Neuropsychologia 71 (2015) 52–63

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Pitch expertise is not created equal: Cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music

Stefanie Hutka [a,b,*], Gavin M. Bidelman [c,d], Sylvain Moreno [a,b]

[a] Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON, Canada M5S 3G3
[b] Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1
[c] Institute for Intelligent Systems, University of Memphis, 365 Innovation Drive., Memphis, TN 38152, USA
[d] School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Ave., Memphis, TN 38105, USA

ARTICLE INFO

Article history:
Received 14 October 2014
Received in revised form
7 March 2015
Accepted 18 March 2015
Available online 19 March 2015

Keywords:
Music
Tone language
ERP
Mismatch negativity
Difference limens

ABSTRACT

Psychophysiological evidence supports a music-language association, such that experience in one domain can impact processing required in the other domain. We investigated the bidirectionality of this association by measuring event-related potentials (ERPs) in native English-speaking musicians, native tone language (Cantonese) nonmusicians, and native English-speaking nonmusician controls. We tested the degree to which pitch expertise stemming from musicianship or tone language experience similarly enhances the neural encoding of auditory information necessary for speech and music processing. Early cortical discriminatory processing for music and speech sounds was characterized using the mismatch negativity (MMN). Stimuli included 'large deviant' and 'small deviant' pairs of sounds that differed minimally in pitch (fundamental frequency, F0; contrastive musical tones) or timbre (first formant, F1; contrastive speech vowels). Behavioural F0 and F1 difference limen tasks probed listeners' perceptual acuity for these same acoustic features. Musicians and Cantonese speakers performed comparably in pitch discrimination; only musicians showed an additional advantage on timbre discrimination performance and an enhanced MMN responses to both music and speech. Cantonese language experience was not associated with enhancements on neural measures, despite enhanced behavioural pitch acuity. These data suggest that while both musicianship and tone language experience enhance some aspects of auditory acuity (behavioural pitch discrimination), musicianship confers farther-reaching enhancements to auditory function, tuning both pitch and timbre-related brain processes.

1. Introduction

There is strong evidence for shared processing in brain regions governing music and language processing (Besson et al., 2011; Maess et al., 2001; Koelsch et al., 2002; Patel 2008; Slevc et al., 2009; Bidelman et al., 2011a). For instance, the processing of musical melody and harmony activate brain areas traditionally associated with language-specific processing, such as Broca's and Wernicke's areas (Maess et al., 2001; Koelsch et al., 2002). Such findings suggest that there is a close association between the underlying neural mechanisms recruited for processing both music and language.

These shared brain structures underlying the processing of music and language raise the question of transfer between domains, i.e., processing in one domain potentially benefiting processing in the other, and vice versa (see Besson et al., 2011 for a review). Ample evidence exists to suggest that musicianship is associated with enhanced language-related processing. For example, musicianship has been associated with language-related perceptual enhancements including formant and voice pitch discrimination (Bidelman and Krishnan, 2010), sensitivity to prosodic cues (Thompson et al., 2004), detecting durational cues in speech (Milovanov et al., 2009), degraded-speech perception (Parbery-Clark et al., 2009a; Bidelman and Krishnan, 2010), second language proficiency (Slevc and Miyake, 2006; Marques et al., 2007), lexical tone identification (Delogu et al., 2006; Lee and Hung, 2008; Delogu et al., 2010), verbal memory (Chan et al., 1998; Franklin et al., 2008) and verbal intelligence (Moreno et al., 2011). Electrophysiological evidence corroborates these findings by demonstrating that musicianship is associated with changes in brain circuitry at both cortical (Pantev et al., 2001; Schon et al., 2004;

---

[*] Correspondence to: ARCT, MA, Rotman Research Institute, Baycrest Centre for Geriatric Care, 3560 Bathurst Street, Toronto, ON, Canada M6A 2E1.
E-mail address: stefanie.hutka@mail.utoronto.ca (S. Hutka).

http://dx.doi.org/10.1016/j.neuropsychologia.2015.03.019
0028-3932/© 2015 Elsevier Ltd. All rights reserved.

Moreno et al., 2009; Chandrasekaran et al., 2009; Marie et al., 2011a,b) and subcortical levels (Musacchia et al., 2008; Wong et al., 2007; Parbery-Clark et al., 2009b; Bidelman et al., 2011a,b, 2014), facilitating both the sensory-perceptual and cognitive control of speech information (Bidelman et al., 2013a). As posited in Moreno and Bidelman (2014), musicians' benefits for speech and language may be mediated by a series of enhancements to both sensory and cognitive mechanisms that operate at multiple levels of processing hierarchy to mediate a range of functions from low-level auditory processing to higher-level cognition. Indeed, Bidelman et al. (2014) have recently shown that musicians have stronger indices of neural coordination between auditory brainstem and cortical activity, which facilitates their enhanced behavioural speech identification. These studies support the notion that musicianship tunes a hierarchy of processing within the auditory system, allowing neural representations to carry more behaviourally relevant information and better support desired behaviours.

A number of neurocognitive models have been proposed to account for the language-domain benefits associated with music training. For example, Patel's (2011) OPERA hypothesis describes why music and language may benefit one another via shared, interactive between-domain processing. Specifically, the constituents of the OPERA framework (Overlap, Precision, Emotion, Repetition, Attention) outline how the coordinated plasticity of music training recruits overlapping language structures (e.g., Broca's area) and increases neural precision of processing following emotional, repetitive, and attentional engagement with music (Patel, 2011). In the expanded OPERA model, Patel (2014) proposed that musicianship can enhance speech processing when music places higher demands than speech on the sensory and cognitive processing mechanisms shared by these two domains. The combination of these higher demands, in conjunction with the emotional rewards, repetition, and attention associated with musicianship, activates neural plasticity, thus changing the neural structures and functions that impact the processing of speech.

As noted by Bidelman et al. (2013a), the OPERA framework does not make any *a priori* assumptions that the music-language association is unidirectional; indeed, the overlap, emotional engagement, repetition, attention, and increased sensory encoding precision are also satisfied by forms of language expertise. As with musicianship, tone language experience (e.g., Mandarin Chinese: Bidelman et al., 2011a,b) has been shown to similarly affect the neural encoding of behaviourally-relevant sound in the brainstem (Bidelman et al., 2011a,b). In addition, studies showing that bilinguals have enhanced cognitive abilities support the notion that, like musicianship, language expertise can boost some perceptual and cognitive abilities (e.g., Bialystok et al., 2005; Bialystok and Depape, 2009; Luk et al., 2010; Gold et al., 2013; Bidelman et al., 2013a).

We directly examined this language-to-music link in our recent study (Bidelman et al., 2013a), which sought to test if listeners with tone language expertise display similar performance to native English-speaking musicians on a number of musical measures, including auditory pitch acuity, music perception, and general cognitive ability. The tone language group was comprised of Cantonese speakers, because exposure to the intricate tonal system of this language closely approximates aspects of pitch exposure gained via musicianship (Bidelman et al., 2013a). For example, the Cantonese tonal inventory consists of six contrastive tones, most of which are level pitch patterns minimally differentiable based on pitch height (Gandour, 1981; Khouw and Ciocca, 2007). More critically, the proximity of tones is on the order of a semitone (e.g., Peng, 2006), paralleling the minimum distance between adjacent pitches found in music (Bidelman et al., 2013a). Given Cantonese speakers' specialisation in perceiving minute changes in steady-state level pitch (Gandour, 1981; Francis et al., 2008), we reasoned that Cantonese listeners would show improvements in both low-level (e.g., pitch discrimination, pitch speed) and higher-level (e.g., pitch memory, melody discrimination) musical abilities relative to native English-speaking nonmusician controls. Though musicians demonstrated superior performance on all auditory tasks, Cantonese speakers performed comparably on perceptual tasks (pitch discrimination and pitch speed), as compared to controls, i.e., (musicians ≈ Cantonese) > nonmusician controls. As tasks became more cognitively demanding (pitch memory, melody discrimination), a gradient effect emerged, such that musicians outperformed Cantonese speakers, who in turn outperformed nonmusician controls (i.e., M > C > NM). These findings provided convincing evidence for a language-to-music association that, when considered with pre-existing literature, supported a bidirectional association between music and language (Bidelman et al., 2013a).

The neural mechanisms subserving these bidirectional relations have yet to be fully explored. Brain substrates underlying cross-domain plasticity from music to language are proposed to be rooted in the interplay between bottom-up and top-down auditory processing. That is, descending, corticofugal projections from the cerebral cortex tune subcortical circuits, while ascending projections from subcortical structures tune cortical circuits (cf. the reverse hierarchy theory of auditory processing, Ahissar et al., 2009; see Patel, 2011 for a discussion of this cortical–subcortical interplay). The influence of tone language experience on musical pitch processing may be shaped by this neural reciprocity. Indeed, tone language speakers (Mandarin Chinese) have enhanced brainstem representation of pitch information (musical pitch intervals and lexical tone, Bidelman et al., 2011a; tuned and detuned musical chords, Bidelman et al., 2011b), comparable to musicians. Understanding the extent to which brain processing in musicians and tone language speakers' show similar neural responses to music and speech would help clarify the perceptual-cognitive transfer effects observed between music and language and potential bidirectionality between domains.

To this end, we directly compared cortical neuroelectric activity elicited by music and speech sounds in native English-speaking musicians (i.e., musical pitch experts), Cantonese speakers (i.e., linguistic pitch experts), and native English-speaking nonmusicians (non-experts with either domain of pitch). Critically, both musicians and Cantonese speakers have similar long-term experience with pitch, but differ in how that expertise was obtained (i.e., musical versus linguistic context). Our musicians and nonmusicians did not have any experience with a tone language, Cantonese or otherwise. This criteria ensured that the musicians did not have (formal) pitch expertise outside of their musical training, and that the non-musician controls did not have any experience with linguistically-based pitch.

As an objective assay of early cortical discrimination of music and speech sounds, we employed the mismatch negativity (MMN), a prominent component of the event-related brain potentials (ERPs). The MMN is a neural index of sensory memory-based detection of auditory change thought to reflect early (i.e., "bottom-up") discrimination processing in the auditory cortices (Naatanen et al., 2007). Previous studies have shown that relative to non-musicians, musicians evoke larger MMN responses to changes in complex sounds (Koelsch et al., 1999; Tervaniemi et al., 2001; Brattico et al., 2006; Fujioka et al., 2004, 2005; Brattico et al., 2009), indicating superior pre-attentive auditory processing (Koelsch et al., 1999). In the current study, comparing MMN in musicians and tone language speakers elicited by contrastive speech and musical sounds allowed us to assess the degree to which these divergent forms of expertise influence early cortical auditory processing related to speech and music sound analysis.

In addition to the MMN, we also examined (i) the P3a and (ii) late discriminative negativity (LDN), a sustained, late-emerging slow-wave component. The P3a sometimes follows the MMN and is characterized by a frontocentrally distributed positive deflection thought to reflect an involuntary attentional switch towards the deviancy (Tervaniemi et al., 2005; for a review, see Escera et al., 2000 and/or Polich, 2007) and/or updating of working memory (Donchin and Coles, 1988; Polich, 2007). Past work using passive paradigms has shown that musicians' P3a in response to sound habituates between blocks, while nonmusicians show enhancement of the P3a between blocks (Seppanen et al., 2012). This difference in short-term plasticity between musicians and nonmusicians suggests that musicianship hones attentional abilities and auditory feature encoding (Seppanen et al., 2012).

In the case of the LDN, there are two common interpretations of its functional role, both of which imply "top-down" influences on auditory processing. Specifically, the LDN has been interpreted as an index of automatic reorienting of attention following the distraction of a deviant sound (Shestakova et al., 2003; Wetzel et al., 2006) and a regulation of higher-order auditory processing that follows the initial change detection reflected by the MMN (Ceponiene et al., 2004; Horvath et al., 2009; see Putkinen et al., 2013). Of interest to the current study, our recent report demonstrated that the LDN is influenced by musical training and language experience (Moreno et al., 2014b).

All participants were tested in two conditions, with either a contrast in musical notes (differing only in pitch) or vowel stimuli (differencing only in first formant frequency) presented in separate blocks. This paradigm allowed us to test for domain-specific effects (e.g., note condition for the musician group; vowel condition for the Cantonese group) and cross-domain effects (e.g., note condition for the Cantonese group; vowel condition for the musician group). Additionally, these stimuli allowed us to examine pitch versus spectral (timbre) discrimination enhancements in music and tone language speakers. We also varied two levels of stimulus difficulty (large versus small sound contrasts) presented in a multiple oddball paradigm (e.g., Naatanen et al., 2004) to determine how music/language expertise impacts different complexities of sound discrimination. In addition to the electrophysiological responses, we obtained behavioural measures of pitch (fundamental frequency, F0) and vowel (first formant, F1) discrimination to measure listeners' perceptual acuity for changes in sound features within the music and speech domain, respectively.

We hypothesised that we would observe enhanced MMN (discrimination), LDN (attentional reorienting), and behavioural discrimination in both the musician and Cantonese groups (relative to nonmusicians), given their extensive experience with pitch and spectral information. Additionally, we expected these neural and behavioral enhancements to extend across domains (i.e., musicians demonstrating enhanced processing of speech and Cantonese speakers demonstrating enhanced processing of musical sounds, relative to controls). These outcomes would suggest that both musicianship and tone language experience are associated with superior pre-attentive as well as top-down auditory processing. They would also be consistent with the notion that music and language expertise tune brain function in a bidirectional fashion (Bidelman et al., 2011a, 2013a).

2. Materials and methods

2.1. Participants

Sixty-seven participants were recruited from the University of Toronto and Greater Toronto Area to participate in this experiment. Four participants' data sets were lost due to technical difficulties or attrition. Of the remaining sixty-three participants, three were deemed outliers (3 standard deviations above the mean) and removed from subsequent analysis. Each participant completed questionnaires to assess linguistic (Li et al., 2006; Wong and Perrachione, 2007) and musical (Bidelman et al., 2013a) background. English-speaking musicians (hereafter referred to as M) (n=21, 14 female) were amateur instrumentalists with at least 8 years of continuous training in Western classical music on their primary instrument ($\mu \pm \sigma$: 15.43 ±6.46 years), beginning at or before the age of 11 ($\mu \pm \sigma$: 7.05 ± 3.32 years). All Ms had formal private or group lessons within the past 5 years, and currently played their instrument(s). See Table S1 of Supplemental material for primary instruments played by the musicians. These inclusion criteria are consistent with similar definitions used in many previous studies examining the neuroplastic effects associated with musicianship (Wong et al., 2007; Chandrasekaran et al., 2009; Parbery-Clark et al., 2009b; Bidelman and Krishnan, 2010; Bidelman et al., 2011a; Cooper and Wang, 2012; Bidelman et al., 2013a).

Critically, English-speaking nonmusicians (hereafter referred to as NM) (n=21, 14 female) had ≤3 years of formal music training on any combination of instruments throughout their lifetime ($\mu \pm \sigma$: 0.81 ± 1.40 years) and had not received formal instruction within the past 5 years. Both Ms and NMs had some exposure to a non-tone second language (M: 90.48%, NM: 66.67%; mainly French or Spanish) but were classified as late learners and/or only mildly fluent in their second language.

Cantonese-speaking participants (hereafter referred to as C) (n=18; 11 female) were classified as late bilinguals, having not received formal instruction in English before the age of ~7 (10.27 ± 5.13 years) (Bidelman et al., 2011a, Chandrasekaran et al., 2009). All participants were born and raised in mainland China or Hong Kong and reported using their native Cantonese on a regular basis (>40% daily use). As with NM participants, Cantonese speakers had minimal musical training throughout their lifetime (0.78 ± 0.94 years) and had not received formal instruction in the past 5 years. Importantly, NM and C did not differ in their minimal extent of music training [$F(1,37)=0.007$, $p=0.935$]. The three groups were closely matched in age (M: 25.24 ± 4.17 years, C: 24.17 ± 4.12 years, NM: 23.38 ± 4.07 years; $F(2,57)=1.075$, $p=0.348$), years of formal education (M: 18.19 ± 3.25 years, C: 16.94 ± 2.46 years; NM: 16.67 ± 2.76 years; $F(2,57)=1.670$, $p=0.198$), and were all right-handed. All participants provided written, informed consent in compliance with an experimental protocol approved by the Baycrest Centre Research Ethics Committee. All were provided financial compensation for their time.

2.2. Cognitive tests

We measured participants' general fluid intelligence and working memory (WM) capacity to rule out differences in cognitive aptitude between groups (e.g., Bidelman et al., 2013a).

2.2.1. Raven's test

General fluid intelligence was measured with Raven's Advanced Progressive Matrices (Raven et al., 1998), which uses exclusively nonverbal material suited for measuring an individual's general cognitive ability without introducing cultural, language, or social bias. Each trial consisted of a 3 × 3 matrix with line drawings depicting abstract patterns in all but the bottom-right cell. Participants were required to select the missing pattern from among 6 to 8 alternatives and were given 10 min to complete the 29-item battery. Items became progressively more difficult over the course of the test. Raw scores (number correct) were recorded and used in subsequent analyses.

2.2.2. Corsi blocks

A digital implementation of the Corsi blocks tapping test (Corsi,

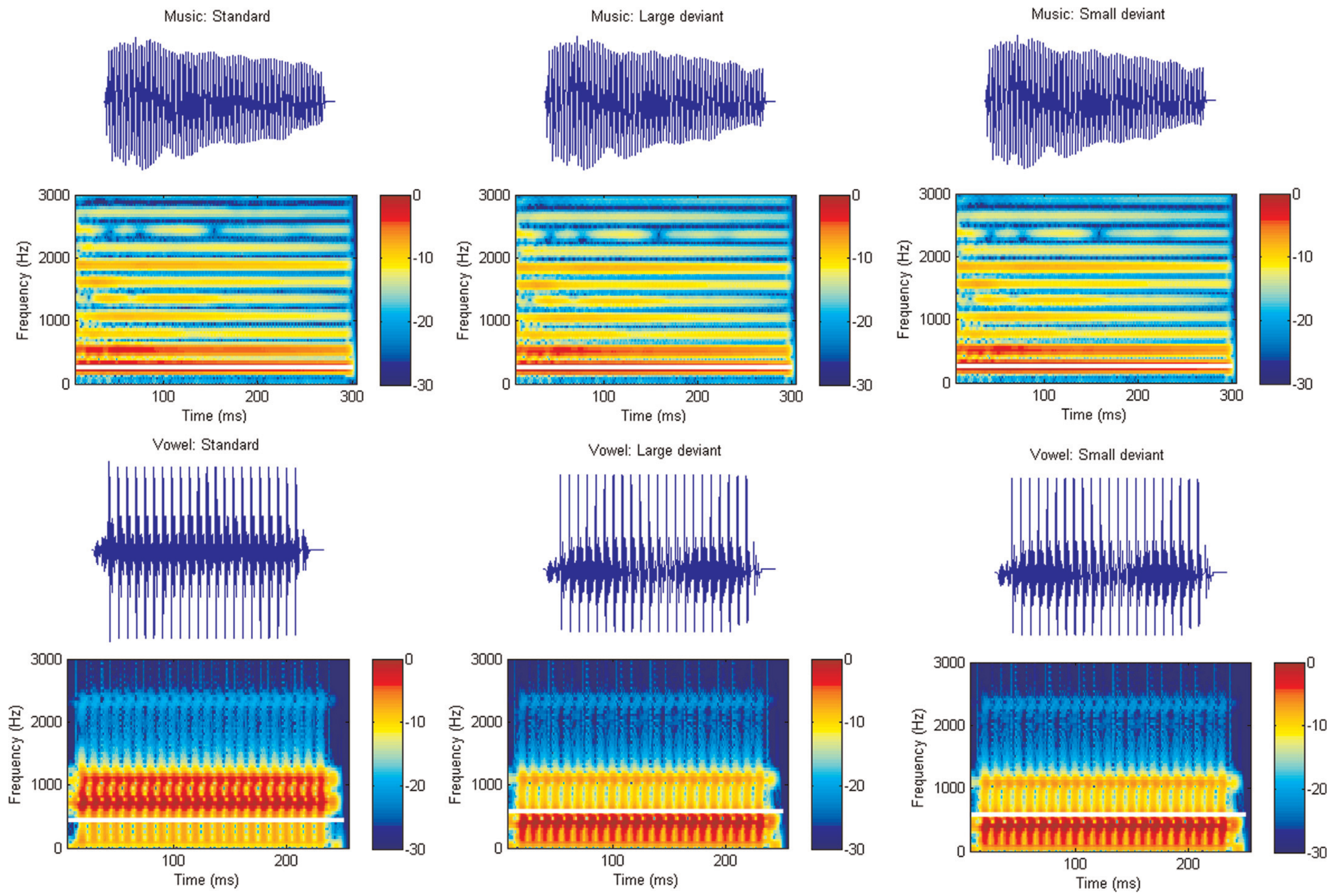

Fig. 1. Spectrograms illustrating the standard, large deviant, and small deviant stimuli for the music (top row) and speech (bottom row) conditions. White lines mark the frequencies of each tone's F0 (music stimuli) and each vowel's F1 (speech stimuli) frequency, respectively.

1972) was used to gauge each individual's nonverbal short-term WM. On each trial, participants saw a 6 × 6 grid of grey squares on the computer screen. A memory sequence was then presented by briefly changing the colour of certain boxes in various locations on the screen. Participants were required to recall the sequence with identical order by clicking on the target boxes. Sequence length gradually increased in set size from 2 to 8 items, becoming progressively harder. Two repetitions were presented for each span length. The longest span-length correctly recalled was used to measure each individual's visual (i.e., non-auditory) WM capacity.

2.3. *Behavioral pitch and timbre difference limens*

Behavioral fundamental frequency difference limens (F0 DLs) and first formant difference limens (F1 DLs) were measured for each participant using three alternative forced choice (3AFC) discrimination tasks (Bidelman and Krishnan, 2010; see Fig. 1A of Bidelman et al., 2013a for a schematic illustration of this task). F0DLs and F1DLs were measured in separate blocks. For F0 DLs, we used complex tones that varied in pitch. Individual tones contained 10 harmonics of the F0, and were 200 ms in duration. F1 DLs were obtained using synthetic speech sounds that varied only in first formant frequency (F1). For these stimuli, the F0 (115 Hz), as well as second (2500 Hz), third (3500 Hz) and fourth (4530 Hz) formants were kept constant across vowels such that only F1 varied.

For a given trial in each task, participants heard three sequential intervals, two containing an identical reference token ($F0_{ref}=220$ Hz for the F0 DL task; $F1_{ref}=300$ Hz for the F1 DL task) and one containing a higher comparison, assigned randomly. The participants' task was to identify which of the three tokens was the 'odd-one-out'.

Discrimination thresholds were measured using a 2-down, 1-up adaptive paradigm that tracks 71% correct performance on the psychometric function (Levitt, 1971). The initial frequency difference between reference and comparison (ΔF) was set at 20% of $F0_{ref}/F1_{ref,}$. Following two consecutive correct responses, ΔF was decreased for the subsequent trial and increased following a single incorrect response. ΔF was varied using a geometric step size factor of two for the first four reversals and was decreased to $\sqrt{2}$ thereafter. Fourteen reversals were measured and the geometric mean of the last eight were used to compute each individual's DL for the run, calculated as the minimum percent change in F0/F1 required to detect a change in the sound (i.e., $\Delta F/F_{nom}$). F0 DLs of two runs were averaged per listener to obtain a final estimate of each individual's F0 discrimination threshold, i.e., the smallest change in musical pitch listeners could reliably detect. Similarly, F1 DLs of two runs were averaged per listener to obtain a final estimate of each individual's F1 discrimination threshold, i.e., the smallest detectable change in speech timbre.

2.4. *EEG task stimuli*

EEGs were recorded using a passive auditory oddball paradigm, consisting of two conditions presented in separate blocks (i.e., music and speech) (Fig. 1).

The presentation order of conditions was counterbalanced across participants. The note condition was comprised of synthesised piano tones, created with Sibelius v7.1.3 and exported as.wav files. The notes consisted of middle C (C4, F0=261.6 Hz), middle C mistuned by an increase of 0.5 semitones (large deviant; 269.3 Hz; 2.9% increase in frequency from standard), and middle C mistuned by an increase of 0.25 semitones cents (small deviant; 265.4 Hz; 1.4% increase in frequency from standard). Tone durations were 300 ms, including 5 ms of rise/fall time in order to reduce spectral splatter in the stimuli.

Speech stimuli consistent of three steady-state vowel sounds (Bidelman et al., 2013b), namely, "oo" as in "book" [ʊ], "aw" as in "pot" [a], and "uh" as in "but" [ʌ], as the standard, large deviant, and small deviant (on the border of categorical perception between the standard and large deviant; Bidelman et al., 2013b), respectively. The duration of each vowel was 250 ms, including 10 ms of rise/fall. The standard vowel had an F1 of 430 Hz, the large deviant 730 Hz (41.1% increase in frequency from standard), and the small deviant 585 Hz (26.5% increase in frequency from standard). Speech tokens contained identical voice fundamental (F0), second (F2), and third (F3) formant frequencies (F0: 100, F2: 1090, and F3: 2350 Hz, respectively), chosen to match prototypical productions from a male speaker (Peterson and Barney, 1952). Speech stimuli were synthesised with a cascade formant synthesiser implemented in MATLAB (The MathWorks) using techniques described by Klatt and Klatt (1990).

Although stimulus duration was different in the speech and note conditions, we were interested in maintaining natural acoustic features and presenting the sound as naturally as possible. The sound onset asynchrony (SOA) was 1000 ms in both conditions so that the stimulus repetition rates (and thus, neural adaptation effects) were comparable for both speech and music ERP recordings. There were a total of 780 trials in each condition including 90 large deviants (12% of the trials), and 90 small deviants (12% of the trials).

2.4.1. EEG recording and data analysis

EEGs were recorded using a 76 channel Biosemi ActiveTwo system (sampling rate=512 Hz) with electrodes placed around the scalp according to standard 10-10 locations (Oostenveld and Praamstra, 2001). During EEG acquisition, all electrodes were referenced to the CMS (Common Mode Sense) electrode, with the DRL (Driven Right Leg) electrode serving as the common ground. Subsequent analyses were performed in EEGLAB (Delorme and Makeig, 2004) and used custom routines coded in MATLAB. Data were re-referenced off-line to the mastoids. Eye movements and artefacts were corrected in the continuous EEG using ICA decomposition in EEGLAB. Excessively noisy channels were interpolated (two nearest neighbour electrodes). Trials with residual voltages exceeding ±500 μV were rejected prior to averaging. The EEG was epoched (−200–1000 ms), baseline-corrected to the pre-stimulus interval, and subsequently averaged in the time domain to obtain ERPs at each electrode site for each response type (standards, deviants) and stimulus condition (musical notes, speech vowels). Grand averaged ERPs were then digitally filtered (0.01–50 Hz, zero-phase response) for response visualisation and quantification.

MMNs were computed by deriving difference waveforms, calculated by subtracting ERPs to the standard stimuli from their corresponding deviant ERPs of the same sequence (i.e., deviant-standard). Fig. S1 of Supplemental material shows the raw ERP traces for standard and deviant stimuli. The presence of the MMN was confirmed by the presence of the response at the mastoids with common average referenced waveforms (Naatanen et al., 2007; see Fig. S2 of Supplemental material). For each participant, MMN amplitudes were measured as the most negative peak in the 100 to 250 ms time window of difference waveforms in a fronto-central electrode cluster (mean of F1, Fz, F2, FC1, FCz, FC2 channels). Similarly, P3a and the LDN were identified in these same channels as the most positive peak in the 200–350 ms time window (P3a) and the mean ERP amplitude in a latency window of

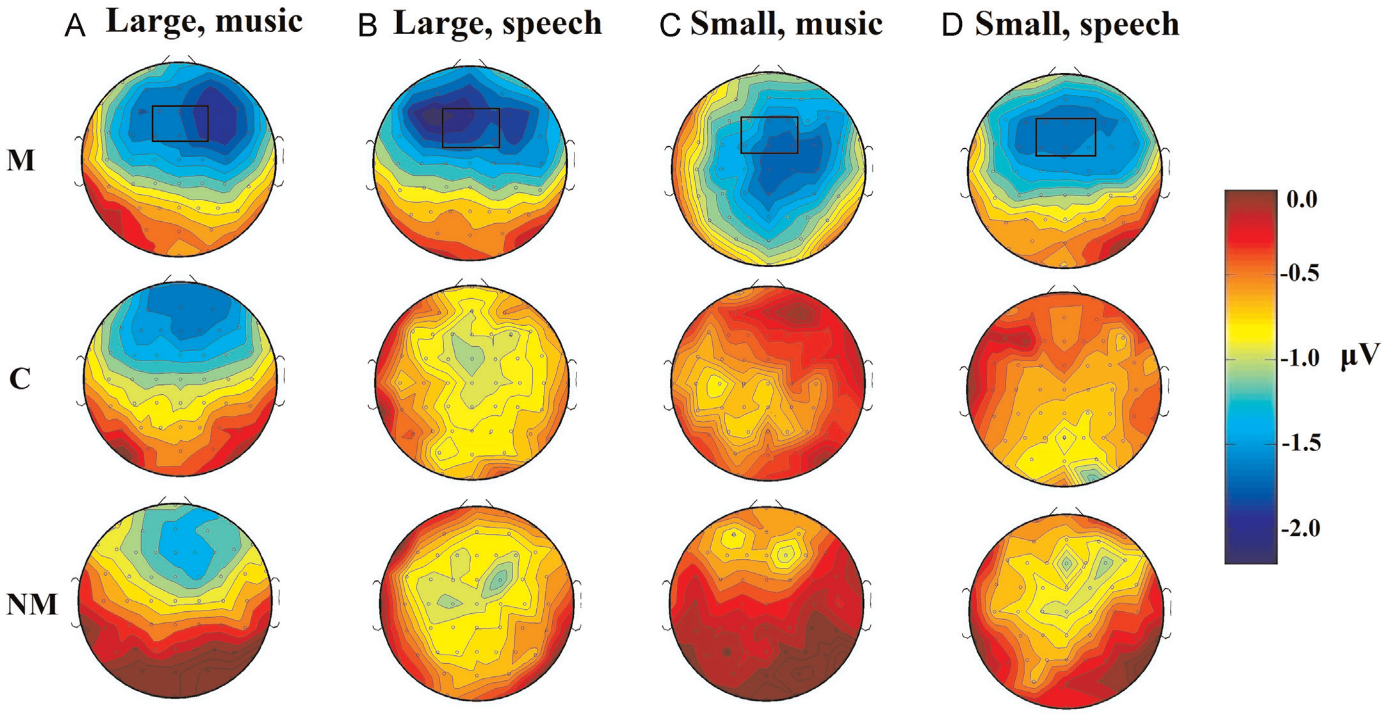

Fig. 2. Scalp topographies for the mismatch negativity (MMN) in the (a) large deviant music condition, (b) large deviant speech condition, (c) small deviant music condition, and (d) small deviant speech condition. The cluster of six electrodes is outlined on the M's topography, as this group drove the significant between-group differences in all conditions. Topographies show mean activation between two time points in each condition, centred around the group mean peak amplitude (190–200 ms for the large deviants; 200–210 ms for the small deviants).

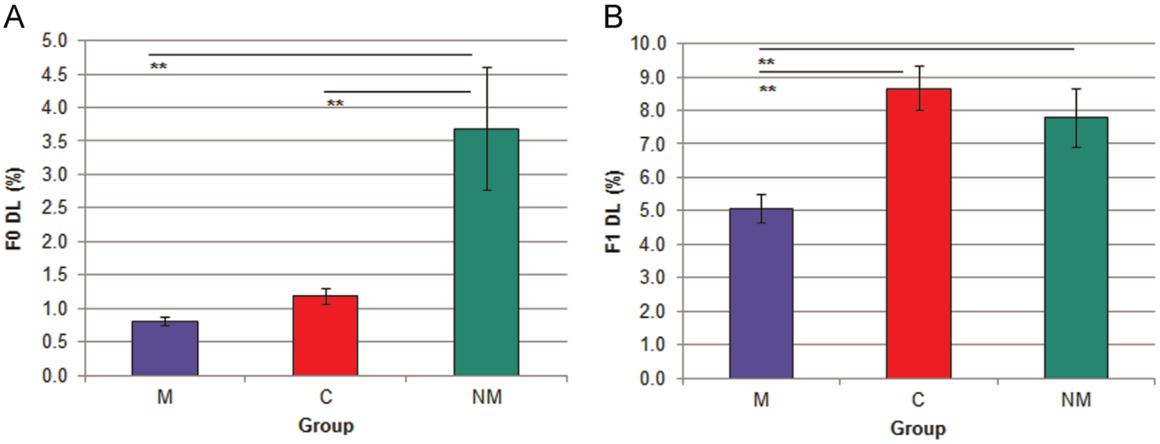

Fig. 3. Behavioral discrimination results. (A): Performance on the fundamental frequency (F0) difference limen (DL) task. Musicians (M) and Cantonese (C) showed superior pitch discrimination performance relative to nonmusician (NM) controls. (B): Performance on the first formant frequency (F1) DL task. Ms showed superior discrimination of the first formant in speech sounds, as compared to C and NM. ** $p \leq 0.01$. Error bars indicate ± 1 *SE*.

300–500 ms (LDN), respectively. All component latencies were selected based on prior research and visual inspection of the waveforms (Luck 2005; Shestakova et al., 2003).

2.5. Procedure

Participants completed the cognitive tests (Raven's Matrices and Corsi blocks), the two difference limens tasks (F0 DL and F1 DL), and then the EEG portions according to a counterbalanced order. During EEG recordings, participants sat in a comfortable chair and watched a silent movie displayed on the computer screen to promote a calm yet wakeful state. Participants were instructed to attend to the movie and ignore the sounds (i.e., passive listening). Auditory stimuli were delivered binaurally from insert earphones (ER-3A) at an intensity of 75 dB SPL. The experimental session lasted approximately 2 h.

2.5.1. Statistical analysis.

A univariate analysis of variance (ANOVA) was run for each cognitive and DL measure. Prior to statistical analyses, F0 and F1 DL values were square-root transformed to satisfy normality and homogeneity of variance assumptions required for parametric statistics.

For each of the MMN, P3a, and LDN, an ANOVA was conducted, with group as the between-subjects factor, and stimulus type (music or speech) and deviant size (small or large) as within-subjects factors. For all analyses, the dependent variable was the amplitude of a cluster of fronto-central channels (average of F1, Fz, F2, FC1, FCz, FC2). For the MMN, laterality effects were also examined, as visual inspection of the scalp topographies indicated the possibility of between-groups and between-condition differences (Fig. 2). To this end, an ANOVA was run, with group as the between-subjects variable, and stimulus type (music or speech), deviant size (small or large), and laterality (left and right)

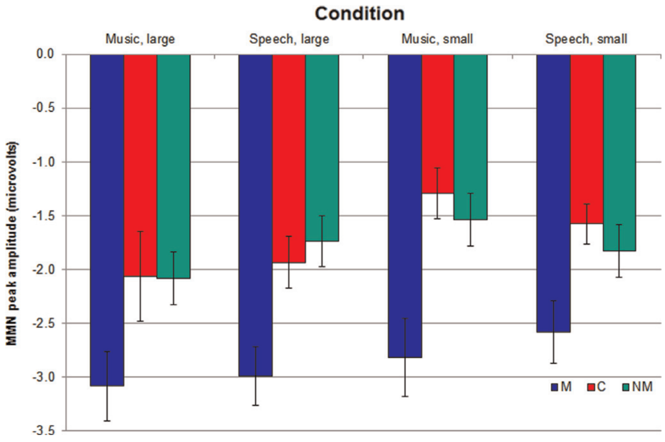

Fig. 5. Mismatch negativity (MMN) peak amplitude between 100ms to 250ms for each condition and group. The peak amplitude is the average peak of six fronto-central channels (F1, Fz, F2, FC1, FCz, FC2). Error bars indicate ± 1 *SE*.

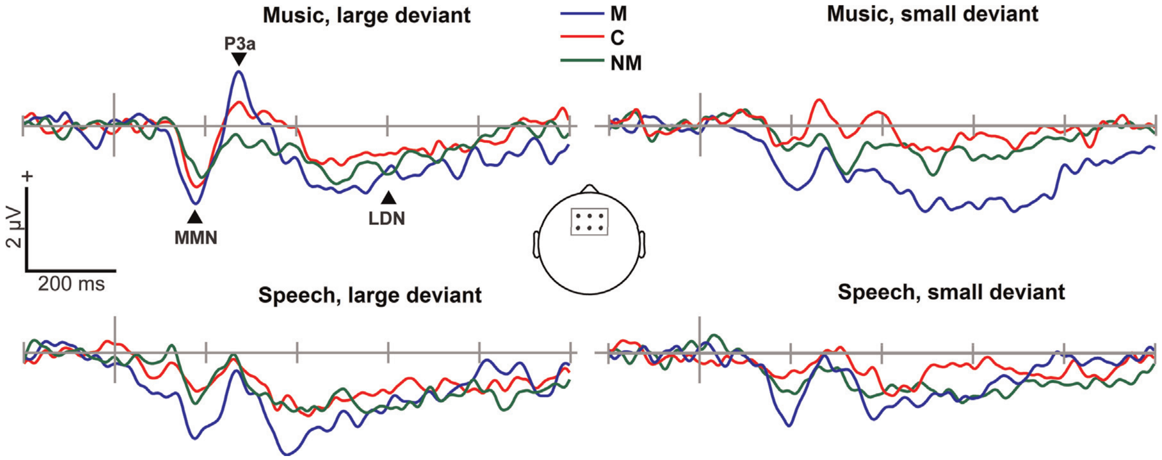

Fig. 4. ERPs difference waves (i.e., MMNs) for each group and condition. Each waveform is an average across six fronto-central channels (inset, F1, Fz, F2, FC1, FCz, FC2).

Table 1
Laterality of MMN amplitude (μV) for each group, stimulus type, and deviant size.

Group	Laterality	Stimulus type	Deviant size	*M*	*SE*
M	Left	Music	Large	−2.578	0.300
			Small	−2.496	0.334
		Speech	Large	−2.856	0.196
			Small	−2.340	0.251
	Right	Music	Large	−3.139	0.269
			Small	−2.513	0.330
		Speech	Large	−2.734	0.221
			Small	−2.395	0.236
C	Left	Music	Large	−1.959	0.374
			Small	−1.266	0.213
		Speech	Large	−1.732	0.226
			Small	−1.404	0.220
	Right	Music	Large	−2.177	0.429
			Small	−1.341	0.197
		Speech	Large	−1.792	0.212
			Small	−1.560	0.216
NM	Left	Music	Large	−1.845	0.193
			Small	−1.539	0.267
		Speech	Large	−1.585	0.224
			Small	−1.651	0.244
	Right	Music	Large	−1.962	0.182
			Small	−1.375	0.257
		Speech	Large	−1.566	0.193
			Small	−1.916	0.225

electrode cluster as within-subjects variables. The left electrode cluster was an average of a subset of left fronto-central electrodes (AF3, F3, and F5); the right cluster was an average of a right fronto-central electrodes (AF4, F4, F6).

Bonferroni corrections were applied to all pairwise contrasts to control for family-wise error ($\alpha=0.05$). When appropriate, the degrees of freedom were adjusted with the Greenhouse-Geisser epsilon (ε). All reported p-values are based on the reduced degrees of freedom, although the original degrees of freedom are reported. Partial eta-squared (η^2_p) was used as a measure of effect size for all ANOVAs. Levene's test was used to test for heterogeneity of variance between groups across all conditions, and within each condition. In all cases, homogeneity of variance assumptions were satisfied for all statistical tests ($ps > 0.05$).

By group, correlations were used to assess the degree to which listeners' auditory neural processing of speech/music predicted their perceptual acuity in each domain. Correlations were performed between all behavioural (F0 DL, F1 DL, Corsi span, and Raven's score) and brain (MMN, P3a, and LDN) measures. A false discovery rate (FDR) procedure (Benjamini and Yekutieli, 2001) was used to correct for multiple correlation tests with a threshold of $a=0.05$. FDR-corrected results are reported.

3. Results

3.1. Cognitive tests

We found no significant group differences in performance on Raven's test [$F(2, 57)=0.395$, $p=0.676$] or Corsi blocks task [$F(2, 59)=0.922$, $p=0.403$], confirming that groups were well-matched in fluid intelligence and WM abilities.

3.2. Difference limens

There was a significant group difference on F0 DL task performance [$F(2,59)=11.914$, $p<0.001$, $\eta^2_\mathrm{p}=0.295$] (Fig. 3A). Pairwise comparisons revealed that Ms performed comparably to C ($p=0.92$) and that both Ms and C outperformed NMs [M versus NM: $p<0.001$; C versus NM: $p=0.003$] (i.e., M=C > NM).

F1 DLs also differed between groups [$F(2,59)=8.937$, $p<0.001$, $\eta^2_\mathrm{p}=0.239$] (Fig. 3B). Pairwise comparisons revealed that Ms outperformed C ($p<0.001$) and NMs ($p=0.011$), achieving better (i.e., lower discrimination thresholds). C performed comparably to NMs ($p=0.776$) (i.e., M > C=NM).

3.3. ERP data

MMN scalp topographies, waveforms, and average peak amplitudes are shown for each group and stimulus condition in Figs. 4 and 5, respectively. Responses to standard and deviant traces are provided in the Supplementary information.

3.4. MMN

M had a larger MMN across all stimulus conditions, relative to C ($p<0.001$) and NM ($p<0.001$; $F(2,57)=15.71$, $p<0.001$, $\eta^2_\mathrm{p}=0.355$). For all three groups, listeners showed larger MMN (i.e., enhanced sound discrimination) for large deviants, as compared to small deviants across both music and speech stimuli [$F(1, 57)=6.453$, $p=0.014$, $\eta^2_\mathrm{p}=0.102$]. All other possible main effects, as well as two- and three-way interactions, were not significant ($p>.05$). These results indicate that M had enhanced early cortical discrimination across both music and speech sounds, relative to C and NM.

3.5. MMN laterality effects

M showed a larger MMN than C and NM across conditions, deviant sizes, and hemispheres (p's <0.001; $F(2,57)=15.61$, $p<0.001$, $\eta p^2=0.354$; see Table 1). Pooling across groups, stimulus type, and deviant size, the MMN was marginally stronger in the right than the left hemisphere [$F(1,57)=3.756$, $p=0.058$, $\eta^2_\mathrm{p}=0.062$]. Pooling across groups, condition, and laterality, large deviants elicited a larger MMN than small deviants [$F(1,57)=14.14$, $p=0.010$, $\eta^2_\mathrm{p}=0.111$].

The interaction of laterality, stimulus type, and deviant size was significant [$F(1,57)=7.91$, $p=0.007$, $\eta^2_\mathrm{p}=0.122$]. This interaction was driven by a main effect of deviant size in the right cluster [$F(1,59)=7.64$, $p=0.008$, $\eta^2_\mathrm{p}=0.115$]. Specifically, large deviants elicited stronger MMNs than small deviants. The interaction of stimulus type and deviant size was also significant [$F(1,59)=5.42$, $p=0.023$, $\eta^2_\mathrm{p}=0.084$]. For the music condition, the large deviants elicited stronger MMNs than small deviants [$F(1,59)=4.32$, $p=0.042$, $\eta^2_\mathrm{p}=0.068$]. All other possible two-, three-, or four-way interactions were not significant, $ps > 0.05$. Collectively, these results indicate that laterality did not differ between groups. Overall, the right cluster elicited a larger MMN for the larger versus the small deviant. The music condition (but not the vowel condition) elicited this same pattern of MMN results across both hemispheres.

3.6. P3a

There was a significant three-way interaction between stimulus type, deviant size, and group [$F(2, 57)=5.59$, $p=0.006$, $\eta^2_\mathrm{p}=0.164$; see Fig. S3 in the Supplemental material]. For Ms, the large deviant had a more positive P3a than the small deviant [$F(1, 20)=3.75$, $p=0.067$, $\eta^2_\mathrm{p}=0.158$]. There was also significant interaction of stimulus type and deviant size in Ms [$F(1, 20)=20.59$, $p<0.001$, $\eta_p^2=0.507$]. Namely, for the music condition, the large deviant had a more positive P3a than for the small deviant [$F(1, 20)=19.37$, $p<0.001$, $\eta_p^2=0.492$]. For the speech condition, the small deviant had a more positive P3a than for the large deviant [$F(1, 20)=5.61$, $p=0.028$, $\eta_p^2=0.219$]. For C, there was a more positive P3a for the music than the speech condition [$F(1, 17)=5.652$, $p=0.029$, $\eta_p^2=0.250$]. For NM, there were no significant main effects or interactions ($p>0.05$). These results indicate that M had stronger involuntary switching of attention for large music deviants, as compared to small music deviants (as indexed by the P3a response), while the opposite pattern was true for the speech condition in Ms. Lastly, C showed stronger involuntary switching of attention for musical sounds (i.e., pitch deviants), as compared to speech sounds, across both deviant sizes.

All other main effects, and two-way interactions including group as a variable, were not significant ($ps>0.05$). There was a significant interaction of stimulus type and deviant size [$F(1, 57)=14.40$, $p<0.001$, $\eta_p^2=0.202$]. Specifically, for the music condition, the large deviant had a more positive P3a than the small deviant [$F(1, 57)=8.90$, $p=0.004$, $\eta_p^2=0.135$] when pooled across groups. For the speech condition, the small deviant had a more positive P3a than the large deviant [$F(1, 57)=5.36$, $p=0.024$, $\eta_p^2=0.086$].

3.7. LDN

There was a significant main effect of group on LDN mean amplitude [$F(2,57)=4.56$, $p=0.015$, $\eta_p^2=0.138$; see Fig. S4 in the Supplemental material]. Specifically, Ms had a more negative LDN than C ($p=0.012$). There was no difference in LDN amplitude between Ms and NMs ($p=0.268$), or C and NMs ($p=0.555$). Pooling across groups and deviant sizes, the speech condition elicited a more negative LDN than the music condition [$F(1,57)=6.48$, $p=0.014$, $\eta_p^2=0.102$]. There was also a significant interaction of stimulus type and deviant size [$F(1,57)=8.436$, $p=0.005$, $\eta_p^2=0.129$]. This was driven by the large deviant eliciting a more negative LDN than the small deviant for the speech condition [$F(1,57)=12.69$, $p=0.001$, $\eta_p^2=0.182$]. These results indicate that across all groups and deviant sizes, the speech condition elicited greater top-down processing/re-orienting for the large deviant as compared to the small deviant. They further indicate that M employed top-down processing/re-orienting to a greater extent than C.

3.8. Correlations

Correlational analyses revealed a significant correspondence between F0 DL and F1 DL thresholds in Ms only ($r=0.61$, $p=0.004$). That is, better pitch discrimination was associated with superior timbre discrimination and vice versa. We found no correlations between behavioural and ERP measures ($ps>0.05$) following FDR correction.

4. Discussion

By comparing cortical MMN responses to music and speech in musicians and tone language speakers, we assessed possible enhancements in auditory neural processing associated with musicianship and tone language experience. Across conditions, musicians showed enhanced MMN as compared to the other groups, suggesting that musicians were better able to automatically discriminate between sounds (i.e., both music and speech), as compared to Cantonese and nonmusician listeners. Surprisingly, Cantonese language experience was not associated with any neurophysiological enhancements in music or speech processing, despite enhanced behavioral acuity for pitch. These findings suggest that different modes of intensive pitch experience, either through music or language in the present case, may have differential effects on the neural mechanisms subserving music and speech sound analysis.

There was no significant interaction of group and stimulus type (i.e., music versus speech) for any of the ERP components. Previously, behavioural melody discrimination tasks have shown that musicians are more sensitive to detecting quarter-semitone changes (i.e., the size of our small music deviants) in a melody, as compared to Cantonese speakers and NMs. For half-semitone changes (i.e., the size of our large music deviants), musicians have outperformed Cantonese speakers, who in turn outperform NMs (Bidelman et al., 2013a). Based on this data, we predicted that musicians would show larger neural discriminatory responses to the small music deviant than other groups. The lack of this effect suggests that perhaps automatic auditory neural processing may be less sensitive than behaviour in picking up changes in pitch and timbre, regardless of experience with pitch. This may have been reflected in the lack of a significant interaction of group and stimulus type.

There were also no differences in component amplitudes between music and speech when collapsing across groups for the MMN and P3a, suggesting that there was not a distinct pattern of brain activity that differentiated music and speech in the present study. For the LDN, speech sound contrasts elicited larger neural responses than for musical stimuli. These findings imply that the recruitment of top-down processing/re-orienting was perhaps more pronounced for changes in stimulus timbre compared to pitch. The former is a less salient acoustic change and may therefore necessitate the engagement of top-down processes not invoked by the more salient changes in pitch. Unlike for stimulus type, participants did differentiate between deviant sizes. Specifically, large deviants typically elicited a more pronounced response (i.e., a more negative MMN and LDN, a more positive P3a). This result is expected, as the small deviant was intended to be more difficult to differentiate from the standard, as compared to large deviants.

4.1. Behavioral auditory benefits of musicianship and language experience

At the behavioural level, we found that musicians and Cantonese participants performed comparably in pitch discrimination as compared to nonmusicians, in agreement with our prior study (Bidelman et al., 2013a). This finding further supports the notion that both musical and linguistic pitch experience similarly improve basic auditory acuity for pitch. However, musicians showed superior F1DLs compared to Cantonese and nonmusicians, suggesting that musicianship also confers benefits to speech processing and increases sensitivity for the timbral characteristics of speech. This finding is consistent with previous behavioral studies that have observed higher perceptual acuity for the timbral characteristics of speech in trained musicians (Chartrand and Belin, 2006; Bidelman and Krishnan, 2010; Bidelman et al., 2014). This result also aligns with a wealth of data supporting a music-to-language association, that is, musicianship benefiting a number of language-related domains (e.g., Chan et al., 1998; Thompson et al., 2004; Delogu et al., 2006, 2010; Slevc and Miyake, 2006; Marques et al., 2007; Franklin et al., 2008; Lee and Hung, 2008; Milovanov et al., 2009; Parbery-Clark et al., 2009a; Bidelman and Krishnan, 2010; Moreno et al., 2011; Bidelman et al., 2014).

We did not observe a behavioural advantage of tone language experience on speech timbre discrimination, as Cantonese participants performed comparably to nonmusicians. These results imply that the auditory benefits of tone languages do not necessarily extend to timbre (i.e., higher-frequency spectral cues) and might be restricted to F0 pitch. This finding may be related to the relative weighting of cues that are perceptually relevant in the Cantonese experience: Pitch is highly relevant at the lexical level, whereas change in the F1 formant is a subtler cue and common to both the English and Cantonese languages. These data align well with the proposition that domain-specific experience (e.g., tone language expertise) can benefit auditory processing in other auditory domains only if the latter exhibits acoustic features that overlap with the demands of an individual's auditory experience (Bidelman et al., 2011b).

Additional evidence supports that different types of experience (i.e., music, linguistic) with a given acoustic cue can similarly benefit perceptual processing of that cue (e.g., Bidelman et al., 2011a). For example, Marie et al. (2012) found that long-term experience with specific parameters of speech, namely duration, extended past the processing of speech sounds, resembling the effects of music expertise. Giuliano et al. (2011) found that native experience with a tone language enhanced pitch discrimination and timing of auditory cortical responses to pitch change. Similarly, Pfordresher and Brown (2009) found that tone language speakers were better able to imitate (via singing) and perceptually discriminate musical pitch. These latter findings suggest that tone language acquisition fine-tunes the processing of critical auditory dimensions in the speech signal, and that this fine-tuning can transfer to non-linguistic domains (Pfordresher and Brown, 2009; Bidelman et al., 2011b, 2013a). More generally, these studies demonstrate that when listeners' auditory experience overlaps with some aspect of the experimental stimuli, enhanced spectral acuity is observed, supporting current behavioural findings.

4.2. *Auditory neurophysiological benefits of linguistic experience and musicianship*

In line with the behavioural speech timbre discrimination data, we found no enhancement in MMN for speech timbre or musical pitch processing in tone language speakers (i.e., $\text{MMN}_\text{C}=\text{MMN}_\text{NM}$). In contrast, we observed robust neural differentiation across all conditions in musicians compared to the two other groups. A more pronounced MMN in musicians across all conditions suggests that musicianship improves auditory cortical processing in a broad manner. The lack of similar auditory neural benefits in Cantonese listeners suggests that pitch and timbral elements of music and speech are not as salient to tone language speakers as they are to musicians (e.g., Bidelman et al., 2011a, 2013a).

In addition to the superior automatic, cortical sound processing observed in musicians (i.e., enhanced MMNs), this group also showed enhanced top-down processing/reorienting (i.e., LDN), as compared to the Cantonese speakers, but not controls. Therefore, these neurophysiological findings only partially corroborate previous studies showing enhanced subcortical (i.e., automatic) auditory responses (e.g., Musacchia et al., 2007, 2008; Wong et al., 2007; Parbery-Clark et al., 2009b, 2011 Bidelman et al., 2011a,b, 2014) and enhanced LDN (i.e., attentional reorienting/higher-order auditory processing) as a result of musical experience (Putkinen et al., 2013). Furthermore, the Cantonese group did not differ significantly from controls (i.e., they did not have a more positive LDN that differentiated them from musicians *and* controls). Future studies could further investigate the auditory LDN in musicians and tone language speakers.

We also did not observe a significant difference between-group difference in P3a amplitude. Past work has found that in passive paradigms, the P3a habituates over time in musicians, while it is enhanced over time in nonmusicians (Seppanen et al., 2012). This habituation suggests that musicianship hones attentional abilities and auditory feature encoding (Seppanen et al., 2012). The current paradigm may not have lasted long enough for such habituation effects to be measured (i.e., approximately 25 min at present, versus a total of 60 min in Seppanen et al., 2012).

Examining the P3a in each group, musicians showed stronger involuntary switching of attention for large music deviants, as compared to small music deviants. The difference between deviant sizes is perhaps surprising, given past evidence that musicians can accurately identify half- and quarter-semitone changes in melody (e.g., Bidelman et al., 2013a). However, the large deviant change is more obvious (i.e., easier to detect) than the small deviant, perhaps accounting for this significant difference. Within the Cantonese group, participants had stronger involuntary switching of attention for musical sounds (i.e., pitch deviants) as indexed by the P3a, as compared to speech sounds, across both deviant sizes. This finding suggests that hearing fundamental frequency changes in a non-linguistic context elicited attention reorientation in Cantonese speakers, who regularly utilise fundamental frequency in a linguistic context. This finding also corroborates literature that has shown that different types of experience with a given acoustic cue can similarly benefit perceptual processing of that cue (e.g., Bidelman et al., 2011a; Marie et al., 2012).

4.3. *Dissociation between neural and perceptual processing of music/speech*

Tone-language speakers' behavioral enhancements for pitch processing were not paralleled in neural enhancements in the present study. This dissociation may be explained in relation to previous work, which has suggested that the engagement of cortical circuitry subserving speech/music percepts depends on the cognitive relevance of the stimulus to the listener (e.g., Halpern et al., 2008; Chandrasekaran et al., 2009; Abrams et al., 2011; Bidelman et al., 2011a,b). For example, in response to musical stimuli, information relayed from subcortical sensory structures engages higher-level cortical mechanisms subserving musical pitch perception in musicians – a process that is not engaged in tone language speakers (Bidelman et al., 2011b). Indeed, strong correlations are observed between brain and behavioural responses to musical chords for musicians but not in listeners lacking musical expertise (i.e., Cantonese and nonmusician participants; Bidelman et al., 2011b). Applying these findings to the present study, enhanced auditory neural processing (as indexed by the MMN) seems to only fully engage higher-level perceptual mechanisms in musicians (rather than in Cantonese participants or nonmusicians). Similarly, it is possible that timbral cues are more salient to musicians as compared to Cantonese or nonmusician participants (e.g., Bidelman et al., 2011a, 2013a) as a result of extensive experience with discriminating high-frequency spectral cues during music training. Higher auditory demands of music relative to language (e.g., Patel, 2011) may thus account for musicians' parallel enhancements in brain and behavioral auditory processing that is not observed in tone language speakers.

One alternative explanation for the differences observed between the neural and behavioural pitch processing data in our tone language group is that measuring mean activation over a cortical patch may not fully represent neural processes underlying their processing of pitch (Hutka et al., 2013). Musicians arguably have a greater range of experience with pitch (e.g., manipulating and producing complex melodies), as compared to tone language speakers. Thus, the information processing capacity for musicians might be different and more robust than that of tone language speakers, manifesting in enhanced neural and behavioual correlates of sound discrimination in the former group. Such effects may be more nuanced for tone language speakers. In future studies, nonlinear methods, such as measuring the brain's information processing capacity over multiple timescales, may have the capacity to better capture how tone language speakers process pitch at the neural versus behavioural level (cf. Hutka et al., 2013).

It is also notable that there was no interaction between laterality and group in any condition, suggesting that the lateralization of pitch and speech processing does not differ between tone language speakers and musicians. However, when collapsing across all other variables, MMN responses were marginally right lateralized. Previous findings show that the right hemisphere is specialized for processing fine spectral features of the auditory input present in musical stimuli, whereas the left hemisphere is more specialized for temporal processing (i.e., for speech perception; see Zatorre et al., 2002 for a review). The current data do not reflect this right lateralization for music or left-lateralization for speech (i.e., no significant interaction of stimulus type and laterality). These results suggest that participants were more focused on the processing fine spectral features of all stimuli, rather than temporal information.

4.4. *Modularity of music and speech processing*

The association between music and speech raises questions about whether language and music rely on independent systems or domain-general processes (Slevc, 2012). Our data suggest that musicians undergo fine-tuning of domain-general processes, such that sound discrimination at the neural and behavioural level is enhanced for processing changes in musical pitch and speech timbre. This aligns with literature that details musicians' enhanced general auditory processing – that is, spectral acuity above and beyond music processing (Kraus and Chandrasekaran, 2010; Moreno and Bidelman, 2014). As for the comparison of musicians and tone language speakers: The differential impact of music versus tone language experience on neurophysiological processing suggests that musicianship and language experience are associated with at least partially divergent neural networks, which are influenced differently by top-down regulation of sensory processes. That is, if the pitch expertise derived from musicianship and tone language experience shared a common neural mechanism, one would have predicted similar enhancements to ERP components in *both* musicians and Cantonese speakers. The findings of Moreno et al. (2014a) support this view of different neural networks in musicianship and linguistic expertise, such that bilinguals and musicians elicited different ERP responses when performing an inhibition task. Though inhibition is an executive function, rather than a perceptual process, the results of Moreno et al. (2014a) suggest that bilingualism and musicianship have differential effects on the neural networks supporting general cognitive skills.

4.5. *Nature versus nurture: Considering the contributions of predispositions versus training to performance*

It is notable to acknowledge that we treat musicianship – that is, training on a musical instrument-and tone language experience as comparable throughout this study. It is important to address the implications of causality inherent to this approach. Most studies that examine the effect of music training on performance make the assumption that training on a musical instrument *causes* improvements on various tasks (e.g., language-related tasks), despite being correlational or quasi-experimental (Schellenberg and Weiss, 2013). Indeed, this point is applicable to this study and most of the literature cited in the present article, with some exceptions (e.g., Thompson et al., 2004; Moreno et al., 2009, 2011; Putkinen et al., 2013; Kraus et al., 2014; Chobert et al., 2014). For example, Chobert et al. (2014) randomly assigned children to either a music or painting training group. The authors measured MMN in response to syllables of differing vowel frequency, duration, and voice onset time at three time points: Before training, after six months of music training, and after 12 months of training. Following twelve months of training, the music group showed larger MMN amplitude for syllabic duration and voice onset time, suggesting that music training enhanced automatic processing of these aspects of speech (i.e., cross-domain plasticity; Chobert et al., 2014).

Though the claim that music training causes improvement is a reasonable assumption, it is also plausible that the opposite causal direction holds – namely, that individuals with poor listening abilities (e.g., as measured by music aptitude tests) may not enroll in long-term musical training, ensuring a positive association between listening abilities and music training (Schellenberg and Weiss, 2013). Similarly, individuals predisposed with high musical aptitude may be more likely to continue with music training (Schellenberg and Weiss, 2013). In addition, individuals who perform well on listening tasks, and are thus enrolled in music lessons, may be self-selected by factors such as high IQ or enhanced cognitive abilities (Schellenberg and Weiss, 2013). In contrast with the pitch expertise associated with extensive music training, Cantonese speakers did not self-select themselves to pursue a tone language, and the pitch expertise required to speak a tone language is inarguably the result of one's direct environment. However, the finding that Cantonese participants demonstrated F0 discrimination comparable to that of musicians (and superior to native-English-speaking nonmusicians), suggests that these tone language effects are indeed experience-dependent.

In an attempt to mitigate the possibility that a third variable, such as intelligence or cognitive abilities, was driving performance in our musically-trained group, we measured fluid intelligence and short-term working memory, and matched participants' educational backgrounds. There were no significant between-group differences found for any of these measures, suggesting that neither intelligence, short-term working memory, nor educational differences were responsible for the effects observed in the current study. These findings help support that the effects observed in our musician group were indeed influenced by music training, rather than pre-existing differences. Nonetheless, it would worthwhile in future studies to asses if the observed effects hold under a longitudinal training study design.

5. Conclusion

The current study tested the degree to which musicianship and tone language experience similarly enhance the neural encoding of auditory information necessary for speech and music processing. Consistent with previous reports (Bidelman et al., 2013a), the present study found that both linguistic pitch experience and musicianship similarly enhance basic pitch discrimination acuity (as measured via F0 DLs). Only musicians showed a benefit in timbral discrimination processing (as measured by F1 DLs), as compared to tone language speakers and nonmusicians. Parallel enhancements of behavioural spectral acuity in early markers of auditory neural processing were only observed in musicians; tone language advantages for pitch discrimination observed behaviorally (e.g., Bidelman et al., 2013a) were not reflected in early cortical MMN responses to pitch changes. Our data suggest that while both extensive music and tone language experience enhance some aspects of auditory acuity (i.e., pitch discrimination), musicianship confers farther-reaching enhancements to auditory function, tuning both pitch and timbre-related brain processes.

Acknowledgements

This work was supported by grants from the Ontario Graduate Scholarship, the GRAMMY® Foundation (USA), and the Natural Sciences and Engineering Research Council of Canada Discovery Grant. We would like to thank Patrick Bermudez for his helpful comments on this draft, Yunjo Lee for her technical assistance, and Courtney Smith, Amanda Chan, Cara Gordon, and Sophia Maguire for assistance with participant testing.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuropsychologia.2015.03.019.

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. Cereb. Cortex 21, 1507–1518. http://dx.doi.org/10.1093/cercor/bhq198.

Ahissar, M., Nahum, M., Nelken, I., Hochstein, S., 2009. Reverse hierarchies and sensory learning. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364, 285–299. http://dx.doi.org/10.1098/rstb.2008.0253.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Ann. Stat. 29, 1165–1188.

Besson, M., Chobert, J., Marie, C., 2011. Transfer of training between music and speech: common processing, attention, and memory. Front Psychol. 2, 1–12. http://dx.doi.org/10.3389/fpsyg.2011.00094.

Bialystok, E., Craik, F.I., Grady, C., Chau, W., Ishii, R., Gunji, A., Pantev, C., 2005. Effect of bilingualism on cognitive control in the Simon task: evidence from MEG. Neuroimage 24, 40–49. http://dx.doi.org/10.1016/j.neuroimage.2004.09.044.

Bialystok, E., Depape, A.M., 2009. Musical expertise, bilingualism, and executive functioning. J. Exp. Psychol. Hum. Percept. Perform. 35, 565–574. http://dx.doi.org/10.1037/a0012735.

Bidelman, G.M., Krishnan, A., 2010. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 1355, 112–125. http://dx.doi.org/10.1016/j.brainres.2010.07.100.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011a. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. 23, 425–434. http://dx.doi.org/10.1162/jocn.2009.21362.

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

Bidelman, G.M., Hutka, S., Moreno, S., 2013a. 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. http://dx.doi.org/10.1371/journal.pone.0060676.

Bidelman, G.M., Moreno, S., Alain, C., 2013b. Tracing the emergence of categorical speech perception in the human auditory system. Neuroimage 79, 201–212. http://dx.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. Eur. J. Neurosci. 40, 2662–2673. http://dx.doi.org/10.1111/ejn.12627.

Brattico, E., Tervaniemi, M., Naatanen, R., Peretz, I., 2006. Musical scale properties are automatically processed in the human auditory cortex. Brain Res. 1117, 162–174. http://dx.doi.org/10.1016/j.brainres.2006.08.023.

Brattico, E., Pallesen, K.J., Varyagina, O., Bailey, C., Anourova, I., Jarvenpaa, M., Eerola, T., Tervaniemi, M., 2009. Neural discrimination of nonprototypical chords in music experts and laymen: an MEG study. J. Cogn. Neurosci. 21, 2230–2244. http://dx.doi.org/10.1162/jocn.2008.21144.

Ceponiene, R., Lepisto, T., Soininen, M., Aronen, E., Alku, P., Naatanen, R., 2004. Event-related potentials associated with sound discrimination versus novelty detection in children. Psychophysiology 41, 130–141. http://dx.doi.org/10.1111/j.1469-8986.2003.00138.x.

Chan, A.S., Ho, Y.C., Cheung, M.C., 1998. Music training improves verbal memory. Nature 396, 128.

Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. Brain Lang. 108, 1–9. http://dx.doi.org/10.1016/j.bandl.2008.02.001.

Chartrand, J.P., Belin, P., 2006. Superior voice timbre processing in musicians. Neurosci. Lett. 405, 164–167. http://dx.doi.org/10.1016/j.neulet.2006.06.053.

Chobert, J., François, 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. http://dx.doi.org/10.1093/cercor/bhs377.

Cooper, A., Wang, Y., 2012. The influence of linguistic and musical experience on Cantonese word learning. J. Acoust. Soc. Am. 131, 4756–4769. http://dx.doi.org/10.1121/1.4714355.

Corsi, P.M., 1972. Human memory and the medial temporal region of the brain [Ph. D. thesis]. Department of Psychology: McGill University.

Delogu, F., Lampis, G., 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. http://dx.doi.org/10.1007/s10339-006-0146-7.

Delogu, F., Lampis, G., Belardinelli, M.O., 2010. From melody to lexical tone: musical ability enhances specific aspects of foreign language perception. Eur. J. Cogn. Psychol. 22, 46–61. http://dx.doi.org/10.1080/09541440802708136.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Donchin, E., Coles, M.G.H., 1988. Is the P300 component a manifestation of context updating? Behav. Brain Sci. 11, 357–374. http://dx.doi.org/10.1017/S0140525 × 00058027.

Escera, C., Alho, K., Schroger, E., Winkler, I., 2000. Involuntary attention and distractibility as evaluated with event-related brain potentials. Audiol. Neurootol. 5, 151–166. http://dx.doi.org/10.1159/000013877.

Francis, A.L., Ciocca, V., Ma, L., Fenn, K., 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. J. Phon. 36, 268–294. http://dx.doi.org/10.1016/j.wocn.2007.06.005.

Franklin, M.S., Moore, K.S., Yip, C.-Y., Jonides, J., Rattray, K., Moher, J., 2008. The effects of musical training on verbal memory. Psychol. Music 36, 353–365. http://dx.doi.org/10.1177/0305735607086044.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004. Musical training enhances automatic encoding of melodic contour and interval structure. J. Cogn. Neurosci. 16, 1010–1021. http://dx.doi.org/10.1162/0898929041502706.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2005. Automatic encoding of polyphonic melodies in musicians and nonmusicians. J. Cogn. Neurosci. 17, 1578–1592. http://dx.doi.org/10.1162/089892905774597263.

Gandour, J., 1981. Perceptual dimensions of tone: evidence from Cantonese. J. Chin. Linguist. 9, 20–36.

Giuliano, R.J., Pfordresher, P.Q., Stanley, E.M., Narayana, S., Wicha, N.Y., 2011. Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Front. Psychol. 2, 1–12. http://dx.doi.org/10.3389/fpsyg.2011.00146.

Gold, B.T., Kim, C., Johnson, N.F., Kryscio, R.J., Smith, C.D., 2013. Lifelong bilingualism maintains neural efficiency for cognitive control in aging. J. Neurosci. 33, 387–396. http://dx.doi.org/10.1523/jneurosci.3837-12.2013.

Halpern, A.R., Martin, J.S., Reed, T.D., 2008. An ERP study of major-minor classification in melodies. Music Percept. 25, 181–191. http://dx.doi.org/10.1525/mp.2008.25.3.181.

Horvath, J., Roeber, U., Schroger, E., 2009. The utility of brief, spectrally rich, dynamic sounds in the passive oddball paradigm. Neurosci. Lett. 461, 262–265. http://dx.doi.org/10.1016/j.neulet.2009.06.035.

Hutka, S., Bidelman, G.H., Moreno, S., 2013. Brain signal variability as a window into the bidirectionality between music and language processing: moving from a linear to a nonlinear model. Front. Psychol. 4, 1–11. http://dx.doi.org/10.3389/fpsyg.2013.00984.

Kraus, N., Slater, J., Thompson, E.C., Hornickel, J., Strait, D.L., Nicol, T., White-Schwoch, T., 2014. Music enrichment programs improve the neural encoding of speech in at-risk children. J. Neurosci. 34, 11913–11918. http://dx.doi.org/10.1523/JNEUROSCI.1881-14.2014.

Khouw, E., Ciocca, V., 2007. Perceptual correlates of Cantonese tones. J. Phon. 35, 104–117. http://dx.doi.org/10.1016/j.wocn.2005.10.003.

Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. Acoust. Soc. Am. 87, 820–857. http://dx.doi.org/10.1121/1.398894.

Koelsch, S., Schroger, E., Tervaniemi, M., 1999. Superior pre-attentive auditory processing in musicians. Neuroreport 10, 1309–1313. http://dx.doi.org/10.1097/00001756-199904260-00029.

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, 956–966. http://dx.doi.org/10.1006/nimg.2002.1154.

Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599–605. http://dx.doi.org/10.1038/nrn2882.

Lee, C.Y., Hung, T.H., 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. J. Acoust. Soc. Am. 124, 3235–3248. http://dx.doi.org/10.1121/1.2990713.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477. http://dx.doi.org/10.1121/1.1912375.

Li, P., Sepanski, S., Zhao, X., 2006. Language history questionnaire: A Web-based interface for bilingual research. Behav. Res. Methods 38, 202–210. http://dx.doi.org/10.3758/bf03192770.

Luck, S.J., 2005. An Introduction to the Event-related Potential Technique. MIT press, Cambridge, MA.

Luk, G., Anderson, J.A., Craik, F.I., Grady, C., Bialystok, E., 2010. Distinct neural correlates for two types of inhibition in bilinguals: response inhibition versus interference suppression. Brain Cogn. 74, 347–357. http://dx.doi.org/10.1016/j.bandc.2010.09.004.

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. http://dx.doi.org/10.1016/s1053-8119(00)90990-x.

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

Marie, C., Delogu, F., Lampis, G., Belardinelli, M.O., Besson, M., 2011b. Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. J. Cogn. Neurosci. 23, 2701–2715. http://dx.doi.org/10.1162/jocn.2010.21585.

Marie, C., Kujala, T., Besson, M., 2012. Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. Cortex 48, 447–457. http://dx.doi.org/10.1016/j.cortex.2010.11.006.

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

Milovanov, R., Huotilainen, M., Esquef, P.A., Alku, P., Valimaki, V., Tervaniemi, M., 2009. The role of musical aptitude and language skills in preattentive duration processing in school-aged children. Neurosci. Lett. 460, 161–165. http://dx.doi.org/10.1016/j.neulet.2009.05.063.

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. Cereb. Cortex 19, 712–723. http://dx.doi.org/10.1093/cercor/bhn120.

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. Psychol. Sci. 22, 1425–1433. http://dx.doi.org/10.1177/0956797611416999.

Moreno, S., Bidelman, G.M., 2014. Examining neural plasticity and cognitive benefit through the unique lens of musical training. Hear. Res. 308, 84–97. http://dx.doi.org/10.1016/j.heares.2013.09.012.

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., Bialystok, E., 2014a. Inhibitory control in bilinguals and musicians: event related potential (ERP) evidence for experience-specific effects. PLoS One 9, e94169. http://dx.doi.org/10.1371/journal.pone.0094169.

Moreno, S., Lee, Y., Janus, M., Bialystok, E., 2014b. Short-term second language and music training induces lasting functional brain changes in early childhood. Child Dev. . http://dx.doi.org/10.1111/cdev.12297

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. USA 104, 15894–15898. http://dx.doi.org/10.1073/pnas.0701498104.

Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear. Res. 241, 34–42. http://dx.doi.org/10.1016/j.heares.2008.04.013.

Naatanen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. Clin. Neurophysiol. 115, 140–144. http://dx.doi.org/10.1016/j.clinph.2003.04.001.

Naatanen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590. http://dx.doi.org/10.1016/j.clinph.2007.04.026.

Oostenveld, R., Praamstra, P., 2001. The five percent electrode system for high-resolution EEG and ERP measurements. Clin. Neurophysiol. 112, 713–719. http://dx.doi.org/10.1016/s1388-2457(00)00527-7.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12, 169–174. http://dx.doi.org/10.1097/00001756-200101220-00041.

Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009a. Musician enhancement for speech-in-noise. Ear Hear. 30, 653–661. http://dx.doi.org/10.1097/aud.0b013e3181b412e9.

Parbery-Clark, A., Skoe, E., Kraus, N., 2009b. Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci. 29, 14100–14107. http://dx.doi.org/10.1523/jneurosci.3256-09.2009.

Parbery-Clark, A., Strait, D.L., Kraus, N., 2011. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. Neuropsychologia 49, 3338–3345. http://dx.doi.org/10.1016/j.neuropsychologia.2011.08.007.

Patel, A.D., 2008. Music, Language, and the Brain. Oxford University Press, Oxford; New Yorkhttp://dx.doi.org/10.1093/acprof:oso/9780195123753.001.0001, xi, 513 p. p..

Patel, A.D., 2011. Why would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis. Front. Psychol. 2, 1–14. http://dx.doi.org/10.3389/fpsyg.2011.00142.

Patel, A.D., 2014. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hear. Res. 308, 98–108. http://dx.doi.org/10.1016/j.heares.2013.08.011.

Peng, G., 2006. Temporal and tonal aspects of Chinese syllables: a corpus-based comparative study of Mandarin and Cantonese. J. Chin. Linguist. 34, 134.

Peterson, G.E., Barney, H.L., 1952. Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24, 175–184. http://dx.doi.org/10.1121/1.1917300.

Pfordresher, P.Q., Brown, S., 2009. Enhanced production and perception of musical pitch in tone language speakers. Atten. Percept. Psychophys. 71, 1385–1398. http://dx.doi.org/10.3758/app.71.6.1385.

Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. 118, 2128–2148. http://dx.doi.org/10.1016/j.clinph.2007.04.019.

Putkinen, V., Tervaniemi, M., Huotilainen, M., 2013. Informal musical activities are linked to auditory discrimination and attention in 2-3-year-old children: an event-related potential study. Eur. J. Neurosci. 37, 654–661. http://dx.doi.org/10.1111/ejn.12049.

Raven, J., Raven, J.C., Court, J.H., 1998. Advanced Progressive Matrices. Harcourt Assessment, San Antonio, TX.

Schellenberg, E.G., Weiss, M.W., 2013. Music and cognitive abilities In: Deutsch, D. (Ed.), The Psychology of music, third ed. Elsevier, Amsterdam, pp. 499–550. http://dx.doi.org/10.1016/b978-0-12-381460-9.00012-2.

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

Seppanen, M., Pesonen, A.K., Tervaniemi, M., 2012. Music training enhances the rapid plasticity of P3a/P3b event-related brain potentials for unattended and attended target sounds. Atten. Percept. Psychophys. 74, 600–612. http://dx.doi.org/10.3758/s13414-011-0257-9.

Shestakova, A., Huotilainen, M., Ceponiene, R., Cheour, M., 2003. Event-related potentials associated with second language learning in children. Clin. Neurophysiol. 114, 1507–1512. http://dx.doi.org/10.1016/s1388-2457(03)00134-2.

Slevc, L.R., 2012. Language and music: sound, structure, and meaning. WIREs Cogn. Sci. 3, 483–492. http://dx.doi.org/10.1002/wcs.1186.

Slevc, L.R., Miyake, A., 2006. Individual differences in second-language proficiency: does musical ability matter? Psychol. Sci. 17, 675–681. http://dx.doi.org/10.1111/j.1467-9280.2006.01765.x.

Slevc, L.R., Rosenberg, J.C., Patel, A.D., 2009. Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. Psychon. Bull. Rev. 16, 374–381. http://dx.doi.org/10.3758/16.2.374.

Tervaniemi, M., Rytkonen, M., Schroger, E., Ilmoniemi, R.J., Naatanen, R., 2001. Superior formation of cortical memory traces for melodic patterns in musicians. Learn. Mem. 8, 295–300. http://dx.doi.org/10.1101/lm.39501.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schroger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. Exp. Brain Res. 161, 1–10. http://dx.doi.org/10.1007/s00221-004-2044-5.

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

Wetzel, N., Widmann, A., Berti, S., Schroger, E., 2006. The development of involuntary and voluntary attention from childhood to adulthood: a combined behavioral and event-related potential study. Clin. Neurophysiol. 117, 2191–2203. http://dx.doi.org/10.1016/j.clinph.2006.06.717.

Wong, P., Perrachione, T.K., 2007. Learning pitch patterns in lexical identification by native English-speaking adults. Appl. Psycholinguist. 28, 565–585, doi: 0.1017/s0142716407070312.

Wong, P.C., 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. http://dx.doi.org/10.1038/nn1872.

Zatorre, R.J., Belin, P., Penhune, V.B., 2002. Structure and function of auditory cortex: music and speech. TiCS 6, 37–46, doi: 101016/S11364-6613(00)01816-7.