

Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians

Gavin M. Bidelman,^{1,2} Michael W. Weiss,³ Sylvain Moreno^{4,5} and Claude Alain^{4,5,6}

¹Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA

²School of Communication Sciences & Disorders, University of Memphis, 807 Jefferson Ave. Memphis, TN 38105, USA

³Department of Psychology, University of Toronto, Mississauga, ON, Canada

⁴Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, ON, Canada

⁵Department of Psychology, University of Toronto, Toronto, ON, Canada

⁶Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

Keywords: auditory event-related potentials, brainstem frequency-following response, categorical speech perception, experience-dependent plasticity, music-to-language transfer effects

Abstract

Musicianship is associated with neuroplastic changes in brainstem and cortical structures, as well as improved acuity for behaviorally relevant sounds including speech. However, further advance in the field depends on characterizing how neuroplastic changes in brainstem and cortical speech processing relate to one another and to speech-listening behaviors. Here, we show that subcortical and cortical neural plasticity interact to yield the linguistic advantages observed with musicianship. We compared brainstem and cortical neuroelectric responses elicited by a series of vowels that differed along a categorical speech continuum in amateur musicians and non-musicians. Musicians obtained steeper identification functions and classified speech sounds more rapidly than non-musicians. Behavioral advantages coincided with more robust and temporally coherent brainstem phase-locking to salient speech cues (voice pitch and formant information) coupled with increased amplitude in cortical-evoked responses, implying an overall enhancement in the nervous system's responsiveness to speech. Musicians' subcortical and cortical neural enhancements (but not behavioral measures) were correlated with their years of formal music training. Associations between multi-level neural responses were also stronger in musically trained listeners, and were better predictors of speech perception than in non-musicians. Results suggest that musicianship modulates speech representations at multiple tiers of the auditory pathway, and strengthens the correspondence of processing between subcortical and cortical areas to allow neural activity to carry more behaviorally relevant information. We infer that musicians have a refined hierarchy of internalized representations for auditory objects at both pre-attentive and attentive levels that supplies more faithful phonemic templates to decision mechanisms governing linguistic operations.

Introduction

Recent behavioral and neuroimaging studies demonstrate that musicianship promotes functional plasticity across multiple sensory modalities benefiting a wide array of perceptual-cognitive abilities (Herholz & Zatorre, 2012; Moreno & Bidelman, 2014). Musicianship has been linked to enrichments in the ability to parse, discriminate and recognize speech (Chartrand & Belin, 2006; Bidelman & Krishnan, 2010), as well as higher phonological awareness (Anvari *et al.*, 2002; Slevc & Miyake, 2006) and second language-learning proficiency (Slevc & Miyake, 2006; Cooper & Wang, 2012). Neurophysiological studies have revealed functional changes in both subcortical (Wong *et al.*, 2007; Musacchia *et al.*, 2008; Bidelman & Krishnan,

2010; Parbery-Clark *et al.*, 2012) and cortical (Shahin *et al.*, 2003; Musacchia *et al.*, 2008) speech processing in musicians. These studies support the notion that brain mechanisms that govern important facets of human communication are primed in musician listeners.

Although acoustically distinct, speech sounds with similar features are typically identified categorically, i.e. they are heard as belonging to one of only a few discrete phonetic classes (Pisoni & Luce, 1987). Categorical perception (CP) emerges early in life (Eimas *et al.*, 1971) and is further modified based on one's native language (Kuhl *et al.*, 1992), suggesting that the neural mechanisms underlying CP are malleable to experiential factors. CP requires a higher-order linguistic abstraction and offers an ideal means to probe how other forms of auditory experience (e.g., musicianship) might alter this fundamental mode of speech perception.

We sought to further characterize musicians' hierarchy of auditory plasticity by directly examining how brainstem and cortical-evoked

Correspondence: Dr G. M. Bidelman, ²School of Communication Sciences & Disorders, as above. E-mail: g.bidelman@memphis.edu

Received 12 December 2013, revised 16 April 2014, accepted 18 April 2014

responses to speech relate to phonetic identification and CP abilities. Enhanced speech identification in musicians may be related to an improved ability to map stimulus features to phonetic meaning, an important requisite of many linguistic skills, including reading, writing and language acquisition (Eimas *et al.*, 1971; Werker & Tees, 1987). Demonstrating that musicianship is indeed associated with improved sound-to-meaning relations would support the notion that music instruction might be used as an effective catalyst for increasing early verbal proficiency (Ho *et al.*, 2003; Moreno *et al.*, 2011). It would also clarify the behavioral consequences of musicians' enhanced brainstem and cortical processing (Musacchia *et al.*, 2008) by demonstrating a direct link between these multi-level neural enhancements and complex speech-listening skills.

To this end, we measured both brainstem and cortical event-related brain potentials (ERPs) to categorically perceived speech sounds in musicians and non-musicians. While previous studies have examined speech representations in the brainstem and cortex (Musacchia *et al.*, 2008; Bidelman *et al.*, 2013b) and how musical experience shapes these neural responses (Musacchia *et al.*, 2008), no study has yet examined music-related benefits to CP, its underlying neural correlates, nor how these multi-level brain enhancements relate to speech perception. Comparing brainstem and cortical speech representations also allows us to reveal the dynamic, hierarchical brain processing that operates on the flow of information from sensory to cognitive facets of the speech network. This allows for a more complete picture of potential music-related plasticity that would be unavailable by recording brainstem or cortical ERPs alone. While a cross-sectional group comparison cannot fully disentangle the effects of 'training' from possible preexisting differences in auditory processing, comparing experts and non-experts is a necessary first step toward future longitudinal studies aiming to establish a causal link between musical training and speech-listening abilities. We hypothesized that musicians' behavioral benefits for speech processing result not only in local signal enhancements within both subcortical and cortical stages of the auditory system (cf., Musacchia *et al.*, 2008) but, critically, a 'coordination' (i.e., interaction) between these lower- and higher-order auditory brain areas subserving speech processing.

Materials and methods

Participants

Twenty-four young adults participated in the experiment: 12 English-speaking musicians (eight female) and 12 non-musicians (eight female). Each participant completed music (Wong & Perrachione, 2007) and language history (Li *et al.*, 2006) questionnaires to assess linguistic and musical background, respectively. Musicians (M) were defined as amateur instrumentalists who had received ≥ 7 years of continuous private instruction on their principal instrument (mean \pm SD; 13.6 ± 4.5 years), beginning prior to age 13 years (7.7 ± 3.5 years; Table 1). Beyond formal private or group lessons, each was currently active in music practice or ensemble engagement. The majority of musicians had advanced musical training (i.e., undergraduate or graduate degrees in music) and practiced on a daily basis. These inclusion criteria are consistent with similar definitions for 'musicians' used in previous studies from our lab and others examining the neuroplastic effects of musicianship on auditory processing (Wong *et al.*, 2007; Chandrasekaran *et al.*, 2009; Parbery-Clark *et al.*, 2009; Zendel & Alain, 2009; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011a; Cooper & Wang, 2012). Requiring musicians to have ≥ 7 years of training also guarantees

long-term experience-dependent plasticity and the potential of observing transfer effects of musicianship to language processing (Chandrasekaran *et al.*, 2009; Bidelman *et al.*, 2011a; Skoe & Kraus, 2012). Non-musicians (NM) had no more than 2 years of self-directed music training (0.4 ± 0.7 years), and had not received instruction within the past 5 years. All participants were right-handed (Oldfield, 1971), exhibited normal hearing (i.e., ≤ 25 dB HL; 500–4000 Hz) and reported no history of neurological disorders. The two groups were also closely matched in age (M: 23.8 ± 4.2 years, NM: 24.8 ± 2.7 years; $t_{22} = 0.63$, $P = 0.54$) and years of formal education (M: 17.3 ± 2.4 years, NM: 17.8 ± 2.0 years; $t_{22} = 0.60$, $P = 0.55$).

All participants spoke Canadian English and had minimal exposure to a second language (L2). Those familiar with a second language (mainly French and Polish) were classified as late bilinguals with low L2 proficiency, i.e. their non-native language was used for no more than 10% of their daily communication. Tone language experience (e.g., Mandarin Chinese, Thai) is known to enhance neural (brainstem/cortical auditory ERPs) and behavioral auditory responses (Chandrasekaran *et al.*, 2009; Bidelman *et al.*, 2011a,b, 2013a). As such, participants with any exposure to a tonal language were excluded from the study to avoid a potential confound of linguistic pitch expertise. All experiments were undertaken with the understanding and written consent of each participant in compliance with the Declaration of Helsinki and a protocol approved by the Baycrest Centre Research Ethics Committee.

Speech vowel continuum

The perception of speech tends to be categorical, such that gradually morphed sounds along a large acoustic continuum are heard as

TABLE 1. Musical demographics of participants

Participant	Instrument(s)	Years of music training	Age of onset (years)
Musicians			
M1	Organ	7	7
M2	Piano/trombone	16	5
M3	Harp	10	13
M4	Violin/piano	13	8
M5	Double bass/bass guitar	11	13
M6	Saxophone	12	6
M7	Piano/guitar	12	8
M8	Cello/piano	17	2
M9	Piano	18	2
M10	Piano/trumpet	12	8
M11	Violin/voice	24	5
M12	Piano/voice	11	12
Mean (SD)		13.6 (4.5)	7.7 (3.5)
Non-musicians			
NM1	Piano	2	12
NM2	Violin	1	12
NM3	Clarinet	1	13
NM4	Piano	1	8
NM5	–	0	–
NM6	–	0	–
NM7	–	0	–
NM8	–	0	–
NM9	–	0	–
NM10	–	0	–
NM11	–	0	–
NM12	–	0	–
Mean (SD)		0.4 (0.7)	11.3 (2.2)*

*The age of onset statistics for non-musicians were computed from the four participants with minimal musical training.

belonging to one of only a few discrete phonetic classes (Liberman *et al.*, 1967; Pisoni, 1973; Harnad, 1987; Pisoni & Luce, 1987). At extreme ends of the continuum, tokens are perceived as having the same phonetic identity or category. Despite equal physical spacing between adjacent stimuli, a sharp change in perception occurs near the midpoint of the continuum where the identity abruptly changes. CP is typically studied using stop consonants that differ minimally in voice-onset time (VOT), i.e. the initial ~ 40 ms containing critical formant transitions into the vowel (Pisoni, 1973; Sharma & Dorman, 1999). Though less salient than stop consonants, steady-state vowels alone can be perceived categorically by simply manipulating individual formant frequencies, prominent acoustic cues that determine speech identity (Pisoni, 1973). Vowels provide critical information to determine what is being said (Assmann & Summerfield, 1990) and thus allow us to assess the neural correlates of these critical cues that characterize speech. In the present study, utilizing vowels also ensured that the entire stimulus contributed to the CP rather than only the initial transient onset (cf., VOT stimuli), thereby maximizing the possibility that ERPs could be used to differentiate phonetic-level information. A five-step synthetic vowel continuum was constructed where 100-ms tokens differed minimally acoustically, but were perceived categorically (Pisoni, 1973; Bidelman *et al.*, 2013b). Tokens contained identical voice fundamental (F0), second (F2) and third formant (F3) frequencies (F0: 100 Hz; F2: 1090 Hz; and F3: 2350 Hz, respectively). The critical stimulus variation was achieved by parameterizing first formant (F1) over five equal steps between 430 and 730 Hz such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/. Acoustic spectrograms of the stimuli are shown in Fig. 1.

Data acquisition and preprocessing

Data acquisition and response evaluation were similar to previous reports from our laboratory (e.g., Bidelman *et al.*, 2013b). Stimuli were delivered binaurally at an intensity of 83 dB SPL through insert earphones (ER-3A). Extended acoustic tubing (50 cm) was used to eliminate electromagnetic stimulus artifact from contaminating neurophysiological responses (Aiken & Picton, 2008; Campbell *et al.*, 2012). The effectiveness of this control was confirmed by the absence of an artifact (and brainstem response) during a control run in which the air tubes were blocked to the ear. Listeners heard 200 randomly ordered exemplars of each token and were asked to label them with a binary response as quickly as possible ('u' or 'a'). The inter-stimulus interval (ISI) was jittered randomly between 400 and 600 ms (20-ms steps, rectangular distribution). An additional 2000 trials (ISI = 150 ms) were then collected in order to detect sub-

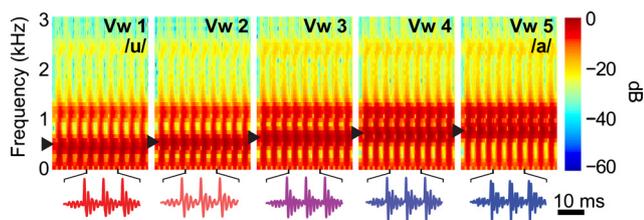


FIG. 1. Categorical speech vowel continuum. Spectrograms of the individual tokens; bottom insets show three periods of individual time-waveforms. First formant frequency was parameterized over five equal steps from 430 Hz to 730 Hz (arrows), such that the resultant stimulus set spanned a perceptual phonetic continuum from /u/ to /a/.

microvolt brainstem ERPs (Bidelman & Krishnan, 2010). A majority of studies demonstrate that early brainstem responses are unaffected by attention (Picton *et al.*, 1971; Picton & Hillyard, 1974; Woods & Hillyard, 1978; Hillyard & Picton, 1979; Galbraith & Kane, 1993; Rinne *et al.*, 2007; Okamoto *et al.*, 2011). Thus, participants watched a self-selected movie with subtitles during brainstem recording to maintain a calm and wakeful state.

Electroencephalograms (EEGs) were recorded differentially between an electrode placed on the high forehead at the hairline (~ Fpz) referenced to linked mastoids. This vertical montage is optimal for recording evoked responses of both subcortical and cortical origin (Musacchia *et al.*, 2008; Krishnan *et al.*, 2012; Bidelman *et al.*, 2013b). Inter-electrode impedance was kept below ≤ 3 k Ω . EEGs were digitized at 20 kHz with a 0.05–3500-Hz passband (NeuroScan SymAmps2). Traces were segmented (cortical ERP: –100 to 600 ms; brainstem ERPs: –40 to 210 ms), baselined to the respective pre-stimulus period, and subsequently averaged in the time domain to obtain ERPs for each condition (EEGLAB; Delorme & Makeig, 2004). Trials $> \pm 50$ μ V were rejected as blinks prior to averaging. Grand averaged evoked responses were then bandpass filtered (80–2500 Hz or 1–30 Hz) to isolate brainstem and cortical ERPs, respectively (Musacchia *et al.*, 2008; Bidelman *et al.*, 2013b).

Behavioral data

Individual vowel identification scores were fit with a two-parameter sigmoid function. We used standard logistic regression: $P = 1 / [1 + e^{-\beta_1(x-\beta_0)}]$, where P is the proportion of trials identified as a given vowel, x the step number along the stimulus continuum, and β_0 and β_1 the location and slope of the logistic fit estimated using non-linear least-squares regression. Comparing parameters between groups revealed possible differences in the location and ‘steepness’ (i.e., rate of change) of the categorical speech boundary as a function of musicianship. Behavioral speech-labeling speeds [i.e., reaction times (RTs)] were computed as the listener’s mean response latency across trials for a given condition. RTs outside 250–1000 ms were deemed outliers and excluded from further analysis (Bidelman *et al.*, 2013b).

ERP response analysis

Brainstem responses

Fast Fourier transforms (FFTs) were computed from the steady-state portion of brainstem time-waveforms (0–100 ms) to assess spectral magnitudes contained in each response. ‘Neural pitch salience’ was then estimated from each spectrum using a harmonic template analysis (for details, see Bidelman & Krishnan, 2009; Supplemental methods). The salience magnitude corresponding to a template F0 of 100 Hz (i.e., the voice pitch of our speech stimuli) was taken as a singular measure of brainstem voice ‘pitch’ encoding for each vowel condition. For quantifying pitch-relevant activity, this type of analysis (incorporating both F0 and its harmonics) is preferable to simply measuring F0 in isolation given the fact that listeners combine information across harmonics to construct a unitary pitch percept (Gockel *et al.*, 2007).

We also quantified ‘voice timbre’ encoding by measuring the F1 magnitude in each brainstem response spectra. F1, the primary cue used in the behavioral task, reflects how well the brainstem transcribes this important timbral feature of speech. F1 magnitudes could not be directly measured from the FFTs because their fre-

quencies are not necessarily harmonics of the F0 (i.e., integer multiple of 100 Hz). To this end, F1 magnitudes were instead quantified from each brainstem ERP as the amplitude of the responses' spectral envelope, computed via linear predictive coding, between 400 and 750 Hz, i.e., the expected F1 range from the input stimulus (Bidelman *et al.*, 2013b; see Supplemental methods for details).

Quantifying both F0 and F1 speech cues provided a means to replicate and extend prior research by showing group difference in both F0 and F1 encoding. Each cue also offered a clear prediction as to the type of neural enhancement we expected to observe. Because all stimuli had the same acoustic F0 frequency and amplitude, we 'did not' expect brainstem F0 salience to be modulated by stimulus condition. However, we 'did' expect F0 encoding to differ between groups, given previously reported musician advantages for voice pitch processing (Wong *et al.*, 2007; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011a). In contrast to F0, brainstem F1 encoding was expected to vary both between stimuli and groups because of the perceptual relevance of F1 to our CP task and the fact that it varied along the stimulus continuum. Stimulus-related changes in F1 but not F0 would further support the notion that these cues are largely independent in the neural encoding and perception of speech (e.g., Bidelman & Krishnan, 2010).

Cortical responses

Peak amplitude and latency were measured for the prominent deflections of the cortical ERPs (Pa, P1, N1, P2, P3) in specific time windows. Pa was taken as the positivity between 25 and 35 ms, P1 as the positivity between 60 and 80 ms, N1 the negative-going trough between 90 and 110 ms, P2 as the positive-going peak between 150 and 250 ms, and P3 as the positivity between 275 and 375 ms (Irimajiri *et al.*, 2005; Bidelman *et al.*, 2013b). While all five waves were quantified, we had specific hypotheses regarding how the N1 and P2 waves would be modulated by changes in speech F1 formant and musicianship. Prior work has shown that of the obligatory ERPs, the N1 and P2 waves are the most sensitive to speech perception tasks (Wood *et al.*, 1971; Alain *et al.*, 2007, 2010; Ben-David *et al.*, 2011; Bidelman *et al.*, 2013b), and prone to the plastic effects of speech sound training (Tremblay *et al.*, 2001; Alain *et al.*, 2007) and long-term musical experience (Shahin *et al.*, 2003; Seppänen *et al.*, 2012). Additionally, our previous study suggested that the neural correlates of CP emerge around the timeframe of N1 and are fully manifested by P2 (Bidelman *et al.*, 2013b). Thus, we focus our primary analyses on these two deflections. Individual N1 and P2 analyses allowed us to further assess how musicianship and vowel stimulus alter each of these early cortical responses. ERP analysis and automated peak selection were performed using custom routines coded in MATLAB[®] 7.12 (The MathWorks, Natick, MA, USA).

Statistical analyses

Unless otherwise specified, two-way, mixed-model ANOVAs were conducted on all dependent variables (SAS[®] 9.3, SAS Institute, Cary, NC, USA). Group (2 levels; M, NMs) functioned as the between-subjects factor; vowel stimulus [5 levels; vowel (vw) 1–5] as the within-subjects factor; subjects nested within group served as a random factor. Tukey–Kramer multiple comparisons controlled Type I error inflation. The significance level was set at $\alpha = 0.05$.

Multiple least-squares regression was also used to determine the extent to which brainstem and cortical ERPs could predict each group's behavioral CP for speech. We constructed a regression model (per group) consisting of both simple main effects as well as an inter-

action term: $\psi_{IDspeed} = \beta_0 + \beta_1 BS_{erp} + \beta_2 C_{erp} + \beta_3 BS_{erp} * C_{erp}$, where ψ represents a listener's behavioral speech classification speed (i.e., RT), BS_{erp} is the magnitude of brainstem encoding, and C_{erp} is the cortical response to speech. β_1 , β_2 , and β_3 represent to-be-estimated scalar coefficients, computed via least-squares analysis, for the weighting of each of these neural factors in the regression model (β_0 = intercept parameter). Regression coefficients were standardized (total variance = 1) to equate the scales between variables and allow us to estimate their individual predictive power on speech identification performance. Adjusted R^2 was used to assess model fits, which increase only if additional terms improve a model more than expected by chance. Additionally, pairwise correlations were used to explore the correspondence between subcortical and cortical speech representations (brainstem: F1 amplitude; cortical: N1 and P2 magnitudes), as well as the link between these brain indices and behavioral speech identification performance.

Results

Behavioral speech identification

Behavioral speech identification functions are shown in Fig. 2. Despite the continuous F1 acoustic change, listeners heard a clear perceptual shift in the phonetic category (/u/ vs. /a/) near token 3. The overall location of the perceptual boundary did not differ between groups [independent samples *t*-test (two-tailed) on β_0 parameter; $t_{22} = 0.47$, $P = 0.65$]. Yet, musicians demonstrated a considerably sharper perceptual boundary (β_1 parameter; $t_{22} = 2.45$, $P = 0.023$) than non-musician listeners (Fig. 2B). When considering musical training as a continuous variable across groups, years of musical training (measured in years post-onset) predicted the steepness of the perceptual boundary (Spearman's: $r_{all} = 0.49$, $P_{all} < 0.001$). That is, perceptual dichotomy between vowel classes was more distinct with increased musical experience when considering all listeners in our sample (Fig. 2C). However, correlations by group were marginal for musicians ($r_M = 0.43$, $P_M = 0.07$) but unreliable for non-musicians ($r_{NM} = -0.13$, $P_{NM} = 0.66$). Thus, analysis by group did not indicate that musical training was associated with improved categorical speech perception in musicians alone.

An ANOVA on speech-labeling speed showed that musicians were faster than non-musicians ($F_{1,22} = 4.21$, $P = 0.04$). The main effect of vowel was also significant ($F_{4,88} = 7.45$, $P < 0.001$), but the interaction between group and vowel was not ($F_{4,88} = 0.10$, $P = 0.98$; Fig. 2D). In both groups, participants were slower at classifying speech tokens near the CP boundary (token 3) relative to others in the continuum (M: $t_{88} = 3.80$, $P < 0.001$; NM: $t_{88} = 3.69$, $P < 0.001$), consistent with previous reports examining speeded vowel classification (Pisoni & Tash, 1974; Bidelman *et al.*, 2013b).

Brainstem ERPs

Figure 3 shows brainstem response time-waveforms and frequency spectra. The ANOVA revealed a main effect of group ($F_{1,22} = 5.22$, $P = 0.032$) on brainstem F0 salience with no main effect of vowel ($F_{1,22} = 0.94$, $P = 0.44$) or group \times vowel interaction ($F_{4,88} = 0.48$, $P = 0.75$; Fig. 3C). For brainstem F1 magnitudes, there were main effects of group ($F_{1,22} = 7.71$, $P = 0.011$) and vowel stimulus ($F_{4,88} = 13.15$, $P < 0.001$) but no interaction ($F_{4,88} = 1.13$, $P = 0.35$; Fig. 3D). The main effect of group for both speech cues indicates stronger subcortical representation of voice pitch (F0) and timbre (F1) information for musicians across the

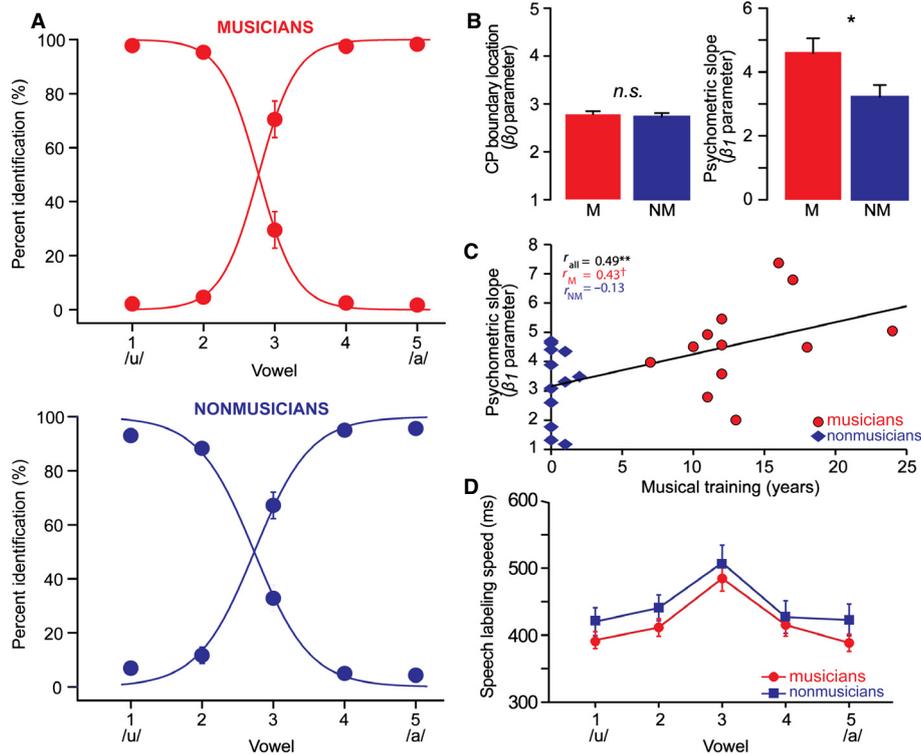


FIG. 2. Perceptual speech classification is enhanced with musical training. (A) Behavioral identification functions. (B) Comparison of the location (left) and 'steepness' (right) of identification functions per group. No difference is observed in the location of the categorical boundary, but musicians (M) obtain steeper identification functions than non-musicians (NM), indicating greater perceptual distinction for the vowels. (C) Across all participants, years of musical training predict speech identification performance. (D) Speech-labeling speed. All listeners are slower to label sounds near the categorical boundary (vw3), but musicians classify speech sounds faster than their non-musician peers. † $P < 0.1$, * $P < 0.05$ and ** $P < 0.01$.

board. In the case of F1, the main effect of vowel stimulus was driven by musicians having larger F1 in response to vw3 and vw4 relative to vw1 ($P = 0.0051$ and $P = 0.042$, respectively). Similarly, supplemental analysis of the temporal phase characteristics of brainstem responses revealed higher trial-to-trial temporal consistency in musicians across the vowel stimuli (Fig. S1). These results indicate a higher temporal precision/coherence in brainstem responses of musicians compared to non-musicians.

Cortical ERPs

Figure 4A shows cortical ERPs elicited by the different vowel stimuli. Consistent with previous studies (e.g., Godey *et al.*, 2001), the early Pa and P1 waves showed considerable variability between listeners. This observation was confirmed by a lack of group difference in both Pa ($F_{1,22} = 0.22$, $P = 0.64$) and P1 ($F_{1,22} = 0.33$, $P = 0.57$) amplitudes. Similarly, the late P3-like deflection was highly variable, and showed neither a consistent group ($F_{1,22} = 1.28$, $P = 0.27$) nor stimulus ($F_{4,88} = 0.92$, $P = 0.45$) effect. In contrast, prominent group and/or stimulus-related differences emerged for the N1 and P2 waves (~100–250 ms). The N1 amplitude did not differ significantly between groups ($F_{1,22} = 0.06$, $P = 0.81$), but varied with vowel stimulus ($F_{4,88} = 10.56$, $P < 0.001$; Fig. 4B). Pairwise comparisons revealed larger N1 amplitude for vw1 (/u/) relative to vw3, vw4 and vw5. Nearing the CP boundary, the N1 wave to vw2 was also greater than vw3. Larger N1 amplitude for stimuli with lower F1 frequency (vw1: /u/) may be related to well-known tonotopic organization of human auditory cortex and the more superficial gen-

erators for low-frequency stimulation (Pantev *et al.*, 1989, 1995; Schreiner & Winer, 2007).

In contrast to N1, the P2 wave was larger in musicians than in non-musicians across the board ($F_{1,22} = 7.43$, $P = 0.012$), but was invariant across vowels ($F_{4,88} = 0.963$, $P = 0.43$; Fig. 4B). These results are consistent with the notion that N1 indexes exogenous acoustic properties of the sound input (Alain *et al.*, 2007; Bidelman *et al.*, 2013b). They also converge with previous studies demonstrating that the neuroplastic effects of long-term auditory training and musicianship are generally larger for P2 relative to N1 (Reinke *et al.*, 2003; Shahin *et al.*, 2003; Seppänen *et al.*, 2012).

Relationships between neural measures and extent of musical training

In musicians, the extent of an individual's self-reported musical training was correlated with measures of both brainstem and cortical speech representations. Years of formal musical training predicted brainstem F0 pitch salience ($r = 0.26$, $P = 0.02$), as well as cortical P2 amplitudes ($r = 0.36$, $P = 0.01$; Bonferroni corrected). No other neurophysiological measures correlated with musical training. The correspondence between musicians' years of musical training and neural measures is consistent with other correlational studies, which suggest that longer trained musicians have more pronounced brainstem/cortical encoding (Musacchia *et al.*, 2008) and perception (Bidelman *et al.*, 2013a) of speech and other complex acoustic sounds. While causality cannot be directly inferred from correlation, these findings suggest that longer musical engagement is at least

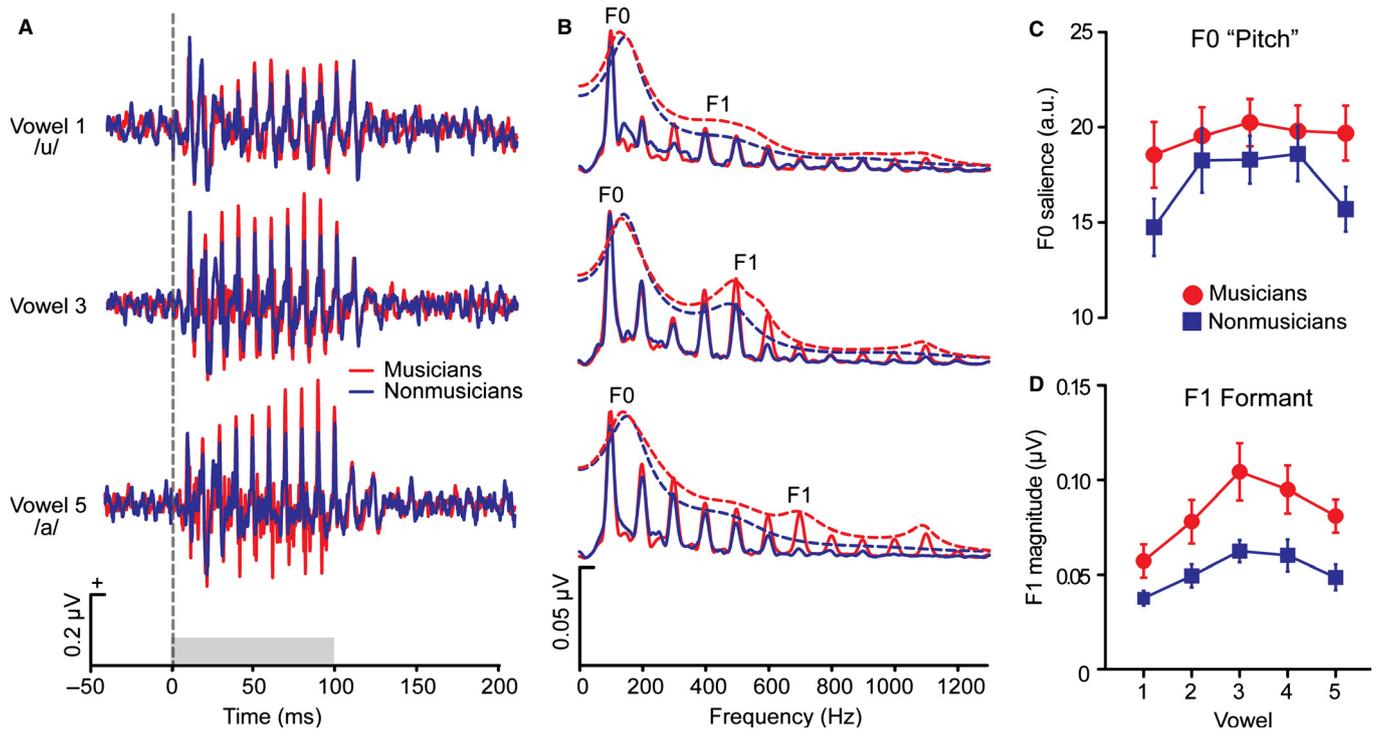


FIG. 3. Musicians have enhanced subcortical-evoked responses to categorical speech. (A) Time-waveforms and (B) corresponding frequency spectra. Musicians (red) have more robust brainstem-evoked responses than non-musicians (blue), indicating enhanced phase-locked activity to the salient spectral cues of speech. Robust energy at the fundamental frequency (100 Hz) and its integer-related harmonics in the response spectra demonstrate robust coding of both voice pitch and timbre information at the level of the brainstem. Response spectral envelopes, computed via linear predictive coding (dotted lines), reveal musicians' increased neural activity near F1, the sole cue for speech identification in the present study. More robust encoding of F0 (C) and F1 (D) in musicians suggests that musicianship improves the brain's encoding of important speech cues.

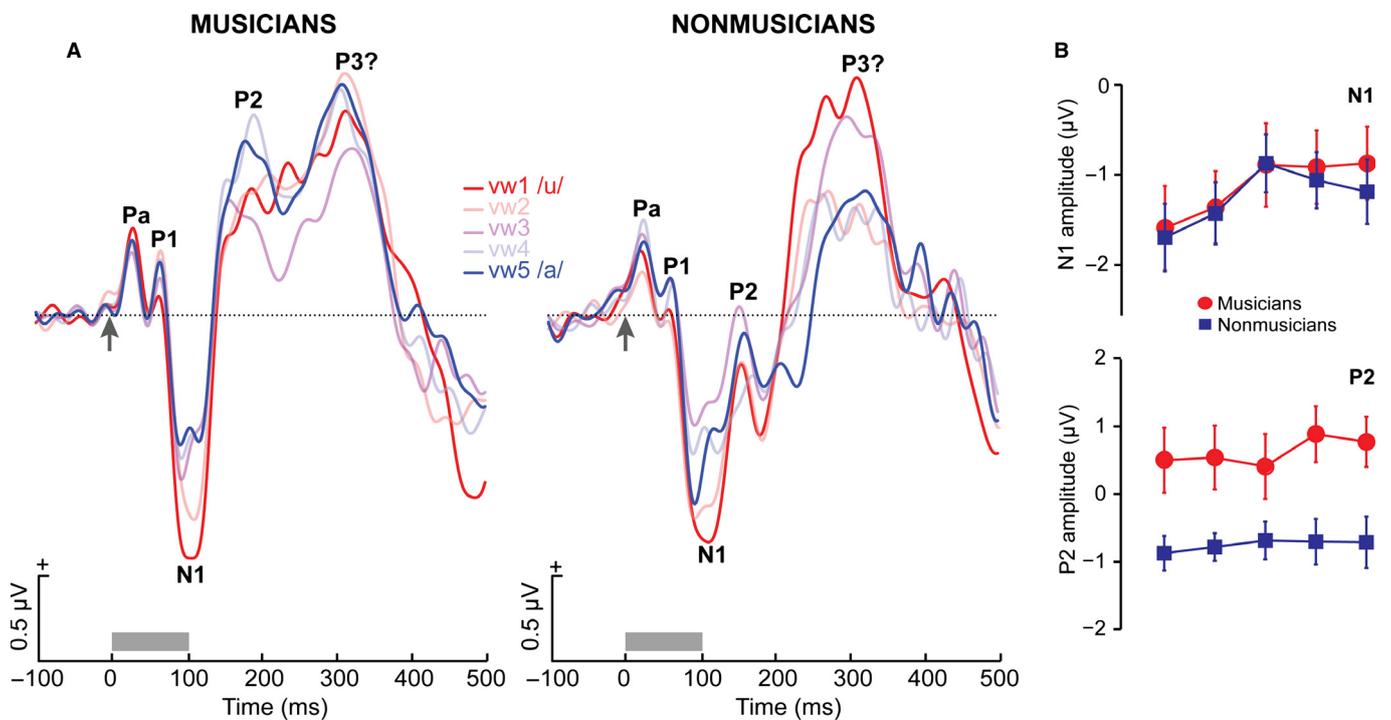


FIG. 4. Musicians have enhanced cortical-evoked responses to categorical speech sounds. (A) Cortical event-related potential (ERP) waveforms. Gray arrows and bars denote the time-locking stimulus. Distinct group differences in response morphology emerge near N1 (~ 100 ms) and persist through the P2 wave (~ 200 ms). (B) N1 and P2 component amplitudes per group. Whereas N1 is modulated by stimulus vowel for both groups, musicians' neuroplastic effects are only apparent in the P2 wave. Larger ERPs in musicians relative to non-musicians suggest that musicianship amplifies the cortical differentiation of speech sounds.

associated with both brainstem and cortical response enhancements for speech sound processing.

Brain-behavior relations

Measures of brainstem F1 amplitude, cortical N1 and P2 amplitudes, and behavioral identification speed were used in further correlational analyses to assess correspondences between both neural (subcortical and cortical) and behavioral measures. Pooling across all responses from the entire vowel continuum, we found a significant relationship between brainstem F1 encoding and cortical P2 amplitudes in musicians but not non-musicians (Fig. 5A; $r_M = 0.40$, $P < 0.001$; $r_{NM} = -0.07$, $P = 0.58$). The brainstem-cortical P2 correlation was stronger for musicians than for non-musicians (Fisher r -to- z transform: $z = 2.90$, $P = 0.0014$). In contrast, there was no relation between brainstem F1 and cortical N1 responses in either group ($r_M = 0.07$, $P = 0.95$; $r_{NM} = 0.22$, $P = 0.45$; not shown). Similarly, cortical-behavioral correlations revealed that P2 amplitude closely predicted listeners' speech identification speeds (Fig. 5B). Larger P2 responses corresponded to faster speech-labeling speeds in musicians ($r_M = -0.37$, $P < 0.001$), but was only marginal in non-musicians ($r_{NM} = -0.22$, $P = 0.09$). Collectively, these findings demonstrate higher correspondence between subcortical, cortical and behavioral speech processing in musically trained individuals.

Multiple least-squares regression was used to determine the extent to which brainstem and cortical ERPs could predict each group's behavioral CP for speech. We used F1 magnitudes for the brainstem regressor; we chose P2 responses for the cortical regressor given that this wave differentiated musician and non-musician groups. Prior to regression analysis, we converted P2 amplitudes to magnitudes (absolute value) to minimize polarity differences that were apparent in the P2 between groups (Fig. 4A). The weighting coefficient (β value) computed for each variable reflects the degree to which that neural measure predicts behavior. The resultant regression function for musicians was: $\psi_{MUSICIAN} = 0.08 + 0.51BS_{erp} - 0.50C_{erp} - 0.18BS_{erp} * C_{erp}$ (bold coefficients denote significant predictor variables, $P < 0.05$), with an overall $adj-R^2 = 0.28$

($P < 0.001$). Thus, brainstem and cortical responses, in addition to their interaction, were robust predictors of musicians' behavioral performance in the CP listening task. This same combination of neural markers was much weaker in predicting behavior for non-musicians: $\psi_{NON-MUSICIAN} = 0.004 - 0.03BS_{erp} - 0.43C_{erp} - 0.20BS_{erp} * C_{erp}$ ($adj-R^2 = 0.14$; $P = 0.01$). Only cortical responses were found to hold significant predictive power for the non-musician group. The higher correspondence between multiple brain responses, their interaction, and perception suggests that musicians may have a sharper interplay and/or coordination between brain areas engaged during speech listening.

Discussion

Previous work has shown that musical expertise enhances the ability to categorize musically relevant sounds, for example pitch intervals and chords (Locke & Kellar, 1973; Siegel & Siegel, 1977; Burns & Ward, 1978; Zatorre & Halpern, 1979; Howard *et al.*, 1992; Klein & Zatorre, 2011), and enhances simple auditory discrimination of important speech cues (Chartrand & Belin, 2006; Moreno *et al.*, 2009; Bidelman & Krishnan, 2010). Here, we link and extend these results by demonstrating a musician advantage in neural and behavioral categorical speech processing, a higher-order linguistic operation. Behaviorally, musicians were faster at categorizing speech tokens and featured a more pronounced (i.e., steeper) boundary between phonetic categories (Fig. 2). These findings suggest that musicians make enhanced sound-to-meaning relations and have heightened phonetic analysis of communicative signals. Musicians' behavioral advantages were accompanied by stronger, more temporally coherent phase-locked brainstem activity to speech sounds than their non-musician peers (Fig 3 and Fig. S1). Complementary results were found at a cortical level, which showed an overall increase in cortical P2 response magnitude relative to non-musicians (Fig. 4). In musicians, brainstem and cortical neurophysiological enhancements (but not behavioral measures; Fig. 2C) were also predicted by their number of years of formal music training; longer music engagement was associated with more robust brain responses to speech. Lastly,

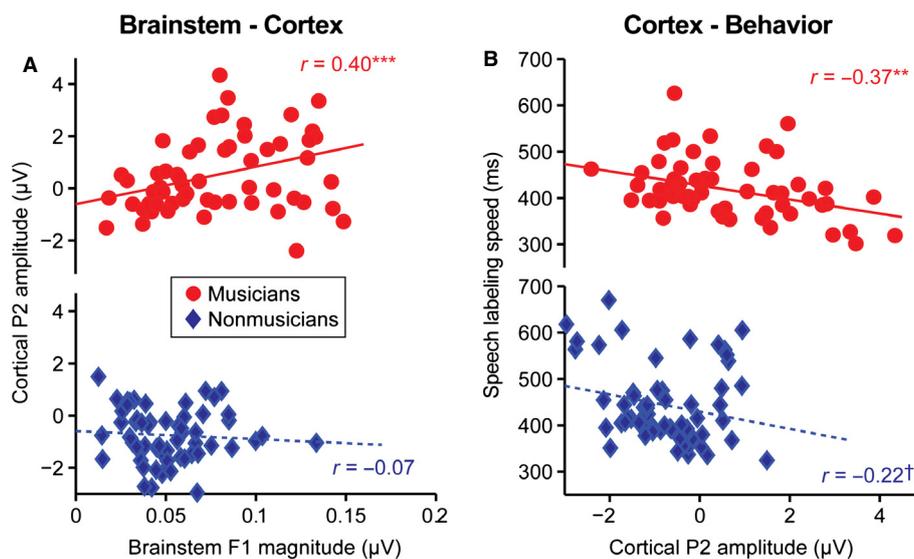


FIG. 5. Brain-behavior relations underlying categorical speech processing. (A) First formant (F1) encoding at the level of the brainstem predicts cortical P2 response amplitudes to speech for musicians (top; Ms) but not non-musicians (bottom; NMs). (B) Cortical P2 amplitudes predict behavioral speech classification speed for Ms but only marginally in NMs; increased P2 corresponds with faster speech identification speed. Dotted regression lines denote non-significant relationships. Data points reflect single-subject responses across all stimuli of the vowel continuum. $^{\dagger}P < 0.1$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

correlations between the multiple brain and behavioral measures revealed that musicians had higher correspondence: (i) between subcortical and cortical responses to speech; and (ii) between neuroelectric brain activity and behavioral speech identification performance (Fig. 5).

Enhanced neurophysiological processing of categorical speech sounds in musicians

Consistent with previous reports comparing brainstem and cortical speech processing (Musacchia *et al.*, 2008), we demonstrate that musicians have enhanced neural representations for speech at multiple levels of the auditory pathway. More importantly, we provide new evidence demonstrating a superiority in musicians' subcortical and cortical encoding of categorically perceived speech sounds. Brainstem ERPs to speech were both more robust and temporally coherent in musicians relative to non-musicians, indicating more precise, time-locked activity to speech at the level of the brainstem (Fig. 3 and Fig. S1). Moreover, both brainstem F0 and cortical P2 neural measures across the speech sound continuum were associated with the years of musicians' formal musical training. These transfer effects corroborate recent neurophysiological studies indicating that musicians' low-level auditory structures carry a more faithful representation of the signal waveform and better extract the important cues of speech, including voice pitch (F0) and timbre cues (F1) (Wong *et al.*, 2007; Parbery-Clark *et al.*, 2009, 2012; Bidelman & Krishnan, 2010; Strait *et al.*, 2013). A more faithful representation of acoustic features at the brainstem may ultimately feed relevant cortical speech mechanisms with a more authentic depiction of speech signal identity (Bidelman *et al.*, 2013b).

By this account, one might expect the earliest experience-dependent modulations to arise cortically in the timeframe of the Pa or P1. While musicians' higher subcortical fidelity converged with an increased cortical responsiveness in the latency range of the P2, we did not observe group differences in the earlier ERP deflections (Pa, P1, N1). Previous work reinforces these inconsistencies, showing both increased (Musacchia *et al.*, 2008) and decreased (Zendel & Alain, 2014) P1 responses in musicians relative to non-musicians. Disparate findings likely result from differences in stimulus features across studies that highly influence the early auditory ERPs. For example, the temporal integration window is shorter for earlier (e.g., Pa, P1) as compared to later (e.g., N1, P2) components of the auditory cortical ERPs (Alain *et al.*, 1997). Thus, stimulus factors (fast presentation rate and longer rise-time) and recording properties (single frontally positioned electrode) used in the present study were suboptimal for the characterization of the early components and, therefore, may have prevented the detectability of potential group differences in these waves. Alternatively, our results may reflect a differential sensitivity of the underlying auditory cortical generators to experience-dependent tuning. Consistent with current findings, previous studies have shown that later components of the auditory ERPs (N1 and P2) are more sensitive to short- and long-term auditory training and speech signals than earlier components (Wood *et al.*, 1971; Shahin *et al.*, 2003; Alain *et al.*, 2007; Alain *et al.*, 2010; Ben-David *et al.*, 2011; Seppänen *et al.*, 2012; Bidelman *et al.*, 2013b). Thus, while musicianship may exert neuroplastic effects in early brainstem processing, when it begins to influence auditory cortical processing is less straightforward.

The link between brainstem and cortical speech encoding may also depend on the specific stimuli and task. Musacchia *et al.* (2008) reported that the relationship between brainstem and cortical speech-evoked responses was generally stronger in musicians

relative to non-musicians. While our data generally agree with these findings, Musacchia *et al.* (2008) noted subcortical–cortical correlations between brainstem F0 amplitude (i.e., voice fundamental encoding) and cortical P1-N1 slope, whereas the current study revealed correlations between brainstem F1 amplitude (a measure of speech timbre encoding) and cortical P2 amplitude. The difference between studies may reflect the use of more transient (Musacchia *et al.*, 2008) vs. sustained (current study) speech tokens that influence brainstem responses. As mentioned previously, stronger correspondence between brainstem and earlier cortical responses might be expected with other stimulus paradigms that optimally evoke the early waves (cf. Musacchia *et al.*, 2008). In the present study, the correlations between brainstem activity and the later cortical waves (P2) were anticipated because we employed a categorical speech perception task, which is known to reflect phonetic-level decisions above and beyond the encoding of acoustic attributes (Chang *et al.*, 2010; Bidelman *et al.*, 2013b). Along these lines, Musacchia *et al.* (2008) also reported a correspondence between brainstem encoding of higher spectral harmonics (3rd - 4th harmonic of a 100 Hz F0) and the latency of the later cortical waves (P2). Coupled with current findings, these results suggest that musicianship provides an overall enhancement of frequency processing in the 300–700-Hz range, a bandwidth critical for the perception of speech timbre, first formant decoding and distinguishing syllables (Assmann & Summerfield, 1990; Bidelman & Krishnan, 2010; Parbery-Clark *et al.*, 2012; Strait *et al.*, 2012). Moreover, musicians' subcortical timbre-related enhancements seem best correlated with later cortical activity in the latency of the N1-P2 time window.

While it is possible that ERP group differences result from heightened attentional abilities (e.g., Strait *et al.*, 2010; Strait & Kraus, 2011), studies demonstrate that musicians' neurophysiological auditory enhancements exist even in the absence of goal-directed attention (Baumann *et al.*, 2008; Musacchia *et al.*, 2008; Parbery-Clark *et al.*, 2009; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011c). Thus, the electrophysiological differences we find between groups are unlikely to result solely from top-down attentional differences between groups. This proposition is supported by our finding of musicians' superior brainstem speech encoding, which inherently reflects pre-attentive auditory processing. While our stimulus paradigm was not specifically designed to elicit neural markers associated with attention (e.g., P3; optimally evoked with oddball paradigms), the presence of a P3-like component (Fig. 4) is suggestive of some attentional reorienting in our task. Yet, consistent with previous studies (e.g., Baumann *et al.*, 2008), we failed to observe a group difference in P3 amplitude, supporting the notion that the effects of musical expertise on speech processing can be separated from those associated with selective attention. Enhancement of the speech signal independent of attention in musicians implies that musical experience might increase neural efficiency associated with processing complex acoustic signals. This would supply higher brain areas a heightened representation of the fine acoustic details in speech without the reliance on potentially sluggish, top-down mediation.

Brainstem and cortical correlates of CP

In both groups, the strength of cortical responses was modulated by the difficulty in perceptually categorizing speech. This notion is supported by the finding that behavioral RTs (reflecting the difficulty of speech identification) and P2 magnitude (reflecting cortical speech processing) covaried inversely with one another (Fig. 5). Increased cortical responses in the timeframe of N1-P2 have been observed with short-term speech training (Alain *et al.*, 2007, 2010; Ben-David

et al., 2011), suggesting that these early cortical waves reflect associative learning and index improvements in speech discrimination with training. Our analysis of the cortical ERPs indicated that the N1 wave was modulated by stimulus acoustics but did not distinguish groups. In contrast, P2 revealed a large group effect but did not covary with the vowel stimulus. These findings lead us to infer that N1 largely indexes exogenous acoustic properties of the sound input, whereas P2 reflects more of the endogenous percept (Alain *et al.*, 2007; Bidelman *et al.*, 2013b), which is more susceptible to the neuroplastic effects of long-term auditory training and musicianship (Shahin *et al.*, 2003; Seppänen *et al.*, 2012). Musicianship reflects the byproduct of extensive, long-term auditory training. Thus, the P2 enhancements and more pronounced behavioral differentiation of phonetic categories observed in our musician cohort may reflect a long-term tuning of the brain mechanisms underlying sound-to-meaning relations (e.g., Seppänen *et al.*, 2012). However, it is important to note that we did not observe a group \times stimulus interaction for any of the neural measures, only strong group effects. Thus, while musicians may outperform non-musicians in behavioral speech sound classification (Fig. 2B) and their brain measures are both amplified and better correlated with their perceptual CP performance (Fig. 5B), musicianship does not seem to alter the normal pattern of speech encoding. That is, we do not find evidence that the neural organization of speech, *per se*, is 'more categorical' in musically trained listeners.

Our results converge with recent reports implicating early cortical ERPs (N1-P2) to CP, as they show distinct modulations with the phonetic (but not acoustic) characteristics of speech phonemes (Chang *et al.*, 2010; Bidelman *et al.*, 2013b). The overall enhancements in the cortical ERPs and finer distinction between prototypical (vw1/vw5) and ambiguous (vw3) speech tokens that we observed in musicians may reflect an augmentation in neural elements subserving sound categorization. Recent functional magnetic resonance imaging studies indicate that behavioral performance in classifying (noise-degraded) speech can be accounted for by the degree of activation in posterior regions of the left superior temporal gyrus (i.e., left planum temporale; Elmer *et al.*, 2012), a putative generator of the P2 wave (Picton *et al.*, 1999). Physical size and functional activation of planum temporale also correlates with a listener's length of musical training (Elmer *et al.*, 2012, 2013). Therefore, the increased ability of musicians to rapidly categorize the phonetic cues of speech may reflect an experience-dependent enhancement in the functional capacity of lateral superior temporal gyrus, planum temporale, and adjacent auditory cortices that generate the aggregate N1-P2 signature and code auditory object recognition.

Intriguingly, we found that musicians' brainstem ERPs showed weaker F1 magnitude for exemplar vowels as compared with the perceptually ambiguous token (e.g. vw1 < vw3; Fig. 3D). At first glance, this encoding selectivity appears to imply categorical organization at the level of the brainstem. It is plausible that musicians' enhanced neural encoding of speech (Figs 3 and 4) may augment and thus reveal subcortical categorical effects unobservable in previous neuroimaging studies examining non-musicians alone (Bidelman *et al.*, 2013b). However, weaker brainstem responses would be expected near the continuum's midpoint relative to prototypical tokens (as we found with the cortical ERPs; Fig. 4). Yet, this is opposite of what we observed; in musicians, brainstem responses were stronger for ambiguous relative to prototypical sounds. While a subcortical origin for CP is an attractive proposition, a more likely explanation for the observed effects relates to response properties of brainstem potentials. Recordings in animals (Ping *et al.*, 2008) and humans (Hoormann *et al.*, 1992) have shown that these ERPs are

maximally evoked by frequencies near \sim 500 Hz. This optimal frequency range roughly coincides with the midpoint of our F1 continuum. It stands to reason then, that musician's increased responses for ambiguous tokens (vw3) might simply reflect the inherent frequency dependence of brainstem potentials rather than subcortical categorical encoding, *per se*. While the neural correlates of CP are clear at a cortical level of processing (current study; Bidelman *et al.*, 2013b), future work is needed to elucidate the impact of experiential factors (e.g., music, language experience) and the potential of categorical neural organization in the brainstem.

Musician superiority in categorical speech perception

Prior work has shown a musician's advantage in the ability to categorize musical pitch intervals/chords (Locke & Kellar, 1973; Siegel & Siegel, 1977; Burns & Ward, 1978; Zatorre & Halpern, 1979; Howard *et al.*, 1992; Klein & Zatorre, 2011). Our study extends this earlier research by showing that musicianship enhances the perceptual categorization of speech. We found faster response times and more dichotomous psychometric boundaries in musically trained listeners (Fig. 2). Notably, musicians and non-musicians did not differ in terms of the 'location' of the boundary, indicating that the perceptual flip between phonetic classes occurred at the same point along the vowel continuum. A similar location in the CP boundary is expected given that all our participants were monolingual English speakers and their phonetic inventories were likely refined in early infancy, long before the onset of music lessons (Kuhl *et al.*, 2008). However, we observed 'steeper psychometric slopes' for vowel classification in musicians, implying sharper acuity for the phonetic features of speech.

In theory, these patterns could result if musically trained listeners placed greater weight on more prototypical sounding tokens than their non-musician peers. Alternatively, a greater degree of phonetic awareness (cf. Anvari *et al.*, 2002) would also be expected if musicianship endows a listener with more refined mental representations of the phonemic inventory of their native vowel-space, as suggested by our EEG data. That is, musical expertise may act to warp or restrict the perceptual space near category boundaries, supplying a more dichotomous decision when classifying sound objects (e.g., Fig. 2B). We argue that musicians' more robust and selective internalized representations for speech across the auditory pathway supply more faithful phonemic templates to these decision mechanisms governing speech sound identification. These results also establish a plausible neurobiological basis to account for musicians' behavioral speech and language benefits observed in this and a growing number of studies (e.g., Anvari *et al.*, 2002; Chartrand & Belin, 2006; Slevc & Miyake, 2006; Bidelman & Krishnan, 2010; Zendel & Alain *et al.*, 2010; Parbery-Clark *et al.*, 2012; Strait *et al.*, 2013).

Hierarchical enhancements to psychophysiological speech processing

Experience-driven reorganization presumably garnered via long-term music engagement appears to engender a higher sensitivity and efficiency in the neural structures engaged during speech perception. The multiple brain enhancements observed in the current study reaffirm the notion that musicianship might exert neuroplastic effects at multiple levels of auditory processing (Schneider *et al.*, 2002; Shahin *et al.*, 2003; Wong & Perrachione, 2007; Musacchia *et al.*, 2008; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011a). Prior studies have shown neural enhancements in musicians at various independent stages of the auditory pathway (for review, see Kraus

& Chandrasekaran, 2010; Herholz & Zatorre, 2012; Moreno & Bidelman, 2014). Unfortunately, in measuring only a single brain response elicited by a single speech token (e.g., Musacchia *et al.*, 2007; Parbery-Clark *et al.*, 2009, 2012; Bidelman & Krishnan, 2010), evidence from earlier research only provides a glimpse of the neurobiological processing of speech and potential interactions between subcortical and cortical levels of auditory processing; they cannot infer potential parallel plasticity nor assess interactions between sensory and cognitive levels of auditory processing that may be differentially shaped by musicianship (e.g., Musacchia *et al.*, 2008). Here, the recording of both brainstem and cortical neuroelectric brain responses in the same listeners reveals that musicians' behavioral advantages in linguistic tasks depend critically on an enhanced coordination (i.e., interaction) between subcortical and cortical neural elements within the speech network. Such coordinated enhancements suggest that musicians maintain a more faithful representation of the original speech signal as it ascends the auditory pathway. In this vein, our results corroborate those of Musacchia *et al.* (2008), who also showed similar correspondence between brainstem and cortical speech ERPs and a stronger coordination of these brain measures in musically trained listeners. However, that study examined only a single stimulus token and did not employ a perceptual speech task. Thus, no direct inference could be made regarding how musicians' improved brainstem/cortical neural processing is reflected behaviorally in speech perception.

By examining brainstem and cortical ERPs elicited during a CP task, our results extend Musacchia *et al.*'s (2008) findings and demonstrate musicians' multiple neural enhancements are directly linked to one another and interact to improve speech-listening behaviors. Correlational analyses revealed that the associations between the various neural indices and subsequent speech perception were generally stronger in musicians (Fig. 5). Multiple regression further showed that in addition to brainstem and cortical activations, musicians' speech perception depended on the 'relationship' (i.e., interaction) between these responses. In contrast, only cortical ERPs predicted perceptual performance for non-musicians; neither brainstem ERPs nor their interaction with cortical ERPs improved behavioral prediction. The significant interaction between musicians' responses implies a stronger reciprocity and perhaps coordination between low- and high-level auditory structures encoding speech information.

That multiple forms/levels of brain activity contribute to musicians' speech behaviors suggests that their brain responses carry more information relevant to speech perception than their non-musician counterparts. Stronger exchange between cortical and subcortical levels would tend to reinforce feed forward and feedback information transfer throughout auditory and non-auditory brain regions (Suga & Ma, 2003; Tzounopoulos & Kraus, 2009; Bajo *et al.*, 2010). An enhanced link between lower- and higher-order brain regions may act to refine signal representations and ultimately sharpen behavioral acuity for speech signals as observed here and in previous reports (Chartrand & Belin, 2006; Bidelman & Krishnan, 2010). Our correlational analyses cannot speak to whether musicians' enhanced brainstem responses 'cause' their enhanced cortical responses, although this is entirely plausible given the well-known corticofugal projections between the auditory cortex and brainstem that act as efferent modulation (Suga & Ma, 2003; Bajo *et al.*, 2010). It is reasonable to infer then that a higher correlation between musicians' neural measures might reflect enhanced functional connectivity between subcortical and cortical hubs of auditory processing. Regardless of the specific biological formulation, our findings suggest that musical experience may engender functional plasticity in a coordinated manner throughout the auditory system.

Limitations and directions for future research

The cross-sectional study design used in the present study cannot rule out the possibility that the observed benefits in musicians results from preexisting group differences (for review, see Alain *et al.*, 2014; Moreno & Bidelman, 2014). For example, certain genetic markers may endow a listener with enhanced auditory recognition abilities (Drayna *et al.*, 2001), ultimately increasing aptitude for musical activities (Ukkola *et al.*, 2009; Park *et al.*, 2012). Alternatively, it is possible that musically savvy individuals differ in behavioral traits such as personality (Corrigall *et al.*, 2013) and/or motivation (McAuley & Tuft, 2011). In this regard, the benefits of musicianship on speech processing observed here and in previous studies (Baumann *et al.*, 2008; Musacchia *et al.*, 2008; Parbery-Clark *et al.*, 2009; Bidelman & Krishnan, 2010; Bidelman *et al.*, 2011b; Skoe & Kraus, 2012) may be partially epiphenomenal in that the advantage may be governed not by musical training *per se*, but by certain genetic and/or behavioral predispositions.

While care must be exercised when drawing causation from correlational data, our findings are corroborated by recent randomized, longitudinal training studies that demonstrate causal, experience-dependent effects of musical training at both behavioral and neurophysiological levels (for reviews, see Herholz & Zatorre, 2012; Moreno & Bidelman, 2014). Changes in brain morphology and physiology subserving auditory processing have been observed following even relatively short-term music lessons (1 year: Fujioka *et al.*, 2006; 2 weeks: Lappe *et al.*, 2008; 15 months: Hyde *et al.*, 2009; 9 months: Moreno *et al.*, 2009). Importantly, these effects remain intact even after controlling for usual confounding factors (e.g., age, socioeconomic background or music listening habits). While our study demonstrates multiple differences in behavioral and neurophysiological speech processing between music expert and non-expert listeners, further longitudinal studies are needed to confirm that these functional advantages also emerge with short- or long-term music training regimens (e.g., Fujioka *et al.*, 2011).

In sum, we infer that musical experience offers an excellent model to understand how human experience differentially shapes multiple levels of brain processing and how neuroelectric brain activity relates to human behavior. Critically, the functional effects of this auditory experience are not independent, but rather produce mutual enhancements to benefit speech recognition abilities. The hierarchy of experience-dependent changes contributes to the widespread auditory and linguistic benefits observed in musicians (Kraus & Chandrasekaran, 2010; Herholz & Zatorre, 2012). Our results ultimately suggest that musicianship tunes the quality of signal representation across functional levels of the auditory system in a way that positively impacts the translation of acoustic speech signals from the periphery to percept.

Supporting Information

Additional supporting information can be found in the online version of this article:

Data S1. Supplemental methods and results detailing the calculation of brainstem F0 neural pitch salience, F1 formant magnitudes, and brainstem response coherence (MSC).

Fig. S1. Temporal phase characteristics of brainstem responses. (A) Polar histograms show the phase-angle distribution across participants; vectors show the mean phase-angle across listeners (measured at the F0 of the response spectra) whose magnitude is proportional to inter-subject coherence. The more focal distribution in musicians suggests less inter-subject variability in this group. (B) Magnitude Squared Coherence (MSC), representing the temporal consistency of

brainstem responses across stimulus presentations. MSC > 0.266 reflects phase-synchrony at a $P = 0.01$ level of significance (Dobie & Wilson, 1989; 1996). Higher MSC in musically trained relative to nonmusician listeners indicates a greater degree of temporal precision in brainstem encoding of speech.

Acknowledgements

The authors thank Yu He for her help in creating the experimental protocol. Portions of this work were supported by Canadian Institutes of Health Research (MOP 106619), the Natural Sciences and Engineering Council of Canada (C.A.), and the GRAMMY Foundation® (G.M.B.). Requests for materials should be addressed to G.M.B.

Abbreviations

CP, categorical perception; EEG, electroencephalogram; ERP, event-related potential; F0, fundamental frequency; F1, first-formant frequency; FFT, fast Fourier transform; ISI, inter-stimulus interval; M, musician; MSC, magnitude squared coherence; NM, non-musician; RT, reaction time; VOT, voice-onset time.

References

- Aiken, S.J. & Picton, T.W. (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hearing Res.*, **245**, 35–47.
- Alain, C., Woods, D.L. & Covarrubias, D. (1997) Activation of duration-sensitive auditory cortical fields in humans. *Electroen. Clin. Neuro.*, **104**, 531–539.
- Alain, C., Snyder, J.S., He, Y. & Reinke, K.S. (2007) Changes in auditory cortex parallel rapid perceptual learning. *Cereb. Cortex*, **17**, 1074–1084.
- Alain, C., Campeanu, S. & Tremblay, K.L. (2010) Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J. Cognitive Neurosci.*, **22**, 392–403.
- Alain, C., Zendel, B.R., Hutka, S. & Bidelman, G.M. (2014) Turning down the noise: the benefit of musical training on the aging auditory brain. *Hearing Res.*, **308**, 162–173.
- 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. *J. Exp. Child Psychol.*, **83**, 111–130.
- Assmann, P.F. & Summerfield, Q. (1990) Modeling the perception of concurrent vowels: the benefit of different fundamental frequencies. *J. Acoust. Soc. Am.*, **88**, 680–697.
- Bajo, V.M., Nodal, F.R., Moore, D.R. & King, A.J. (2010) The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.*, **13**, 253–260.
- Baumann, S., Meyer, M. & Jancke, L. (2008) Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *J. Cognitive Neurosci.*, **20**, 2238–2249.
- Ben-David, B.M., Campeanu, S., Tremblay, K. & Alain, C. (2011) Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology*, **48**, 797–807.
- Bidelman, G.M. & Krishnan, A. (2009) Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.*, **29**, 13165–13171.
- 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. (2011a) Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cognitive Neurosci.*, **23**, 425–434.
- 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 Cognition*, **77**, 1–10.
- Bidelman, G.M., Krishnan, A. & Gandour, J.T. (2011c) Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur. J. Neurosci.*, **33**, 530–538.
- 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.
- Bidelman, G.M., Moreno, S. & Alain, C. (2013b) Tracing the emergence of categorical speech perception in the human auditory system. *NeuroImage*, **79**, 201–212.
- Burns, E.M. & Ward, W.D. (1978) Categorical perception - phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J. Acoust. Soc. Am.*, **63**, 456–468.
- Campbell, T., Kerlin, J.R., Bishop, C.W. & Miller, L.M. (2012) Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hearing*, **33**, 144–150.
- 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.
- Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M. & Knight, R.T. (2010) Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.*, **13**, 1428–1432.
- Chartrand, J.P. & Belin, P. (2006) Superior voice timbre processing in musicians. *Neurosci. Lett.*, **405**, 164–167.
- Cooper, A. & Wang, Y. (2012) The influence of linguistic and musical experience on Cantonese word learning. *J. Acoust. Soc. Am.*, **131**, 4756–4769.
- Corrigan, K.A., Schellenberg, E.G. & Misura, N.M. (2013) Music training, cognition, and personality. *Front. Psychol.*, **4**, 222.
- Delorme, A. & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *J. Neurosci. Meth.*, **134**, 9–21.
- Drayna, D., Manichaikul, A., de Lange, M., Snieder, H. & Spector, T. (2001) Genetic correlates of musical pitch recognition in humans. *Science*, **291**, 1969–1972.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971) Speech perception in infants. *Science*, **171**, 303–306.
- 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.
- Elmer, S., Hanggi, J., Meyer, M. & Jancke, 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.
- Fujioka, T., Ross, B., Kakigi, R., Pantev, C. & Trainor, L.J. (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, **129**, 2593–2608.
- Fujioka, T., Mourad, N., He, C. & Trainor, L.J. (2011) Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals. *Clin. Neurophysiol.*, **122**, 43–51.
- Galbraith, G.C. & Kane, J.M. (1993) Brainstem frequency-following responses and cortical event-related potentials during attention. *Percept. Motor Skills*, **76**, 1231–1241.
- Gockel, H.E., Moore, B.C., Carlyon, R.P. & Plack, C.J. (2007) Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *J. Acoust. Soc. Am.*, **121**, 373–382.
- Godey, B., Schwartz, D., de Graaf, J.B., Chauvel, P. & Liégeois-Chauvel, C. (2001) Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin. Neurophysiol.*, **112**, 1850–1859.
- Harnad, S.R. (1987) *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, New York, NY.
- Herholz, S.C. & Zatorre, R.J. (2012) Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*, **76**, 486–502.
- Hillyard, S.A. & Picton, T.W. (1979) Event-related brain potentials and selective information processing in man. In Desmedt, J.E. (Ed.), *Progress in Clinical Neurophysiology*. Karger, Basel, pp. 1–52.
- Ho, Y., Cheung, M. & Chan, A. (2003) Music training improves verbal but not visual memory: cross sectional and longitudinal explorations in children. *Neuropsychology*, **17**, 439–450.
- Hoormann, J., Falkenstein, M., Hohnsbein, J. & Blanke, L. (1992) The human frequency-following response (FFR): normal variability and relation to the click-evoked brainstem response. *Hearing Res.*, **59**, 179–188.
- Howard, D., Rosen, S. & Broad, V. (1992) Major/Minor triad identification and discrimination by musically trained and untrained listeners. *Music Percept.*, **10**, 205–220.
- Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C. & Schlaug, G. (2009) The effects of musical training on structural brain development: a longitudinal study. *Ann. NY Acad. Sci.*, **1169**, 182–186.
- Irimajiri, R., Golob, E.J. & Starr, A. (2005) Auditory brain-stem, middle- and long-latency evoked potentials in mild cognitive impairment. *Clin. Neurophysiol.*, **116**, 1918–1929.

- Klein, M.E. & Zatorre, R.J. (2011) A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, **49**, 878–887.
- Kraus, N. & Chandrasekaran, B. (2010) Music training for the development of auditory skills. *Nat. Rev. Neurosci.*, **11**, 599–605.
- Krishnan, A., Bidelman, G.M., Smalt, C.J., Ananthakrishnan, S. & Gandour, J.T. (2012) Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, **50**, 2849–2859.
- Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255**, 606–608.
- Kuhl, P.K., Conboy, B.T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M. & Nelson, T. (2008) Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. T. Roy. Soc. B.*, **363**, 979–1000.
- Lappe, C., Herholz, S.C., Trainor, L.J. & Pantev, C. (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. *J. Neurosci.*, **28**, 9632–9639.
- Li, P., Sepanski, S. & Zhao, X. (2006) Language history questionnaire: a web-based interface for bilingual research. *Behav. Res. Methods*, **38**, 202–210.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychol. Rev.*, **74**, 431–461.
- Locke, S. & Kellar, L. (1973) Categorical perception in a non-linguistic mode. *Cortex*, **9**, 355–369.
- McAuley, J.D. & Tuft, S. (2011) Musician advantages in music perception: an issue of motivation, not just ability. *Music Percept.*, **28**, 505–518.
- Moreno, S. & Bidelman, G.M. (2014) Understanding neural plasticity and cognitive benefit through the unique lens of musical training. *Hearing Res.*, **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. *Cereb. Cortex*, **19**, 712–723.
- 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.
- 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.
- Musacchia, G., Strait, D. & Kraus, N. (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing Res.*, **241**, 34–42.
- Okamoto, H., Stracke, H., Bermudez, P. & Pantev, C. (2011) Sound processing hierarchy within human auditory cortex. *J. Cognitive Neurosci.*, **23**, 1855–1863.
- Oldfield, R.C. (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, **9**, 97–113.
- Pantev, C., Hoke, M., Lutkenhoner, B. & Lehnertz, K. (1989) Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science*, **246**, 486–488.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G. & Elbert, T. (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroen. Clin. Neuro.*, **94**, 26–40.
- Parbery-Clark, A., Skoe, E. & Kraus, N. (2009) Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.*, **29**, 14100–14107.
- Parbery-Clark, A., Tierney, A., Strait, D.L. & Kraus, N. (2012) Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience*, **219**, 111–119.
- Park, H., Lee, S., Kim, H.J., Ju, Y.S., Shin, J.Y., Hong, D., von Grothuss, M., Lee, D.S., Park, C., Kim, J.H., Kim, B., Yoo, Y.J., Cho, S.I., Sung, J., Lee, C., Kim, J.I. & Seo, J.S. (2012) Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *J. Med. Genet.*, **49**, 747–752.
- Picton, T.W. & Hillyard, S.A. (1974) Human auditory evoked potentials. II. Effects of attention. *Electroen. Clin. Neuro.*, **36**, 191–199.
- Picton, T.W., Hillyard, S.A., Galambos, R. & Schiff, M. (1971) Human auditory attention: a central or peripheral process? *Science*, **173**, 351–353.
- Picton, T.W., Alain, C., Woods, D.L., John, M.S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J. & Trujillo, N.J. (1999) Intracerebral sources of human auditory-evoked potentials. *Audiol. Neuro-otol.*, **4**, 64–79.
- Ping, J., Li, N., Gaibraith, G., Wu, X. & Li, L. (2008) Auditory frequency-following responses in rat ipsilateral inferior colliculus. *NeuroReport*, **19**, 1377–1380.
- Pisoni, D.B. (1973) Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept. Psychophys.*, **13**, 253–260.
- Pisoni, D.B. & Luce, P.A. (1987) Acoustic-phonetic representations in word recognition. *Cognition*, **25**, 21–52.
- Pisoni, D.B. & Tash, J. (1974) Reaction times to comparisons within and across phonetic categories. *Percept. Psychophys.*, **15**, 285–290.
- Reinke, K., He, Y., Wang, C. & Alain, C. (2003) Perceptual learning modulates sensory evoked response during vowel segregation. *Cognitive Brain Res.*, **17**, 781–791.
- Rinne, T., Stecker, G.C., Kang, X., Yund, E.W., Herron, T.J. & Woods, D.L. (2007) Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *NeuroReport*, **18**, 1311–1314.
- 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. *Nat. Neurosci.*, **5**, 688–694.
- Schreiner, C.E. & Winer, J.A. (2007) Auditory cortex mapmaking: principles, projections, and plasticity. *Neuron*, **56**, 356–365.
- Seppänen, M., Hamalainen, J., Pesonen, A.-K. & Tervaniemi, M. (2012) Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Front. Hum. Neurosci.*, **43**, 1–13.
- Shahin, A., Bosnyak, D.J., Trainor, L.J. & Roberts, L.E. (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.*, **23**, 5545–5552.
- Sharma, A. & Dorman, M.F. (1999) Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J. Acoust. Soc. Am.*, **106**, 1078–1083.
- Siegel, J.A. & Siegel, W. (1977) Absolute identification of notes and intervals by musicians. *Percept. Psychophys.*, **21**, 143–152.
- Skoe, E. & Kraus, N. (2012) A little goes a long way: how the adult brain is shaped by musical training in childhood. *J. Neurosci.*, **32**, 11507–11510.
- Slevc, R.L. & Miyake, A. (2006) Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.*, **17**, 675–681.
- Strait, D.L. & Kraus, N. (2011) Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front. Psychol.*, **2**, 113.
- 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 Res.*, **261**, 22–29.
- Strait, D.L., Chan, K., Ashley, R. & Kraus, N. (2012) Specialization among the specialized: auditory brainstem function is tuned in to timbre. *Cortex*, **48**, 360–362.
- Strait, D.L., O'Connell, S., Parbery-Clark, A. & Kraus, N. (2013) Musicians' enhanced neural differentiation of speech sounds arises early in life: developmental evidence from ages 3 to 30. *Cereb. Cortex*, doi: 10.1093/cercor/bht103. [Epub ahead of print].
- Suga, N. & Ma, X. (2003) Multiparametric corticofugal modulation and plasticity in the auditory system. *Nat. Neurosci.*, **4**, 783–794.
- Tremblay, K., Kraus, N., McGee, T., Ponton, C. & Otis, B. (2001) Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hearing*, **22**, 79–90.
- Tzounopoulos, T. & Kraus, N. (2009) Learning to encode timing: Mechanisms of plasticity in the auditory brainstem. *Neuron*, **62**, 463–469.
- Ukkola, L.T., Onkamo, P., Raijas, P., Karma, K. & Jarvela, I. (2009) Musical aptitude is associated with AVPR1A-Haplotypes. *PLoS One*, **4**, e5534.
- Werker, J.F. & Tees, R.C. (1987) Speech perception in severely disabled and average reading children. *Can. J. Psychol.*, **41**, 48–61.
- Wong, P.C. & Perrachione, T.K. (2007) Learning pitch patterns in lexical identification by native English-speaking adults. *Appl. Psycholinguist.*, **28**, 565–585.
- 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.
- Wood, C.C., Goff, W.R. & Day, R.S. (1971) Auditory evoked potentials during speech perception. *Science*, **173**, 1248–1251.
- Woods, D.L. & Hillyard, S.A. (1978) Attention at the cocktail party: Brainstem evoked responses reveal no peripheral gating. In Otto, D.A. (Ed.) *Multidisciplinary Perspectives in Event-Related Brain Potential Research (EPA 600/9-77-043)*. U.S. Government Printing Office, Washington, DC, pp. 230–233.
- Zatorre, R. & Halpern, A.R. (1979) Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Percept. Psychophys.*, **26**, 384–395.
- Zendel, B.R. & Alain, C. (2009) Concurrent sound segregation is enhanced in musicians. *J. Cognitive Neurosci.*, **21**, 1488–1498.
- Zendel, B.R. & Alain, C. (2012) Musicians experience less age-related decline in central auditory processing. *Psychol. Aging*, **27**, 410–417.
- Zendel, B.R. & Alain, C. (2014) Enhanced attention-dependent activity in the auditory cortex of older musicians. *Neurobiol. Aging*, **35**, 55–63.