Research paper

Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians

Gabriella Musacchia [a,*], Dana Strait [a,b], Nina Kraus [a,c]

[a] Auditory Neuroscience Laboratory, Department of Communication Sciences, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA
[b] School of Music, Northwestern University, Evanston, IL 60208, USA
[c] Departments of Neurobiology and Physiology and Otolaryngology, Northwestern University, Evanston, IL, USA

ARTICLE INFO

ABSTRACT

Musicians have a variety of perceptual and cortical specializations compared to non-musicians. Recent studies have shown that potentials evoked from primarily brainstem structures are enhanced in musicians, compared to non-musicians. Specifically, musicians have more robust representations of pitch periodicity and faster neural timing to sound onset when listening to sounds or both listening to and viewing a speaker. However, it is not known whether musician-related enhancements at the subcortical level are correlated with specializations in the cortex. Does musical training shape the auditory system in a coordinated manner or in disparate ways at cortical and subcortical levels? To answer this question, we recorded simultaneous brainstem and cortical evoked responses in musician and non-musician subjects. Brainstem response periodicity was related to early cortical response timing across all subjects, and this relationship was stronger in musicians. Peaks of the brainstem response evoked by sound onset and timbre cues were also related to cortical timing. Neurophysiological measures at both levels correlated with musical skill scores across all subjects. In addition, brainstem and cortical measures correlated with the age musicians began their training and the years of musical practice. Taken together, these data imply that neural representations of pitch, timing and timbre cues and cortical response timing are shaped in a coordinated manner, and indicate corticofugal modulation of subcortical afferent circuitry.

1. Introduction

Playing music is a cognitively complex task that requires, at minimum, sensations from the sound he or she is playing, the sight of sheet music and the touch of the instrument to be utilized and integrated. Proficiency at doing so accumulates over years of consistent training, even in cases of high innate talent. Not surprisingly, instrumental musicians exhibit behavioral and perceptual advantages over non-musicians in music-related areas such as pitch discrimination (Tervaniemi et al., 2005) and fine motor control skills (Kincaid et al., 2002). Musicians have also shown perceptual improvements over non-musicians in both native and foreign linguistic domains (Magne et al., 2006; Marques et al., 2007). It is thought that neural plasticity related to musical training underlies many of these differences (Hannon and Trainor, 2007).

Highly-trained musicians exhibit anatomical, functional and event-related specializations compared to non-musicians. From an anatomical perspective, musicians have more neural cell bodies (grey matter volume) in auditory, motor and visual cortical areas of the brain (Gaser and Schlaug, 2003) and have more axonal projections that connect the right and left hemispheres (Schlaug et al., 1995). Not surprisingly, professional instrumentalists, compared to amateurs or untrained controls, have more activation in auditory areas such as Heschel's gyrus (Schneider et al., 2002) and the planum temporale (Ohnishi et al., 2001) to sound. Musical training also promotes plasticity in somatosensory regions; with string players demonstrating larger areas of finger representation than untrained controls (Elbert et al., 1995). With regard to evoked potentials (EPs) thought to arise primarily from cortical structures, musicians show enhancements of the P1–N1–P2 complex to pitch, timing and timbre features of music, relative to non-musicians (Pantev et al., 2001). Trained musicians show particularly large enhancements when listening to the instruments that they themselves play (Munte et al., 2003; Pantev et al., 2003). Musicians' cortical EP measures are also more apt to register fine-grained changes in complex auditory patterns and are more sensitive to pitch and interval changes in a melodic contour than non-musicians (Fujioka et al., 2004a; Pantev et al., 2003). Moreover, musician-related plasticity is implicated in these and other studies because enhanced cortical EP measures have been correlated to the length of musical training or musical skill.

* Corresponding author. Tel.: +1 847 491 2465; fax: +1 847 491 2523.
E-mail address: g-musacchia@northwestern.edu (G. Musacchia).

Recent studies from our laboratory have suggested that playing a musical instrument also "tunes" neural activity peripheral to cortical structures (Musacchia et al., 2007; Wong et al., 2007). These studies showed that evoked responses thought to arise predominantly from brainstem structures were more robust in musicians than in non-musician controls. The observed musician-related enhancements corresponded to stimulus features that may be particularly important for processing music. One such example is observed with the frequency following response (FFR), which is thought to be generated primarily in the inferior colliculus and consists of phase-locked inter-spike intervals occurring at the fundamental frequency (F0) of a sound (Hoormann et al., 1992; Krishnan et al., 2005). Because F0 is understood to underlie the percept of pitch, this response is hypothesized to be related to the ability to accurately encode acoustic cues for pitch. Enhanced encoding of this aspect of the stimulus would clearly be beneficial to pitch perception of music. Accordingly, our previous studies demonstrated larger peak amplitudes at F0 and better pitch tracking in musicians relative to non-musicians. Another example was observed with Wave δ (~8 ms post-acoustic onset) of the brainstem response to sound onset, which has been hypothesized to be important for encoding stimulus onset (Musacchia et al., 2006, 2007). Stimulus onset is an attribute of music important for denoting instrument attack and rhythm, and therefore it is perhaps unsurprising that we observed earlier Wave δ responses in musicians than non-musicians. More importantly, FFR and Wave δ enhancement in musicians was observed with both music and speech stimuli and was largest when subjects engaged multiple senses by simultaneously lip-reading or watching a musician play. This suggests that while these enhancements may be motivated by music-related tasks, they are pervasive and apply to other stimuli that possess those stimulus characteristics.

A key point to be noted regarding prior EP studies showing musician-related enhancements is that none have attempted to relate enhancements in measures thought to arise from brainstem structures (e.g., the FFR) with measures thought to arise largely from cortical regions (e.g., P1, N1 and P2 potentials). One crucial piece of information that could be gleaned from this approach would be that we may be able to determine which stimulus features are relevant to cortical EP enhancements in musicians. Such determinations could be made because musician-related enhancements in brainstem responses correspond to representations of specific stimulus features (e.g. pitch, timing and timbre).

The implications of these data could be strengthened considerably if the EP data were also correlated with performance on music-related behavioral tasks. Previous work has suggested that short and long-term experience with complex auditory tasks (e.g. language, music, auditory training) may shape subcortical circuitry likely through corticofugal modulation of sensory function (Kraus and Banai, 2007; Russo et al., 2005; Song et al., 2008; Krishnan et al., 2005). Correlations between measures of brainstem and cortical EPs that coincide with improved performance on a musical task would provide support for the notion that specific neural elements are recruited to perform a given task, and that such selections are mediated in a top-down manner through experience (e.g., Reverse Hierarchy Theory; Ahissar and Hochstein, 2004), presumably via reciprocal cortical–subcortical interactions. Although Reverse Hierarchy Theory (RHT) has been used to consider visual cortical function, it is our view that this mechanism applies to subcortical sensory processing and that the application of its principles can explain the malleability of early sensory levels.

The idea of a cognitive-sensory interplay between subcortical and cortical plasticity is not new, and theories of learning increasingly posit a co-operation between "bottom-up" ad "top-down" plasticity [for review, see (Kral and Eggermont, 2007)]. Galbraith was one of the first to recognize that human brainstem function is sensitive to cognitive states and functions and can be modulated by selective auditory attention (Galbraith et al., 2003), and when reaction times to auditory stimuli are shorter (Galbraith et al., 2000). The FFR is also selectively activated when verbal stimuli are consciously perceived as speech (Galbraith et al., 1997) and is larger to a speech syllable than to a time-reversed version of itself (Galbraith et al., 2004). In addition, several lines of research suggest that subcortical activity is enhanced in people who have had protracted linguistic (Krishnan et al., 2005; Xu et al., 2006) or musical training (Musacchia et al., 2007; Wong et al., 2007) and degraded in people with certain communication disorders (Banai et al., 2005; Russo et al., in press). Malleability of the human brainstem response is not restricted to lifelong training, however, as short-term auditory training has also been shown to enhance the FFR in children and adults (Russo et al., 2005; Song et al., 2008). Physiological work in animals demonstrates that improved signal processing in subcortical structures is mediated by the corticofugal system during passive and active auditory exposure (Yan and Suga, 1998; Zhou and Jen, 2000). Prior anatomical findings suggest several potential routes that propagate action potentials from the auditory cortex to subcortical centers such as the medial geniculate body and inferior colliculus (IC) (Kelly and Wong, 1981; Saldana et al., 1996; Huffman and Henson, 1990). Consistent with this notion of reciprocal cortical–subcortical interaction, the current work investigates the relationship between experience and the representation of stimulus features at the sensory and cortical level.

In order to examine the relationship between evoked-potentials and experience, we recorded simultaneous brainstem and cortical EPs in musicians and non-musician controls. Because previous data showed that musician-related effects extend to speech and multisensory stimuli, the speech syllable "da" was presented in three conditions: when subjects listened to the auditory sound alone, when the subjects simultaneously watched a video of a male speaker saying "da", and when they viewed the video alone. Our analysis focused on comparing measures of the speech-evoked brainstem response that have been previously reported as enhanced in musicians with well-established measurements of cortical activity (e.g., the P1–N1–P2 complex). Thus, we were particularly interested in the representation of the timing of sound onset, pitch and timbre in the brainstem response. By correlating these neurophysiological measures and comparing them to behavioral scores on tests of musical skill and auditory perception, we were able to establish links between brainstem measures, cortical measures and behavioral performance and to show which relationships were strengthened by musical training.

2. Materials and methods

2.1. Subjects

Participants in this study consisted of 26 adults (mean age 25.6 ± 4.1 years, 14 females) with normal hearing (<15 dB HL pure-tone thresholds from 500 to 4000 Hz). We assume that all listeners had similar audiometric profiles because we are unaware of any data suggesting that normal-hearing musicians have a different audiometric profile than normal-hearing non-musicians. Participants were selected to have normal or corrected-to-normal vision (Snellen Eye Chart, 2001) and no history of neurological disorders. All participants gave their informed consent before participating in this study in accordance with the Northwestern University Institutional Review Board regulations. Subjects categorized as musicians (N = 14) were self-identified, began playing an instrument before the age of five, had 10 or more years of musical experience, and practiced more than three times weekly for four or more hours consistently over the last 10 years. Controls (N = 12) were categorized by the failure to meet the musician criteria.

2.2. Musical aptitude measures

We administered two in-house measures of auditory and musical skill: *Seashore's Test of Musical Talents* (Seashore, 1919) and Colwell's *Musical Achievement Test* (MAT-3) (Colwell, 1970). Seashore's test consists of six subtests: Pitch, Rhythm, Loudness, Time, Timbre and Tonal Memory. Each subtest is a two-alternative forced choice auditory discrimination task that asks listeners to judge whether the second sound (or sequence) is different from the first. Because of its use of pure and complex sine waves, and the method of evaluation, the Seashore battery of listening tests is widely-understood to measure basic psychoacoustic skills rather than actual musical aptitude. The MAT-3 consists of five subtests and was designed as an entrance exam for post-secondary instrumental students. Accordingly, some MAT-3 tests were too advanced for the non-musicians. We administered MAT-3 subtests of Tonal Memory and Solo Instrument Recognition (I) to all subjects. Musicians were also given MAT-3 tests of Melody Recognition, Polyphonic Chord Recognition and Ensemble Instrument Recognition (II). Introductory verbal instruction was provided at the start of each test and subtest, with musical examples for each question provided via a portable stereo system. Bivariate correlation tests among tests of musical skill and neurophysiological measures were conducted and independent *t*-tests between groups were conducted to determine the extent of musician-related differences.

2.3. Stimuli and recording procedure

Stimuli were presented binaurally via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL) while the subject sat in a comfortable chair centered 2.3 m from a 15.2 cm × 19.2 cm projection screen. The speech syllable "da" was presented in three conditions: (1) when subjects heard the sound alone and simultaneously watched a captioned video (A); (2) when, instead of a captioned movie, subjects viewed a video token of a male speaker saying "da" simultaneously (AV); and (3) when subjects viewed the video of the speaker without sound (V). The synthesized speech syllable (Klatt Software, 1980) was 350 ms in duration with a fundamental frequency of 100 Hz. F1 and F2 of the steady state were 720 Hz and 1240 Hz, respectively. Video clips of a speaker's face saying "da" were edited to 850 ms durations (FinalCut Pro 4, Apple Software). When auditory and visual stimuli were presented together, the sound onset occurred 460 ms after the onset of the first video frame. The acoustic onset occurred synchronously with the visual release of consonant closure.

Stimuli were presented in 12 blocks of 600 stimulus repetitions with a 5-min break between blocks (Neurobehavioral Systems Inc., 2001). Each block consisted of either A, V or AV stimuli, with modality of presentation order randomized across all subjects. Auditory stimuli were presented at 84 dB SPL in alternating polarities. This presentation level insured that the signal was clearly audible and well above threshold to all subjects. To control for attention, subjects were asked to silently count the number of target stimuli they saw or heard and to report that number at the end of each block. Target stimuli were slightly longer in duration than the standards (auditory target = 380 ms, visual target = 890 ms) and occurred 4.5 ± 0.5% of the time. Performance accuracy was measured by counting how many targets the subject missed (error%).

2.4. General neurophysiology recording procedure

Electroencephalographic (EEG) data were recorded from Ag–AgCl scalp electrode Cz (10–20 International System, earlobe reference, forehead ground) with a filter passband of 0.5–2000 Hz and a sampling rate of 20 kHz (Compumedics, El Paso, TX, USA). Following acquisition, the EEG data were highpass and lowpass filtered offline to emphasize brainstem or cortical activity, respectively (see below).

Although there is ample evidence that generators in brainstem structures figure prominently in what we refer to as the "brainstem" and "cortical" responses, it is worth noting that these far-field evoked potentials do not reflect the activity of brainstem or cortical structures exclusively. Because far-field responses record the sum of all neuroelectric activity, higher-level activity (e.g. thalamic, cortical) may be concomitantly captured to some degree in both the onset and FFR measures and vice-versa. Neural generators that contribute to the human brainstem response have been identified primarily through simultaneous surface and intracranial recordings of responses to clicks during neurosurgery (Hall, 1992; Jacobson, 1985). The cochlear nucleus, the superior olivary complex, the lateral lemniscus, and the inferior colliculi have been shown to predominantly contribute to the first five transient peaks (Waves I–V, ~1–6 ms post-acoustic onset) recorded from the scalp. Pure tones and complex sounds evoke the FFR which is thought to primarily reflect phase-locked activity from the inferior colliculus (Smith et al., 1975; Hoormann et al., 1992; Krishnan et al., 2005). Moreover, the FFR can emerge at latency of ~6 ms, which precedes the initial excitation of primary auditory cortex (~12 ms) (Moushegian et al., 1973; Celesia, 1968). Finally, and perhaps most convincingly, cryogenic cooling of the IC greatly decreases or eliminates the FFR (Smith et al., 1975). Despite this evidence, it is possible that evoked FFR activity may reflect concomitant cortical activity after cortical regions have been activated (e.g. ~12 ms). At longer latencies, the FFR most likely reflects a mix of afferent brainstem activity, cortically modulated efferent effects, and synchronous cortical activity. According to these data and for the sake of parsimony and accord with previous studies, we utilize the terms "brainstem" and "cortical" in this study to denote high- and low-pass filtered EP responses, respectively.

2.5. Brainstem response analysis

After acquisition, a highpass filter of 70 Hz was applied to the EEG data. Typically, this type of passband is employed to emphasize the relatively fast and high-frequency neural activity of putative brainstem structures. After filtering, the data were epoched from −100 to 450 ms, relative to acoustic onset. A rejection criterion of ±35 μV was applied to the epoched file so that responses containing high myogenic or extraneous activity above or below the criterion were excluded. The first 2000 epochs that were not artifact-rejected from each condition (A, V, AV) were then averaged for each individual. We then assessed measures of the brainstem response that reflect stimulus features that have been shown to differ between musicians and non-musicians. Brainstem onset response peak, Wave δ, was picked from each individual's responses, yielding latency and amplitude information. The FFR portion of the brainstem response was submitted to a fast Fourier transform (FFT). Strength of pitch encoding was measured by peak amplitudes at F0 (100 Hz) and timbre representation by peak amplitudes at harmonics H2 (200 Hz), H3 (300 Hz), H4 (400 Hz), and H5 (500 Hz) as picked by an automatic peak-detection program. Because we assessed measures that have previously been shown to differ between musicians and non-musicians, we used one-tailed independent *t*-tests to assess group differences in brainstem response measures.

2.6. Cortical response analysis

EEG data were lowpass filtered offline at 40 Hz. This passband is employed to emphasize the relatively slow and low-frequency neural activity of putative cortical origin. Responses were epoched

and averaged with an artifact rejection criterion of ±65 μV and the first 2000 artifact-free sweeps were averaged in each condition. Cortical response peaks (P1, N1, P2 and N2) were chosen from each subject's averages, providing amplitude and latency information. Strength of neural synchrony in response to a given stimulus was assessed by P1–N1 and P2–N2 peak-to-peak slopes.

2.7. Description of brainstem and cortical responses

The brainstem response to a speech syllable mimics stimulus characteristics with high fidelity (Johnson et al., 2005; Kraus and Nicol, 2005; Russo et al., 2004). The beginning portion of the brainstem response to speech (~0–30 ms) encodes the onset of sound in a series of peaks, the first five of which are analogous to responses obtained in hearing clinics with click or tone stimuli (e.g. Waves I–V) (Hood, 1998). With this stimulus, a large peak is also typically observed at ~8–12 ms, called Wave δ (Musacchia et al., 2006; Musacchia et al., 2007). Other laboratories have demonstrated similar relationships between the temporal characteristics of tonal stimuli in the human brainstem response (Galbraith and Doan, 1995; Galbraith et al., 1997, 2003, 2004; Krishnan et al., 2005; Akhoun et al., 2008). In the current study, we restricted our peak latency and amplitude analyses to Wave δ because it was the only brainstem peak to sound onset that previously differed between musicians and non-musicians. The voiced portion of the speech syllable evokes an FFR, which reflects neural phase-locking to the stimulus F0. Fig. 1 shows the grand average brainstem responses of musicians and non-musicians in A and AV conditions. The grand average FFTs are shown in insets.

Grand average cortical responses are shown in Fig. 2. Speech stimuli, presented in either the A or AV condition, elicited four sequential peaks of alternating positive and negative polarity and are labeled P1, N1, P2, and N2, respectively. As is typically observed in cortical responses to sound, these components occurred within ~75–250 ms post-acoustic stimulation (Hall, 1992).

To investigate relationships between musical training and brainstem and cortical processing, Pearson's r correlations were
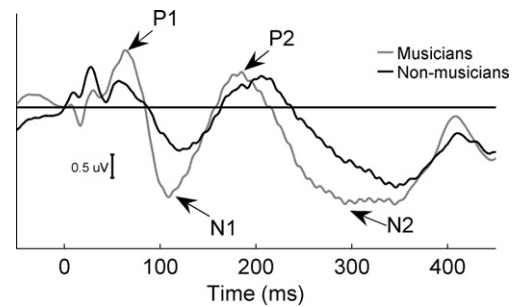
Fig. 2. Musician and non-musician grand average cortical responses to speech in the AV condition. The speech syllable "da" in both A and AV conditions elicited four sequential peaks of alternating positive and negative deflections labeled P1, N1, P2, and N2, respectively. The slope between P1 and N1 was calculated to assess the synchrony of positive to negative deflection in the early portion of the cortical response. Peaks of cortical activity were earlier and larger in musicians (grey) than in non-musicians (black). In addition, P1–N1 slope was steeper in musicians compared to non-musicians. Similar effects were seen in the A condition.

run between all measures of musicianship and brainstem and cortical responses.

3. Results

3.1. Differences between musicians and non-musicians

As has been shown in previous studies, musicians had more robust encoding of speech periodicity in the FFR. Musicians had larger F0 peak amplitudes, in both the A ($t = 2.42$, $p = 0.012$) and AV conditions ($t = 2.33$, $p = 0.015$, compared to non-musicians. Group differences were also observed on measures of timbre representation ($t_{H3} = 2.00$, $p = 0.029$; $t_{H4} = 1.784$, $p = 0.045$; $t_{H5} = 1.767$, $p = 0.045$) and onset timing ($t\delta_{Latency} = 1.95$, $p = 0.032$) in the AV condition.

Overall, P1 and N1 peaks were earlier and larger in the musician group (Fig. 2). Musicians had larger amplitudes at P1 in the A
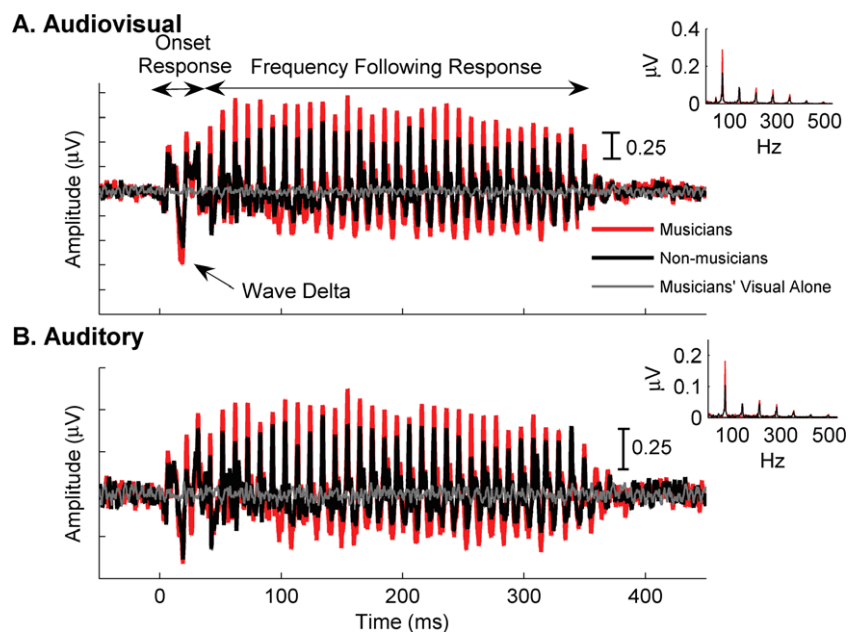
Fig. 1. Grand average brainstem responses to speech. (A) Musicians (red) have more robust responses than non-musicians (black) in the Audiovisual (Panel A) condition. Initial peaks of deflection (0–30 ms) represent the brainstem response to sound onset. Wave Delta of the response to sound onset are noted. The subsequent periodic portion reflects phase-locking to stimulus periodicity (frequency following response). Seeing a speaker say "da" elicited little brainstem activity, as illustrated by the musician's Visual Alone grand average (grey). Non-musicians showed the same type of visual response, but for clarity, are not shown. (B) The same musician-related effect is observed in the Auditory condition. Frequency spectra of the group averages, as assessed by Fast Fourier Transforms, are inset in each panel.

($t = 2.106$, $p = 0.046$) and AV ($t = 3.001$, $p = 0.006$) conditions and at N1 in the AV condition ($t = 2.099$, $p = 0.047$). P1–N1 slope, our measure of early aggregate cortical timing, was steeper in musicians compared to non-musicians for both the A ($t = 2.90$, $p = 0.01$) and AV conditions ($t = 5.01$, $p < 0.001$). Later components, as measured by P2 and N2 latency and P2–N2 slope, did not differ between groups.

Perceptual test scores showed that musicians scored better than non-musicians on both the Seashore and MAT tests of tonal memory (MAT-3: $M_{musicians} = 18.07$, $M_{non-musicians} = 12.25$, $t = 4.50$, $p < 0.001$; Seashore: $M_{musicians} = 97.86$, $M_{non-musicians} = 87.78$, $t = 3.44$, $p = 0.002$).

3.2. Relationships between brainstem and cortical measures

Among brainstem response measures that differ between musicians and non-musicians, periodicity encoding correlated with measures of P1–N1 slope and P2 and N2 latency most consistently (Table 1, Fig. 3). Across all subjects ($n = 26$), larger F0 peak amplitudes of the brainstem response were associated with steeper cortical P1–N1 slopes in both A ($r = -0.47$, $p = 0.02$) and AV ($r = -0.50$, $p = 0.01$) conditions. F0 amplitude in the AV condition also correlated with measures of later cortical peaks, P2 and N2 ($r_{P2} = -0.49$, $p = 0.01$; $r_{N2} = -0.44$, $p = 0.02$), such that larger F0 amplitudes were associated with earlier latencies. Correlations between F0 amplitude and P2–N2 latencies did not reach statistical significance in the A condition. Taken together, correlations between F0, P1–N1 slope and P2 latency suggest that faithful and robust representation of F0 is associated with pervasively faster cortical timing.

Measures of harmonic encoding correlated more specifically to later cortical peak timing. H3 peak amplitude correlated with P2 latency ($r = -0.40$, $p = 0.04$) in the AV condition and H4 peak amplitude correlated with N2 latency in the A condition ($r = 0.42$, $p = 0.03$) (Table 2).

Table 3 shows that P1–N1 slope in the AV condition and N2 peak latency in the A condition correlated with brainstem onset timing (Wave δ latency). That is, steeper P1–N1 slope and earlier N2 latency correlate with earlier brainstem onset responses.

In order to determine whether the relationships between brainstem and cortical measures were stronger in musicians than non-musicians, we conducted an heterogeneity of regression line test on values from pairs of measures that significantly correlated across all subjects. Results from these tests indicated that the regression-line slopes differed between the musicians and non-musicians for the F0 and P1–N1 slope relationship in the A condition ($F = 8.61$, $p < 0.01$). Examination of the within-group correlation values for F0 and P1–N1 slope revealed that the musicians had a stronger correlation than non-musicians ($r = -0.70$ vs. $r = 0.13$, respectively). This same trend, though not significant, was seen in the AV condition ($r = -0.42$ for musicians vs. $r = 0.05$ for non-musicians).
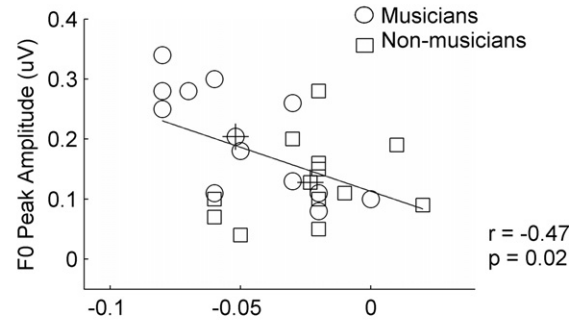

Fig. 3. Relationship between P1–N1 slope and FFR encoding of pitch cues. (A) Peak amplitude of the fundamental frequency (F0) correlated negatively with P1–N1 slope, indicating an association of larger F0 amplitude with steeper P1–N1 slope. Overall, musicians (circles) had larger F0 amplitudes and steeper slopes than non-musicians (squares). (B) This relationship was also observed in the Audiovisual condition. Group means (crossed symbols) show that musicians have larger F0 amplitudes and steeper P1–N1 slopes than non-musicians in both stimulus conditions.

Table 2
Pearson correlation coefficients for relationships between FFR harmonic encoding and late EP measures across all subjects

| | | Brainstem harmonic encoding | |
		A H4 amplitude	AV H3 amplitude
Cortcial	P1–N1 slope	−0.05	−0.27
	P2 latency	0.21	−0.40[*]
	N2 latency	0.42[*]	−0.36

[*] $p < 0.05$.

Table 3
Pearson correlation coefficients for relationships between peaks of the ABR to sound onset and late EP measures across all subjects

| | | ABR onset timing | |
		A delta latency	AV delta latency
Cortcial	P1–N1 slope	0.01	0.51[**]
	P2 latency	0.3	0.26
	N2 latency	0.50[**]	0.18

[**] $p < 0.01$.

3.3. Relationships between perceptual scores and neurophysiological measures

P1–N1 slope related to perceptual measures of tonal memory from the MAT-3 and Seashore tests (Table 4). In both tests, subjects were presented with two successive sequences of tones and asked to choose which tone was different in the second sequence they

Table 1
Pearson correlation coefficients for relationships between measures of FFR periodicity and late EP measures in all subjects

| | | FFR periodicity encoding | |
		A F0 amplitude	AV F0 amplitude
Cortcial	P1–N1 slope	−0.47[*]	−0.50[**]
	P2 latency	0.24	−0.49[*]
	N2 latency	0.21	−0.44[*]

[*] $p < 0.05$.
[**] $p < 0.01$.

Table 4
Pearson correlation coefficients for relationships between cortical measures and perceptual scores across all subjects

		Cortical	
		A P1–N1 slope	AV P1–N2 slope
Test scores	Loudness	0.18	−0.16
	Timbre	0.15	0.16
	SEA tonal mem	−0.34	−0.47[*]
	MAT tonal mem	−0.43[*]	−0.50[*]

[*] $p < 0.05$.

Table 5
Pearson correlation coefficients for relationships between brainstem response measures and perceptual scores across all subjects

		Brainstem response	
		Timing	Harmonic
		A delta latency	A H2 amplitude
Test scores	Loudness	−0.41[*]	0.25
	Timbre	−0.04	0.47[*]
	SEA tonal mem	−0.37	0.23
	MAT tonal mem	−0.36	0.11

[*] $p < 0.05$.

heard. The Seashore test presented pure tones, while the MAT-3 consisted of musical notes played on the piano. Across the entire subject population, standardized tonal memory scores correlated with P1–N1 slope measures for both tests in both modalities (A: $r_{MAT-3} = −0.43$, $p = 0.03$; AV: $r_{MAT-3} = −0.50$, $p = 0.01$; $r_{SEA} = −0.47$, $p = 0.02$). Correlations between neurophysiological and behavioral measures were observed between Wave δ latency and Seashore's loudness subtest ($r = −0.41$, $p = 0.04$), as well as between H2 peak amplitude and Seashore's test of timbre discrimination ($r = 0.47$, $p = 0.02$) in the A condition (Table 5).

Heterogeneity of regression slope analysis showed that the relationship between Seashore tonal memory scores and AV P1–N1 slope differed between musicians and non-musicians ($F = 4.99$, $p < 0.05$). Within group correlations showed that musicians had stronger correlations between these measures than non-musicians ($r = −0.52$ vs. $r = 0.11$, respectively).

3.4. Relationships between neurophysiological measures and extent of musical training

Subject history reports of musical training were assessed only for individuals in the musician group. Therefore, analysis of musical training measures was restricted to musicians. F0 amplitude, onset response latency and P1–N1 slope correlated with years of consistent musical practice while measures of harmonic representation correlated with the age that musicians began their training. Consistent practice among musicians was measured by the self-reported number of years, within the last ten, each player practiced his or her instrument (>3 times per week for >4 h per day). This
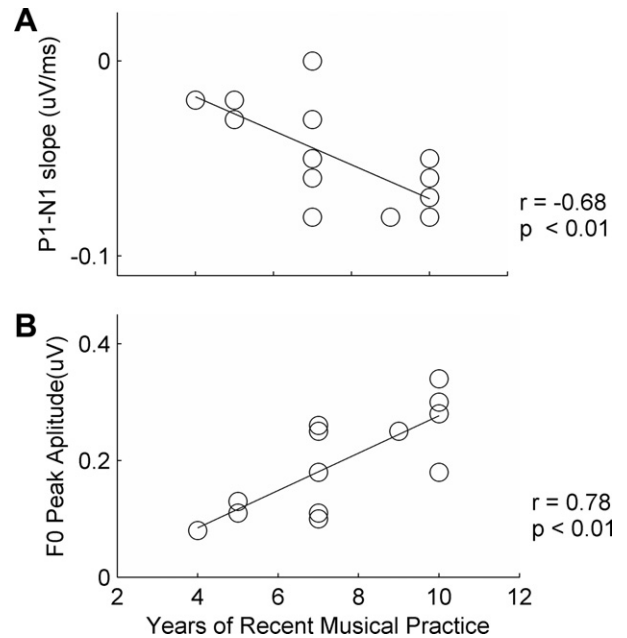

Fig. 4. Relationships between neurophysiological measures and musical training in musicians. (A) More years of consistent musical practice were associated with steeper P1–N1 slope values in the Auditory condition ($r = −0.68$, $p = 0.007$). (B) Years of consistent musical practice also correlated with brainstem measures of F0 amplitude in the Auditory and Audiovisual conditions ($r_A = 0.78$, $p = 0.001$; $r_{AV} = 0.72$, $p = 0.003$). Only data from the Auditory condition are depicted in panel B.

measure of musical training strongly correlated with F0 amplitude, onset response latency and P1–N1 slope in both modalities (Table 6, Fig. 4). More years of consistent musical practice was associated with larger F0 peak amplitudes in both conditions ($r_A = 0.79$, $p = 0.001$; $r_{AV} = 0.72$, $p = 0.003$). This measure of musical training also correlated with Wave δ latency, such that earlier latencies were associated with more years of practice. Similarly, more years of consistent practice was associated with steeper P1–N1 slopes in the A condition ($r = −0.68$, $p = 0.007$). The age that musicians began playing correlated negatively with timbre representation, as measured by H3 and H4 peak amplitude in the A condition (Table 6). That is, earlier beginning age was associated with larger harmonic peak amplitudes ($r_{H3} = −0.60$, $p = 0.047$; $r_{H4} = −0.63$, $p = 0.02$).

4. Discussion

4.1. Musician-related plasticity and corticofugal modulation

The first picture that emerges from our data is that recent musical training improves one's auditory memory and shapes composite (P1–N1) and pitch-specific encoding (F0) in a co-coordinated manner. Our EP and behavior correlations suggest that complex auditory task performance is related to the strength of the P1–N1 response. Both the Seashore and MAT-3 Tonal Memory

Table 6
Pearson correlation coefficients for relationships between EP measures and two metrics of musical experience

	Brainstem					Cortical
	Timing	Periodicity		Harmonics		Timing
	AV delta latency	A F0 amplitude	AV F0 amplitude	A H3 amplitude	A H4 amplitude	A P1–N1 slope
Age began	0.27	−0.41	−0.26	−0.60[*]	−0.63[*]	0.37
Musical practice	−0.72[**]	0.79[**]	−0.72[**]	0.36	0.40	−0.68[*]

[*] $p < 0.05$.
[**] $p < 0.01$.

tests require listeners to hold a sequence of pitches in memory and identify pitch differences in a second sequence. Scores from both tests correlated with P1–N1 slopes in both A and AV modalities such that steeper slopes were associated with higher scores. Not surprisingly, these measures are affected by musicianship: instrumental musicians performed better on the tests and had steeper P1–N1 slopes than non-musicians. Our P1–N1 results corroborate previous work showing that that musical training is associated with earlier and larger P1–N1 peaks (Fujioka et al., 2004b).

However, it was not only the individual tests and measures that were musician-related. Musicians had a statistically stronger *correlation* between this set of brain and behavior measures than non-musicians. While it is well-known that trained musicians outperform untrained controls and have more robust evoked-potentials than non-musicians, our data show that the accord, or relationship, between brain and behavior is also improved in musicians. Our data steer us one step further, however. Because steeper P1–N1 slopes are associated with more years of musical training, we can speculate that the accord between brain and behavior is strengthened with consistent years of musical training.

Interestingly, variance in the P1–N1 slope measure is also explained by peak amplitude of the fundamental frequency in the FFR across all subjects. This indicates that robust, frequency-specific representations of a sound's pitch are vital to later, composite measures of neural activity. F0 amplitude, like P1–N1 slope, also varies positively with years of consistent musical training. Taken together, the P1–N1, FFR, and Tonal Memory correlations imply that the high cognitive demand of consistent musical training improves auditory acuity and shapes composite and frequency-specific encoding in a coordinated manner.

We can interpret these data in terms of corticofugal mechanisms of plasticity. Playing music involves tasks with high cognitive demands, such as playing one's part in a musical ensemble, as well as detailed auditory acuity, such as monitoring ones intonation while playing the part. It is conceivable that the demand for complex organization and simultaneously detail-oriented information engages cortical mechanisms that are capable of refining the neural code at a basic sensory level. This idea is consistent with models of perceptual learning that involve perceptual weighting with feedback (Nosofsky, 1986). In this case, attention to pitch-relevant cues would increase the perceptual weighting of these dimensions. Positive and negative feedback in the form of harmonious pitch cues and auditory beats could shift the weighting system to represent the F0 more faithfully. Our theory also comports with the Reverse Hierarchy Theory (RHT) of visual learning (Ahissar, 2001). The RHT suggests that goal-oriented behavior shapes neural circuitry in "reverse" along the neural hierarchy. Applied to our data, this would suggest that the goal of accurately holding successive pitches in auditory memory would first tune complex encoding mechanisms (e.g. cortical), followed by a "backward" search for increased signal-to-noise ratios of pitch related features in sensory systems (e.g. brainstem). Indeed, this interpretation has been invoked by Kraus and colleagues to interpret subcortical changes in subcortical function associated with short-term training and lifelong language and music experience in language-compromised, typical listeners and auditory experts (e.g. Kraus and Banai, 2007; Banai et al., 2007; Song et al., 2008; Wong et al., 2007; Musacchia et al., 2007). Finally, recent models also suggest that top-down guided plasticity may be mediated by sensory-specific memory systems. Instead of being generated by prefrontal and parietal memory systems, it is thought that sensory memory is directly linked to the sensory system used to encode the information (Pasternak and Greenlee, 2005). In this way, enhancements at the sensory encoding level would increase the probability of creating accurate sensory memory traces.

With respect to our other evoked-potential measures, the second concept to emerge is the relationship between auditory discrimination of fine-grained stimulus features, such as timbre, the neural representation of those features in the FFR and the age at which musical training began. Seashore's test of timbre is a two-alternative forced choice procedure that asks subjects to discriminate whether a second sound differs (in perceived timbre tonality) from the first. Timbre is widely understood to be the sound quality which can distinguish sounds with the same pitch and loudness (e.g., the quality a trumpet versus a violin). Acoustic differences such as harmonic context and sound rise time give rise to this perception (Erickson, 1978). In contrast to the previous case where behavior was linked with later, cortical EPs, behavioral scores on timbre discrimination were directly related to harmonic components of the FFR. Specifically, larger H3 and H4 amplitudes were associated with better timbre scores across all subjects. However, like the relationship between behavior and later cortical peak components, the representation of harmonics does seem to simply distinguish musicians as a group from non-musicians because the amplitude of H3 and H4 was positively correlated with the age at which musical training began.

One interpretation of these data is that tasks requiring auditory discrimination of subtle stimulus features depend more heavily upon stimulus-specific encoding mechanisms. Consistent with theories of corticofugal modulation, it is possible that cognitive demands of timbre discrimination tasks progressively tune sensory encoding mechanisms related to harmonic representation. In this case, lifelong experience distinguishing between instruments may strengthen the direct link between the sensory representation of harmonic frequencies and the perception that they subserve. It is important to note that cortical EPs are not completely bypassed in timbre perception. H3 and H4 amplitude in the A condition correlate to P2 and N2 peak latency, respectively. Consequently, these cortical response components may be related to the encoding of these subtle stimulus features. Perhaps this is one of the reasons that timbre discrimination, anecdotally, takes longer to perceive than the pitch of a note. A similar type of mechanism may underlie the correlation between Wave δ latency and loudness discrimination ability. However, the functional relationship between response to sound onset and the perception of a sound's amplitude is less transparent, although response timing is a common neural reflection of sound intensity (Jacobson, 1985).

4.2. The continuum between expert and impaired experience

The current study shows how extensive musical training strengthens the relationship between measures of putatively low- and high-levels of neural encoding. On the other end of the experience continuum, previous data in school-aged children indicate that the strength of these relationships can be weakened in the language-impaired system. In normal-learning children, Wible and colleagues demonstrated a relationship between brainstem response timing and cortical response fidelity to signals presented in background noise, which learning-impaired (LI) children fail to show (Wible et al., 2005). The normal pattern of hemispheric asymmetry to speech was also disrupted in LI children with brainstem response abnormalities (Abrams et al., 2006). In addition, children with brainstem response timing deficits showed reduced cortical sensitivity to acoustic change (Banai et al., 2005). Taken together with those findings in language-impaired systems, the current findings suggest a continuum of cohesive brainstem-cortical association that can be disrupted in impaired populations and strengthened by musical training.

5. Conclusion

Overall, our data indicate that the effects of musical experience on the nervous system include relationships between brainstem and cortical EPs recorded simultaneously in the same subject to seen and heard speech. Moreover, these relationships were related to behavioral measures of auditory perception and were stronger in the audiovisual condition. This implies that musical training promotes plasticity throughout the auditory and multisensory pathways. This includes encoding mechanisms that are relevant for musical sounds as well as for the processing of linguistic cues and multisensory information. This is in line with previous work which has shown that experience which engages cortical activity (language, music, auditory training) shapes subcortical circuitry, likely through corticofugal modulation of sensory function. That is, brainstem activity is affected by lifelong language expertise (Krishnan et al., 2005), its disruption (reviewed in Banai et al., 2007) and music experience (Musacchia et al., 2007; Wong et al., 2007) as well as by short term training (Russo et al., 2005; Song et al., 2008; Russo et al., 2005). Consistent with this notion of reciprocal cortical–subcortical interaction, the current work shows a relationship between sensory representation of stimulus features and cortical peaks. Specifically, we find that musical training tunes stimulus feature-specific (e.g. onset response/FFR) and composite (e.g. P1–N2) encoding of auditory and multi-sensory stimuli in a coordinated manner. We propose that the evidence for corticofugal mechanisms of plasticity [e.g. (Suga and Ma, 2003)] as well as the theories that these data drive (Ahissar and Hochstein, 2004), combined with theories of music acquisition and training [e.g. (Hannon and Trainor, 2007)] provide a theoretical framework for our findings. Further research is needed to determine directly how top-down or bottom-up mechanisms may contribute to music-related plasticity in the cortical/subcortical auditory pathway axis. Experiments, such as recording the time course of brainstem-cortical interactions could prove to be especially fruitful in this area.

Acknowledgements

References

Abrams, D.A., Nicol, T., Zecker, S.G., Kraus, N., 2006. Auditory brainstem timing predicts cerebral asymmetry for speech. J. Neurosci. 26 (43), 11131–11137.

Ahissar, M., 2001. Perceptual training: a tool for both modifying the brain and exploring it. Proc. Natl. Acad. Sci. USA 98 (21), 11842–11843.

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8 (10), 457–464.

Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clin. Neurophysiol. 119 (4), 922–933.

Banai, K., Abrams, D., Kraus, N., 2007. Sensory-based learning disability: insights from brainstem processing of speech sounds. Int. J. Audiol. 46 (9), 524–532.

Banai, K., Nicol, T., Zecker, S.G., Kraus, N., 2005. Brainstem timing: implications for cortical processing and literacy. J. Neurosci. 25 (43), 9850–9857.

Celesia, G.G., 1968. Auditory evoked responses. Intracranial and extracranial average evoked responses. Arch. Neurol. 19, 430–437.

Colwell, R., 1970. Musical Achievement Test 3 and 4. Interpretive Manual. Follet Education Corpopration.

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., Taub, E., 1995. Increased cortical representation of the fingers of the left hand in string players. Science 270 (5234), 305–307.

Erickson, R., 1978. Sound Structure in Music. University of California Press, Berkeley.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004a. Musical training enhances automatic encoding of melodic contour and interval structure. J. Cogn. Neurosci. 16 (6), 1010–1021.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004b. Musical training enhances automatic encoding of melodic contour and interval structure. J. Cogn. Neurosci. 16 (6), 1010–1021.

Galbraith, G.C., Doan, B.Q., 1995. Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. Int. J. Psychophysiol. 19 (3), 203–214.

Galbraith, G.G., Jhaveri, S.P., Kuo, J., 1997. Speech-evoked brainstem frequency-following responses during verbal transformations due to word repetition. Electroencephalogr. Clin. Neurophysiol. 102, 46–53.

Galbraith, G.C., Chae, B.C., Cooper, J.R., Gindi, M.M., Ho, T.N., Kim, B.S., Mankowski, D.A., Lunde, S.E., 2000. Brainstem frequency-following response and simple motor reaction time. Int. J. Psychophysiol. 36 (1), 35–44.

Galbraith, G.C., Olfman, D.M., Huffman, T.M., 2003. Selective attention affects human brain stem frequency-following response. Neuroreport 14 (5), 735–738.

Galbraith, G.C., Amaya, E.M., de Rivera, J.M., Donan, N.M., Duong, M.T., Hsu, J.N., Tran, K., Tsang, L.P., 2004. Brain stem evoked response to forward and reversed speech in humans. Neuroreport 15 (13), 2057–2060.

Gaser, C., Schlaug, G., 2003. Brain structures differ between musicians and non-musicians. J. Neurosci. 23 (27), 9240–9245.

Hall, J.W.I.I., 1992. Handbook of Auditory Evoked Responses. Allyn and Bacon, Needham Heights, MA.

Hannon, E.E., Trainor, L.J., 2007. Music acquisition: effects of enculturation and formal training on development. Trends Cogn. Sci. 11 (11), 466–472.

Huffman, R.F., Henson Jr., O.W., 1990. The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. Brain Res. Rev. 15, 295–323.

Hood, L.J., 1998. Clinical Applications of the Auditory Brainstem Response. Singular, San Diego.

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. Hear. Res. 59 (2), 179–188.

Jacobson, J.T., 1985. The Auditory Brainstem Response. College-Hill Press, San Diego.

Johnson, K.L., Nicol, T.G., Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. Ear Hear. 26 (5), 424–434.

Kral, A., Eggermont, J.J., 2007. What's to lose and what's to learn: development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Res. Rev. 56, 259–269.

Kelly, J.P., Wong, D., 1981. Laminar connections of the cat's auditory cortex. Brain Res. 212, 1–15.

Kincaid, A.E., Duncan, S., Scott, S.A., 2002. Assessment of fine motor skill in musicians and nonmusicians: differences in timing versus sequence accuracy in a bimanual fingering task. Percept. Motor Skill. 95 (1), 245–257.

Kraus, N., Banai, K., 2007. Auditory Processing Malleability: focus on language and music. Curr. Dir. Psychol. Sci. 16 (2), 105–109.

Kraus, N., Nicol, T., 2005. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. Trends Neurosci. 28 (4), 176–181.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. 25 (1), 161–168.

Magne, C., Schon, D., Besson, M., 2006. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. J. Cogn. Neurosci. 18 (2), 199–211.

Marques, C., Moreno, S., Luis, C.S., Besson, M., 2007. Musicians detect pitch violation in a foreign language better than nonmusicians: behavioral and electrophysiological evidence. J. Cogn. Neurosci. 19 (9), 1453–1463.

Munte, T.F., Nager, W., Beiss, T., Schroeder, C., Altenmuller, E., 2003. Specialization of the specialized: electrophysiological investigations in professional musicians. Ann. N.Y. Acad. Sci. 999, 131–139.

Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr. Clin. Neurophysiol. 35, 665–667.

Musacchia, G., Sams, M., Nicol, T., Kraus, N., 2006. Seeing speech affects acoustic information processing in the human brainstem. Exp. Brain Res. 168 (1–2), 1–10.

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 (40), 15894–15898.

Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. J. Exp. Psychol. Gen. 115 (1), 39–61.

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., Katoh, A., Imabayashi, E., 2001. Functional anatomy of musical perception in musicians. Cereb. Cortex 11 (8), 754–760.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12 (1), 169–174.

Pantev, C., Ross, B., Fujioka, T., Trainor, L.J., Schulte, M., Schulz, M., 2003. Music and learning-induced cortical plasticity. Ann. N.Y. Acad. Sci. 999, 438–450.

Pasternak, T., Greenlee, M.W., 2005. Working memory in primate sensory systems. Nat. Rev. Neurosci. 6 (2), 97–107.

Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. Clin. Neurophysiol. 115 (9), 2021–2030.

Russo, N.M., Nicol, T.G., Zecker, S.G., Hayes, E.A., Kraus, N., 2005. Auditory training improves neural timing in the human brainstem. Behav. Brain Res. 156 (1), 95–103.

Russo, N.M., Bradlow, A.R., Skoe, E., Trommer, B.L., Nicol, T., Zecker, S., Kraus, N., in press. Deficient brainstem encoding of pitch in children with autism spectrum disorders. Clin. Neurophys.

Schlaug, G., Jancke, L., Huang, Y., Staiger, J.F., Steinmetz, H., 1995. Increased corpus callosum size in musicians. Neuropsychologia 33 (8), 1047–1055.

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 (7), 688–694.

Seashore, C.E., 1919. The Psychology of Musical Talent. Silver, Burdett and Company.

Saldana, E., Feliciano, M., Mugnaini, E., 1996. Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of the intracollicular projections. J. Comp. Neurol. 371, 15–40.

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472.

Song, J.H., Skoe, E., Wong, P.C., Kraus, N., 2008. Plasticity in the adult human brainstem following short-term linguistic training. J. Cogn Neurosci. March 27. [Epub ahead of print].

Suga, N., Ma, X., 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. Nat. Rev. Neurosci. 4 (10), 783–794.

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), 1–10.

Wible, B., Nicol, T., Kraus, N., 2005. Correlation between brainstem and cortical auditory processes in normal and language-impaired children. Brain 128 (Pt 2), 417–423.

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 (4), 420–422.

Xu, T., Krishnan, A., Gandour, J.T., 2006. Specificity of experience-dependent pitch representation in the brainstem. Neuroreport 17 (15), 1601–1605.

Yan, W., Suga, N., 1998. Corticofugal modulation of the midbrain frequency map in the bat auditory system. Nat. Neurosci. 1 (1), 54–58.

Zhou, X., Jen, P.H., 2000. Brief and short-term corticofugal modulation of subcortical auditory responses in the big brown bat, *Eptesicus fuscus*. J. Neurophysiol. 84 (6), 3083–3087.