Behavioral/Systems/Cognitive

Selective Subcortical Enhancement of Musical Intervals in Musicians

Kyung Myun Lee,[1,2] **Erika Skoe,**[2,3] **Nina Kraus,**[2,3,4,5] **and Richard Ashley**[1,6]

[1]Bienen School of Music, [2]Auditory Neuroscience Laboratory, and Departments of [3]Communication Sciences, [4]Neurobiology and Physiology, [5]Otolaryngology, and [6]Cognitive Science, Northwestern University, Evanston, Illinois 60208

By measuring the auditory brainstem response to two musical intervals, the major sixth (E3 and G2) and the minor seventh (E3 and F#2), we found that musicians have a more specialized sensory system for processing behaviorally relevant aspects of sound. Musicians had heightened responses to the harmonics of the upper tone (E), as well as certain combination tones (sum tones) generated by nonlinear processing in the auditory system. In music, the upper note is typically carried by the upper voice, and the enhancement of the upper tone likely reflects musicians' extensive experience attending to the upper voice. Neural phase locking to the temporal periodicity of the amplitude-modulated envelope, which underlies the perception of musical harmony, was also more precise in musicians than nonmusicians. Neural enhancements were strongly correlated with years of musical training, and our findings, therefore, underscore the role that long-term experience with music plays in shaping auditory sensory encoding.

Introduction

With long-term musical experience, the musician's brain has shown functional and structural adaptations for processing sound (Pantev et al., 2001; Gaser and Schlaug, 2003; Peretz and Zatorre, 2005). Prior investigations into the neurological effects of musical experience have mainly focused on the neural plasticity of the cortex (Shahin et al., 2003; Trainor et al., 2003; Kuriki et al., 2006; Rosenkranz et al., 2007; Lappe et al., 2008), but recent studies have shown that neural plasticity also extends to the subcortical auditory system. This is evidenced by enhanced auditory brainstem response (ABR) phase locking to fundamental pitch and the harmonics of the fundamental and by earlier response latencies in subcortical responses to musical, linguistic, and emotionally valent nonspeech sounds (Musacchia et al., 2007, 2008; Wong et al., 2007; Strait et al., 2009) (for review, see Kraus et al., 2009). Because playing an instrument and listening to music involves both high cognitive demands and auditory acuity, these subcortical enhancements may result from corticofugal (top–down) mechanisms. Here, we hypothesize that, as a result of this dynamic cortical-sensory interplay, such subcortical enhancements should be evident in the behaviorally relevant aspects of sound. To investigate this, we examined musicians' and nonmusicians' responses to musical intervals. Our results show that musicians have selective neural enhancements for behaviorally relevant components of musical intervals, namely the upper tone and components that reflect the interaction of the two tones (combination tones and temporal envelope periodicity).

The musical interval, in which two tones are played simultaneously, plays a fundamental role in the structure of music because the vertical relation of two concurrent tones underlies musical harmony, the most distinctive element of music. Acoustically, the perception of two concurrent tones is not the simple summation of both tones. When two tones are played simultaneously, the two tones interact, and their phase-relationships produce the perception of combination tones that are not physically present in the stimulus (Moore, 2003). Combination tones have percepts corresponding to the frequencies, $f_1 - k(f_2 - f_1)$, where f_1 and f_2 denote frequencies of two tones ($f_1 < f_2$) and k is a positive integer (Smoorenburg, 1972). Combination tones are derived from the distortion products (DPs) generated by the nonlinear behavior of the auditory system (Robles et al., 1991; Large, 2006). DPs have been extensively studied, but most of the focus has been at the level of the cochlea (Kim et al., 1980). Less is known about human subcortical processing of harmonically complex musical intervals (Fishman et al., 2000, 2001; Tramo et al., 2001; Larsen et al., 2008). Our current research explores this topic by measuring the brainstem responses to consonant and dissonant intervals in musically trained and control subjects. Previous studies have demonstrated the existence of nonlinear components (including $f_2 - f_1$ and $2f_1 - f_2$) in the human brainstem response to intervals composed of pure tones (f_1 and f_2) (Greenberg et al., 1987; Chertoff and Hecox, 1990; Rickman et al., 1991; Chertoff et al., 1992; Galbraith, 1994; Krishnan, 1999; Pandya and Krishnan, 2004; Elsisy and Krishnan, 2008). We, therefore, expected that distortion products would occur in response to harmonically rich musical intervals.

As a relay station for transmitting information from the inner ear to the cortex, the brainstem plays a role in the unconscious, sensory processing of stimuli. The neural activity of brainstem nuclei can be measured by recording the evoked auditory brainstem response (Hood, 1998). A major neural generator of the
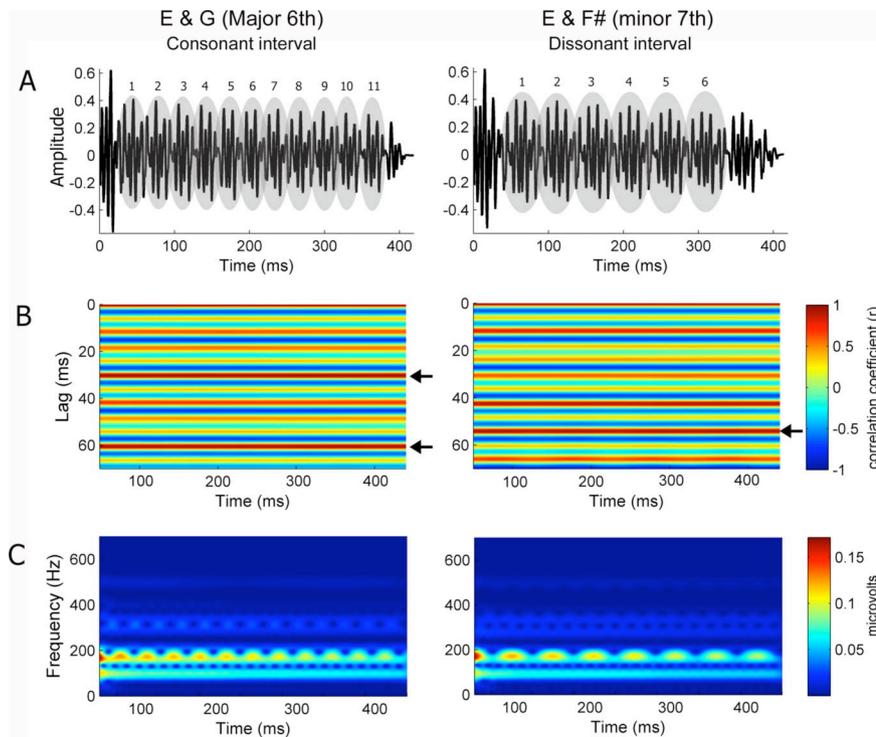
Figure 1. Consonant and dissonant musical stimuli (left and right, respectively). **A**, Temporal envelopes. Numbers indicate the cycle of the amplitude modulation. **B**, Autocorrelograms. The time indicated on the *x*-axis refers to the midpoint of each 150 ms time bin analyzed. Color indicates the degree of periodicity, and the arrows indicate the highest periodicity for each stimulus. The consonant interval shows the highest periodicity at 30.25 ms cycle, whereas the dissonant interval shows it at 54.1 ms. **C**, Spectrograms.

ABR is the midbrain inferior colliculus (IC) (Worden and Marsh, 1968; Moushegian et al., 1973; Smith et al., 1975; Hoormann et al., 1992; Chandrasekaran and Kraus, 2009). IC neurons are capable of phase locking to stimulus periodicities up to 1000 Hz (Schreiner and Langner, 1988; Langner, 1992). Electrophysiological responses elicited in the human brainstem represent the spectral and temporal characteristics of the stimulus with remarkable fidelity (Russo et al., 2004; Krishnan et al., 2005; Banai et al., 2009). Here, we focus on the spectral resolution and the temporal precision of the auditory brainstem response, to compare how musicians and nonmusicians represent musical intervals.

Materials and Methods

Subjects. Twenty-six adults participated in this study. All subjects reported no audiologic or neurologic deficits, had normal click-evoked auditory brainstem response latencies, and had binaural audiometric thresholds at or below 20 dB HL for octaves from 125 to 4000 Hz. Subjects completed a questionnaire that assessed musical experience in terms of beginning age, length, and type of performance experience. Subjects were divided into musicians and nonmusicians according to their years of musical experience: 10 subjects (seven females and three males; mean age, 25.8 years; six pianists, two violinists, and two vocalists) were categorized as "musicians," with 10 or more years of musical training that began at or before the age of 7, and 11 subjects (five females and six males; mean age, 23.5 years) were categorized as "nonmusicians," with <3 years of musical training. The other five subjects (four females and one male; mean age, 23.2), who failed to meet either of these criteria, were categorized as "amateur musicians," and their data were used only in the correlation analyses. Informed consent was obtained from all subjects. The Institutional Review Board of Northwestern University approved this research.

Stimuli. Two types of musical intervals were presented: one consonant and one dissonant. The consonant interval was a major sixth consisting of E3 (166 Hz) and G2 (99 Hz), and the dissonant interval was a minor seventh composed of E3 (166 Hz) and F#2 (93 Hz). Thus, both intervals have a common upper tone, E3. The stimuli were 400 ms in length (Fig. 1*A*), and the timbre was an electric piano sound (Fender Rhodes recorded from a digital synthesizer). The lower tone of each interval, G2 and F#2, was presented 10 ms later than the upper tone, E3, so that the upper and lower tones would be heard as being distinct from one another.

Procedure. The two musical intervals, E3/G2 and E3/F#2, were presented in separate testing blocks with block order alternated across subjects. The stimuli were binaurally presented through insert earphones (ER3; Etymotic Research) at an intensity of ~70 dB sound pressure level (Neuroscan Stim; Compumedics) with alternating polarities to eliminate the cochlear microphonic. Interstimulus interval ranged from 90 to 100 ms. During testing, subjects watched a muted movie of their choice with subtitles. Data collection followed procedures outlined in Wong et al. (2007). After two blocks of musical intervals, responses to three single tones (E3, G2, and F#2) were collected, to rule out the existence of combination tones in the single-tone condition and facilitate the identification of the fundamental frequency (f_0) and harmonics in the interval conditions.

Responses were collected using Scan 4.3 Acquire (Neuroscan; Compumedics) with four Ag-AgCl scalp electrodes, differently recorded from Cz (active) to linked earlobes references, with the forehead ground. Contact impedance was <5 kΩ for all electrodes. For the musical intervals, two sub blocks of ~3000 sweeps per sub block were collected at each stimulus polarity with a sampling rate of 20 kHz. For the single tones, 1000 sweeps per tone were collected. Filtering, artifact rejection, and averaging were performed off-line using Scan 4.3 (Neuroscan; Compumedics). Reponses were bandpass filtered from 20 to 2000 Hz (12 dB/oct roll off), and trials with activity greater than ±35 μV were considered artifacts and rejected, such that the final number of sweeps was 3000 ± 100. Responses of alternating polarities were added together to isolate the neural response by minimizing the stimulus artifact and cochlear microphonic (Gorga et al., 1985).

Analysis. The ABR represents the spectral and temporal characteristics of the stimulus by the neural interspike interval (fundamental frequency and harmonics), the latency (time), and the amplitude envelope of the response. To evaluate the spectral composition of the response, fast Fourier transform analysis was performed over the frequency-following response (FFR), the most periodic portion of response (50–350 ms). The initial 50 ms was excluded to avoid the potential confound of the nonperiodic attack information that relates to the stimulus timbre. For each subject, average spectral response amplitudes were computed over 5-Hz-wide bins surrounding the f_0 and harmonics of each tone, as well as the combination tones. Table 1 displays frequency regions of interest for each interval. For each frequency bin, an independent samples *t* test was performed between musicians and nonmusicians. For bins showing significant group differences, Pearson's correlations between musical experience and spectral amplitudes were calculated.

Second, a running autocorrelation analysis was used to evaluate the temporal envelope of the stimuli and responses, relating to the periodicity of amplitude modulation created by the interaction of two tones. The resulting running-autocorrelogram (lag vs time) graphically represents signal periodicity over the course of a time-varying waveform. Autocorrelation was performed on 150 ms bins, and the maximum (peak) autocorrelation value (expressed as a value between −1 and 1) was recorded for each bin, with higher values indicating more periodic time frames

Table 1. Frequencies of f_0, harmonics, and combination tones examined in the brainstem response to musical intervals

Consonant interval[a]		Dissonant interval[b]	
Frequency	Component	Frequency	Component
32 Hz	$(2f_1 - f_2)$	20 Hz	$(2f_1 - f_2)$
67 Hz	$(f_2 - f_1)$	53 Hz	$(2f_2 - 3f_1)$
99 Hz	G f_0	73 Hz	$(f_2 - f_1)$
131 Hz	$(3f_1 - f_2)$	93 Hz	F# f_0
166 Hz	E f_0	113 Hz	$(3f_1 - f_2)$
198 Hz	G H_2	166 Hz	E f_0
233 Hz	$(2f_2 - f_1)$	186 Hz	F# H_2
265 Hz	$(f_2 + f_1)$	239 Hz	$(2f_2 - f_1)$
297 Hz	G H_3	259 Hz	$(f_2 + f_1)$
332 Hz	E H_2	279 Hz	F# H_3
364 Hz	$(f_2 + 2f_1)$	332 Hz	E H_2
396 Hz	G H_4	352 Hz	$(f_2 + 2f_1)$
431 Hz	$(2f_2 + f_1)$	372 Hz	F# H_4
463 Hz	$(f_2 + 3f_1)$	425 Hz	$(2f_2 + f_1)$
498 Hz	E H_3	445 Hz	$(f_2 + 3f_1)$
530 Hz	$(2f_2 + 2f_1)$	498 Hz	E H_3
664 Hz	E H_4	518 Hz	$(2f_2 + 2f_1)$
		664 Hz	E H_4

Combination tones generated by two tones of each interval are denoted by parentheses. f_2 denotes upper tone, and f_1 denotes lower tone. The brainstem response represented difference tone $f_2 - f_1$, cubic difference tone $2f_1 - f_2$, and sum tones (e.g., $f_2 + f_1$, $f_2 + 2f_1$, and $f_2 + 3f_1$).
[a]f_1 = G2 (99 Hz), f_2 = E3 (166 Hz).
[b]f_1 = F#2 (93 Hz), f_2 = E3 (166 Hz).

(150 total bins, 1 ms interval between the start of each successive bin). The strength of the envelope periodicity was calculated as the average of the autocorrelation peaks (maximum r values) across the 150 bins for each subject. The sharpness of phase locking was calculated from the autocorrelation function as the distance (i.e., time shift) between the average maximum r (max r) and max $r -0.3$. The autocorrelograms of the stimuli (Fig. 1B, left panel) show that the consonant interval has the highest average periodicity ($r = 0.99$) at the 30.25 ms cycle. The temporal envelope of the dissonant interval is a bit more complicated than that of the consonant interval, as there are two dominant periodicities; the highest average periodicity ($r = 0.96$) occurs at 54.1 ms, and the second highest ($r = 0.93$) occurs at 42.45 ms (Fig. 1B, right panel).

Results
Encoding of musical intervals in the brainstem
The brainstem response faithfully represented the fundamental frequencies (f_0 s) and harmonics of the two tones composing each interval (Fig. 2). Specifically, the f_0 s of the individual tones were encoded with larger amplitudes than the harmonics, reflecting the relative amplitudes in the stimuli (Fig. 2). However, unlike the spectra of the stimuli, the brainstem response amplitude to f_0 of the upper tone (E) was smaller than the lower tone in the consonant interval (G, t test, $p < 0.001$) and the dissonant interval (F#, t test, $p < 0.01$), reflecting the low-pass characteristics of auditory brainstem responses (Schreiner and Langner, 1988).

The individual response spectra show prominent peaks not only at the frequencies of f_0 and harmonics but also at the frequencies that do not physically exist in the spectra of stimuli. Most of these components were found to correspond with the frequencies of combination tones produced by the interaction of the single tones in each interval (Table 1). To rule out the possibility that these combination tones are the result of acoustic or electric artifacts arising from presentation or collection setup, we presented the original stimuli through the Neuroscan and Etymotic equipment into a Bruel & Kjaer 2-CC coupler and recorded the output. This recorded output waveform did not show any spectral component corresponding to the combination tones.

Differences between musicians and nonmusicians
Significant group differences were found in the spectral analysis of the FFR period. For the consonant interval (Fig. 3A), musicians showed significantly larger amplitudes for the harmonics of the upper tone E (t test, H_2, $p < 0.01$; H_3, $p < 0.05$; H_4, $p < 0.05$) and combination tones, $f_2 + 2f_1$ and $f_2 + 3f_1$ (t test, $p < 0.05$; $p < 0.05$). Figure 3B displays the amplitudes of these frequencies with significant group differences indicated. Furthermore, the number of years of musical training was significantly correlated with the amplitude of each of the frequencies, such that the longer a person has been playing, the larger the amplitude (Fig. 3C). This analysis included the amateur group (total, 26) and showed that the second harmonic of E (332 Hz) and $f_2 + 3f_1$ (463 Hz) are the frequencies most positively correlated with the years of musical training (Fig. 3D).

For the dissonant interval (Fig. 4A), consistent with the result for the consonant interval, the amplitudes of harmonics of the upper tone E were significantly larger in musicians (t test, H_2, $p < 0.01$; H_3, $p < 0.05$; H_4, $p < 0.05$), whereas f_0 was not (Fig. 4A,B). Among combination tones, the amplitudes of $f_2 + 2f_1$ and $f_2 + 3f_1$ were larger in musicians ($p < 0.01$; $p < 0.01$). A positive correlation with the years of musical training was also obtained for the amplitudes of the second and fourth harmonic of E and sum tones, $f_2 + 2f_1$ and $f_2 + 3f_1$ (Fig. 4C). In addition, musicians showed larger amplitudes for frequencies that are not present in the stimulus and neither harmonics nor combination tones, such as 384, 393, and 404 Hz ($p < 0.01$; $p < 0.01$; $p < 0.01$), and interestingly, the amplitudes of 393 and 404 Hz were also positively correlated with years of musical training ($r = 0.573$, $p < 0.01$; $r = 0.544$, $p < 0.01$). In contrast to the consonant interval, the dissonant interval exhibited frequencies for which nonmusicians showed significantly larger amplitudes than musicians, namely the third harmonic of the lower tone F# ($p < 0.05$) and $3f_1 - f_2$ ($p < 0.05$). The amplitude of each of these two frequencies was negatively correlated with the years of musical training (F#, H_3, $r = -0.456$, $p < 0.05$; $3f_1 - f_2$, $r = -0.561$, $p < 0.01$).

A striking finding is that group differences occurred mostly for the upper tone (E) and not for the lower tone in both intervals (Fig. 5). This difference was not found even in the response to the single E (only the H_3 of E showed significant group difference at the 0.05 level). To examine the effect of tone position in detail, the amplitudes of the harmonic components of the upper and lower tones were examined by a repeated-measures ANOVA with one between-subject factor (group: musicians, nonmusicians) and three within factors (interval: consonant, dissonant; tone position: upper, lower; and harmonic component: f_0, H_2, H_3, H_4). A significant interaction effect between group and tone position was found ($F = 19.7$, $p < 0.001$): this underscores the finding that for both intervals, musicians showed larger amplitudes than nonmusicians only for the upper tone. There was also a significant triple interaction between tone position, harmonic component, and group ($F = 3.6$, $p < 0.05$). This interaction was attributable to musicians having larger amplitudes only for the harmonics of the upper tone: H_2, H_3, H_4, but not for the f_0. Given that in the spectra of the stimuli the harmonics of the lower tone have physically higher amplitudes than those of the upper tone (Fig. 2), this result suggests that musicians represent intervals with enhanced upper tones.

Both consonant and dissonant intervals have periodic temporal envelopes (Fig. 1) that can be heard as beats and relate to a perceptual dimension of roughness versus "smoothness." This percept results from the periodic amplitude variations that arise when two tones with frequencies close to one another are played
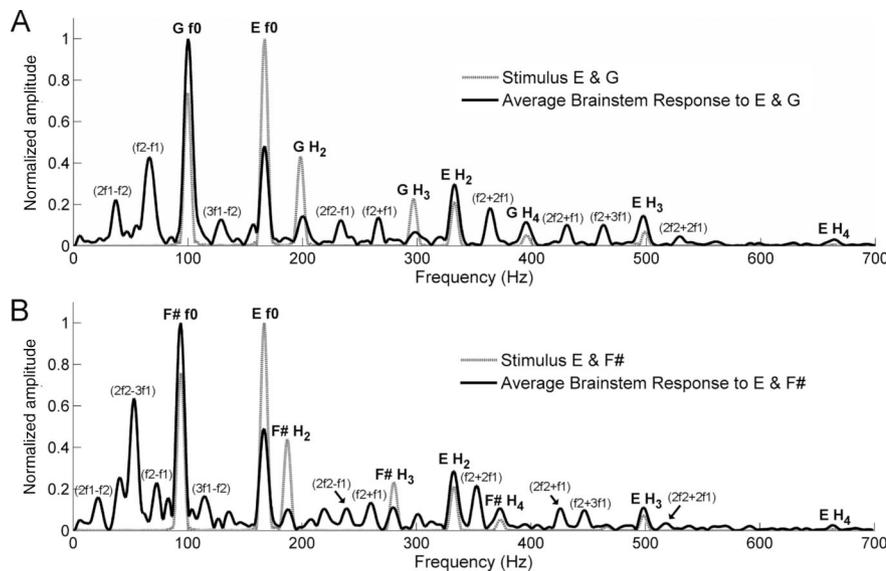
Figure 2. Stimulus and response spectra for the consonant (**A**) and the dissonant (**B**) intervals. The response spectrum displays the average of all 26 subjects. The stimulus and response spectral amplitudes are scaled relative to their respective maximum amplitudes. Parentheses denote combination tones that do not exist in stimuli. f_1 denotes the lower tone and f_2 denotes the upper tone of each interval. See also Table 1.
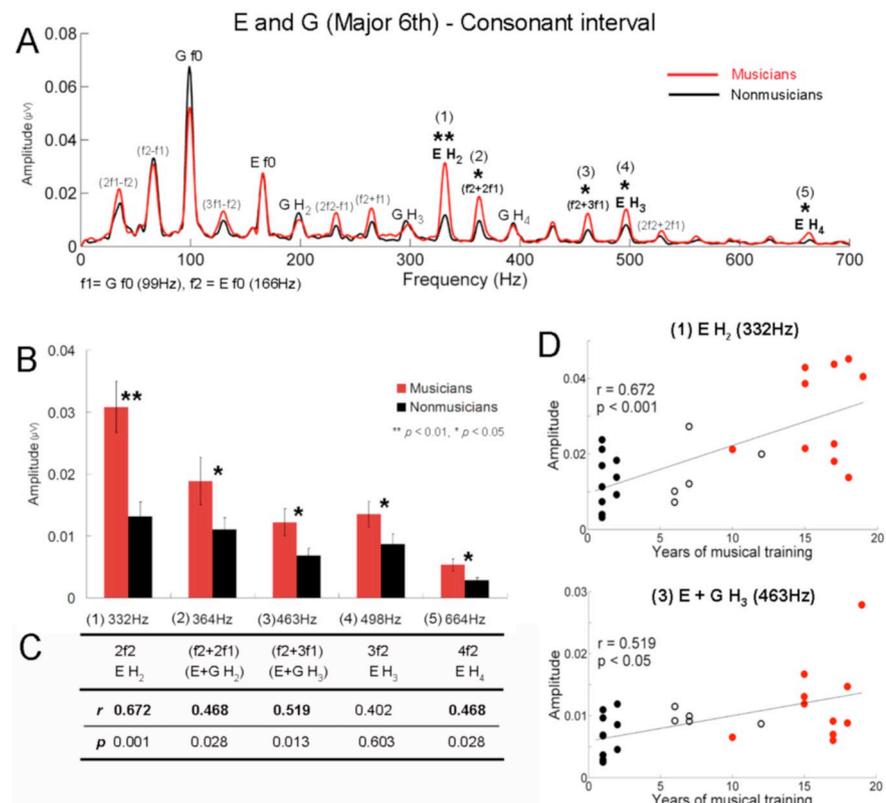
Figure 3. Musicians show heightened responses to the harmonics of the upper tone and sum tones in the consonant interval. **A**, Grand average spectra for musicians (red) and nonmusicians (black) for the consonant interval. f_1 and f_2 denote G and E. **B**, Amplitudes of frequencies showing significant group differences. Error bars represent ±1 SE. **C**, Pearson's correlations (*r* and *p*) between amplitudes and years of musical training. Significant correlations appear in bold. **D**, Individual amplitudes of the second harmonic of E (top) and $f_2 + 3f_1$ ($= E f_0 + G H_3$) (bottom) as a function of years of musical training for all subjects ($n = 26$) including the amateur group. Filled black circles represent nonmusicians, open black circles amateur musicians, and red circles represent musicians.

simultaneously (Rossing, 1990). To examine how the brainstem response represents the envelope periodicity of the stimulus, we evaluated the temporal regularity of the FFR using the autocorrelational analysis. The running autocorrelograms for musician

and nonmusician groups are plotted in Figures 6A and 7A, with red indicating the highest autocorrelations (periodicity). For the consonant interval, both groups showed the highest autocorrelation at 30.2 ms, the dominant envelope periodicity of the stimulus. There was no significant group difference in the strength of the envelope periodicity (*t* test, $p = 0.39$; musicians, 0.787; nonmusicians, 0.816). However, the morphology of the autocorrelation function was found to differ between the two groups. This difference is evident in Figure 6A: the band of color ~30.2 ms is sharper (i.e., narrower) for the musicians and broader for the nonmusicians, suggesting that the phase-locked activity to the temporal envelope is more accurate (i.e., sharper) in musicians than nonmusicians (cf. Krishnan et al., 2005). The sharpness of the autocorrelation function showed a significant group difference (*t* test, $p < 0.05$; musicians, 0.395; nonmusicians, 0.521), i.e., the width was sharper in musicians (Fig. 6A). Moreover, there was a significant correlation between years of musical training and the sharpness ($r = -0.52$, $p < 0.05$), such that the longer an individual has been practicing music, the sharper the function (Fig. 6B). Of particular interest is that nonmusicians showed strong periodicities not only at intervals of 30.2 ms but also every 10 ms. This 10 ms period corresponds to 99 Hz (the lower tone, G). Thus, it is assumed that the periodicity at 30.2 ms for the nonmusicians is driven in part by the robust neural phase locking to the period of the lower tone, G. Thus, to isolate the periodicity at 30.2 ms, we calculated the change in periodicity from 30.2 ms (r_1) to 10 ms (r_2) using ($r_1 - r_2)/(r_1 + r_2)$. Using this metric, we found a significant group difference between musicians and nonmusicians: musicians showed higher values than nonmusicians (ANOVA, $F = 6.7$, $p < 0.05$). In addition, these values were positively correlated with years of musical training ($r = 0.454$, $p < 0.05$).

In the autocorrelogram of the response to the dissonant interval, the two highest periodicities of both musicians and nonmusicians matched the pattern in the stimulus autocorrelation: musicians with 54.15 ms ($r = 0.678$) and 42.4 ms ($r = 0.550$) and nonmusicians with 54 ms ($r = 0.706$) and 42.6 ms ($r = 0.611$). Group differences were not significant. Autocorrelograms of the two groups are illustrated in Figure 7A (left, musicians; right, nonmusicians). Here, musicians also exhibited sharper phase locking to the more dominant period of the stimulus, 54.1 ms (*t* test, $p < 0.01$; musicians, 0.453; nonmusicians, 0.647). In addition, the sharpness of the autocor-

relation function was correlated with years of musical experience ($r = -0.505$, $p < 0.05$); the longer an individual has been practicing music, the sharper the phase locking (Fig. 7B).

In summary, this study revealed three main results. First, musicians showed heightened responses to the harmonics of the upper tone of musical intervals. Second, the brainstem represents the combination tones produced by two simultaneously presented musical tones with musicians showing enhanced responses for specific combination tones (sum tones). Third, musicians showed sharper phase locking to the temporal envelope periodicity generated by the phase relationships (i.e., amplitude modulation) of the two tones of the musical intervals.

Discussion

Musicians' enhancement of the upper tone in the brainstem

By comparing the brainstem response to musical intervals between musicians and nonmusicians, the present study revealed that harmonic components of an upper tone in musical intervals are significantly enhanced in musicians. This is an unexpected result from an acoustic standpoint, given that the harmonics of the upper tone have lower intensities than those of the lower tone (Fig. 2). Furthermore, because of the low-pass filter properties of the brainstem, we would have predicted better phase locking to the lower note. However, previous behavioral and electrophysiological research has demonstrated results consistent with our study. For example, behavioral studies have shown that changes occurring in the upper voice are more easily detected than those in the middle or lower voice (Palmer and Holleran, 1994; Crawley et al., 2002). Furthermore, electrophysiological studies showed larger and earlier mismatch negativity magnetic field (MMNm) responses for deviations in the upper voice melody than in the lower voice (Fujioka et al., 2005, 2008). Musicians have also been found to have more enhanced MMNm than nonmusicians. This has been attributed to cortical reorganization resulting from long-term training because the source of MMN is located mainly in the auditory cortex (Fujioka et al., 2005). Our study is the first to show that this upper tone dominance at the cortical level extends to the subcortical sensory system. Given the strong correlation between the length of musical training (years) and the extent of subcortical enhancements of the upper tone, we suggest that the subcortical representation on the upper tone is weighted more heavily as musical experience becomes more extensive. In music,
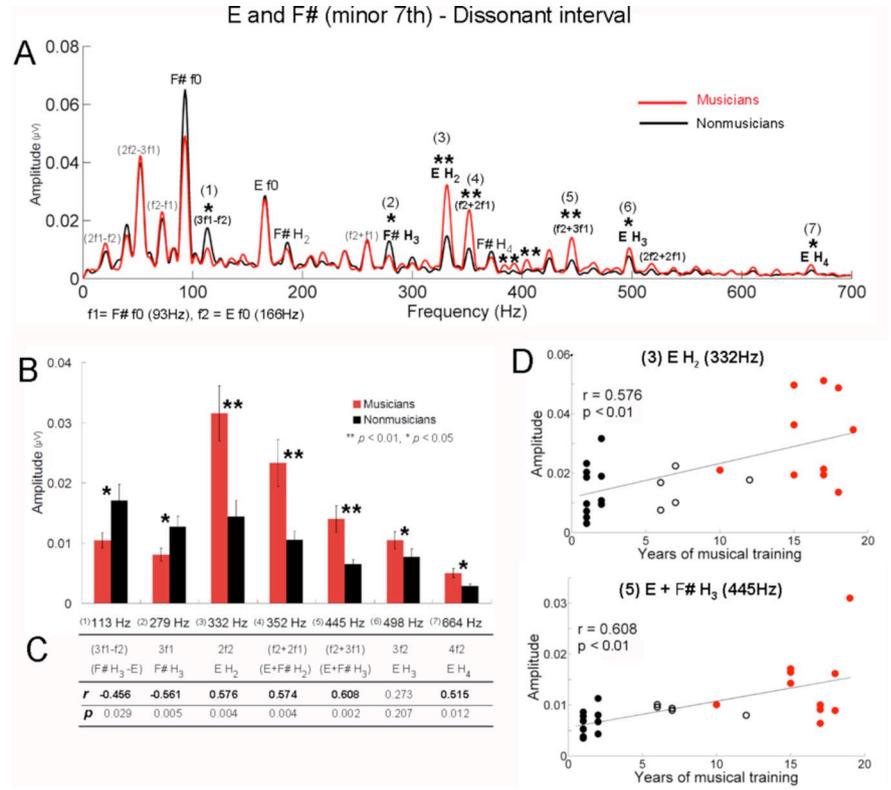

Figure 4. In the dissonant interval, musicians show heightened responses to the harmonics of the upper tone and sum tones, whereas nonmusicians show enhanced responses to the third harmonic of the lower tone and $3f_1 - f_2$. **A**, Grand average spectra for musicians (red) and nonmusicians (black) for the dissonant interval. f_1 and f_2 represent F# and E. **B**, Amplitudes of frequencies showing significant group differences. Error bars display ± 1 SE. **C**, Pearson's correlations (r and p) between amplitudes and years of musical training. Significant correlations appear in bold. **D**, Individual amplitudes of the second harmonic of E (top) and $f_2 + 3f_1$ ($= E f_0 + F\# H_3$) (bottom) as a function of years of musical training for all subjects ($n = 26$) including the amateur group. Filled black circles represent nonmusicians, open black circles represent amateur musicians, and red circles represent musicians.
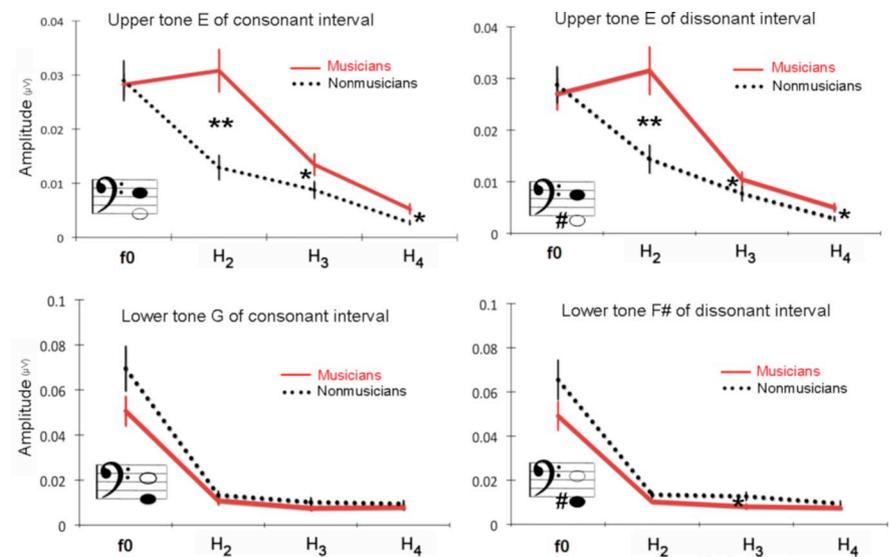

Figure 5. Amplitudes of f_0 and three harmonics of the upper and lower tones in the consonant (left) and dissonant (right) intervals. Musicians show heightened responses to the harmonics of the upper tone.

the main melodic theme is most often carried by the upper voice. To increase the perceived salience of the melody voice, performers play a melody louder than other voices and ~30 ms earlier (melody lead) (Palmer, 1997; Goebl, 2001). Thus, long-term experience with actively parsing out the melody may fine-tune the
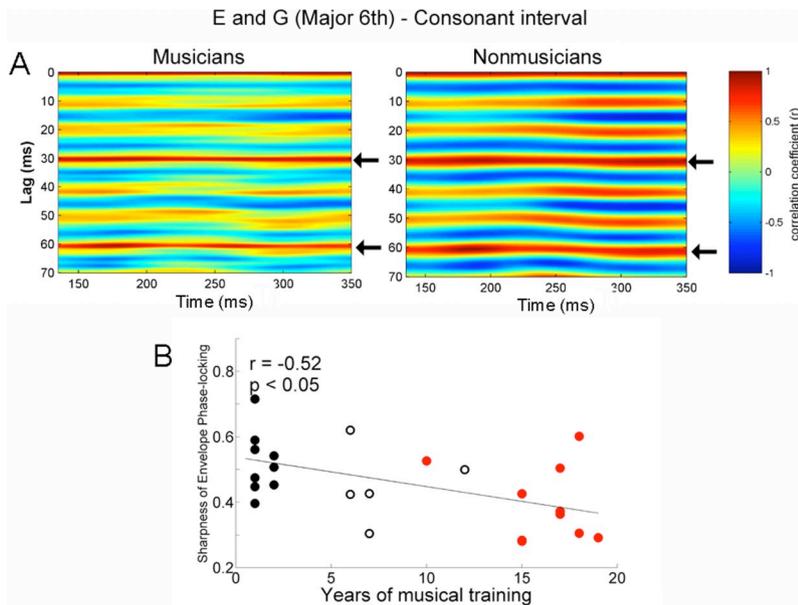
Figure 6. **A**, Autocorrelograms of the response of musicians (left) and nonmusicians (right) for the consonant interval. The time indicated on the *x*-axis refers to the midpoint of each 150 ms time bin analyzed. Color indicates the degree of periodicity. The band of color ~30.2 ms (indicated by arrows) is narrower (i.e., sharper phase locking) for the musicians than nonmusicians. This pattern is repeated at 60.4 ms, i.e., 30.2 ms ×2. **B**, Sharpness of envelope phase locking for the consonant interval as a function of years of musical training for all subjects (*n* = 26) including the amateur group. The longer an individual has been practicing music, the sharper the phase locking (i.e., narrower autocorrelation peak at 30.2 ms).
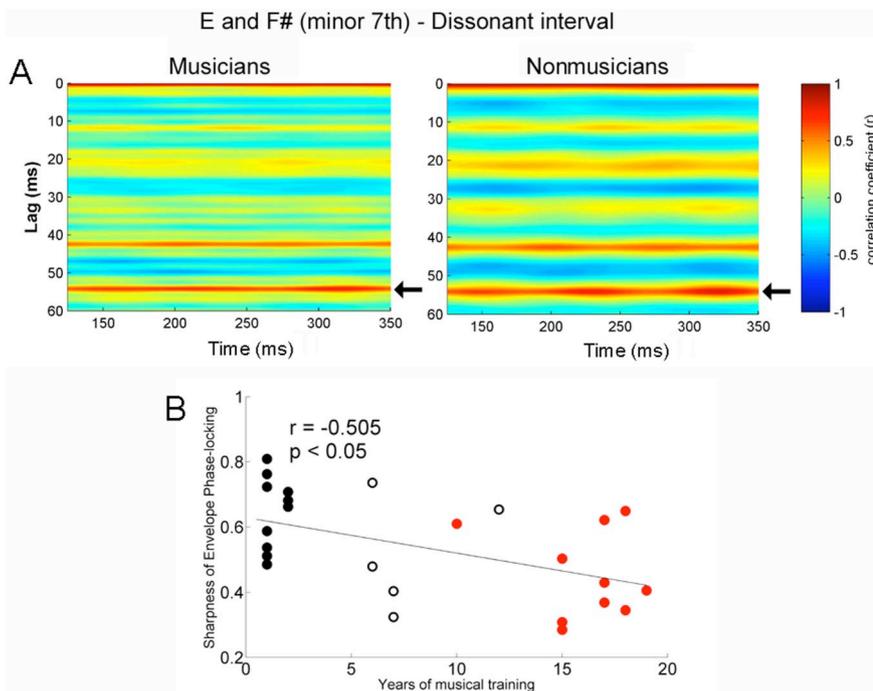
Figure 7. **A**, Autocorrelograms of the response of musicians (left) and nonmusicians (right) for the dissonant interval. The time indicated on the *x*-axis refers to the midpoint of each 150 ms time bin analyzed. Color indicates the degree of periodicity. The band of color ~54.1 ms (indicated by arrows) is sharper for the musicians and broader for nonmusicians. **B**, Sharpness of envelope phase locking for the dissonant interval as a function of years of musical training for all subjects (*n* = 26) including the amateur group. The longer an individual has been practicing music, the sharper the phase locking (i.e., narrower autocorrelation peak at 54.1 ms).

brainstem representation of the upper tones of musical intervals. Our results also underscore the role that long-term experience plays in shaping basic auditory encoding and the link between sensory function and repeated exposure to behaviorally relevant signals (Saffran, 2003; Saffran et al., 2006). This tuning of the

subcortical auditory system is likely mediated by the corticofugal pathway, a vast track of descending efferent fibers that connect the cortex and lower structures (Suga et al., 2002; Winer, 2005; Kral and Eggermont, 2007; Luo et al., 2008). In the animal model, the corticofugal system works to fine-tune subcortical sensory receptive fields of behaviorally relevant sounds by linking learned representation and the neural encoding of the physical acoustic features (Suga et al., 2002). As has been previously postulated by our group, the brainstem of musicians may be shaped by the similar corticofugal mechanisms (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009) (for review, see Kraus et al., 2009). Consistent with this corticofugal hypothesis and observations of experience-dependent sharpening of primary auditory cortex receptive fields (Fritz et al., 2007; Schreiner and Winer, 2007), we maintain that subcortical enhancements do not result simply from passive, repeated exposure to musical signals or pure genetic determinants. Instead, the refinement of auditory sensory encoding is driven by a combination of these factors and behaviorally relevant experiences such as life-long music making. This idea is reinforced by correlational analysis showing that subcortical enhancements vary as a function of musical experience.

The largest group-effects were found for H_2 but not f_0 of the upper tone in each interval. This result is consistent with previous brainstem research using Mandarin tones in which Chinese and English speakers showed the largest group difference on H_2 (Krishnan et al., 2005). This H_2 advantage over the f_0 could be explained by the behavioral relevance of H_2. Musical sounds are complex tones composed of harmonics, and the timbre or "sound quality" depends primarily on the harmonic components. Thus, long-term musical experience involving timbre-oriented training may lead musicians to be more sensitive to the timbre-related aspects of sounds. In fact, previous studies (Pantev et al., 2001; Margulis et al., 2009) showed that musicians' cortical responses are enhanced for timbres of the instrument they have been trained on, suggesting that the neural representation of timbre is malleable with training. This cortical plasticity likely extends to the brainstem resulting in the enhancement of timbre-related aspects of sound, harmonics. The group difference is possibly magnified at H_2 because among the harmonics, H_2 has the highest spectral energy in the stimulus.

Further research is needed to examine how the representation of the upper and lower tone is different between consonant and dissonant intervals in the brainstem response. In this study, overall re-

sponse patterns were similar between the consonant and the dissonant intervals. However, in the dissonant condition, nonmusicians did show a greater response for a combination tone and a harmonic of the lower tone than did musicians. To gain a deeper insight into the different neural representation of consonant and dissonant intervals, we are conducting similar research using major second and perfect fifth. This may shed light on the unexpected results of nonmusicians having larger amplitudes for the lower notes only in the dissonant condition.

Combination tones in the brainstem

This study shows that for musical intervals, the brainstem response represents combination tones derived from the DPs generated by the nonlinear behavior of the auditory system (Robles et al., 1991). At the level of the cochlea, the concurrent presentation of two tones produces distortion products which result from the nonlinearity of outer hair cell motion (Robles et al., 1991; Rhode and Cooper, 1993). Distortion product otoacoustic emissions (DPOAEs) can be measured using small microphones placed inside the ear canal, and the magnitude of the cubic difference tone ($2f_1 - f_2$) plays a role in the evaluation of hearing by providing a noninvasive tool to assess the integrity of the normal active process of the cochlea (Kemp, 2002). We observed a neural correlate of this DPOAE at $2f_1 - f_2$ in the spectra of the brainstem responses. There have been several studies demonstrating the presence of $2f_1 - f_2$ in the FFRs to two pure tones (Chertoff and Hecox, 1990; Rickman et al., 1991; Chertoff et al., 1992; Pandya and Krishnan, 2004; Elsisy and Krishnan, 2008). However, we are the first to document this phenomenon using ecologically valid stimuli, namely musical intervals composed of complex tones. Moreover, we observed $2f_1 - f_2$ occurring at low frequencies, such as 32 Hz for the consonant and 20 Hz for the dissonant interval. Because of the presence of background noise in the ear canal, DPOAEs at frequencies <1000 Hz are difficult to measure. Therefore, this result suggests that the FFR could complement the less reliable DPOAE at low frequencies (Pandya and Krishnan, 2004; Elsisy and Krishnan, 2008).

Because we recorded far-field potentials, our data cannot conclusively resolve whether these DPs in the FFR reflect cochlear and/or central nonlinearities. On the one hand, DPs measured in the FFR may be a response to DPs originally created by the mechanical properties of the cochlea (McAlpine, 2004; Pandya and Krishnan, 2004; Elsisy and Krishnan, 2008). Alternatively, the DPs may be created during neural processing of the signal. Given the size of the generating potentials, the contribution of midbrain structures to the FFR likely overshadows any response being picked up from the auditory periphery. Nevertheless, the enhancements evident in the musician FFR likely do not arise peripheral to the brainstem; to our knowledge, there is little evidence of experience-dependent plasticity at the level of the auditory nerve or cochlea (Perrot et al., 1999; Brashears et al., 2003), whereas there is ample evidence to support experience-dependent plasticity in the midbrain (Suga and Ma, 2003; Knudsen, 2007). Another piece of evidence supporting the central origin of our DPs in the FFR is that the amplitude of the cochlear distortion product decreases rapidly as the frequency ratio (F_R) between the two stimulating tones increases (Goldstein, 1967). In this study, we used F_Rs of 1.6 (166 Hz/99 Hz) and 1.7 (166 Hz/93 Hz), and very few studies have demonstrated cochlear distortion products with such wide F_Rs (Knight and Kemp, 1999, 2000, 2001; Dhar et al., 2005). Another possibility is that $2f_1 - f_2$ in the FFR reflects the amplitude modulation frequency of the $2f_1$ and f_2 components of the stimulus, given that inferior colliculus neu-

rons readily respond to the frequency of the amplitude modulation (Hall, 1979; Cariani and Delgutte, 1996; Joris et al., 2004; Aiken and Picton, 2008). The spectrograms of the stimuli show an obvious amplitude modulation between $2f_1$ and f_2 (Fig. 1C). Specifically, in the consonant interval, $2f_1$ (198 Hz) and f_2 (166 Hz) produce an amplitude modulation of the 31 ms cycle (1 s/31 ms = 32.2 Hz, which corresponds to 198 − 166 Hz) and in the dissonant interval, $2f_1$ (186 Hz) and f_2 (166 Hz) are modulated at the 50 ms cycle (1 s/50 ms = 20 Hz, which corresponds to 186 − 166 Hz). Further research is needed to explore the origin of distortion products by comparing DPOAE and the brainstem response in the same individual (Dhar et al., 2009).

Musicians' accurate neural phase locking to the temporal envelope periodicity of the interval

By comparing the periodicity of the stimulus and response waveforms, this study observed that the brainstem response represents the temporal envelope of musical intervals. This finding is consistent with previous research showing that the inferior colliculus phase-locks to the envelope and frequency of amplitude modulation of sound (Hall, 1979; Cariani and Delgutte, 1996; Joris et al., 2004; Aiken and Picton, 2008). When we compared musicians and nonmusicians, we found that musicians showed more accurate representation of the envelope periodicity of the stimulus than nonmusicians. In music, the temporal envelope of an interval is a critical cue for determining the harmonic characteristic of the interval because the sensory consonance and dissonance of an interval is related to the sensation of beats and roughness generated by amplitude modulation occurring at intervals of 4 to 50 ms (Helmholtz, 1954). Our study shows that musicians' sensory systems have been refined to process this musically important periodicity. This heightened precision of neural phase locking after long-term experience is supported by research on the neural encoding of speech sounds. For example, Krishnan et al. (2005) found Mandarin Chinese speakers have sharper phase locking for linguistically relevant Mandarin pitch contours than English speakers. By showing a correlation between years of musical training and the sharpness of the neural phase locking to the envelope periodicity, this study confirms that long-term musical experience that includes selective attention to the harmonic relation of concurrent tones modulates the subcortical representation of the behaviorally relevant features of musical sound, namely the periodicity of the amplitude modulation, which underlies musical harmony. Previous studies have shown that temporal coding properties of cortical neurons can be modified by learning (Joris et al., 2004), and we show the plasticity for the temporal coding extends to the subcortical system.

Conclusion

By measuring the brainstem responses to musical intervals of musicians and nonmusicians, we found that musicians have specialized sensory systems for processing behaviorally relevant aspects of sound. Specifically, musicians have heightened responses to the harmonics of the upper tone, a feature often important in melody perception. In addition, the acoustic correlates of consonance perception (i.e., temporal envelope) are more precisely represented in the subcortical responses of musicians. The role of long-term musical experience in shaping the subcortical system is reinforced by the strong correlation between the length of musical training and the neural representation of these stimulus features. By demonstrating that cortical plasticity to behaviorally relevant aspects of musical intervals extends to the subcortical system, this study supports the notion that subcortical tuning is

driven, at least in part, by top–down modulation by the corticofugal system.

References

Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. Hear Res 245:35–47.

Banai K, Hornickel JM, Skoe E, Nicol T, Zecker S, Kraus N (2009) Reading and subcortical auditory function. Cereb Cortex. Advance online publication. Retrieved April 20, 2009. doi:10.1093/cercor/bhp024.

Brashears SM, Morlet TG, Berlin CI, Hood LJ (2003) Olivocochlear efferent suppression in classical musicians. J Am Acad Audiol 14:314–324.

Cariani PA, Delgutte B (1996) Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J Neurophysiol 76:1717–1734.

Chandrasekaran B, Kraus N (2009) The scalp-recorded brainstem response to speech: neural origins. Psychophysiology, in press.

Chertoff ME, Hecox KE (1990) Auditory nonlinearities measured with averaged auditory evoked potentials. J Acoust Soc Am 87:1248–1254.

Chertoff ME, Hecox KE, Goldstein R (1992) Auditory distortion products measured with averaged auditory evoked potentials. J Speech Hear Res 35:157–166.

Crawley EJ, Acker-Mills BE, Pastore RE, Weil S (2002) Change detection in multi-voice music: the role of musical structure, musical training, and task demands. J Exp Psychol Hum Percept Perform 28:367–378.

Dhar S, Long GR, Talmadge CL, Tubis A (2005) The effect of stimulus-frequency ratio on distortion product otoacoustic emission components. J Acoust Soc Am 117:3766–3776.

Dhar S, Abel R, Hornickel J, Nicol T, Skoe E, Zhao W, Kraus N (2009) Exploring the relationship between physiological measures of cochlear and brainstem function. Clin Neurophysiol. Advance online publication. Retrieved April 20, 2009. doi:10.1016/j.clinph.2009.02.172.

Elsisy H, Krishnan A (2008) Comparison of the acoustic and neural distortion product at 2f1–f2 in normal-hearing adults. Int J Audiol 47:431–438.

Fishman YI, Reser DH, Arezzo JC, Steinschneider M (2000) Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness. J Acoust Soc Am 108:235–246.

Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, Arezzo JC, Howard MA, Steinschneider M (2001) Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. J Neurophysiol 86:2761–2788.

Fritz JB, Elhilali M, Shamma SA (2007) Adaptive changes in cortical receptive fields induced by attention to complex sounds. J Neurophysiol 98:2337–2346.

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2005) Automatic encoding of polyphonic melodies in musicians and nonmusicians. J Cogn Neurosci 17:1578–1592.

Fujioka T, Trainor LJ, Ross B (2008) Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. Neuroreport 19:361–366.

Galbraith GC (1994) Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr Clin Neurophysiol 92:321–330.

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

Goebl W (2001) Melody lead in piano performance: expressive device or artifact? J Acoust Soc Am 110:563–572.

Goldstein JL (1967) Auditory nonlinearity. J Acoust Soc Am 41:676–689.

Gorga M, Abbas P, Worthington D (1985) Stimulus calibration in ABR measurements. In: The auditory brainstem response (Jacobsen J, ed), pp 49–62. San Diego: College-Hill.

Greenberg S, Marsh JT, Brown WS, Smith JC (1987) Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. Hear Res 25:91–114.

Hall JW 3rd (1979) Auditory brainstem frequency following responses to waveform envelope periodicity. Science 205:1297–1299.

Helmholtz H (1954) On the sensations of tone. New York: Dover.

Hood LJ (1998) Clinical applications of the auditory brainstem response. San Diego: Singular.

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 59:179–188.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541–577.

Kemp DT (2002) Otoacoustic emissions, their origin in cochlear function, and use. Br Med Bull 63:223–241.

Kim DO, Molnar CE, Matthews JW (1980) Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure. J Acoust Soc Am 67:1704–1721.

Knight RD, Kemp DT (1999) Relationships between DPOAE and TEOAE amplitude and phase characteristics. J Acoust Soc Am 106:1420–1435.

Knight RD, Kemp DT (2000) Indications of different distortion product otoacoustic emission mechanisms from a detailed f1, f2 area study. J Acoust Soc Am 107:457–473.

Knight RD, Kemp DT (2001) Wave and place fixed DPOAE maps of the human ear. J Acoust Soc Am 109:1513–1525.

Knudsen EI (2007) Fundamental components of attention. Annu Rev Neurosci 30:57–78.

Kral A, Eggermont JJ (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.

Kraus N, Shoe E, Parbery-Clark A, Ashley R (2009) Experience-induced malleability in neural encoding of pitch, timbre and timing: implications for language and music. Ann N Y Acad Sci, in press.

Krishnan A (1999) Human frequency-following responses to two-tone approximations of steady-state vowels. Audiol Neurootol 4:95–103.

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 25:161–168.

Kuriki S, Kanda S, Hirata Y (2006) Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. J Neurosci 26:4046–4053.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799–1822.

Lappe C, Herholz SC, Trainor LJ, Pantev C (2008) Cortical plasticity induced by short-term unimodal and multimodal musical training. J Neurosci 28:9632–9639.

Large EW (2006) A generic nonlinear model for auditory perception. In: Auditory mechanisms: processes and models (Nuttall AL, Ren T, Gillespie P, Grosh K, de Boer E, eds), pp 516–517. Singapore: World Scientific.

Larsen E, Cedolin L, Delgutte B (2008) Pitch representations in the auditory nerve: two concurrent complex tones. J Neurophysiol 100:1301–1319.

Luo F, Wang Q, Kashani A, Yan J (2008) Corticofugal modulation of initial sound processing in the brain. J Neurosci 28:11615–11621.

Margulis EH, Mlsna LM, Uppunda AK, Parrish TB, Wong PC (2009) Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum Brain Mapp 30:267–275.

McAlpine D (2004) Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. J Neurophysiol 92:1295–1311.

Moore BCJ (2003) An introduction to the psychology of hearing. Boston: Academic.

Moushegian G, Rupert AL, Stillman RD (1973) Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr Clin Neurophysiol 35:665–667.

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

Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. Hear Res 241:34–42.

Palmer C (1997) Music performance. Annu Rev Psychol 48:115–138.

Palmer C, Holleran S (1994) Harmonic, melodic, and frequency height influences in the perception of multivoiced music. Percept Psychophys 56:301–312.

Pandya PK, Krishnan A (2004) Human frequency-following response correlates of the distortion product at 2f1–f2. J Am Acad Audiol 15:184–197.

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

Peretz I, Zatorre RJ (2005) Brain organization for music processing. Annu Rev Psychol 56:89–114.

Perrot X, Micheyl C, Khalfa S, Collet L (1999) Stronger bilateral efferent influences on cochlear biomechanical activity in musicians than in non-musicians. Neurosci Lett 262:167–170.

Rhode WS, Cooper NP (1993) Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae. Hear Res 66:31–45.

Rickman MD, Chertoff ME, Hecox KE (1991) Electrophysiological evidence of nonlinear distortion products to two-tone stimuli. J Acoust Soc Am 89:2818–2826.

Robles L, Ruggero MA, Rich NC (1991) Two-tone distortion in the basilar membrane of the cochlea. Nature 349:413–414.

Rosenkranz K, Williamon A, Rothwell JC (2007) Motorcortical excitability and synaptic plasticity is enhanced in professional musicians. J Neurosci 27:5200–5206.

Rossing TD (1990) The science of sound. Reading, MA: Addison-Wesley.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clinical Neurophysiology 115:2021–2030.

Saffran JR (2003) Musical learning and language development. Ann N Y Acad Sci 999:397–401.

Saffran JR, Werker J, Werner L (2006) The infant's auditory world: hearing, speech, and the beginnings of language. In: Handbook of child psychology (Kuhn D, Siegler M, eds), pp 58–108. New York: Wiley.

Schreiner CE, Winer JA (2007) Auditory cortex mapmaking: principles, projections, and plasticity. Neuron 56:356–365.

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J Neurosci 23:5545–5552.

Smith JC, Marsh JT, Brown WS (1975) Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr Clin Neurophysiol 39:465–472.

Smoorenburg GF (1972) Combination tones and their origin. J Acoust Soc Am 52:615–632.

Strait DL, Kraus N, Skoe E, Ashley R (2009) Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. Eur J Neurosci 29:661–668.

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

Suga N, Xiao Z, Ma X, Ji W (2002) Plasticity and corticofugal modulation for hearing in adult animals. Neuron 36:9–18.

Trainor LJ, Shahin A, Roberts LE (2003) Effects of musical training on the auditory cortex in children. Ann N Y Acad Sci 999:506–513.

Tramo MJ, Cariani PA, Delgutte B, Braida LD (2001) Neurobiological foundations for the theory of harmony in western tonal music. Ann N Y Acad Sci 930:92–116.

Winer JA (2005) Decoding the auditory corticofugal systems. Hear Res 207:1–9.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci 10:420–422.

Worden FG, Marsh JT (1968) Frequency-following (microphonic-like) neural responses evoked by sound. Electroencephalogr Clin Neurophysiol 25:42–52.