and in line with the multi-dimensional view that language, like music, is processed in interaction with other cognitive functions, results also point to the activation of parts of Broca's areas in tasks requiring verbal and non-verbal working memory and/or executive functions (Schulze et al., 2011). Thus, although the functional role of Broca's area is still under debate, current evidence suggests that it may be part of a large prefrontal network, that includes language-specific as well as general cognitive functions (Fedorenko, Duncan, & Kanwisher, 2012; Hagoort, 2014; Schulze et al., 2011).

Wernicke's Area: From Language Comprehension to the Fine Temporal Structure of Speech

The same general picture emerges when considering Wernicke's area [part of the planum temporale (PT) in the left superior temporal gyrus (STG)], that was initially taken to play a major role in language comprehension. There is indeed clear evidence that temporal regions are crucial for speech processing. At the same time, there is also growing evidence for the functional diversity of these brain regions, with sub-areas sub-serving both linguistic and non-linguistic functions (see Liebenthal, Desai, Humphries, Sabri, & Desai, 2014 for results of a large meta-analysis). Directly related to our concerns, recent results have highlighted the influence of musicianship on various regions of the temporal lobe that are involved in speech processing.

Jäncke and collaborators (2002) showed that the left PT is specialized for decoding phonetic features, in particular the fine-grained temporal structure of speech such as voice-onset-time. Interestingly, subsequent research revealed that the left PT is similarly activated both by speech and by non-speech sounds (Zaehle, Wüstenberg, Meyer, & Jäncke, 2004) and more activated in professional musicians than in non-musicians when the phonetic categorization task is difficult (Elmer, Meyer, & Jäncke, 2012). Moreover, enhanced phonetic discrimination in musicians was positively correlated with the cortical surface area of the left PT (Elmer, Hänggi, Meyer, & Jäncke, 2013) and with increased structural connectivity between the right and left PT (Elmer, Hänggi, & Jäncke, 2016), providing evidence that long-term intensive musical training is associated with anatomical and functional changes in speech-specific brain regions such as the PT.

This is not to say, however, that the PT or, more generally, the superior temporal lobe is only involved in temporal speech processing. For instance, results of intra-cranial recordings in epileptic patients (Sammler et al., 2013) revealed activation of the left and right superior temporal lobes in response to violations of syntactic structures in both music and speech. Thus, while these data again provided evidence for shared neuroanatomical regions for syntactic processing in both domains, they were also somewhat surprising in showing a predominance of frontal regions for music harmonic processing and of temporal regions for linguistic syntax processing.

Temporal Processing in Music and Speech

Results of two recent studies are revealing regarding the issue of shared vs. distinct neural substrates in music and speech (Abrams et al., 2011; Rogalsky, Rong, Saberi, & Hickok, 2011). In both experiments, results of fMRI data showed similar activations of frontal and temporal regions of both hemispheres when processing the temporal structure of sentences and melodies, thereby arguing against a simple dichotomy between the left hemisphere for language and the right hemisphere for music and providing support for shared neural resources between music and speech. However, as noted by Rogalsky and collaborators (2011): "activation overlap does not necessarily imply computational overlap or even the involvement of the same neural systems at a finer-grained level of analysis" (p. 3846; see also Peretz, Vuvan, Lagrois, & Armony, 2015). Indeed, results using a finer-grained approach based on multivariate pattern analysis showed that the two types of stimuli elicited different patterns of spatial activity in music and speech. Thus, the authors concluded that temporal structure is encoded differently within the two domains and that distinct cortical networks are activated.

However, while both experiments examined the processing of temporal structures in music and speech, results were quite different in the two studies. In Abrams et al. (2011), activations in response to natural music and speech stimuli were found in Broca's and surrounding areas as well as in the left superior and middle temporal gyri, but the pattern of activation associated with temporal disruptions was different for music and speech. By contrast, using different rates of presentation of melodic phrases and jabberwocky sentences, Rogalsky et al. (2011) found no activation of Broca's area but increased activity in the temporal lobe, with some differences for melodic phrases and for jabberwocky sentences. Thus, confronting these results clearly showed that differences in task design and in participants (non-musicians in Abrams et al., 2011 vs. participants with different amount of musical training in Rogalsky et al., 2011), as well as differences in stimuli (e.g., sentences had a semantic content in the Abrams et al., 2011 study but not in Rogalsky et al., 2011) led to notable differences in the processing of temporal structures in music and speech.

Music-Specific Brain Regions?

Some answers to this question can be found in a recent experiment by Angulo-Perkins and collaborators (2014). These authors used fMRI, different types of natural stimuli (speech, human vocalizations (non-speech), musical excerpts played at the violin, at the piano and at the synthetic piano, environmental sounds and monkey vocalizations) and a task that was irrelevant to the aims of the study (detect the presentation of pure tones) to determine whether listening to music activates specific areas of the temporal lobe. Based on the results of a subsidiary analysis showing larger activation in the right planum polare for the violin excerpts than for speech, the authors concluded that "the planum polare may be involved in the processing of specific temporal characteristics that are inherent to music" (p. 136). However, the evidence for music-specific regions in this study was quite weak first because no region showed stronger activation for music than for speech in the main analysis (i.e., when considering both violin and piano musical excerpts and excluding non-speech vocalizations) and second because the bilateral anterior regions of the superior temporal gyrus (aSTG) responded similarly to music and human vocalizations (speech and non-speech) compared to environmental sounds and monkey vocalizations.

Nevertheless, it is interesting that converging results (music activating the planum temporale and planum polare bilaterally) were very recently reported by Norman-Haignere, Kanwisher, & McDermott (2015) when non-musician participants performed a task that, as in Angulo-Perkins et al. (2014), was irrelevant to the goal of the study (i.e., to detect changes in sound level). Moreover, these authors tried to overcome limitations of previous studies by using a large number of stimuli (165 sounds) that varied along different dimensions (speech, environmental sounds, guitar, pop or rock songs). Results showed that six components explained 80% of the variance in the data. Most importantly, while components 1 to 4 responded to all sound categories, component 5 (located in the STG) responded to sounds categorized as speech and component 6 (located in the planum temporale and planum polare) to sounds categorized as music. These results thus provided evidence for distinct regions involved in some aspects of speech and music processing with a nice fit with previous findings of Angulo-Perkins et al. (2014). However, results again showed weak music selectivity possibly because the neural populations involved in music processing overlapped with the other neuronal populations identified by components 1 to 5. In sum, results of these two studies again provided evidence for both shared and distinct neural substrates for music and speech.

Interim Summary

The literature comparing the brain structures involved in music and language processing is very rich and we only reviewed a few recent examples. However, we hope that these examples clearly illustrate that what we currently know is both complex and exciting. The complexity comes from behavioral and anatomical variability (within and between subjects), from the diversity of tasks and stimuli used to compare similar levels of processing in music and language (and how this similarity is defined), and from using diverse experimental methods (e.g., behavior, brainstem or cortical evoked potentials, intra-cranial recordings, fMRI) and procedures for data analysis.

These results are exciting because they strongly contribute to our understanding of how language and music are implemented in the human brain. Importantly, when considered in detail, results of each one of the experiments reported above (Abrams et al., 2011; Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015; Rogalsky et al., 2011) showed evidence for both shared and distinct neural networks involved in music and speech processing depending upon the level and type of analysis. These results are thus compatible with the proposed complementarity of the previously described cascade and multi-dimensional interpretations.

General Conclusion

Taken together, experimental results related to transfer effects contribute to the evolution of our understanding of human cognition. Theoretically, they demonstrate that the computations necessary for language and music processing are not performed independently from other cognitive functions and that language and music are not modular and encapsulated systems (Fodor, 1983). In line with current views in the neuroscience of language (Friederici & Singer, 2015), one way to reconcile the seemingly contradictory findings reported above is to consider the brain as a dynamic system (see also Loui,& Przysinda, this volume) in which "the functionality of a region is co-determined by the network of regions in which it is embedded at particular moments in time" (p. 136; Hagoort, 2014), this being determined by the stimuli, the task demands and the cognitive, emotional, motivational state of the individual. Such a dynamic view of the brain functional organization is based on the idea that neuronal networks can be "coordinated anew on each occasion" (p. 331; Friederici & Singer, 2015).

Pragmatically, evidence that long-term intensive musical training is associated with anatomical and functional changes in various brain regions involved in speech processing (e.g., left planum temporale; Elmer et al., 2012; right planum temporale and bilateral planum polare, Angulo-Perkins et al., 2014; Norman-Haignere et al., 2015) should encourage the use of musical training in the classroom and of music therapy as a remediation method for patients with language deficits. More generally, if common activation for music and speech also reflects the activity of higher-order cognitive networks sustaining attention, memory (working memory, short-term and long-term memory), executive functions (inhibition, updating and switching), emotion and motivation, musical training can also be beneficial for patients with attention and memory problems.

In sum, cognitive neuroscience is a very dynamic field and a lot of data are accumulating that sometimes show converging evidence on the functional specialization of brain regions for music and language processing and sometimes show diverging results. The field may be soon ready for a paradigmatic revolution in the Kuhnian sense (Kuhn, 1970), from current static localized networks to dynamic size-varying networks underlying brain high-level functioning.

Acknowledgments

We would like to thank Eduardo Martinez-Montès for relevant comments on this chapter and the Labex Brain and Language Research Institute (BLRI; ANR-11-LABX-0036) for continuous support. ED is supported by a doctoral fellowship from the BLRI and MBa by a doctoral fellowship from the French Ministry of Research and Education.

Notes

1. We use language as a general term but most of the time we refer to speech.
2. See Asaridou and McQueen (2013) for the influence of language experience on music abilities
3. Different tasks are typically used to test for cognitive flexibility (e.g., switching task), inhibitory control (e.g., Stroop task, Simon task; Flanker task; Go/No-Go tasks and stop-signal task), working memory (e.g., backward digit task; n-back task), short-term memory (forward digit span; nonword repetition) and long-term memory (delayed recognition, cued-recall or free recall tests).
4. For instance, /pa1/ low tone with a short vowel means "to find" and /pa:1/ low tone with a long vowel means "forest."

Core Reading

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (in press). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience.*

Friederici, A.D., & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences*, *19*(6), 329–338.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A Meta-Analysis. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.01777

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron*, *88*(6), 1281–1296.

Further References

Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., & Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex (New York, NY: 1991)*, *21*(7), 1507–1518.

Angulo-Perkins, A., Aubé, W., Peretz, I., Barrios, F. A., Armony, J. L., & Concha, L. (2014). Music listening engages specific cortical regions within the temporal lobes: differences between musicians and non-musicians. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *59*, 126–137.

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*(2), 111–130.

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

Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Frontiers in Human Neuroscience*, 7. doi:10.3389/fnhum.2013.00926

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology*, 2. doi:10.3389/fpsyg.2011.00094

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor Readers: A Comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, 7(2), 113–123.

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(3), 1240–1249.

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

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652–2658.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). 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*, *24*(4), 956–967.

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), 4756–4769.

Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: A relation mediated by executive functions. *Music Perception* (2), 195.

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.

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, *221*(1), 331–344.

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, *49*(10), 2812–2821.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, *22*(3), 650–658.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology: CB*, *22*(21), 2059–2062.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, MA: MIT press.

Francois, C., Chobert, J., Besson, M., & Schon, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038–2043.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*(5), 1083–1094.

Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., & Besson, M. (2016). Music and dyslexia: A new musical training method to improve reading and related disorders. *Frontiers in Psychology*, 7. doi:10.3389/fpsyg.2016.00026

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

Jäncke, L., Wüstenberg, T., Scheich, H., & Heinze, H.-J. (2002). Phonetic perception and the temporal cortex. *NeuroImage*, *15*(4), 733–746.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, *47*(2), 735–744.

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, *1252*(1), 100–107.

Kuhn, T. S. (1970). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Liebenthal, E., Desai, R. H., Humphries, C., Sabri, M., & Desai, A. (2014). The functional organization of the left STS: A large scale meta-analysis of PET and fMRI studies of healthy adults. *Frontiers in Neuroscience*, 8. doi:10.3389/fnins.2014.00289

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: Musical expertise enhances the recognition of emotions in speech prosody. *Emotion (Washington, D.C.)*, *11*(5), 1021–1031.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A.D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*(5), 540–545.

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), 2701–2715.

Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*(8), 1858–1866.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science*, *22*(11), 1425–1433.

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*, *19*(3), 712–723.

Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proceedings of the National Academy of Sciences*, *112*(40), 12522–12527.

Patel, A.D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, *308*, 98–108.

Peretz, I., Vuvan, D., Lagrois, M.-E., & Armony, J.L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1664), 20140090–20140090.

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: A longitudinal study. *Frontiers in Psychology*, 3. doi:10.3389/fpsyg.2012.00572

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. *The Journal of Neuroscience*, *31*(10), 3843–3852.

Sahin, N.T., Pinker, S., Cash, S.S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*(5951), 445–449.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Grigutsch, M., Huppertz, H.-J., Knoesch, T. R., Wellmer, J., Widman, G., Eljer, C.E., Friederici, A.D., & Schulze-Bonhage, A. (2013). Co-localizing linguistic and musical syntax with intracranial EEG. *NeuroImage*, *64*, 134–146.

Schellenberg, E.G. (2011). Examining the association between music lessons and intelligence. *British Journal of Psychology (London, England: 1953)*, *102*(3), 283–302.

Schulze, K., Zysset, S., Mueller, K., Friederici, A.D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*(5), 771–783.

Slevc, L.R., & Miyake, A. (2006). Individual differences in second-language proficiency does musical ability matter? *Psychological Science*, *17*(8), 675–681.

Wong, P.C.M., & Perrachione, T.K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4).

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), 420–422.

Zaehle, T., Wüstenberg, T., Meyer, M., & Jäncke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. *European Journal of Neuroscience*, *20*(9), 2447–2456.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLOS ONE*, *9*(6) e99868. doi:10.1371/journal.pone.0099868

CHAPTER 10

ONE STEP BEYOND: MUSICAL EXPERTISE AND WORD LEARNING

STEFAN ELMER, EVA DITTINGER,
AND MIREILLE BESSON

10.1 A Cortical Framework of Speech and Language Processing

10.1.1 The Faculty of Language and its Neural Substrate

Speech and language processing constitute a uniquely human faculty that can be distinguished from other forms of communication in the animal kingdom. In fact, even though it has been proposed that different species possess the faculty of language in a broad sense (i.e. mimic, gesture, olfactory cues, etc.), only human language is characterized by a recursive structure (Fitch, 2010), the latter referring to the ability to produce an infinite number of phrases from a finite number of entities (i.e. phonemes and words). Until now, different evolutionary theories have attempted to explain the possible origin of speech and language processing in human beings, ranging from the expansion of brain size relative to body weight, genetics, brain asymmetries, anatomical characteristics of the larynx, and mirror neurons, to cultural and societal aspects (among others) (Fitch, 2010; Fitch & Reby, 2001; Hauser, 2002). From a linguistic perspective, speech and language processing can be subdivided into different subsystems including phonology, syntax, semantics, morphology, and pragmatics. Phonology refers to knowledge of the sound structure, syntax deals with the rules governing the combination and the order of words in a sentence, semantics addresses the meaning of single words and sentences, morphology is concerned with the structure of words, whereas pragmatics examines language in contexts (e.g. discourse, inference, interaction). Furthermore, depending on intonations and stresses, semantics can be influenced.

In the last decades, both lesion studies and neuroimaging techniques have fundamentally contributed to a better understanding of the cortical organization of speech and language processing. For more than a century, the classical Wernicke-Lichtheim-Geschwind model (Boland, 2014) was considered an accurate representation. This simplistic model was purely based on lesion studies with patients suffering from receptive (i.e. Wernicke's), expressive (i.e. Broca's), or conduction aphasia, and postulated that Broca's area is crucial for language production, whereas Wernicke's area subserves language comprehension functions. In addition, the arcuate fasciculus, a major fibre bundle connecting posterior superior temporal regions (i.e. Brodmann's area 42, Wernicke's area) with Broca's region (i.e. Brodmann's area 44 and 45, pars opercularis and triangularis), was recognized to mediate information exchange between these areas. Even though this historical neurological model enabled the description of a variety of aphasic symptoms, recognizing the contribution of left-sided perisylvian areas to perception and articulation, it is reductive, oversimplified, and relies on brain lesions instead of healthy functioning.

Since that time, several branches of research have fundamentally contributed to improve and ameliorate the cartography of speech and language processing in the human brain. In analogy to the visual system (Milner & Goodale, 2008; Miskin & Ungerleider, 1982), current models conjointly postulate the involvement of two parallel, bidirectional, and hierarchically-organized processing streams stretching from the auditory-related cortex toward the temporal pole (ventral stream) and the frontal lobe (dorsal stream) (Rauschecker & Scott, 2009), and meshing at two points of convergence, namely in the posterior supratemporal plane and in the ventral part of the frontal cortex (see Figure 10.1). However, these models diverge somewhat in the description of the processes supported by the two processing streams, as well as in their linguistic and neurological conceptualization. In turn, we will summarize some of the most popular frameworks of cortical speech and language processing proposed by Hickok and Poeppel (2007), Bornkessel-Schlesewski and Schlesewski (2013), Friederici (2009, 2011, 2012), and Hagoort (2014). It is important to

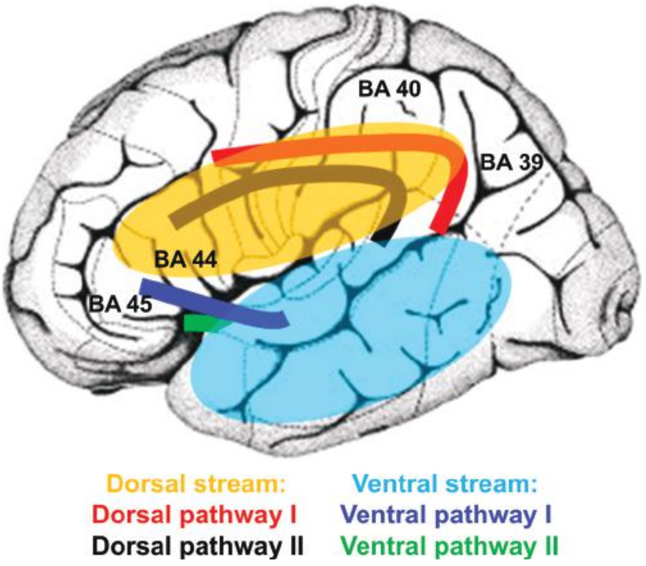


FIGURE 10.1 Ventral and dorsal streams of language processing.

Table 10.1 Summary of current models of cortical speech and language processing proposed by Hickok and Poeppel (2007), Bornkessel-Schlesewski and Schlesewski (2013), Friederici (2009, 2011, 2012), and Hagoort (2014)

Authors	Type of model	Short description
Hickok and Poeppel	Dual-stream model	The ventral stream is responsible for sound-to-meaning mapping, whereas the dorsal one supports sensory-motor mapping mechanisms and articulation. The ventral stream is bilaterally organized, whereas the dorsal one is lateralized to the left hemisphere.
Bornkessel-Schlesewsky and Schlesewsky	Dual-stream model	The ventral stream operates in a time-independent manner by activating and unifying conceptual schemata necessary for creating units of increased complexity and enabling semantic integration. Otherwise, the dorsal stream subserves general time-dependent processes and is engaged in segmenting the input into prosodic words, combining these elements into a syntactic structure, as well as in assessing them into actioncentred representations.
Friederici	Dual-stream model	This model postulates a dorsal and a ventral processing stream that is compatible with the underlying white matter architecture. Within this framework, the dorsal and ventral streams are responsible for more than one isolated function, and based on the underlying white matter architecture, each of them can be segregated into two subpathways (i.e. pathway I & II).
Hagoort	Dynamic and cognitive model	This model acts on the assumption that for central aspects of language processing the neural infrastructure is shared between comprehension and production systems in the form of dynamic networks, and that this neural substrate is not language- specific. Frontal, temporal, and parietal brain regions are differentially recruited based on task-related network characteristics, meaning that the functional role of a specific brain area is influenced by the other regions of the network depending on information type, processing demands, and cognitive control.

mention that, besides these models (outlined in Table 10.1), there are a multiplicity of other frameworks that are not discussed in this chapter.

10.1.2 Current Models of Speech and Language Processing

The dual-stream model proposed by Hickok and Poeppel (2007) relies on evidence from both lesion- and task-related neuroimaging studies, and postulates a ventral stream that processes speech signals for comprehension and a dorsal one that maps acoustic speech signals to frontal lobe articulatory networks. When speech signals reach the auditory-related cortex, the model already postulates a division of labour between the two hemispheres (Giraud et al., 2007; Zatorre & Belin, 2001), and highlights a relative specialization of the left hemisphere for the processing of transient and fast-changing acoustic cues (i.e. segmental, time windows of about 25 ms, 40 Hz). By contrast, the right-sided counterpart is more sensitive to slow acoustic modulations and frequency information (i.e. suprasegmental, time windows of about 250 ms, 4 Hz). Even though this relative processing asymmetry is controversial (Santoro et al., 2014; Overath et al., 2015), it has previously been associated with a differential spacing between microcolumns, myelination (Harasty et al., 2003; Seldon, 1981), as well as with asymmetric spontaneous neural oscillations in the theta (right-sided asymmetry) and gamma (left-sided asymmetry) frequency range (Giraud et al., 2007). These specific oscillations have been proposed to play an important role in 'packing' the multitime speech signal (i.e. phonemes, words, etc.) into units of the appropriate temporal granularity (Giraud & Poeppel, 2012) necessary for further processing steps along the ventral and dorsal streams.

In the Hickok and Poeppel (2007) model, the speech signal, after acoustic analysis, spreads along a bilaterally distributed ventral route engaged in meaning extraction (i.e. lexical-semantic processing) through a cascade of hierarchical processes. The posterior supratemporal plane and the posterior part of the superior temporal sulcus (STS) act as an interface between spectrotemporal and phonological processes. At the next hierarchical level, phonological information is mapped onto lexical representations in memory by recruiting posterior, middle, and inferior portions of the temporal lobe. Finally, information converges on a combinatorial network situated in the proximity of the temporal pole and supporting lexical-semantic integration, sentence-level processing, as well as syntactic and semantic nesting. In contrast to the bilateral organization of the ventral route (with a slight left-hemispheric bias), the dorsal stream is strongly left-dominant, maps sensory and phonological representations onto articulatory motor representations in the frontal cortex, and constitutes a bridge between the speech perception and production systems. The dorsal pathway originates from the posterior supratemporal plane, runs through a sensory-motor interface at the parieto-temporal junction, and projects to Broca's area as well as to the dorsal part of the premotor cortex. The model proposed by Hickok and Poeppel is anatomically well-defined. However, it only focuses on speech and not generally on language processing, and lacks a certain linguistic depth, especially regarding syntactic and lexical-semantic processes.

Recently, Bornkessel-Schlesewsky and Schlesewsky (2013) presented an alternative dual-stream model that attempts to unify neurobiological assumptions and linguistic sentence comprehension. Similar to Hickok and Poeppel, the authors postulate the engagement of bidirectional ventral and dorsal processing streams, however with important computational differences regarding time dependence. The ventral stream projects from auditory core areas along the superior temporal plane toward the anterior temporal cortex and ventral frontal cortex, whereas the dorsal stream runs, via a relay station situated in the inferior parietal lobe, to the inferior frontal cortex. In their framework, the authors abstain from a conceptual dichotomy between comprehension and articulation and instead propose time-dependent processes along the dorsal stream and time-independent ones along the ventral stream. Thereby, it is assumed that the ventral stream operates in a time-independent manner by activating and unifying conceptual schemata necessary for creating units of increased complexity and enabling semantic integration. Furthermore, the model postulates that the ventral stream enables word-level semantic information, as well as phrase-structure comprehension, by the activation and unification of actor-event schemata (e.g. who and what) that are actor-centred (i.e. focus on persons or objects responsible for a certain events) and category neutral (i.e. nouns or verbs). Unification occurs by integrating one schema (e.g. who) into another one (e.g. what). According to the same model, the dorsal stream does not specifically subserve articulation and repetition but, rather, general time-dependent processes. Therefore, this stream supports both speech production and comprehension and is engaged in segmenting the input into prosodic words, combining these elements into a syntactic structure, as well as in assessing them in actioncentred representations (i.e. who is responsible for a certain event).

Finally, it is important to mention that even though the two processing streams converge in the ventral part of the frontal cortex, this brain region is not assumed to support specific linguistic functions but, rather, accommodates action planning and general executive functions such as verbal and non-verbal memory, inhibitory control, switching, and updating. This framework fundamentally contributes to a better understanding of the cortical implementation of linguistic processes (especially phonology, semantics, and syntax). Otherwise, the model is anatomically vaguely defined and does not explicitly address processing asymmetries along the ventral and dorsal streams.

An alternative dual-stream model is the one proposed by Friederici (2009, 2011, 2012). This model is based on sentence processing, is anatomically as well as linguistically well-defined, and reconciles precise cortical cartography with the underlying white matter pathways. Within this framework, the dorsal and ventral streams are responsible for more than one isolated function, and based on the underlying white matter architecture, each of them can be segregated into two subpathways. The dorsal stream constitutes one fibre bundle connecting the superior temporal cortex to the premotor cortex (via the superior longitudinal fasciculus, dorsal pathway I) as well as by a second pathway linking the temporal cortex to pars opercularis (via the arcuate fasciculus, dorsal pathway II). By contrast, the ventral stream relies on the fibre bundle running from the anterior temporal cortex to pars triangularis (via the extreme capsule, ventral pathway I) as well as on the connection between the anterior supratemporal regions and the frontal operculum (via the uncinate fasciculus, ventral pathway II) (see Figure 10.1).

Initial acoustic and phonological analyses involve the primary and secondary auditory cortex, from where activity spreads along the ventral stream to anterior and posterior supratemporal regions. In successive processing steps, initial phrase structure building is conjointly analysed by the left anterior temporal cortex and the frontal operculum (i.e. ventral pathway II). Successively, semantic, grammatical, and thematic relations are processed in a parallel manner. Semantic analyses are supported by middle-posterior areas of the superior and middle temporal gyrus, as well as by pars triangularis and orbitalis (i.e. ventral pathway I). Otherwise, syntactically complex sentences are dependent upon pars opercularis and the posterior temporal cortex (i.e. dorsal pathway II). Furthermore, within this framework, the anterior part of the temporal lobe supports both semantic and syntactic processing and subserves combinatorial processes. Finally, the connection between the posterior supratemporal plane and the premotor cortex (i.e. dorsal pathway I) promotes auditory-to-motor mapping mechanisms, whereas the fibre bundle bridging the posterior temporal cortex and pars opercularis (i.e. dorsal pathway I) supports syntactic processes (especially when sentences are complex). Prosodic information is assumed to be predominantly processed in the right hemisphere and integrated with left-hemispheric syntactic information through the posterior part of the corpus callosum (i.e. isthmus).

The model proposed by Hagoort (2014) overlaps with the models described above in that it posits the engagement of temporal, parietal, and frontal brain regions as the constitutional entities of speech and language processing. However, this model acts on the assumption that for central aspects of language processing, the neural infrastructure is shared between comprehension and production systems in the form of dynamic networks, and that this neural substrate is not language-specific (see also Friederici & Singer, 2015). Within this framework, brain regions situated along the ventral and dorsal streams are differentially recruited based on task-related network characteristics, meaning that the functional role of a specific brain area is influenced by the other regions of the network depending on information type (i.e. phonological, syntactic, and semantic), processing demands, and cognitive control mechanisms.

This perspective has been implemented in a 'Memory, Unification, and Control' (MUC) model that postulates that regions situated in distributed networks in the temporal and inferior parietal cortex generally subserve mnemonic representations (i.e. phonological word forms, morphological information, and the syntactic templates associated with nouns, verbs, and adjectives). Otherwise, frontal regions (including Broca's area) are crucially involved in unification operations by generating larger structures (i.e. phonologic, semantic, and syntactic) from the templates retrieved from memory. Finally, the model also posits that 'memory' and 'unification' mechanisms are hierarchically subordinated to higher executive control mechanisms that are executed by the dorsolateral prefrontal cortex, the anterior cingulate cortex, as well as by the parietal attention system. This model is compatible with studies on non-human primates militating that diverse perceptual and cognitive functions are based on similar neural mechanisms, leading to suggestions of a general rather than a language-specific intrinsic organization of the human brain (Rauschecker & Scott, 2009; Friederici & Singer, 2015).

10.2 PERCEPTUAL AND COGNITIVE DEMANDS ON SPEECH- AND LANGUAGE-LEARNING MECHANISMS

The investigation of speech- and language-learning mechanisms in infants provides an empirically-based framework for better comprehending the perceptual and cognitive processing underlying this uniquely human faculty. Currently, it is known that the auditory cortex starts functioning at about 24 weeks of gestation and that shortly after birth, infants are characterized by some left-hemispheric language specialization (Dehaene-Lambertz & Spelke, 2015; Perani et al., 2011). These functional-anatomical constraints, in association with neural commitment (Kuhl, 2004), facilitate language-learning mechanisms in infants, including the perception and discrimination of vowels, consonants, phonetic contrasts, and stress patterns in words. Nowadays, it is also recognized that a part of the dorsal stream (i.e. dorsal pathway I) linking the auditory cortex with the premotor cortex is observable shortly after birth, whereas a second pathway connecting the auditory cortex with Broca's area (i.e. dorsal pathway II) matures only later during development (Brauer et al., 2013).

Speech constitutes a concatenated acoustic signal whose parameters (e.g. pitch and envelope) not only vary in time but also depend on the talker (e.g. gender), speech rate, as well as on the context (e.g. loudness of the environment). Consequently, before acquiring single words, infants have to learn to decrypt the 'speech code' by figuring out the composition of the phonetic categories of a specific language. This mechanism, called 'categorical perception', is (at least partially) mediated by the 'magnet effect', a phenomenon where prototypical phonetic representations stored in memory attract surrounding deviant sounds (Kuhl, 2004). A commonplace example of this effect can be observed in English infants who learn that /r/ and /l/ pertain to different phonetic categories, whereas Japanese children treat these two phonemes as equivalent because they are not lexically contrastive in Japanese.

Aside from categorical perception, a further important phenomenon is speech segmentation (i.e. the ability to extract meaningful sounds from continuous speech). The recognition of words' boundaries is at least partly based on 'statistical learning'—an implicit faculty that enables infants to analyse statistical distributions and relationships between speech sounds (Kuhl, 2004; Saffran, 2003). For example, within the German language it is more probable (i.e. there is a higher transitional probability) that the consonant /r/ follows the consonants /t/ and /p/ than /z/. Importantly, infants also strongly rely on prosodic cues (e.g. linguistic stress on the first syllable in German) for the segmentation of a continuous speech signal into different subunits and to identify potential word candidates.

Certainly, speech- and language-learning mechanisms are also strongly influenced by social factors that enable infants to be attracted to infant-directed speech (e.g. motherese) by providing enriched referential information through action-based forms (Kuhl, 2007). In this context, there is evidence showing that language-learning mechanisms rely on the functional contribution of phylogenetically older subcortical reward systems (Péron et al., 2016) possibly involved in reinforcing human motivation to learn a language (Ripollés et al., 2014).

One of the most distinctive differences between infants' and adults speech and language acquisition is that, in the latter, several of the processes described above are established and the brain has already committed to the mother tongue. In fact, in adults, the neural circuits underlying speech processing are fully developed and prefrontal brain regions supporting higher cognitive functions (i.e. attention, memory, planning, inhibition, etc.) and explicit learning strategies have reached a maturational ceiling (Gogtay et al., 2004). On the other hand, a common experience, such as being exposed to a new language in a foreign country, brings to light several analogies between speech and language acquisition in infants and adults. Similar to infants, adults have to learn to recognize the phonetic repertoire of a foreign language as well as to segregate continuous speech into subunits in order to recognize words' boundaries and to identify single words. Depending on the phonetic overlap between native and foreign language (e.g. Indo-European, Asian, or neo-Latin), the acquisition process can be either facilitated or hindered. A further point is that not only the phonetic properties of a language but also its spectrotemporal attributes, as well as its syntactic complexity, have an influence on the learning process. This is, for example, the case for tonal (e.g. Mandarin or Cantonese Chinese) and quantitative languages (e.g. Finnish) where phonemes vary in pitch (i.e. rising or falling), temporal extension of the vowel (i.e. short or long), or even by a combination of such spectral and temporal attributes that contribute to differences in word meaning (e.g. Thai).

Currently, there is a significant amount of literature addressing differential aspects of speech- and language-learning mechanisms in both adults and children, ranging from the articulation of foreign speech sounds, categorical perception, speech segmentation, and word learning, to the implicit or explicit acquisition of syntactic knowledge. Even though all these studies cannot be discussed in detail here, it is important to emphasize that the neural circuits underlying different aspects of language-learning mechanisms are the same as those described in Section 10.1, 'A Cortical Framework of Speech and Language Processing'. Therefore, we will only provide a few examples of some of these studies.

Recently, López-Barroso and colleagues (2013) measured a sample of adult participants who performed an artificial language-learning task consisting of segmenting and learning single pseudo-words presented in the form of concatenated speech. Results demonstrated that word-learning ability was related to increased functional and structural connectivity between the left auditory cortex and Broca's region (dorsal pathway). In other studies, Golestani and colleagues reported that French participants who more accurately learnt to discriminate (Golestani and Zatorre, 2004) or pronounce (Golestani and Pallier, 2007) non-native phonetic contrasts were characterized by increased brain activity in left perisylvian areas as well as by enhanced grey-matter density in brain regions supporting speech articulation, respectively. Finally, previous electrophysiological studies on vocabulary learning were able to demonstrate lexical-semantic facilitation effects (i.e. as reflected by increased N400 amplitudes) after only a few hours of training (McLaughlin et al., 2004; Perfetti et al., 2005).

In the next section, we will introduce professional musicians, as well as children undergoing short- or long-term music training, as a vehicle for better understanding the mutual interdependence between perception and cognition during different aspects of speech and language learning. Thereby, we will draw a bridge between functional and structural training-related brain changes, perceptual and cognitive benefits, and several aspects of language learning.

10.3 Music to Speech Transfer Effects

Compared to language research, the neuroscience of music is a relatively new field that has also led to fascinating discoveries. This is at least partly based on the fact that, while all normally-developing children end up being language experts, not all human beings are professional musicians. In fact, much has been learned about the anatomo-functional organization of the brain and about brain plasticity by studying the musician's brain and the impact of long-term music training on different perceptual and cognitive functions (Elbert et al., 1995; Jäncke, 2012; Münte et al., 2002; Schlaug et al., 1995; Sluming et al., 2002; Schneider et al., 2002).

Here, we focus on transfer effects, defined as the influence of training in one domain on the level of performance in another domain. Specifically, we address transfer effects from music training to several aspects of language processing (in a broad sense that includes speech processing), to cross-modal integration, and to executive functions (see Asaridou & McQueen, 2013, for the influence of linguistic experience on music processing). The results described in this section were obtained using different methodologies: behavioural measurements, electrophysiological recordings at the level of the brainstem and at the cortical level (event-relatedpotentials—ERPs), and functional magnetic resonance imaging (fMRI). First, we review the growing evidence for transfer effects; then, we consider two main interpretations of such transfers; and finally, we address the questions of the influence of music training or of genetic predispositions for music, and whether music and language processing rely on shared or distinct neural substrates.

10.3.1 Growing Evidence for Transfer Effects

Results of many experiments have demonstrated the positive influence of music training on speech perception (Besson et al., 2011; Kraus & Chandrasekaran, 2010). For instance, there is clear evidence that music training influences the segmental processing of speech sounds (consonants, vowels, and syllables) (see Appendix 10.1 and Audio 10.1 for examples of consonant-vowel syllables) at multiple levels of the auditory system from the brainstem (Bidelman & Krishnan, 2010; Musacchia et al., 2007; Wong et al., 2007) to cortical regions (Bidelman et al., 2011; Chobert et al., 2014; Bidelman & Alain, 2015; Elmer et al., 2012; Meyer et al., 2012; Ott, 2011). Music training also positively influences pitch processing in tonal languages such as Mandarin Chinese and Thai, in which pitch variations in vowels change the meaning of words (Wong et al., 2007; Bidleman et al., 2011; Alexander et al., 2005; Bidelman et al., 2013; Lee & Hung, 2008). At the suprasegmental level (couple of syllables, words, and sentences), results have shown that musicians are typically more sensitive than non-musicians to linguistic and emotional prosody (i.e. speech melody and rhythm) (Cason & Schön, 2012; Lima & Castro, 2011; Ma & Thompson, 2015; Magne et al., 2006; Marques et al., 2007; Moreno et al., 2009; Schön et al., 2004; Thompson et al., 2004, 2012; and for contrastive results, Trimmer & Cuddy, 2008), as well as to the timbre of human voices (Chartrand & Belin, 2006).

As reviewed in Section 10.2, categorical perception and speech segmentation are the cornerstones of speech perception. In this context, Bidelman and collaborators (2013) demonstrated an influence of music training on the categorical perception of speech sounds (/u/ to /a/ continuum) at the cortical level: the P2 component of the ERPs was sensitive to between-categories' phonetic boundaries defined by psychometric functions. By contrast, this effect was not significant at the brainstem level. Results of subsequent experiments also showed that younger (Bidelman et al., 2014) and older musicians (Bidelman & Alain, 2015) were faster and showed steeper boundaries between phonetic categories in a vowel categorization task than non-musicians. Increased auditory sensitivity may thus be one of the driving forces behind enhanced categorical perception and enhanced speech processing in musicians.

Speech segmentation is also fundamental to speech comprehension. This is clearly exemplified when learning a foreign language that is perceived as a continuous stream of nonsense words. François and colleagues (2013) used a longitudinal approach in children, over a period of two school years, during which the children were trained in music or in painting (45 minutes, twice a week in the first year and once a week in the second year). Children first listened to 5 minutes of an artificial, continuous, sung language in which syllables varied in their transitional probability (as previously described) and was higher within three syllabic items (hence considered as familiar) than between two consecutive items (hence considered as unfamiliar). Children were then asked which of two items was most familiar. At the behavioural level, implicit recognition of familiar and unfamiliar items steadily increased over the course of the two years of music training—but not of painting training. At the cortical level, and similarly to adults (François & Schön, 2011), only the music-trained children were characterized by a fronto-central negative component that was larger to unfamiliar than to familiar items. Thus, this longitudinal study demonstrated that music training improved speech segmentation.

Certainly, transfer from music to language is by no means limited to low-level speech processing, such as categorical perception or speech segmentation, but has also been shown to extend to higher-level language processing. For example, based on the idea that both music and language are structured sequences of events that unfold in time, several studies have investigated the influence of music training on syntactic processing in adults (Fitzroy & Sanders, 2012) and in children (Janus et al., 2016; Jentschke & Koelsch, 2009).. Jentschke and Koelsch (2009) compared the ERPs to violations of linguistic and musical syntax in musically trained and untrained 10–11-year-old children. The electrophysiological markers of both types of violations were larger in the former group. Not surprisingly, musically trained children were more sensitive to harmonic structure than children without such training. What was more surprising is that they also showed more comprehensive knowledge of the syntactic structure of sentences, possibly through faster implicit syntactic processing and/or a more efficient use of the prosodic and rhythmic cues that constrain syntactic constructions (Roncaglia-Denissen et al., 2013; Schmidt-Kassow & Kotz, 2009).

A hotly debated issue in the literature is whether the influence of music training on different abilities is causally linked to music training or rather results from genetic predispositions for music. Cross-sectional studies comparing (professional) musicians and non-musicians, children or adults, do not allow for this issue to be addressed since, as pointed out by Schellenberg (2004), correlation is not causality. To our knowledge, the only way to test causality in humans is to use a test–training–retest longitudinal approach that compares two groups of non-musicians (children or adults)—one group trained with music and the other group trained with an equally interesting activity, such as painting or cooking. Participants are pseudo-randomly assigned to one of the two groups, thereby ensuring that no between-group differences on the different tests of interest are found before training. If musically trained participants outperformed painting-/cooking-trained participants in the retest session, this is evidence that the type of training strongly influenced the results. This approach has been successfully used to demonstrate the influence of music training on the perception of pitch variations in sentence context (Moreno et al., 2009) and on the preattentive processing of the temporal aspects of speech (Chobert et al., 2014) and on speech segmentation (Francois et al., 2013), as previously mentioned.

10.3.2 Interpretations of Transfer Effects

Having summarized some of the evidence for music to language transfer effects, we now turn to the most important question: *How can we explain music to speech processing transfer effects?* Two main interpretations, that we refer to as the cascade and multidimensional hypotheses, have been proposed in the literature.

Following the cascade hypothesis, transfer effects arise because speech and music are auditory signals relying on the same acoustic parameters (i.e. duration, frequency, intensity, and timbre). As musicians are highly trained in perceiving the acoustic structure of sounds, sound encoding is facilitated not only in music but also in speech. Thus, enhanced perceptual encoding and categorization of speech sounds in musicians facilitates higher levels of speech processing. Let us take the example of novel word learning, that will be considered in detail later. If a learner is able to differentiate the subtle acoustic features of different phonemes, he/she may form a more precise phonological representation of the new word. Consequently, it will be easier to associate such a phonological representation with the corresponding word meaning than a less distinct one. This bottom-up interpretation may explain why musicians are more sensitive to the spectro-temporal aspects of speech processing at the segmental and suprasegmental levels, as already reviewed, as well as to other aspects of speech perception (e.g. speech in noise perception, speech segmentation, sentence syntactic structure).

The multidimensional hypothesis is based on the fact that music training is multidimensional. Playing a musical instrument involves auditory and visual perception (the notes on the score), visuo-auditory-motor integration (transforming visual notes into sounds through movements), selective and divided attention (focusing attention on one's own instrument and dividing attention between the different instruments of the orchestra), and motor control (adapting posture and fine distal movements). Playing a musical piece also requires memory (most musicians play by heart), executive functions (switching between visual and auditory codes), inhibitory control (withholding a movement to play at the right moment and updating information), and emotion (as translated into the interpretation of the musical piece). Since professional musicians are at their advantage in these different functions, they may outperform non-musicians when these functions are necessary for the task at hand.

Evidence for this multidimensional hypothesis is accumulating from several results showing transfer from music training to cross-modal integration (Lee & Noppeney, 2011; Pantev et al., 2009; Chen et al., 2008) and to executive functions. Executive functions are defined as top-down processes that control behaviour, and typically include selective attention, working memory (WM), short- and long-term memory, and cognitive control (inhibitory control, cognitive flexibility, updating), although this is still a matter of controversy (Diamond, 2013). In this respect, musicians have been shown to be more efficient at audio-motor learning than non-musicians (Barrett et al., 2013; Lahav et al., 2007; Mathias et al., 2015), possibly because they use different integration strategies, with musicians relying more on auditory and non-musicians more on visual information (Paraskevopoulos et al., 2012, 2014, 2015). In addition, there is evidence that adult musicians outperform non-musicians in WM tasks based on musical stimuli (George & Coch, 2011; Schulze & Koelsch, 2012; Schulze, Mueller, et al., 2011; Schulze, Zysset, et al., 2011; Pallesen et al., 2010; Schulze et al., 2012; Williamson et al., 2010), even though the influence of music training on verbal memory is more controversial (Williamson et al., 2010; Brandler & Rammsayer, 2003; Chan et al., 1998). Importantly, WM, short-term memory, and long-term memory are tightly intertwined, and more work is clearly needed to disentangle the different components of executive functions (Franklin et al., 2008; Jakobson et al., 2008). Research in children also demonstrated that music training can have an influence on executive functions (cognitive flexibility, processing speed, inhibitory control, non-verbal intelligence) as well as on short-term and long-term memory (Janus et al., 2016; Moreno et al., 2011; Bergman Nutley et al., 2014; Roden et al., 2012; Zuk et al., 2014).

In summary, results of a number of studies suggest that music training in adults and in children positively influences several aspects of language processing, executive functions, WM, as well as short-term and long-term memory. Importantly, differences between adult musicians or children with music training and controls are generally larger for the most demanding tasks, when most resources are needed to perform the task at hand (Besson et al., 2011; Diamond, 2013). However, in line with a dynamic and interactive view of human cognition, results also showed that these different functions are intrinsically linked. For instance, Diamond (2013) reported that the training of task-switching abilities transferred to verbal and non-verbal WM, inhibition (Stroop interference), and reasoning tasks. Better understanding of these intricate relationships is an exciting aim of future research, keeping in mind that we need to use well-controlled experimental designs, standardized tests when they are available (e.g. forward digit span for short-term memory, backward digit span for WM), and data-analysis methods that allow controlling for the effects of the many different factors that can also influence the results.

10.3.3 Transfer Effects and Shared Neural Networks

Finally, one current hotly debated topic is whether the music to language and cognition transfer effects already reviewed are supported by *shared neural networks*. In fact, the question of whether brain networks involved in music and language processing are similar or different is a long-standing question that recently received new answers.

Let us take the example of syntactic processing and Broca's area. Using fMRI, early results demonstrated that Broca's area, considered as specifically involved in the processing of linguistic syntactic structures for over a century (Berwick et al., 2013; Friederici et al., 2006), was also activated when processing musical structures (Koelsch et al., 2002; Levitin & Menon, 2003; Maess et al., 2001; Tillmann et al., 2003; Vuust et al., 2006). These results provided evidence that processing syntax in music and language relied on shared neural substrates. This conclusion was further supported by recent results of Abrams et al. (2011) and Rogalsky et al. (2011) showing that similar activations of frontal and temporal regions in both hemispheres are activated by temporal violations in linguistic sentences and melodies. However, results based on a more fine-grained approach— multivariate pattern analysis—showed that the two types of stimuli elicited spatially distinct activity. Thus, based on these results, the authors concluded that temporal structure is encoded differently within the two domains and that distinct cortical networks are activated. Interestingly, brain structures in these networks involved the voice-selective areas identified by Belin and collaborators (2000) and the speech-selective component that emerged from the hypothesis-free voxel decomposition method recently used by Norman-Haignere and colleagues (2015).

In summary, through the example of Broca's and temporal areas, the most studied brain structures in the neuroscience of language, we have seen evidence both for shared and for distinct networks involved in processing syntax and temporal structures in music and language. Importantly, evidence is tightly linked to the specific aspects of music and language that are compared and to the methods chosen for analysis. Depending upon the temporo-spatial resolution of the method, upon the characteristics of the stimuli, and upon the task at hand, results may show overlap of brain regions involved in music and language processing

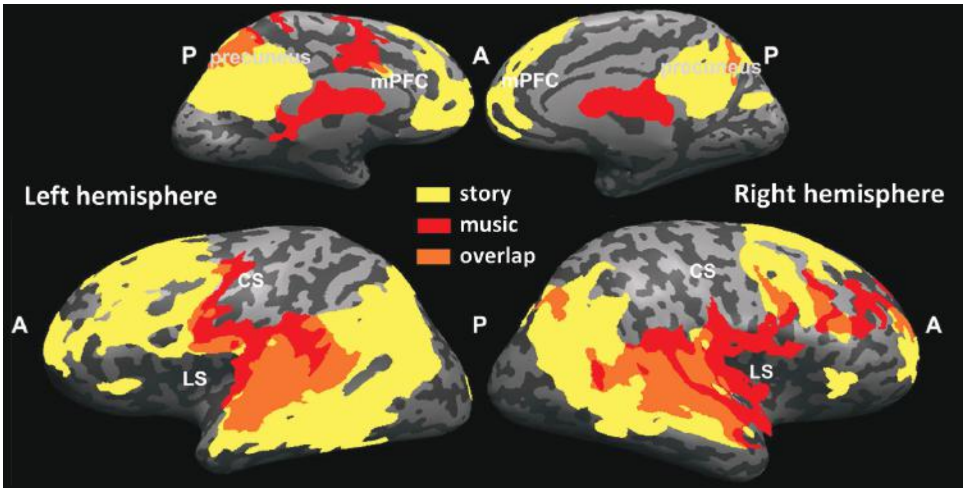


FIGURE 10.2 Maps of brain activation whilst attentively listening to an intact story (yellow) or an intact musical excerpt (red) are shown for experienced pianists. Story listening evoked reliable responses in the temporo-parietal junction, angular gyrus, inferior frontal gyrus, lateral and medial prefrontal areas, and orbitofrontal cortex. Reliable responses to music were found in the lateral sulcus, pre-central gyrus, and middle frontal gyrus. Overlapping regions of reliable responses to both stimuli (orange) were evident in early auditory areas along the superior temporal gyrus (STG). mPFC = middle pre-frontal cortex, A = anterior, P = posterior, CS = central sulcus, LS = lateral sulcus.

Adapted from Farbood M.M. et al,'The neural processing of hierarchical structure in music and speech at different timescales', *Frontiers in Neuroscience*, Volume 9, Issue 157, Copyright © 2015 Farbood et al., doi: 10.3389/fnins.2015.00157, under the terms of the Creative Commons Attribution Licence (CC BY 4.0).

or distinct local networks involved in specific aspects of language and music processing (see Figure 10.2). There is no doubt that tremendous progress in our understanding of the language–music relationship will be made in the years to come by using finer-grained analyses of the spatio-temporal dynamics of brain networks in well-controlled experiments. Moreover, language and music are complex human functions that are not processed independently of other cognitive and emotional functions. As a consequence, we consider the cascade and the multidimensional interpretations of transfer effects as strongly complementary. Finally, as we will see in the following section, the ERPs method, that allows us to continuously record on-line changes in brain activity associated to the stimuli and task at hand, has also provided interesting results regarding the spatio-temporal dynamics of music to language transfer effects.

10.4 Music Training and Word Learning

In previous parts of this chapter, we considered word learning as an example of a multidimensional task relying on both perceptual and cognitive functions. Therefore, in this section, we specifically examine meaning acquisition of novel words, and we illustrate why word learning constitutes a wonderful opportunity to study the influence of music training on semantic processing, one of the key features of language.

When it comes to learning the meaning of novel words, the learner has to focus attention to the stimuli in order to discriminate spectral and temporal phonetic contrasts, build new phonological representations, and associate these representations with meaning by recruiting working-, short-term-, episodic-, and semantic-memory processes. Finally, initial word representations have to be consolidated to build longer-lasting and more robust forms of these representations.

Based on its high temporal resolution, the ERPs method has been frequently used to capture the dynamics of word learning. Thereby, the building up of initial word representations has been shown to be reflected by the rapid emergence of a frontally distributed N400, a negative-going ERP component that develops between 300 and 600 ms after stimulus presentation. For instance, McLaughlin and colleagues (2004) were able to show increased N400 amplitudes in native English speakers after 14 hours of training with French words. Perfetti and colleagues (2005) revealed similar results after only 45 minutes of learning the meaning of low-frequency words.

Finally, in the case of learning a novel word's meaning from highly constraining meaningful contexts, Batterink and Neville (2011) showed the integration of such novel meanings into semantic networks after ten repetitions, Mestres-Missé and colleagues (2007) demonstrated the rapid development of an N400 after only three exposures to such words, and Borovsky and collaborators (2010) even after a single exposure to the words. In conclusion, while learning the meaning of novel words may seem to be slow and laborious, initial word representations can be built up within short training sessions and after only a few repetitions depending on the context in which novel words are presented.

In Section 10.3, we provided evidence for transfer from music to a variety of levels of language processing, including the perception of acoustic-phonetic parameters, segmentation, phonology, and syntax. To go one step beyond, we examined whether professional music training also facilitates semantic processing. We tested the hypothesis that professional music training facilitates word learning, designing an ecologically valid series of experiments aimed at tracking the electrophysiological dynamics of phonological categorization, semantic acquisition, as well as semantic retrieval (Dittinger et al., 2016) (see Figure 10.3). Specifically, two groups of adult French speakers, comprising fifteen professional musicians and fifteen non-musicians, performed first a phonological categorization task, consisting of identifying nine natural Thai monosyllabic words containing either a simple voicing contrast, a tonal, a vowel length, or an aspiration contrast. Importantly, two (/ba/ and /pa/) out of these nine words were part of the French phonemic repertoire and therefore simple to categorize. By contrast, the other seven words contained contrasts which are linguistically irrelevant for French speakers, but relevant for quantitative or tonal languages (i.e. vowel length, pitch, and aspiration contrasts that are lexically relevant in Thai), resulting in more difficult categorization tasks (Dittinger et al., 2018). Following the categorization task, participants learned the meaning of these nine words through picture–word associations during a word-learning phase of about 6 minutes. Then, participants were tested for training success by asking them if a presented picture–word pair matched or mismatched the previously learned association (i.e. matching task). Moreover, to determine whether word learning was restricted to the picture–word pairs learned during the training phase or whether the meaning of the newly learned words was already integrated into semantic networks so that

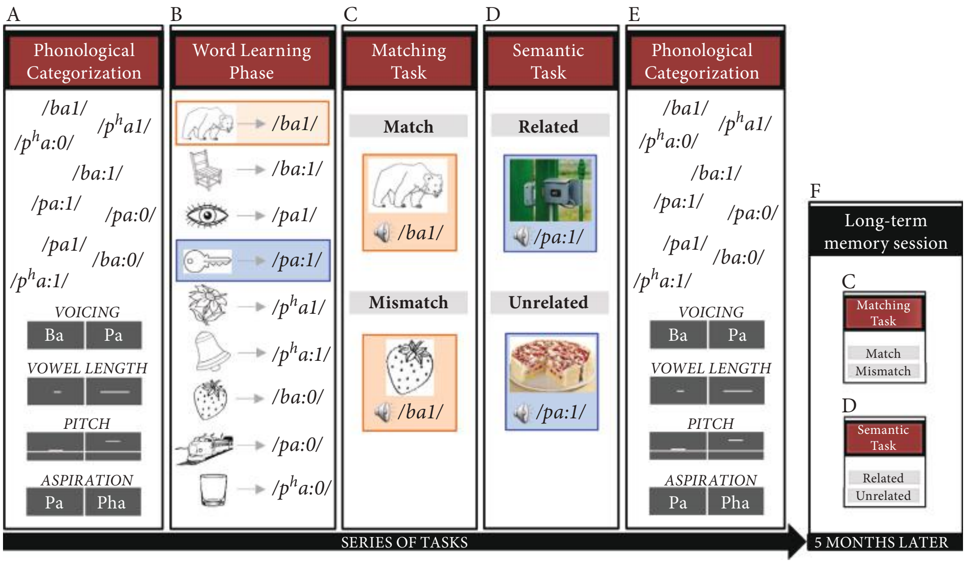


FIGURE 10.3 Participants performed a series of tasks. First, in the phonological categorization task (A), nine natural Thai monosyllabic words had to be categorized based on voicing, vowel length, pitch, or aspiration contrasts. Second, in the word-learning phase (B), each word was paired with its respective picture. Third, in the matching task (C), the words were presented with one of the pictures, either matching or mismatching the previously learned associations. Fourth, in the semantic task (D), the words were presented with novel pictures that were either semantically related or unrelated to the novel words. Fifth, participants again completed the four subtasks of the phonological categorization task (E). Finally, participants came back 5 months after the main session to perform again the matching and semantic tasks (F).

priming effects generalized to new pictures, participants performed a semantic task during which novel pictures that had not been seen in the former task were presented in combination with the previously learned words. They were asked to decide whether the picture and the word were semantically related or unrelated. Finally, participants were behaviourally retested 5 months after the main experimental session, to assess first, how long rapidly installed word representations can last, and second, whether professional music training influences long-term memory—two aspects that had not been investigated before.

The originality of this series of experiments is that EEG was simultaneously recorded in all these tasks (except in the long-term-memory session). This allowed us to follow the temporal dynamics of word learning from the early stages of word categorization and initial word encoding to subsequent stages of word-meaning retrieval once the novel words had been integrated into pre-existing semantic networks. Thus, we aimed at studying different processes underlying word learning that have previously only been explored in isolation in single experiments, to gain a more complete and integrated view of word learning. Since, as already noted, several studies have evidenced a positive influence of music not only on auditory perception, but also on attention, audiovisual integration abilities, as well as memory functions, we expected that professional musicians would be at an advantage to learn these novel words compared to non-musicians.

In line with this hypothesis, results showed that professional musicians learned the meaning of novel words more efficiently than controls, and this result was supported by both behavioural and electrophysiological data. Behaviourally, musicians outperformed non-musicians in word categorization and, as expected, group differences were particularly large for the tonal and aspiration contrasts (Dittinger et al., 2018). While these contrasts are not part of the French phonetic repertoire, the ability to discriminate them is inevitable for the acquisition of several foreign languages. For meaning acquisition, musicians performed similarly to non-musicians in the matching task, but outperformed non-musicians in the semantic task. These results were taken as evidence that musicians had already better integrated the novel words' meanings into semantic networks (i.e. enabling them to generalize the knowledge to novel pictures). Importantly, these behavioural results were supported by group differences in electrophysiological markers. In line with the development of a frontal N400 during novel-word encoding (McLaughlin et al., 2004; Perfetti et al., 2005; Batterink &Neville, 2011; Mestres-Misse et al., 2007; Borovsky et al., 2010, 2012), all participants showed enhanced N400s over frontal scalp sites after the first half (i.e. only 3 minutes) of the word-learning phase. However, only musicians showed additional N400 increases over left centro-parietal scalp regions after the second half of the word-learning phase, suggesting that musicians were faster in encoding word meaning and integrating novel words into existing semantic networks.

During the test phase (i.e. matching and semantic tasks), musicians were characterized by a typical centro-parietal N400 effect (Kutas & Federmeier, 2011) resulting from larger N400 amplitudes for unexpected (i.e. mismatching or unrelated) than for expected (i.e. matching or related) conditions (see Figure 10.4). By contrast, the N400 was still frontally distributed in non-musicians. In summary, while both groups showed the typical electrophysiological marker of word learning (i.e. the frontal N400) during the learning phase, only musicians showed semantic priming effects during the test phase that were similar to those typically found for known words. In line with this conclusion, a correlation between musical aptitudes and the amplitude of the semantic N400 effect was found for musicians, but not for non-musicians, thereby clearly pointing to a relationship between musicality and word learning. Finally, in the behavioural retest 5 months after the main session, musicians remembered more words compared to non-musicians, thereby showing evidence for long-lasting word representations and an influence of music training on verbal long-term memory.

In a second step, data from this word-learning experiment was reanalysed by means of functional connectivity (Dittinger et al., 2017). Functional connectivity is defined as the statistical association or dependency among two or more anatomically distinct functional time series (Friston et al., 1996). Functional connectivity is a useful method for studying functional relationships between regions as a function of expertise. In the first part of this chapter, we reviewed some of the current models of speech processing that converge on the view that two main processing streams, the ventral (i.e. sound-to-meaning mapping) and the dorsal (i.e. sound-to-articulation mapping) pathways, are involved in speech processing (Friederici, 2009). Specifically, the dorsal pathway relies on a fibre tract corresponding to the superior longitudinal fasciculus (SLF), sprawling from inferior parietal and superior-posterior temporal brain regions towards Broca's area and promoting auditory-to-motor mapping mechanisms. Recently, Lopez-Barroso and colleagues (2013) showed that word learning was correlated with the strength of functional and structural connectivity between Broca's and Geschwind's territory (including the angular gyrus (AG) and the supramarginal

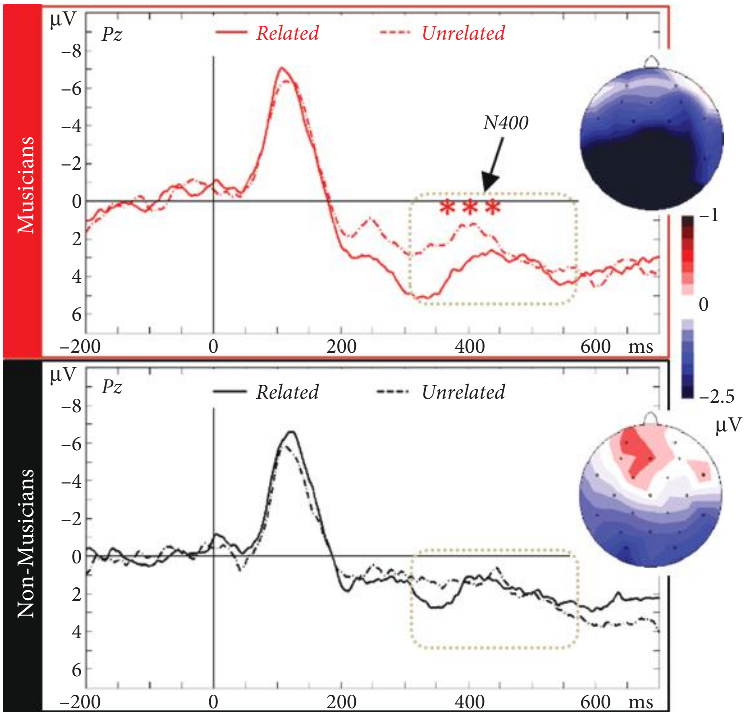


FIGURE 10.4 Semantic task. ERPs recorded at parietal sites (Pz) are overlapped for semantically related (solid lines) and unrelated (dotted lines) words, separately for musicians (red) and non-musicians (black). Time in milliseconds is in abscissa, the amplitude of the effects in microvolts is in ordinate, time zero corresponds to word onset, and negativity is plotted upwards. The grey dotted rectangles represent the typical N400 latency window, and the level of significance of the related vs unrelated difference in the two groups is represented by asterisks (with $^{***}p < 0.001$). Topographic voltage distribution maps of the unrelated minus related differences in musicians and non-musicians are illustrated for the N400 component and voltage values are scaled from −2.5 to +1.0 μV.

gyrus (SMG)) in the left hemisphere. Moreover, there is evidence that the functional-structural architecture of the left dorsal processing stream is influenced by professional musical training (Halwani et al., 2011; Klein et al., 2016; Oechslin et al., 2009).

Based on these results, we investigated functional connectivity between AG/SMG (region of interest (ROI) 1, Brodmann area (BA) 39/40) and Broca's area (ROI 2, BA 44/45) in the three tasks related to word learning previously described (i.e. word learning phase, matching tasks, and semantic tasks) and compared patterns of connectivity between musicians and the non-musician controls. Specifically, we evaluated non-linear functional connectivity by using lagged coherence, that is, a measure of the variability of time differences between two signals (e.g. coming from ROI 1 and from ROI 2) in a specific frequency band (Lehmann et al., 2006; Thatcher, 2012). We focused on theta (4–7 Hz) oscillations based on previous literature evidencing that theta reflects neuronal communications over long-range circuits and is a reliable frequency band to examine mnemonic processes (Ward, 2003).

As expected, results revealed increased left-hemispheric functional connectivity in musicians compared to controls, but only in the semantic task. In addition, this increased connectivity was correlated with the cumulative number of training years. Results were interpreted as showing facilitated feed-forward and feed-backward exchanges between AG/SMG and Broca's area in musicians, thereby facilitating the rehearsal and learning of novel words in musicians. These results are in line with previous results (López-Barroso et al., 2013; Klein et al., 2016) and with the group differences in ERPs and behaviour described previously. Furthermore, the data indicates a relationship between the musicians' superiority in word learning and the temporal alignment of neural oscillations in the theta frequency band in the left dorsal stream.

Taken together, ERPs and functional connectivity revealed two main findings. First, word learning was reflected in the spatio-temporal dynamics of the N400 component: while initial word learning was reflected by frontally increasing N400s, centro-parietal N400s only developed once the novel words started being integrated into the pool of well-known words. This difference in scalp distribution may reflect different cognitive processes and clearly points to distinct neural generators. For instance, frontal N400s to novel words are compatible with results showing that prefrontal and temporal brain regions are associated with the maintenance of novel information in working or short-term memory (Hagoort, 2014) and with the initial building up of episodic memory traces (Rodriguez-Fornells et al., 2009). By contrast, centro-parietally distributed N400s are in line with results showing that semantic representations are possibly stored in the left inferior parietal cortex (Catani & Ffytche, 2005) and/or in the left temporal lobe (Geranmayeh et al., 2015). Furthermore, our experimental design allowed us to track the fast spatio-temporal dynamics of word learning that were characterized by a shift in N400 distribution from frontal to parietal networks after only 3–6 minutes of training. These results open new perspectives for further research on brain plasticity and word learning. In addition, we provide the first behavioural evidence for the longevity of these rapidly established word representations, highlighting that even rapid brain plasticity can have long-lasting consequences.

Second, our results revealed that word learning was facilitated by professional music training. How can we account for such a transfer and what could be the implications of these results? As already mentioned in Section 10.3.2, two main interpretations have been proposed to explain why musicians learn novel words more efficiently than non-musicians. The first one, in terms of cascading effects, claims that enhanced auditory perception facilitates word learning in musicians. Such an interpretation would be in line with bottom–up accounts of transfer effects. Support for this interpretation was provided by Wong and Perrachione (2007) and by Cooper and Wang (2012) who showed that both tone pitch identification and musical aptitudes were significantly correlated with word-learning success in adult English speakers. To directly test for the causality of these effects, Cooper and Wang (2013) trained English non-musicians on the perception of Cantonese tones, and results demonstrated that enhanced perception at the tone level significantly improved word learning. By contrast, the multidimensional interpretation acknowledges potential top–down influences on word learning, as well as interactions between the acoustic properties of sounds, task demands, and expertise of the listener.

While our experimental design does not allow for the disentangling of the cascade and multidimensional accounts, the present results clearly reveal that music training influences the semantic level of language processing, thereby going one step beyond previous studies on transfer effects. Certainly, further studies are needed to replicate these results, possibly with children and older adults, and to disentangle the respective contribution of perceptive and cognitive functions to word learning, thereby possibly lifting the exciting secret of why musicians seem to be better at learning novel words.

10.5 Conclusion

The issue of music to speech transfer effects has generated great interest in the scientific community, as well as in the lay public, probably because music and speech are fascinating domains. The multidimensional aspects of music and speech—how they both rely on perceptual, cognitive, emotional, and motor processes through multiple interactions—are important new avenues for future research. Similarly, much more needs to be done to fully understand how these two abilities are implemented in the brain and whether they rely on shared or distinct neural resources. Based on the current state of knowledge, it is clear that results are tightly linked to the tasks and stimuli that are presented and to the methods that are used for data analysis. Also clear is that music exerts a profound influence on the brain's structural and functional organization, thanks to brain plasticity. It is worth noting in this respect that recent results demonstrated changes in brain electric activity in less than 3 minutes in a novel word-learning task (Dittinger et al., 2017). Taken together, these results open exciting new perspectives for the rehabilitation of patients (children, young adults, and older adults) with various neurological or psychiatric deficits. In this respect, music training may have a strong societal impact.

Acknowledgements

We are grateful for the help provided by the Labex BLRI (ANR-11-LABX-0036), supported by the French National Agency for Research (ANR), under the programme 'Investissements d'Avenir' (ANR-11-IDEX-0001-02). ED is supported by a doctoral fellowship from the BLRI. The authors declare no competing financial interests.

References

Abrams, D. A. *et al.* (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cereb. Cortex* **21**, 1507–1518.

Alexander, J. A., Wong, P. C., & Bradlow, A. R. (2005). *Lexical tone perception in musicians and non-musicians*. Paper presented at 9th European Conference on Speech Communication and Technology, Lisbon, Portugal (pp. 397–400).

Asaridou, S. S. & McQueen, J. M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. *Front. Psychol.* **4**, article 321.

Barrett, K. C., Ashley, R., Strait, D. L., & Kraus, N. (2013). Art and science: how musical training shapes the brain. *Front. Psychol.* **4**, article 713.

Batterink, L. & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *J. Cogn. Neurosci.* **23**, 3181–3196.

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

Bergman Nutley, S., Darki, F., & Klingberg, T. (2014). Music practice is associated with development of working memory during childhood and adolescence. *Front. Hum. Neurosci.* 7, 926.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* **17**, 89–98.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Front. Psychol.* **2**, article 94.

Bidelman, G. M. & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J. Neurosci.* **35**, 1240–1249.

Bidelman, G. M. & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* **1355**, 112–125.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain Cogn.* 77, 1–10.

Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage* **79**, 201–212.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur. J. Neurosci.* **40**, 2662–2673.

Boland, J. E. (2014). A history of psycholinguistics: the pre-Chomskyan era. *Historiographia Linguistica* 41(1), 168–175.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2013). Reconciling time, space and function: a new dorsal-ventral stream model of sentence comprehension. *Brain Lang.* **125**, 60–76.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Lang. Learn. Dev.* **8**, 278–302.

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition* **116**, 289–296.

Brandler, S. & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychol. Music* **31**, 123–138.

Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain Lang.* **127**, 289–295.

Cason, N. & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia* **50**, 2652–2658.

Catani, M. & Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain J. Neurol.* **128**, 2224–2239.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature* **396**, 128.

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

Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *J. Cogn. Neurosci.* **20**, 226–239.

Chobert, J., Francois, C., Velay, J.-L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb. Cortex* **24**, 956–967.

Cooper, A. & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *J. Acoust. Soc. Am.* **134**, EL133–EL139.

Cooper, A. & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.* **131**, 4756–4769.

Dehaene-Lambertz, G. & Spelke, E. S. (2015). The infancy of the human brain. *Neuron* **88**, 93–109.

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* **64**, 135–168.

Dittinger, E. *et al.* (2016). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *J. Cogn. Neurosci.* 28(10), 1584–1602.

Dittinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2017). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Hum. Brain Mapp.* **39**, 722–734.

Dittinger, E., D'império, M., & Besson, M. (2018). Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *Eur. J. Neurosci.* 1–13. doi: 10.1111/ejn.13939

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science* **270**, 305–307.

Elmer, S., Meyer, M., & Jancke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb. Cortex* **22**, 650–658.

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., & Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Front. Neurosci.* **9**, 157.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T. & Reby, D. (2001). The descended larynx is not uniquely human. *Proc. R. Soc. B Biol. Sci.* **268**, 1669–1675.

Fitzroy, A. B. & Sanders, L. D. (2012). Musical expertise modulates early processing of syntactic violations in language. *Front. Psychol.* **3**, 603.

Francois, C., Chobert, J., Besson, M., & Schon, D. (2013). Music training for the development of speech segmentation. *Cereb. Cortex* **23**, 2038–2043.

Francois, C. & Schön, D. (2011). Musical expertise boosts implicit learning of both musical and linguistic structures. *Cereb. Cortex* **21**, 2357–2365.

Franklin, M. S. *et al.* (2008). The effects of musical training on verbal memory. *Psychol. Music* **36**, 353–365.

Friederici, A. D. (2009). Pathways to language: fiber tracts in the human brain. *Trends Cogn. Sci.* **13**, 175–181.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiol. Rev.* **91**, 1357–1392.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.* **16**, 262–268.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 2458–2463.

Friederici, A. D. & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends Cogn. Sci.* **19**, 329–338.

Friston, K. J., Frith, C. D., Fletcher, P., Liddle, P. F., & Frackowiak, R. S. (1996). Functional topography: multidimensional scaling and functional connectivity in the brain. *Cereb. Cortex* **6**, 156–164.

George, E. M. & Coch, D. (2011). Music training and working memory: an ERP study. *Neuropsychologia* **49**, 1083–1094.

Geranmayeh, F., Leech, R., & Wise, R. J. S. (2015). Semantic retrieval during overt picture description: left anterior temporal or the parietal lobe? *Neuropsychologia* **76**, 125–135.

Giraud, A.-L. & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517.

Giraud, A.-L. *et al.* (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127–1134.

Gogtay, N. *et al.* (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8174–8179.

Golestani, N. & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cereb. Cortex* **17**, 929–934.

Golestani, N. & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *NeuroImage* **21**, 494–506.

Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Curr. Opin. Neurobiol.* **28**, 136–141.

Halwani, G. F., Loui, P., Rüber, T., & Schlaug, G. (2011). Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. *Front. Psychol.* **2**, 156.

Harasty, J., Seldon, H. L., Chan, P., Halliday, G., & Harding, A. (2003). The left human speech-processing cortex is thinner but longer than the right. *Laterality* **8**, 247–260.

Hauser, M. D. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402.

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Percept. Interdiscip. J.* **26**, 41–55.

Jäncke, L. (2012). The relationship between music and language. *Front. Psychol.* **3**, article 123.

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *J. Exp. Child Psychol.* **144**, 84–97.

Jentschke, S. & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage* **47**, 735–744.

Klein, C., Liem, F., Hänggi, J., Elmer, S., & Jäncke, L. (2016). The 'silent' imprint of musical training. *Hum. Brain Mapp.* **37**, 536–546.

Koelsch, S. *et al.* (2002). Bach speaks: a cortical 'language-network' serves the processing of music. *NeuroImage* **17**, 956–966.

Kraus, N. & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev. Neurosci.* **11**, 599–605.

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

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

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* **62**, 621–647.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* **27**, 308–314.

Lee, C.-Y. & Hung, T.-H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* **124**, 3235–3248.

Lee, H. & Noppeney, U. (2011). Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1441–1450.

Lehmann, D., Faber, P. L., Gianotti, L. R. R., Kochi, K., & Pascual-Marqui, R. D. (2006). Coherence and phase locking in the scalp EEG and between LORETA model sources, and microstates as putative mechanisms of brain temporo-spatial functional organization. *J. Physiol. Paris* **99**, 29–36.

Levitin, D. J. & Menon, V. (2003). Musical structure is processed in 'language' areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage* **20**, 2142–2152.

Lima, C. F. & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emot. Wash. DC* **11**, 1021–1031.

López-Barroso, D. *et al.* (2013). Word learning is mediated by the left arcuate fasciculus. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13168–13173.

Ma, W. & Thompson, W. F. (2015). Human emotions track changes in the acoustic environment. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14563–14568.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nat. Neurosci.* **4**, 540–545.

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. *J. Cogn. Neurosci.* **18**, 199–211.

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. *J. Cogn. Neurosci.* **19**, 1453–1463.

Mathias, B., Palmer, C., Perrin, F., & Tillmann, B. (2015). Sensorimotor learning enhances expectations during auditory perception. *Cereb. Cortex* **25**, 2238–2254.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nat. Neurosci.* **7**, 703–704.

Mestres-Misse, A., Rodriguez-Fornells, A., & Munte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cortex* **17**, 1858–1866.

Meyer, M., Elmer, S., & Jäncke, L. (2012). Musical expertise induces neuroplasticity of the planum temporale. *Ann. N.Y. Acad. Sci.* **1252**, 116–123.

Milner, A. D. & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia* **46**, 774–785.

Mishkin, M. & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* **6**, 57–77.

Moreno, S. *et al.* (2009). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cereb. Cortex* **19**, 712–723.

Moreno, S. *et al.* (2011). Short-term music training enhances verbal intelligence and executive function. *Psychol. Sci.* **22**, 1425–1433.

Münte, T. F., Altenmüller, E., & Jäncke, L. (2002). The musician's brain as a model of neuroplasticity. *Nat. Rev. Neurosci.* **3**, 473–478.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15894–15898.

Norman-Haignere, S., Kanwisher, N. G., & McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* **88**, 1281–1296.

Oechslin, M. S., Imfeld, A., Loenneker, T., Meyer, M., & Jäncke, L. (2009). The plasticity of the superior longitudinal fasciculus as a function of musical expertise: a diffusion tensor imaging study. *Front. Hum. Neurosci.* **3**, 76.

Ott, C. (2011). Processing of voiced and unvoiced acoustic stimuli in musicians. *Front. Psychol.* **2**, article 195.

Overath, T., McDermott, J. H., Zarate, J. M., & Poeppel, D. (2015). The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. *Nat. Neurosci.* **18**, 903–911.

Pallesen, K. J. *et al.* (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE* **5**, e11120.

Pantev, C., Lappe, C., Herholz, S. C., & Trainor, L. (2009). Auditory-somatosensory integration and cortical plasticity in musical training. *Ann. N.Y. Acad. Sci.* **1169**, 143–150.

Paraskevopoulos, E., Kraneburg, A., Herholz, S. C., Bamidis, P. D., & Pantev, C. (2015). Musical expertise is related to altered functional connectivity during audiovisual integration. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12522–12527.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2012). Musical expertise induces audiovisual integration of abstract congruency rules. *J. Neurosci.* **32**, 18196–18203.

Paraskevopoulos, E., Kuchenbuch, A., Herholz, S. C., & Pantev, C. (2014). Multisensory integration during short-term music reading training enhances both uni- and multisensory cortical processing. *J. Cogn. Neurosci.* **26**, 2224–2238.

Perani, D. *et al.* (2011). Neural language networks at birth. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16056–16061.

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1281–1292.

Péron, J., Frühholz, S., Ceravolo, L., & Grandjean, D. (2016). Structural and functional connectivity of the subthalamic nucleus during vocal emotion decoding. *Soc. Cogn. Affect. Neurosci.* **11**, 349–356.

Rauschecker, J. P. & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724.

Ripollés, P. *et al.* (2014). The role of reward in word learning and its implications for language acquisition. *Curr. Biol. CB* **24**, 2606–2611.

Roden, I., Kreutz, G., & Bongard, S. (2012). Effects of a school-based instrumental music program on verbal and visual memory in primary school children: a longitudinal study. *Front. Psychol.* **3**, article 572.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philos. Trans. R. Soc. B Biol. Sci.* **364**, 3711–3735.

Rogalsky, C., Rong, F., Saberi, K., & Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* **31**, 3843–3852.

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., & Kotz, S. A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PloS One* **8**, e56000.

Saffran, J. R. (2003). Statistical language learning mechanisms and constraints. *Curr. Dir. Psychol. Sci.* **12**, 110–114.

Santoro, R. *et al.* (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* **10**, e1003412.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychol. Sci.* **15**, 511–514.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science* **267**, 699–701.

Schmidt-Kassow, M. & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J. Cogn. Neurosci.* **21**, 1693–1708.

Schneider, P. *et al.* (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* **5**, 688–694.

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* **41**, 341–349.

Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Percept. Interdiscip. J.* **29**, 255–267.

Schulze, K. & Koelsch, S. (2012). Working memory for speech and music. *Ann. N.Y. Acad. Sci.* **1252**, 229–236.

Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *Eur. J. Neurosci.* **33**, 189–196.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* **32**, 771–783.

Seldon, H. L. (1981). Structure of human auditory cortex. I. Cytoarchitectonics and dendritic distributions. *Brain Res.* **229**, 277–294.

Sluming, V. *et al.* (2002). Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. *NeuroImage* **17**, 1613–1622.

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Dev. Neuropsychol.* **37**, 476–496.

Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 19027–19032.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emot. Wash. DC* **4**, 46–64.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Ann. N.Y. Acad. Sci.* **999**, 209–211.

Trimmer, C. G. & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emot. Wash. DC* **8**, 838–849.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension activates language areas (BA47). *NeuroImage* **31**, 832–841.

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends Cogn. Sci.* **7**, 553–559.

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem. Cogn.* **38**, 163–175.

Wong, P. C. M. & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.* 28(4), 565–585.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**, 420–422.

Zatorre, R. J. & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* **11**, 946–953.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLOS ONE* **9**, e99868.

APPENDIX 10.1

AUDITORY STIMULI

The four consonant-vowel (CV) syllables (two natural German consonant-vowel syllables and two reduced-spectrum analogues; see Audio 10.1) were used in three previous publications in order to assess putative advantages of musicians in processing fast-changing phonetic cues. These stimuli consisted of the German CV syllables /ka/ (voiceless initial consonant) and /da/ (voiced initial consonant) as well as of its reduced-spectrum analogues. The duration of the syllables was about 350 ms, and the voice-onset time (VOT) of /da/ and /ka/ was approximately 13 ms and 53 ms, respectively. For the reduced-spectrum analogues, spectral information was removed from the CV syllables by replacing the frequency-specific information in a broad frequency region with band-limited white noise (band 1: 500–1500 Hz, band 2: 2500–3500 Hz). Amplitude and temporal cues were preserved in each spectral band, resulting in double-band-pass filtered noise with temporal CV-amplitude dynamics. A detailed description of the stimuli can be found here:

Elmer, S., Meyer, M., & Jäncke, L. (2012). Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cerebral Cortex*, **22**, 650–658.

Elmer, S., Hänggi, J., Meyer, M., & Jäncke, L. (2013). Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. *Cortex*, **49**, 2812–2821.

Elmer, S., Hänggi, J., & Jäncke, L. (2016). Interhemispheric transcallosal connectivity between the left and right planum temporale predicts musicianship, performance in temporal speech processing, and functional specialization. *Brain Structure and Function*, **221**, 331–344.

How music training influences language processing: Evidence against informationnal encapsulation

Mireille Besson[1], Eva Dittinger[1, 2] & Mylène Barbaroux[1]
[1]Laboratoire de Neurosciences Cognitives CNRS & Aix-Marseille University, Marseille
[2] Brain and Language Research Institute, CNRS & Aix-Marseille University, Marseille

ABSTRACT

To investigate the modularity of language processing and, specifically, the question of whether the language module is informationally encapsulated, many experiments examined the impact of music expertise and music training on the language system (phonology, semantics and syntax). Finding positive evidence would argue against language as an independent ability isolated from other cognitive abilities. We first review the evolution of global or "massive" modularity, as advocated by Fodor in his influential book (1983), to reduced local modularity, (Fodor, 2003). We then consider experimental data relevant to these issues: the emerging picture favors the view that music abilities, as well as other cognitive abilities (attention, memory, executive functions) influence language processing. These influences are seen in behavior as well as in the complex brain networks that sustain behavior. In sum, evidence is accumulating supporting the idea that the language system is not independent from other cognitive abilities.

Keywords: Language, music, modularity

Comment l'entrainement musical influence le traitement du langage : Preuves contre l'encapsulation informationnelle

RÉSUMÉ

Afin de tester la modularité du traitement du langage et, plus spécifiquement, la question de savoir si le module du langage est informationnellement encapsulé, de nombreuses expériences ont eu pour but d'étudier l'influence de l'expertise musicale et de l'apprentissage de la musique sur le traitement linguistique, notamment aux niveaux phonologique, sémantique et syntaxique. Une influence positive démontrerait que le

---

[1] Corresponding author: Laboratoire de Neurosciences Cognitives, UMR 7291, CNRS & Aix, Marseille Université Pôle 3C, Case C, Centre St Charles, 3 Place Victor Hugo, 13331 Marseille Cedex 3, France, Email: mireille.besson@univ-amu.fr

langage n'est pas une fonction isolée des autres fonctions cognitives. Nous considérons d'abord l'évolution d'une conception modulaire globale ou "massive", défendue par Fodor dans un livre paru en 1983 et qui a fortement influencé les recherches en sciences et neurosciences du langage, vers une conception plus locale et réduite de la modularité, comme revue par Fodor en 2003. Nous décrivons ensuite des données expérimentales qui montrent que les habiletés musicales, comme d'autres fonctions cognitives (attention, mémoire, fonctions exécutives), influencent le traitement du langage au niveau comportemental, aussi bien qu'au niveau des réseaux cérébraux complexes qui sous-tendent les comportements. Ainsi, de nombreux résultats sont en accord avec l'idée que le langage n'est pas indépendant des autres fonctions cognitives.

Mots-clés : Langage, musique, modularité

1. INTRODUCTION

Language, one of the most human ability[2], has long been considered as relying on specific, dedicated processes. However, in recent years, neurobiological models have emerged that emphasize the highly dynamic and distributed aspects of language processing. (Friederici & Singer, 2015; Hagoort, 2014; Hickok & Poeppel, 2007). Nevertheless, how these processes are implemented in the brain, how neuronal assemblies communicate together to allow us producing and comprehending language remains a major mystery. In the first section of this review, we consider the question of whether language is a domain-specific or domain-general cognitive ability in view of the evolution of the concept of modularity from massive modularity in Fodor (1983) to reduced local modularity in Fodor (2000/2003). Interestingly, the evolution of these philosophical considerations parallels progress in our understanding of the anatomo-functional organization of the brain developed in cognitive and computational neuroscience, from the old idea of one function-one structure (e.g., Gall, 1835 and the Classic "Wernicke-Lichtheim-Geschwind" model) to many functions-one structure and many structures for one function (Park & Friston, 2013). In the second section, we focus on work considering the impact of music expertise and music training on language processing to test the following hypothesis: if language is a domain-specific and informationally-encapsulated ability, music expertise should have no influence on the various computations

---

[2] We consider that the language ability is subserved by different functions (categorical perception, speech segmentation etc…) and one aim is to determine whether (some of) these functions are also involved in music abilities.

involved in language processing. As we will see, many results in the literature allow us to reject this hypothesis.

2. MASSIVE MODULARITY VS LOCAL MODULARITY

In his book "The modularity of mind" (1983) that strongly influenced research in cognitive neuroscience for many years, Fodor defined a system as modular if it possesses a number of specific properties: (1) domain specific, (2) mandatory, (3) with limited access to the mental representations that they compute, (4) fast, (5) informationally encapsulated, (6) with 'shallow' outputs, (7) with characteristic and specific breakdown patterns, and (8) with a characteristic pace and sequencing of their ontogeny. In this framework[3], a module is a processing device that only uses the information available in its own innate database without being influenced by anything else (the extreme example of a modular system is a reflex). Importantly, Fodor clearly insisted that 'informational encapsulation is at the heart of modularity' (2003, p. 107) because this is how modules can be functionally specified. In this context, the "language organ" (Pinker, 1994) was typically considered as a modular system that included several hierarchically organized sub-modules dedicated to phonetic, phonologic, semantic, syntactic, pragmatic aspects of language processing, each sub-module doing its own computations without influence from other computational levels and transmitting the results of its own computations to the next hierarchical level.

Over the years, numerous results have accumulated against this view[4] that led Fodor to revise his conception of modularity in his less-known book, "The mind doesn't work this way" (2000/2003)[5], in which he considers that only local systems can be modular. Global systems are not modular

---

[3] The basic assumption underlying modularity, as defined by Fodor (1983), is that cognitive mental processes are computational (i.e. thinking is a form of computation). Cognitive processes are specific (logico-algebric) computations (i.e. formal — non-semantic- operations) on mental representations (i.e. the relationship between the world and the mind) that are structured syntactically (i.e. they obey an ensemble of rules that define the relationship between the different elements). However, this basic assumption has also been called into question by Fodor (2000).

[4] For instance, results typically highlighted strong interactions between different levels of language processing. To take only one example among many, Dehaene et al (2010) showed that learning to read increased the level of activation in brain regions involved in phonological processing (in the planum temporale) when listening to speech.

[5] Fodor (2000/2003) book was also possibly written in response to Pinker (1997) « How the mind works », W.W. Norton & Company, New York, London.

because they use all the information available within the entire system to compute their function (see also Hagoort, 2014). Applied to language, this implies that the language system is not informationally encapsulated because it does not work independently of other cognitive functions that provide relevant information (e.g., attention, memory…). Of course, the next problem is then to define local and global systems. Marr (1982) advocated the view that global systems can be decomposed into a collection of modular, independent, and specialized sub-processes[6]. In line with this view, Park & Friston (2013) more recently proposed a model that may possibly reconcile the modular and non-modular theoretical frameworks: "Brain functions can be characterized by local integration within segregated modules for specialized functions and global integration of modules for perception, cognition, and action". Thus, while micro neural circuits are possibly characterized by a modular architecture, macroscopic brain networks are non-modular and highly interactive, with a pattern of interactions that dynamically change over time as a function of context and task-demands ("functional integration among segregated brain areas" Friston, 2011). Importantly, Park & Friston (2013) proposed that at the level of the module (micro-neural circuits), functional connectivity (defined as "as statistical dependencies among remote neurophysiological events") is closely related to the underlying structural connectivity (defined as "the anatomical connections usually estimated using fiber tractography from diffusion tensor imaging, DTI"). Thus, at this level, there is possibly a one-function-to-one-structure mapping with the idea that structural networks are constraining functional networks. However, at the level of global systems such as language, there are many function-structure relationships so that a neuronal architecture can be involved in diverse cognitive functions and a global system such as language may rely on different brain structures.

Let's consider, for instance, Broca's area. It has long been considered that this region, in the left inferior frontal cortex (for the problems posed by its precise localization, see Tremblay & Dick, 2016), was specific to speech production and to processing of syntactic structures. However, there is now clear evidence that Broca's area is also activated by the processing of phonological, lexical and semantic information (Sahin, Pinker, Cash, Schomer, Halgren, 2009). Moreover, results also point to the activation of parts of Broca's areas in tasks requiring verbal and non-verbal working memory and/or executive functions (Schulze et al., 2011) as well as in

---

[6] It is important to note that this view was criticized as being open-ended with no empirical failure point: any processes can be decomposed into more refined sub-processes (Van Orden, Pennington, & Stone, 2001).

tasks based on the processing of musical syntax (Maess, Koelsch, Gunter & Friederici, 2001, see below). The same general picture also emerges when considering Wernicke's area (part of the Planum Temporale (PT) in the left Superior Temporal Gyrus (STG), with similar localization problems as for Broca's area; see Tremblay & Dick, 2016 and DeWitt & Rauschecker, 2013), that was initially taken to play a major role in language comprehension. Temporal regions are clearly crucial for speech processing but there is also growing evidence for their functional diversity, with sub-areas implied both in linguistic and in nonlinguistic functions (see Liebenthal, Desai, Humphries, Sabri, & Desai, 2014 for results of a large meta-analysis). Taken together, these results are important for at least two reasons. First, they strongly call into question the one structure-one function mapping between Broca's or Wernicke'areas and speech production, syntax processing and speech comprehension. Broca's or Wernicke's area do not appear to be the brain structures hosting a speech production module, a syntactic module or a speech comprehension module that would carry on its own computations autonomously, without being influenced by other levels of language processing or by other cognitive functions. Although the localization and functional role of Broca's and Wernicke's areas are still under debate (Tremblay & Dick, 2016), current evidence suggests that they may be part of a large prefrontal-temporal network, that includes language-specific as well as general cognitive functions that closely interact with one another (Fedorenko, Duncan, & Kanwisher, 2012; Hagoort, 2014; Schulze et al., 2011).

Second, the repeated finding that Broca's area is not only involved in syntactic processing but also in phonological, lexical and semantic processing more generally argues against the idea that syntactic constructions are processed independently from lexico-semantic information, or put the other way around, these findings argue in favor of the idea that syntactic constructions have meaning. Showing that Broca area is activated by syntactic as well as by lexico-semantic processing is indeed compatible with linguistic theories issued from Cognitive and Construction Grammar (e.g., Goldberg, 1995, Langacker, 1991) that consider that there is no clear separation between syntax and lexico-semantic information. As clearly stated by Goldberg (1995, p.7): "In Construction Grammar, no strict division is assumed between the lexicon and syntax. Lexical constructions and syntactic constructions differ in internal complexity, […] but both lexical and syntactic constructions are essentially the same type of declaratively represented data structure: both pair form with meaning." For instance, two apparently similar grammatical constructions may induce differences in meaning: in "*Bill* *sent a walrus to Joyce*" the accent is on the salience of the path due to the use of "to" but in "*Bill sent Joyce a walrus*", the accent is on the salience of the possessive relationship due to the use of a ditransitive construction. Another example: "*Sally baked her sister a cake*", Sally baked a cake with the intention to give it to her sister and this intention is inferred from the grammatical construction (not present in the verb "to bake")[7]. Within this framework, we previously showed that the same syntactic incongruity (an intransitive verb followed by a direct object) was processed differently depending upon the semantic context of the sentence. Thus, while there was no significant differences between the correct sentence (.g., "L'ennemi a préparé un complot", 'The enemy prepared a scheme') and the syntactically incongruous sentence with congruent semantics (e.g., *L'ennemi a conspiré (INTR) un complot, 'The enemy conspired a scheme'), the differences observed in behavior (percent errors) and in electrophysiological measures (N400 and P600 components) were significant when the semantics of the sentence did not help resolve the syntactic incongruity (e.g., *L'ennemi a déjeuné (INTR) un complot", *'The enemy lunched a scheme' (Magne, Besson & Robert, 2014). To end this short and incomplete section on the relationships between syntax and semantics, we refer the interested reader to the book by Tomasello (1998) "The new psychology of language" in which Langacker position is clearly summarized: "The ultimate goal is not to create a mathematically coherent grammar that normatively parses the linguistic universe into grammatical and ungrammatical sentences but rather to detail the structured inventory of symbolic units that make up particular natural languages" (p.xiii).

3. HOW MUSIC EXPERTISE AND MUSIC TRAINING INFLUENCE LANGUAGE PROCESSING

To further test whether language is a domain-specific and informationally-encapsulated ability, we now focus on the specific issue of the influence of music expertise and music training[8] on the various computations involved in language processing. The central hypothesis is

---

[7] We are thankful to Dr Stéphane Robert for providing us with these examples.

[8] Music expertise refers to musicians that are already experts in the domain (i.e., professional musicians). By contrast, music training refers to participants that are currently being trained in music and who are not yet musicians.

that if the language system is an informationnaly encapsulated module, it should be impermeable to the influence of other cognitive abilities. Consequently, if we can demonstrate that musical abilities, as well as other cognitive abilities (attention, memory, executive functions, not reviewed here) impact language processing (categorical perception, segmental and supra-segmental processes, phonology, syntax and semantics), these findings would argue against one of the main characteristics of a modular system: encapsulation. Below we review evidence showing that these influences can be seen in behavior as well as in the complex brain networks that sustain behavior.

Categorical perception is fundamental to speech perception by allowing listeners to categorize continuous acoustic changes in the speech signal into discrete phonetic categories. Bidelman and collaborators (2014) first demonstrated that music training in both younger and older musicians is associated with more efficient vowel categorical perception (/u/ to /a/ continuum) as reflected by behavioral measures, more precise phase-locking of brainstem responses and increased amplitude of cortical evoked responses to relevant speech cues. Moreover, Habib and collaborators (2016) tested the efficacy of a newly developed cognitivo-music training method to improve categorical perception in children with dyslexia. They showed a normalization of the identification and discrimination of inter-categorical boundary on a /ba/-/pa/ continuum after intensive music training clustered on three consecutive days as well as after distributed music training over a 6 weeks period. Thus, increased auditory sensitivity in musicians and in children with music training is possibly one of the driving forces behind enhanced categorical perception in musicians.

Speech segmentation is also fundamental to speech comprehension as is clearly exemplified when learning a foreign language that is first perceived as a continuous stream of nonsense words. François and colleagues (2013) used a longitudinal approach over two school-years with 8 years old children to examine the impact of music compared to painting training on the ability to extract “words” from a continuous stream of meaningless sung syllables. Implicit recognition of meaningless words steadily increased over the two years of music but not of painting training and this was associated with modulations of a frontally distributed N400 component (FN400).

At the segmental level (consonants, vowels and syllables), music training is correlated to the discrimination of Mandarin tones in native English-speakers (Lee & Hung, 2008) and of lexical tones in Italian speakers (Delogu, Lampis & Belardinelli, 2010). Moreover, Marie, Delogu, Lampis, Belardinelli & Besson (2011) showed that, at the cortical level, both lexical tone discrimination (as reflected by the N200/N300 component) and higher-order decision processes (as reflected by the P300b component) were more efficient in musicians than in non-musicians. Very recently, Dittinger et al (2018) tested the impact of musical expertise on the categorization of syllables that did (/p/, /b/) or did not (/ph/) belong to the French repertoire. Professional musicians outperformed non-musicians when the task required the discrimination of non-native phonemes, and the difference between native and non-native phonemes, as reflected by the N200 and P300 components, was larger in musicians than in non-musicians. Finally, in the study mentioned above, Habib et al (2016) also showed that music training improved the perception of syllabic duration in children with dyslexia.

At the supra-segmental level (words, sentences, discourse), early research showed that adult musicians and children with music training are more sensitive than non-musicians to linguistic prosody (e.g., final pitch rise in sentences; see Besson and coll. 2011, for review) and to emotional prosody (Lima & Castro, 2011; but see Trimmer & Cuddy, 2008, for contrastive results). Follow-up studies by Moreno and coll. (2009) used a longitudinal approach to compare non-musician children (8-12 years old) before and after 6 months of music or painting training. Results showed enhanced perception of prosodic intonation only in the music group, together with better reading abilities of complex words.

At the phonological level, there is evidence that musical abilities are predictive of phonological skills in children (Anvari, Trainor, Woodside & Levy, 2002) and in adults (Slevc & Miyake, 2006). These results, based on a cross-sectional approach, are in line with those of a longitudinal study with 6-7 years old children showing that two months of rhythm-based training produced roughly comparable enhancements on a variety of standardized tests of phonological processing than an equivalent amount of training of phonological skills (Bhide, Power & Goswami, 2013). They are also in line with the conclusions of an interesting meta-analysis of longitudinal studies conducted by Gordon, Fehd & McCandliss (2015) showing that music training significantly improved phonological awareness skills, even if the effect sizes were small. By contrast, these analyses also revealed that the evidence for an impact of music training on reading had not yet been convincingly demonstrated.

At the syntactic level, Jentschke & Koelsch (2009) reported that, as expected, violations of harmonic structures elicited larger cortical responses in musically trained children than in controls. More surprisingly, violations of linguistic structures also elicited the largest effects in musically trained children, which was taken as evidence that the automatic processing of linguistic syntactic structures developed faster in children with than without music training, possibly because syntax relies on similar processes in both music and language or because musically-trained children made better use of the prosodic and rhythmic cues that constrain syntactic constructions (Cason & Schön, 2012). Results by Gordon et al (2014) also support this interpretation: children with stronger rhythmic abilities also showed higher grammatical competence, as measured by their ability to produce sentences with relevant grammatical constructions. Taken together, these results are in line with early findings using fMRI and showing that Broca’s area is activated when processing music as well as linguistic syntax (Maess et al, 2001), that is, when processing structured sequences of events that unfold in time, independently of whether these events form linguistic sentences or musical phrases.

Finally, at the semantic level, Dittinger and collaborators, recently provided evidence that novel word learning was faster and more efficient in professional musicians (Dittinger et al, 2016), in children with music training (Dittinger et al, 2017) and, although the differences were smaller, in older musicians (Dittinger et al, submitted.) than in controls. Specifically, participants were first asked to learn the meaning of monosyllabic Thaï words through picture-word associations (see Figure 1). A frontal N400 component (FN400), taken as an index that words had acquired meaning (Mestres-Misse, Rodriguez-Fornells & Münte, 2007), developed after only 3 minutes of training which clearly reflected fast brain plasticity. Importantly, the N400 developed faster in musicians than in non-musicians (see Figure 2). To test whether participants had learned the associations, a matching task was used in which picture-word pairs were presented that matched or mismatched those learned in the learning phase. Finally, to test for semantic generalization, new pictures (not seen before in the experiment) were presented that were semantically related or unrelated to the newly-learned words (see Figure 1). In both the matching and semantic tasks, the N400 effect (i.e., the difference between mismatching/unrelated and matching/related words) was larger in musicians than in non-musicians (see Figure 3). As typically found in the literature, the N400 effect was larger over centro-parietal regions in musicians but it was more frontally distributed in non-musicians. This was taken as evidence that musicians were more efficient at integrating the meaning of novel words into semantic networks than non-musicians and this was in line with their higher level of performance in the semantic task.

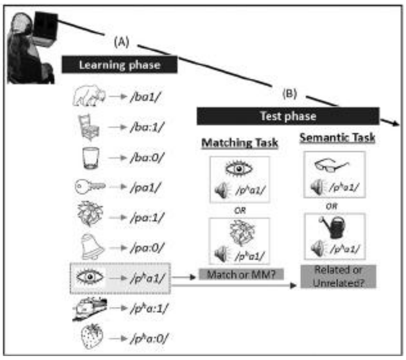


Figure 1. Schematic representation of the experimental procedure that included a learning phase and a test phase (see text for details).

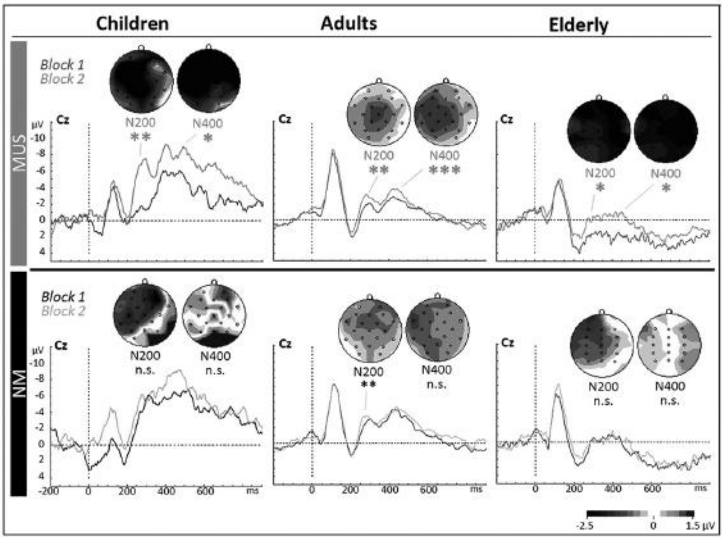


Figure 2. Event-Related Potentials (ERPs) recorded in the learning phase, when children with music training, professional young musicians and elderly musicians (top row) and non-musician controls (bottom row) learned the meaning of novel words through picture-word associations. ERPs recorded at the central site (Cz) to the words are compared between the first and the second learning blocks and the differences are significant for the N200 and N400 components in musicians but not in non-musicians. Topographical maps are presented to illustrate the differences in scalp distribution of these components between Block 1 and Block 2 of the learning phase. On this figure and on the next one, time is on the abscissa (in milliseconds, ms) and the amplitude of the effects is on the ordinate (in microvolts, µV).

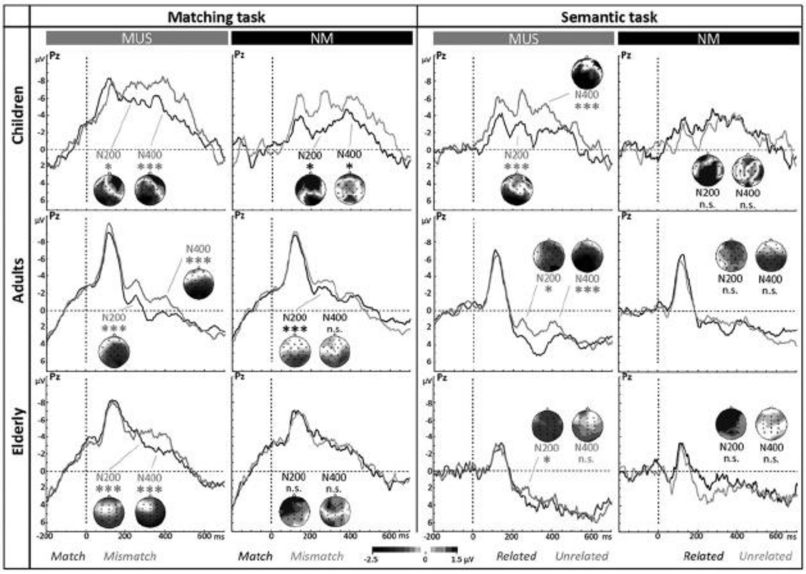


Figure 3. ERPs recorded in the test phase, when children with and with no music training (top row), professional young-adult musicians and non-musician adults (middle row) as well as elderly musicians and elderly controls (bottom row) and non-musician controls (bottom row) performed the matching task (left columns) and the semantic task (right columns). ERPs recorded at the parietal site (Pz) are compared between words that matched or mismatched the picture association learned previously in the matching tasks as well as between words that were semantically related or unrelated to new pictures. In both tasks and in the three groups of participants, effects on the N200 and N400 components were overall significantly larger in musicians than in controls. Topographical maps are presented to illustrate the scalp distribution of the effects.

How can we account for these effects and more generally for the influence of music training at various levels of language processing? Two main interpretations, the cascade and multi-dimensional interpretations, have been proposed. Following the cascade interpretation (bottom-up), increased sensitivity to low-level acoustic parameters such as pitch or duration, that are common to music and speech, drives the influence of music training at different levels of language processing (e.g., phonetic, phonologic, prosodic, syntactic and semantic; Wong, Skoe, Russo, Dees, & Kraus, 2007; Besson et al, 2011; Dittinger et al, 2016). In other words, because musicians perceive speech sounds better than non-musicians, they are more sensitive to prosodic cues such as pitch and rhythm and they form more accurate phonological representations. This, in turn (hence the cascade), increases the stability of lexical representations and facilitates the construction of syntactic structures, word learning and semantic processing.

Following the multi-dimensional interpretation (top-down), that shares several aspects with the OPERA hypothesis proposed by Patel (2014), both language and music are processed in interaction with other cognitive, emotional and motor abilities. For instance, playing a music instrument is a multi-dimensional ability that requires auditory and visual perception (auditory processing of sounds and visual processing of notes on the score), auditory-visuo-motor integration (playing and hearing the notes visually presented on the score), selective and divided attention (focusing attention on one's own instrument and dividing attention between the different instruments of the orchestra and the conductor…), motor control (posture and fine movements of the hands, elbow, lips…), memory (playing musical pieces by heart is common practice in professional musicians), cognitive control (executive functions, such as cognitive flexibility, inhibitory control and working memory, see Zuk, Benjamin, Kenyon & Gaab, 2014) and emotion (as reflected by the interpretation of the musical piece). Thus, it may come as no surprise that extensive training of these different abilities in musicians, from auditory perception to cognitive control, facilitates various levels of language processing. In this respect the cascade and multi-dimensional hypotheses are complementary with both bottom-up and top-down processes probably at play to various degrees in most experimental designs. Results of speech in noise perception experiments also support this view, by showing that the fidelity of the brainstem response is correlated with the ability to hear speech in noise that, in turn, is correlated with auditory working memory, thereby pointing to both bottom-up and top-down influences between the subcortical and cortical levels of speech processing (e.g., Kraus, Strait & Parbery-Clark, 2012).

Directly related to the issue raised in this second section, the results reviewed above clearly demonstrate that the various language processing levels that have been examined so far (categorical perception, speech segmentation, phonology, syntax and semantics) are not impermeable to the influence of music expertise and music training. It is therefore unlikely that language functions as an informationally encapsulated module, independently from other cognitive abilities.

4. CONCLUSION

Science is a never-ending process and more experiments are clearly needed to better understand the influence of music expertise and music training on language processing at the behavioral level, by testing groups of participants with wide range of musical abilities, by using standardized tests when they are available and by trying to control for the effects of the many different factors that can influence the results (the task at hand, socio-economic status, bilingualism…) using different statistical analyses based, for instance, on structural equation modelling. At the brain level, approaches based on structural and functional connectivity and, may be more importantly, on effective connectivity as derived from computational modelling, are very promising avenues for future research.

Another issue that we did not address here is whether the many associations between music expertise and the various levels of language processing reviewed above are causally linked to music training. Indeed, some authors have argued that children with good music aptitudes (Swaminatha & Schellenberg, 2017) or with higher intelligence (IQ, Schellenberg, 2011) are more likely to take music lessons than children with lower music aptitudes or IQ scores. While this may be the case, results of longitudinal studies have also shown that music training can causally influence language processing. For instance, non-musician children trained with music for six or 12 months showed improved pre-attentive processing of segmental cues (vowel duration and voice-onset-time; Chobert et al, 2014) as well as increased perception of prosodic cues and better reading of complex words (Moreno et al., 2009) compared to children trained with painting. Taken together, these results and others suggest that both nature and nurture contribute to the strong influence of music training on language processing.

Based on the above, we would like to point out what we consider as two major advances in the sciences of language. First, as a global system, language can no longer be considered as an autonomous “mental organ” (Pinker, 1994) relying on specific, dedicated processes (Fodor, 1983)**. Rather, current evidence favors the view that language comprehension and production are processed in interaction with other cognitive abilities such as attention, memory, emotions and actions (Fodor, 2000).** As nicely written in the article by Tremblay & Dick (2016) that motivated this special issue: “As a field, we need to study the interactions between language and other functional systems in order to fully understand the neurobiological underpinning of human language and language disorders, and the degree to which it is dependent upon various other cognitive, sensorimotor and emotional processes, all of which must come together to put language into action."

Second, while we are still far from understanding the mind-brain relationship and how neuronal assemblies can produce thoughts, language and music…, advances in brain imaging methods allow us to go beyond the localizationist idea of one-to-one mapping between structures and functions (Gall, 1835; Broca, 1863) to address the much more complex, but probably more realistic view of many functions for one structure and many structures for one function (Park & Friston, 2013). This is also reflected in the evolution of the neurobiological models of language, from static boxes connected with stable arrows to dynamic brain networks, *à la Facebook*, with the strength of connections, nodes and hubs strongly and rapidly fluctuating over time to consider all available information at particular moment in time (Friederici & Singer, 2015; Hagoort, 2014; Hickok & Poeppel, 2007). The convergence in the evolution of philosophical, linguistics and cognitive neuroscience approaches makes us very optimist that a "true" interdisciplinarity is born that will go beyond the old debates to open new perspectives on the complex mind/brain problem.

REFERENCES

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of experimental child psychology*, *83*(2), 111-130.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: common processing, attention, and memory. *Frontiers in psychology*, *2*, 94.

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, *7*(2), 113-123.

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*(4), 2662-2673.

Broca, P. (1863). *Localisations des fonctions cérébrales. Siège de la faculté du langage articulé.* Bulletin de la Société d»Anthropologie 4: 200–208.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*(11), 2652-2658.

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.

DeWitt, I., & Rauschecker, J. P. (2013). Wernicke's area revisited: parallel streams and word processing. *Brain and language*, *127*(2), 181-191.

Dittinger E, Valizadeh SA, Jäncke L, Besson M, Elmer S. (submitted). Testing Relationships between Music Training and Word Learning across the Lifespan by using Cross-Sectional and Comparative Electrophysiological Approaches in Children, Adults, and Elderlies. *NeuroImage*.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. *Journal of cognitive neuroscience*, *28*(10), 1584-1602.

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in human neuroscience*, *11*, 233.

Dittinger, E., D'império M. & Besson, M. (2018). Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *European Journal of Neuroscience, 1-13.*

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059-2062.

Fodor, J. (2000/2003). *The mind doesn't work this way*. Cambridge MA: MIT Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge MA: MIT Press.

François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, *23*(9), 2038-2043.

Friederici, A. D., & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences*, *19* (6), 329–338.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain connectivity*, *1*(1), 13-36.

Gall, F. J. (1835). *On the functions of the brain and of each of its parts: With observations on the possibility of determining the instincts, propensities, and talents, or the moral and intellectual dispositions of men and animals, by the configuration of the brain and head (Vol. 1).* Marsh, Capen & Lyon.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in psychology*, *6*, 1777.

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2014). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*(4), 635-644.

Habib, M., Lardy, C., Desiles, T., Commeiras, C., Chobert, J., & Besson, M. (2016). Music and dyslexia: a new musical training method to improve reading and related disorders. *Frontiers in psychology*, *7*, 26.

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

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

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage*, *47*(2), 735-744.

Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Annals of the New York Academy of Sciences*, *1252*(1), 100-107.

Langacker, R. W. (1991). Cognitive grammar, in *Linguistic theory and grammatical description*. F. Droste and J. Joseph (eds), 275-306. Amsterdam: John Benjamins.

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), 3235-3248.

Liebenthal, E., Desai, R. H., Humphries, C., Sabri, M., & Desai, A. (2014). The functional organization of the left STS: a large scale meta-analysis of PET and fMRI studies of healthy adults. *Frontiers in neuroscience*, *8*, 289.

Lima, C. F., & Castro, S. L. (2011). Speaking to the trained ear: musical expertise enhances the recognition of emotions in speech prosody. *Emotion*, *11*(5), 1021.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature neuroscience*, *4*(5), 540-545.

Magne, C., Besson, M., & Robert, S. (2014). Context influences the processing of verb transitivity in French sentences: more evidence for semantic– syntax interactions. *Language and Cognition*, *6*(2), 181-216.

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), 2701-2715.

Marr, D. (1982). *Vision*. San Fransico, CA: Freeman.

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2006). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*(8), 1858-1866.

Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: more evidence for brain plasticity. *Cerebral Cortex*, *19*(3), 712-723.

Park, H. J. & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science, 342*, 579-588

Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing research*, *308*, 98-108.

Pinker, S. (1994). *The language instinct.* Harper Perennial Ed., 1995.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, *326*(5951), 445-449.

Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuro-architecture of verbal and tonal working memory in nonmusicians and musicians. Human brain mapping, 32(5), 771-783.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675-681.

Tomasello, M. (1998). *The new psychology of language*. London: Erlbaum.

Tremblay, P., & Dick, A. S. (2016). Broca and Wernicke are dead or moving past the classic model of language neurobiology. *Brain and language*, *162*, 60-71.

Trimmer, C. G., & Cuddy, L. L. (2008). Emotional intelligence, not music training, predicts recognition of emotional speech prosody. *Emotion*, *8*(6), 838.

Wong, P. C., 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.

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS one*, *9*(6), e99868.

Acknowledgments : We thank the CNRS, the Labex BLRI (Grant/Award Number: ANR-11-LABX-0036) and the Institut Convergence ILCB (Grant/Award Number: ANR-16-CONV-0002) for continuous support.

Curriculum Vitae

Eva Dittinger was born in Melk (Austria), where she also attended high school. She obtained her diploma (Abitur) in 2006, receiving the highest overall grade ("mit ausgezeichnetem Erfolg"). She started an undergraduate degree in Food Science and Biotechnology at the University of Natural Resources and Life Sciences in Vienna (Austria) in 2006. The European Union supported her in 2009 with an Erasmus Student Fellowship for studying at the University of Kuopio (Finland), and in 2010 with an Erasmus Internship Fellowship for a research internship at the Max Planck Institute for Molecular Genetics in Berlin (Germany). After having graduated from the University of Natural Resources and Life Sciences in Vienna, she moved to Italy and pursued her studies with a Master's degree in Molecular Biotechnology at the University of Turin (Italy), for which she was rewarded a scholarship from Austria (Top-Stipendium) for undergraduate students abroad. Eva received the Master of Science degree with the highest distinction (cum laude) in 2014. Subsequently, she was supported by the European Union with a Leonardo da Vinci Fellowship, allowing her to do a 5-months internship in Marseille (France), where in the following she continued her PhD under the supervision of Mireille Besson and Mariapaola D'Imperio, funded by the Brain and Language Research Institute (BLRI, ANR-11-LABX), a French institute of excellence.

Publications

Dittinger, E., Korka, B., & Besson, M. (submitted). Behavioral and electrophysiological evidence for better long-term memory in professional musicians, and its contribution to novel word learning.

Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (submitted). Testing Relationships between Music Training and Word Learning across the Lifespan by using Cross-Sectional and Comparative Electrophysiological Approaches in Children, young Adults, and older Adults.

Besson, M., **Dittinger, E.** & Barbaroux, M. (in press). How music training influences language processing: evidence against informational encapsulation. In "*Revue année psychologique*", Eds P. Colé and C. Pattamadilok, Humensis.

Elmer, S., **Dittinger, E.**, & Besson, M. (in press). One step beyond: musical expertise and word learning. In "*The handbook of voice perception*", Eds S. Frühholz and P. Belin, Oxford University Press.

Dittinger, E., D'Imperio, M., & Besson, M. (2018) Enhanced neural and behavioral processing of a non-native phonemic contrast in professional musicians. *European Journal of Neuroscience*. doi: 10.1111/ejn.13939. [Epub ahead of print]

Besson, M., Barbaroux, M., & **Dittinger, E.** (2017). Music in the brain: Music and language processing. In "*The Routledge Companion to Music Cognition*", Eds R. Ashley and R. Timmers, Routledge/ Taylor & Francis group.

Dittinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2017). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Human Brain Mapping*. doi: 10.1002/hbm.23877. [Epub ahead of print]

Dittinger, E., Chobert, J., Ziegler, J., & Besson, M. (2017). Fast brain plasticity during word learning in musically trained children. *Frontiers in Human Neuroscience*, 11, 233.

Dittinger, E., & Besson, M. (2016). Word leaning in children and adult musicians. International Journal of Psychophysiology, 108:45. doi: 10.1016/j.ijpsycho.2016.07.151

Besson, M., Jäncke, L., Elmer, S., Nan, Y., Tervaniemi, M., & **Dittinger, E.** (2016). Language, music and the brain: from theoretical to clinical approaches. International Journal of Psychophysiology. doi: 10.1016/j.ijpsycho.2016.07.150

D'Imperio, M., **Dittinger, E.**, & Besson, M. (2016). "Prosodie et intonation : notions de base et données neuro-psycholinguistiques". In "*Traité de neurolinguistique*", Eds S. Pinto and M. Sato, De Boeck-Supérieur.

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and word meaning acquisition: from faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, 28, 1584-1602.

“Where words fail, music speaks.”

Hans Christian Andersen