

## COGNITIVE NEUROSCIENCE

# Musical experience and neural efficiency – effects of training on subcortical processing of vocal expressions of emotion

Dana L. Strait,<sup>1,2</sup> Nina Kraus,<sup>1,3,4,5</sup> Erika Skoe<sup>1,3</sup> and Richard Ashley<sup>2,6</sup>

<sup>1</sup>Auditory Neuroscience Laboratory, Frances Searle Building, 2240 Campus Drive, Evanston, IL 60208 2952, USA

<sup>2</sup>School of Music, Northwestern University, Evanston, IL, USA

<sup>3</sup>Department of Communication Sciences, Northwestern University, Evanston, IL, USA

<sup>4</sup>Department of Neurobiology and Physiology, Northwestern University, Evanston, IL, USA

<sup>5</sup>Department of Otolaryngology, Northwestern University, Evanston, IL, USA

<sup>6</sup>Program of Cognitive Science, Northwestern University, Evanston, IL, USA

**Keywords:** auditory brainstem response, auditory plasticity, emotion, evoked potentials, training

## Abstract

Musicians exhibit enhanced perception of emotion in speech, although the biological foundations for this advantage remain unconfirmed. In order to gain a better understanding for the influences of musical experience on neural processing of emotionally salient sounds, we recorded brainstem potentials to affective human vocal sounds. Musicians showed enhanced time-domain response magnitude to the most spectrally complex portion of the stimulus and decreased magnitude to the more periodic, less complex portion. Enhanced phase-locking to stimulus periodicity was likewise seen in musicians' responses to the complex portion. These results suggest that auditory expertise engenders both enhancement and efficiency of subcortical neural responses that are intricately connected with acoustic features important for the communication of emotional states. Our findings provide the first biological evidence for behavioral observations indicating that musical training enhances the perception of vocally expressed emotion in addition to establishing a subcortical role in the auditory processing of emotional cues.

## Introduction

The perception of emotion in speech is fundamental to the skillful communication needed for our everyday social and professional interactions. Deficiencies in the perception of emotion in speech are characteristic of various disorders, such as autism spectrum disorders, resulting in psychological and social isolation (Van Lancker *et al.*, 1989). Vocal conveyance of emotion largely relies on paralinguistic elements, defined as the non-verbal components of spoken communication (Grandjean *et al.*, 2006). Dependent on fast-changing acoustic features involved in the perception of pitch, timing and timbre (Juslin & Laukka, 2003; Scherer, 2003), emotion detection requires fine-grained sensitivity to spectral and temporal components of paralinguistic cues.

Emotion perception in both speech and music relies on shared acoustic (Scherer, 2003) and neural mechanisms (Nair *et al.*, 2002; Price *et al.*, 2005), suggesting that extensive experience in one domain may lend perceptual benefits to the other. Musical experience has been shown to enhance sensitivity to emotion in speech in both children and adults, with musicians more accurately identifying emotions expressed

in speech samples (Thompson *et al.*, 2004; Dmitrieva *et al.*, 2006). Enhanced sensitivity in musicians is not surprising given that musicians must attend to the detailed acoustic properties of sound on a daily basis, monitoring and manipulating their output to express precisely defined musical intentions. Due to its spectral complexity, wide frequency range and large variations in tone durations, music serves as an extremely effective vehicle for auditory training (Saunders, 1996; Zatorre *et al.*, 2002).

Musicians' heightened sensitivity to emotion in speech may be related to well-documented structural and functional reorganization at cortical levels ( Schlaug, 2001; Schneider *et al.*, 2002; Gaser & Schlaug, 2003; Hutchinson *et al.*, 2003; Costa-Giomi, 2005; Schlaug *et al.*, 2005; Bangert & Schlaug, 2006; Shahin *et al.*, 2007). In fact, professional musicians have a stronger capacity for neural plasticity – even for functions not related to musical tasks (Ragert *et al.*, 2004). Subcortical influences of musical training have also been observed, including stronger phase-locking to fundamental pitch and earlier onsets in evoked responses to linguistic and musical sounds with limited acoustic complexity (Musacchia *et al.*, 2007, 2008; Wong *et al.*, 2007). These observations suggest that auditory expertise, here demonstrated by musicians, results in subcortical sensory processing malleability of two acoustic properties shared by language and music (pitch and timing). These acoustic properties, along with time-varying

Correspondence: Dr N. Kraus, <sup>1</sup>Auditory Neuroscience Laboratory, as above.  
E-mail: nkraus@northwestern.edu

Received 9 October 2008, revised 6 December 2008, accepted 8 December 2008

harmonic structures (timbre), contribute to the perception of emotion in both speech and music and are fundamental to human communication (Juslin & Laukka, 2003). Through our use of a complex emotionally charged stimulus, we were able to explore musicians' subcortical sensitivity to acoustic features fundamental to human communication, including features not previously shown to be affected by musical training.

Subcortical involvement in the auditory processing of emotion has not yet been established but may be explored through investigations of the auditory brainstem response (ABR). The ABR reflects spectral and temporal acoustic features, such as those that contribute to the auditory perception of emotion, with remarkable fidelity (Kraus & Nicol, 2005). Recent evidence indicates the existence of a fast subcortical auditory–limbic pathway (Marsh *et al.*, 2002; Brandao *et al.*, 2005; Macedo *et al.*, 2005) that may be analogous to the well-documented and evolutionarily ancient subcortical–limbic pathway in the visual system (Johnson, 2005). Exploration of subcortical responses to emotional sounds via the ABR may advance our understanding of subcortical function in responses to emotionally charged auditory cues.

In order to gain a better understanding for how musical experience influences the neural processing of affective speech-related events, we recorded evoked brainstem potentials to an emotionally charged human vocal sound – an infant's unhappy cry. We expected musicians to demonstrate enhanced subcortical fidelity to fine-grained acoustic features fundamental to the perception of emotion (pitch, timing and timbre) when compared with non-musicians. Enhanced pitch and timbre processing would be evidenced by musicians' stronger neural phase-locking to stimulus periodicity, with enhancements in timing indicated by musicians' earlier responses to transient components of the stimulus. Overall, we aimed to provide a biological basis for musicians' enhanced perception of emotion in speech by investigating the contribution of subcortical mechanisms to the processing of vocally communicated emotional states.

## Materials and methods

### Subjects

Subjects consisted of 30 adults between the ages of 19 and 35 years (mean age 24.7 years, 18 females). All subjects gave their informed consent in accordance with the Northwestern University Institutional Review Board guidelines based on The Belmont Report (1979). Subjects were right-handed and had audiograms demonstrative of normal hearing thresholds (<15 dB HL pure-tone thresholds for octave frequencies from 125 to 8000 Hz), as measured by both air and bone conduction. Integrity of the auditory brainstem was assessed using both click and speech (/da/) stimuli presented with the BioMARK (Biological Marker of Auditory Processing) recording protocol (Biologic, a Natus Company, Mundelein, IL, USA, 2006). Normal responses to the click and the /da/ were prerequisites for inclusion.

Subjects completed a questionnaire that addressed family history and musical experience. The family history portion confirmed that subjects had no familial neurological or ontological abnormalities. The musical experience portion included questions pertaining to duration and intensity of musical practice, performance consistency, teaching experience, formal training and participation in collaborative ensembles, such as orchestras, ensembles or choirs.

To investigate effects of different aspects of musical training, we grouped musicians according to two criteria: 'musicians by onset age' (MusAge,  $n = 11$ ) and 'musicians by years' (MusYrs,  $n = 15$ ). MusAge subjects had begun musical training at or before age 7 years,

whereas MusYrs subjects had received more than 10 years of consistent musical training. Consistency was defined by a minimum of three hourly practice sessions per week in addition to group or private music lessons. Because of the nature of our criteria, some musicians fell into both groups ( $n = 9$ ). Non-musicians (NonMus) in both cases were categorized by failure to meet those criteria. When subjects were grouped by age of musical experience onset, NonMus  $n = 19$ , whereas  $n = 15$  when the grouping relied on the number of years of musical practice.

### Procedure and stimulus

Auditory-evoked responses were elicited by a complex vocal stimulus derived from an emotionally salient auditory scenario synthesized and behaviorally evaluated by the Center for the Study of Emotions and Attention (University of Florida, Gainesville, FL, USA; sound file 278 from 0.39 to 0.64 s; Bradley & Lang, 2007). The sound from which we extracted our stimulus had been rated as highly salient and of strong negative valence according to CSEA raters. Consisting of an infant's unhappy cry, the stimulus can be described in terms of sections that exhibit varying levels of acoustic complexity. For a definition and illustration of our use of the qualifier 'acoustic complexity', see the Results and Fig. 1. In order to confirm emotional valence and salience of the stimulus, participants were asked to rate it according to the self-assessment manikin (SAM), a scale used to assess both the perceptual salience and valence of expressed emotion (Bradley & Lang, 1994).

Auditory-evoked potentials were recorded from each subject. The stimulus was presented binaurally at 82 dB sound-pressure level in alternating polarities via insert earphones (ER-3; Etymotic Research, Elk Grove Village, IL, USA) using the software NEUROSCAN STIM 2 (Neuroscan; Compumedics, Charlotte, NC, USA) with an interstimulus interval of 78.6 ms. During the recording session participants watched a nature movie with muted sound to promote stillness and reduce electrical artifact. Alertness during the recording sessions was assessed by quizzes designed to test how well individuals attended to the movie. Correct answers resulted in additional monetary compensation beyond the standard hourly payment. Electrical components of evoked neural responses were recorded using Ag-AgCl scalp electrodes and the NEUROSCAN ACQUIRE 4.3 recording system. Responses were recorded at a 20 000 Hz sampling rate from Cz (active), with a linked-earlobe reference and the forehead as ground. Responses to 5200 presentations of the stimulus were recorded, lasting between 30 and 35 min.

### Data processing and analysis

Data were off-line filtered using Neuroscan Edit 4.3 from 70 to 2000 Hz with a 12 dB per octave roll off. These filtering characteristics align with phase-locking capabilities of auditory brainstem nuclei while minimizing lower-frequency cortical activity (Akhoun *et al.*, 2008). The recording site (Cz) with earlobe as reference and forehead as ground, as we use here, has been shown to reflect a response of central brainstem origin (Galbraith, 1994; Galbraith *et al.*, 2000). For a further review of the origins of the ABR and the low-pass filter characteristics of the response, see Song *et al.* (2008). The filtered recordings were epoched using a –50 to 300 ms time window, with the stimulus acoustic onset at 0 ms. Trials with activity greater than  $\pm 35 \mu\text{V}$  were rejected as artifacts and excluded from the final averaged response. Responses from alternating polarities were summed to minimize contributions from the cochlear microphonic

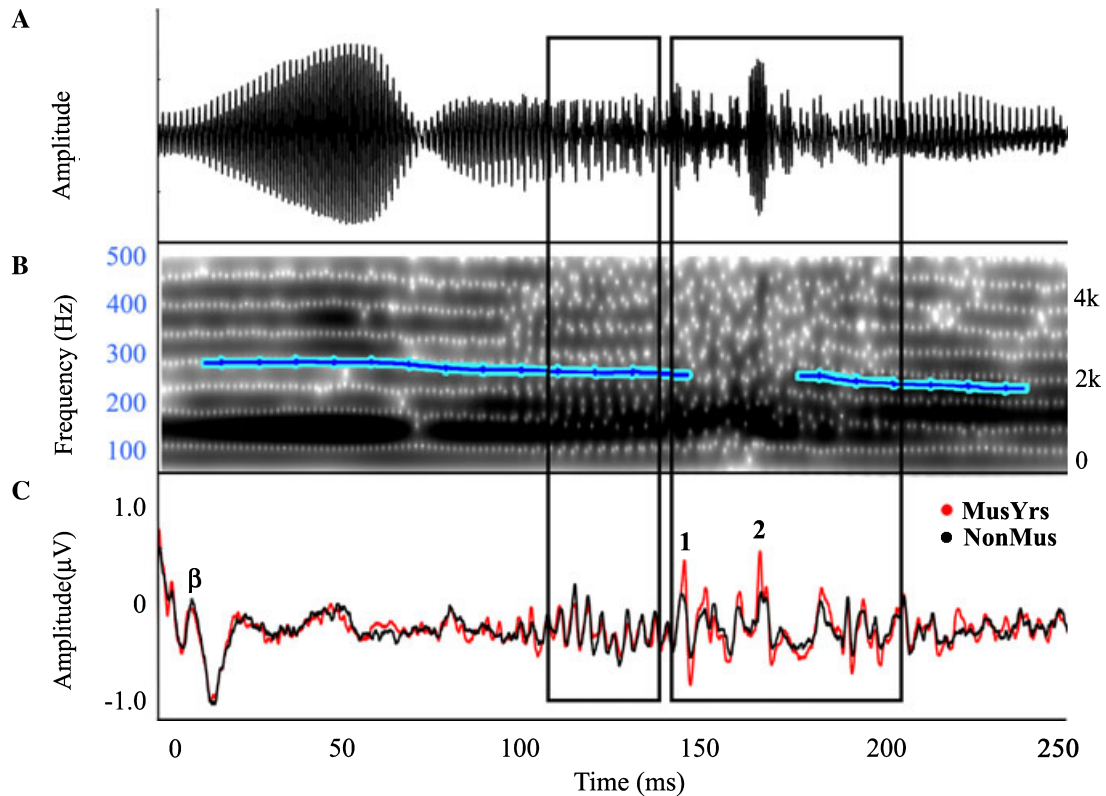


FIG. 1. Stimulus and grand average response waveforms. Response waveforms have been shifted back in time ( $\sim 7$  msec) to align the stimulus and response onsets. Boxes delineate two stimulus subsections and the corresponding brainstem responses. The first subsection (112–142 ms) corresponds to the periodic portion and the second (145–212 ms) corresponds to the more complex portion. (A) Stimulus time-amplitude waveform. (B) Stimulus spectrogram. The stimulus  $F_0$  is superimposed as a highlighted line ( $\sim 280$  Hz, left axis) with higher frequency spectral components plotted between white dotted lines (right axis). Although the  $F_0$  is detectable during the first section, the greater acoustic complexity of the second section results in the inability of the sound analyzing software (Praat) to track the  $F_0$ . The harmonics are likewise more aperiodic. (C) The averaged responses of MusYrs and NonMus. Major peaks ( $\beta$ , 1 and 2) are labeled above the waveform.

response, an electrical potential generated by activity of sensory receptors in the cochlea (Teas, 1993).

For final analyses we distilled the stimulus and corresponding neural responses down to two segments, referred to as the periodic and complex portions. This distillation was based on the following reasons: (1) in the stimulus, the periodic and complex portions were both internally consistent; (2) these two acoustically contrastive regions occurred sequentially with no intervening signal; and (3) like the stimulus, the recorded responses could also be segregated into multiple sections. The responses to the periodic and complex stimulus portions were largely reflective of the morphology of the stimulus waveform itself (Fig. 1) in that responses to the periodic portion exhibited periodic waveforms whereas responses to the complex portion were reflective of the transient amplitude changes in the stimulus. In contrast to other regions, the responses to the periodic and complex portions were highly replicable across subjects and were reliably above the noise floor. The noise floor was determined by calculating rectified mean amplitudes (RMAs) for each section and comparing these to the RMA of the noise floor (i.e. neural activity prior to the onset of sound). This comparison yielded a signal to noise ratio ( $SNR = RMA_{\text{section}} / RMA_{\text{noise floor}}$ ). Only response sections that consistently yielded an SNR of 1.4 or higher were included in subsequent analyses.

To provide a gross measure of response amplitude, we calculated RMAs for the periodic, complex and neighboring response regions. The stimulus periodicity in the sections surrounding the periodic and complex portions was above the phase-locking capabilities of brainstem nuclei (Hall, 1979), resulting in comparatively smaller

response amplitudes. Because free responses to these sections were not above the noise floor, we do not address these sections.

Neural responses to the complex stimulus portion resulted in a series of peaks that aligned in the time-domain with the amplitude bursts in the stimulus. Peaks were identified in NEUROSCAN EDIT 4.3 and labeled 1–5. In the response waveforms (Fig. 1C), time zero refers to the time at which the stimulus was presented and peak latency was calculated relative to that point. In order to provide more time-specific measures of response amplitude, peak amplitudes were calculated relative to the response baseline (zero volts). Therefore, all peaks provided both amplitude and latency data. Onset peaks, occurring between 7 and 20 ms, were also selected and labeled with Greek letters  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ . The largest and most replicable peaks are labeled in Fig. 1, including the onset peak  $\beta$  and peaks 1 and 2, with individual waveforms differing from mean latencies by no more than 0.6 ms.

Spectral components of the neural responses were extracted using the fast Fourier Transform (FFT). Amplitudes were recorded for three spectral peaks, corresponding to stimulus' fundamental frequency ( $F_0$ ) and spectral components (labeled  $H_2$  and  $H_3$ ) that represent the maxima within given frequency ranges ( $F_0$ : 280–305 Hz;  $H_2$ : 470–485 Hz;  $H_3$ : 570–585 Hz). These spectral peaks were interpreted to correspond to representations of pitch ( $F_0$ ) and timbre ( $H_2$ ,  $H_3$ ). Due to the variance in frequency over time of our stimulus, FFTs were calculated for response time bins corresponding to the periodic and complex stimulus subportions, demarcated in Fig. 1 (the corresponding response time ranges are 120–142 ms and 145–212 ms). To



increase spectral resolution to 1 Hz, response segments were zero-padded to 20 000 points after applying a hanning window. All frequency analyses were carried out via MATLAB 7.5.0 routines (The Mathworks, Natick, MA, USA). Normality for all data was confirmed by the Kolmogorov–Smirnov test for equality.

Regression analyses supported the grouping of musicians into two subgroups. We consequently restricted our analyses to amplitude and latency domains for the MusYrs grouping and to the frequency domain for the MusAge grouping. Accordingly, spectral peak amplitudes of  $F_0$ ,  $H_2$  and  $H_3$  were subjected to a two-group one-way ANOVA according to the MusAge grouping and RMA values, peak amplitudes and peak latencies were subjected to a two-group one-way ANOVA according to the MusYrs grouping. Correlation analyses between neural response characteristics and age of training onset or duration of musical practice were performed on all subjects with musical training histories ( $n = 20$ ).

## Results

Contrary to our initial hypotheses, no convincing differences between musically trained and untrained participants were observed when musicians were grouped into a single category. Regression analyses supported the grouping of musicians into two subgroups: whereas MusYrs grouping was predicted best by amplitude and latency measures ( $P < 0.001$ ; MusAge  $P < 0.06$ ), the MusAge grouping was predicted best by frequency encoding measures ( $P < 0.05$ ; MusYrs  $P < 0.40$ ). We consequently restricted our analyses to amplitude and latency domains for the MusYrs grouping, and to the frequency domain for the MusAge grouping.

Both MusYrs and MusAge musicians' responses exhibited enhancements and economy that were intricately connected with the time-varying acoustic features of the stimulus. Enhancements, reflected by larger time- and frequency-domain response magnitudes, were most evident in musicians' responses to the most complex portions of the sound, with economy (smaller amplitudes) seen in their responses to the periodic portion (Grabner *et al.*, 2006). Earlier peak latencies in musician responses reflect faster synchronous neural responses to the temporal characteristics of the stimulus, thus serving as a measure of response timing enhancement.

Figure 1 shows both the stimulus (Fig. 1A) and average responses for the MusYrs and NonMus (Fig. 1C), with boxes defining two acoustically different sections referred to in this manuscript as the periodic and complex portions, respectively. These same qualifiers will be used when referring to the brainstem responses to these stimulus sections. To account for neural response lag, the average responses (Fig. 1C) are shifted back in latency by approximately 7 ms so that the response and stimulus onsets are in visual alignment, maximizing visual coherence between the two signals. Figure 1B shows a tracing of the stimulus' fundamental frequency ( $F_0$ ) in bold (as calculated by PRAAT; Boersma & Weenink, 2008), giving rise to the perception of pitch, with other spectral components that contribute to the perception of vocal timbre marked between white dotted lines. The first area of the stimulus, from 112 to 142 ms, is characterized by greater periodicity. The second area of the stimulus, from 145 to 212 ms, is characterized by greater complexity, including fast-changing transient features such as amplitude bursts. Such increased complexity is evidenced by the scattered spectral components (aperiodic white dots) and loss of  $F_0$  tracking (highlighted line). Acoustic characteristics constituting complexity include deviation in pitch from the  $F_0$  (0.5% in the periodic portion and 2.7% in the complex portion), harmonic jitter (1.27% and 2.03%) and SNR (13.46

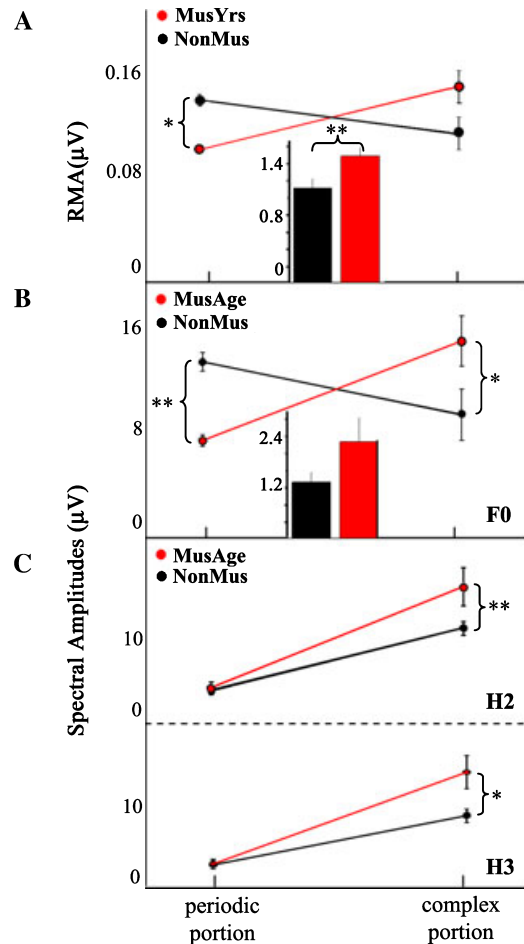


FIG. 2. Interactions between responses of MusYrs/MusAge and NonMus to the periodic and complex portions of the stimulus. (A) A group  $\times$  portion interaction was observed between MusYrs and NonMus RMAs in responses to the periodic and complex stimulus portions ( $F_{1,28} = 6.04$ ,  $P < 0.02$ ). While there is a significant increase in RMAs for the MusYrs responses to the complex portion relative to the periodic portion ( $t_{28} = 4.70$ ,  $P < 0.0001$ ), NonMus responses do not change ( $t_{28} = 0.03$ ,  $P < 0.99$ ). The inset figure represents response RMA ratios of the complex to the periodic portion (Ratio =  $RMA_{\text{complex}}/RMA_{\text{periodic}}$ ), further confirming MusYrs to show more differentiated responses between the two sections than NonMus ( $F_{1,28} = 9.58$ ,  $P < 0.005$ ). (B) A group  $\times$  portion interaction was observed between MusYrs and NonMus  $F_0$  encoding in responses to the periodic and complex stimulus portions ( $F_{1,28} = 5.39$ ,  $P < 0.03$ ). Compared with NonMus, MusAge show enhanced representation of the fundamental periodicity ( $F_0$ ) for the complex portion ( $F_{1,28} = 7.04$ ,  $P < 0.03$ ) and decreased representation in the periodic portion ( $F_{1,28} = 5.04$ ,  $P < 0.01$ ). The inset figure represents  $F_0$  response ratios of the complex to the periodic portion (Ratio =  $F_{0 \text{ complex}}/F_{0 \text{ periodic}}$ ). Relative to the NonMus group, MusAge show enhanced  $F_0$  encoding in responses to the complex relative to the periodic portion ( $F_{1,28} = 3.56$ ,  $P < 0.06$ ). (C) MusAge also show enhanced encoding of frequencies above the  $F_0$  ( $H_2$ ,  $F_{1,28} = 7.95$ ,  $P < 0.01$ ;  $H_3$ ,  $F_{1,28} = 6.16$ ,  $P < 0.02$ ) in responses to the complex portion of the stimulus. Error bars represent one standard error, \* $P < 0.05$  and \*\* $P < 0.01$ .

and 6.82 dB), as calculated by PRAAT. Such acoustic features contributed to perceived negative emotional valence ( $t_{28} = 2.62$ ,  $P > 0.01$ ; Mean = 3.7, SD = 1.6) and moderate emotional saliency ( $t_{28} = 4.45$ ,  $P < 0.0001$ ; Mean = 5.4, SD = 2.1), as indicated by SAM ratings of the subjects. The scales ranged from 1 to 9 in both cases, with 1 being low (saliency) or negative (valence).

The response RMAs to the periodic and complex stimulus portions are plotted in Fig. 2A. A  $2 \times 1$  ANOVA with group (MusYrs/NonMus) and stimulus portion (periodic/complex) as independent variables

and RMA as the dependent variable revealed a significant interaction between group and response portions ( $F_{1,28} = 6.04$ ,  $P < 0.02$ ). This indicates that MusYrs and NonMus have differentiated responses to the two sections. In order to further define this differentiation, we computed the difference in RMAs between MusYrs and NonMus responses to these two regions, computed as an amplitude ratio between the responses to the periodic and the complex portions (Ratio =  $RMA_{\text{complex}}/RMA_{\text{periodic}}$ ;  $F_{1,28} = 9.58$ ,  $P < 0.005$ ), plotted as the inset in Fig. 2A. The amplitudes of the two sections are illustrated in the larger graph, showing MusYrs to have smaller responses to the periodic portion than NonMus ( $F_{1,28} = 4.68$ ,  $P < 0.03$ ), with relatively larger responses to the complex portion. Whereas the MusYrs within-group RMAs differ between the periodic and complex portions (paired samples  $t$ -test,  $t_{28} = 4.70$ ,  $P < 0.0001$ ), NonMus do not ( $t_{28} = 0.025$ ,  $P < 0.99$ ). Peak amplitudes confirm MusYrs to have larger responses to the complex portion than NonMus (peak 1:  $F_{1,28} = 10.25$ ,  $P < 0.003$ ; peak 2:  $F_{1,28} = 4.88$ ,  $P < 0.03$ ; for a response waveform illustrating the location of these peaks, refer to Fig. 1C).

Within the complex portion there was also a positive correlation between peak amplitudes and years of consistent musical practice across all individuals with musical experience, regardless of group status ( $n = 20$ ; Fig. 3A; peak 1: Pearson's  $r = 0.454$ ,  $P < 0.04$ ). In regards to response timing, peak latency-related timing enhancements were observed in MusYrs responses, even as early as the response

onset (Fig. 4; onset peak  $\beta$ :  $F_{1,28} = 4.82$ ,  $P < 0.04$ ; peak 1:  $F_{1,28} = 8.72$ ,  $P < 0.006$ ).

Compared with non-musicians, subjects who began training prior to the age of 7 years (MusAge) showed enhanced representations of frequencies important for the perception of pitch and timbre ( $F_0$ ,  $H_2$ ,  $H_3$ ) in response to the complex stimulus portion, observed through FFT-derived spectral amplitudes. Again, these differences were found to be connected with acoustic characteristics of the auditory input. A  $2 \times 1$  analysis of variance with group (MusAge/NonMus) and stimulus portion (periodic/complex) as independent variables and portion  $F_0$  encoding as the dependent variable revealed a significant interaction between group and response portion ( $F_{1,28} = 7.04$ ,  $P < 0.01$ ). This indicates that MusAge and NonMus have differentiated responses to the two sections. MusAge showed smaller representations of the  $F_0$  in response to the periodic portion of the stimulus than NonMus, but larger  $F_0$  amplitudes to the complex portion (Fig. 2B; periodic:  $F_{1,28} = 7.04$ ,  $P < 0.01$ ; complex:  $F_{1,28} = 5.04$ ,  $P < 0.03$ ). The inset figure demonstrates the interaction between  $F_0$  encoding in MusAge and NonMus responses to these two regions as a ratio (Ratio =  $F_{0\text{complex}}/F_{0\text{periodic}}$ ;  $F_{1,28} = 3.563$ ,  $P < 0.06$ ).  $F_0$  amplitudes in response to the complex portion correlated with the age of onset of musical practice, with individuals who began at an earlier age showing larger  $F_0$  representations (Fig. 3B: Pearson's  $r = -0.500$ ,  $P < 0.03$ ).

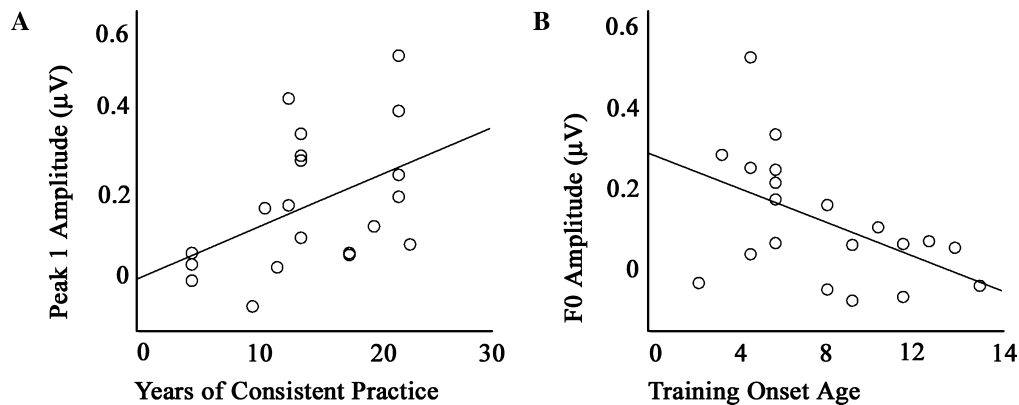


FIG. 3. Correlations between measures of musical experience and subcortical response characteristics to the complex stimulus portion. (A) Peak amplitudes correlate with the number of years of consistent musical practice (peak 1: Pearson's  $r = 0.454$ ,  $P < 0.04$ ), whereas (B) the age in years at which individuals began musical training correlate with fundamental frequency encoding (Pearson's  $r = -0.500$ ,  $P < 0.03$ ). The more years an individual has consistently practiced his/her instrument(s), the larger their peak amplitudes and the earlier age an individual began musical training, the more robust their  $F_0$  encoding. All individuals with musical experience are represented in both correlations.

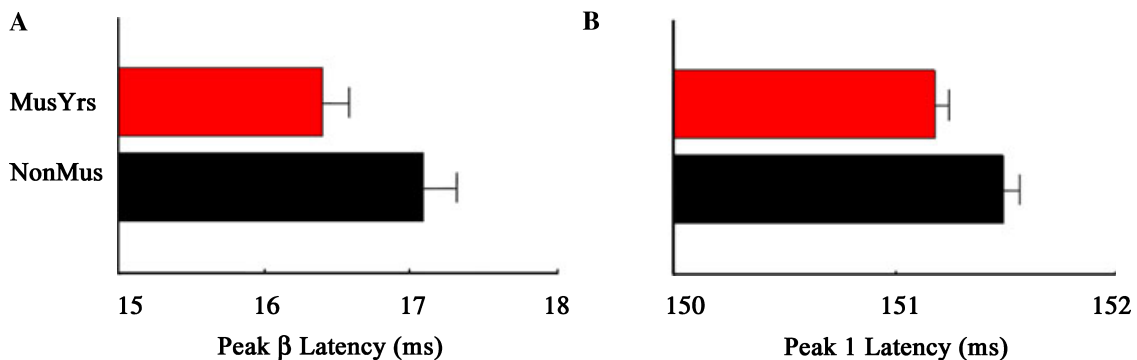


FIG. 4. Peak latencies for MusYrs and NonMus. In comparison to NonMus, MusYrs show earlier peak latencies for the onset peak  $\beta$  (A) and Peak 1 of the complex portion (B) (onset peak  $\beta$ :  $F_{1,28} = 4.82$ ,  $P < 0.04$ ; peak 1:  $F_{1,28} = 8.72$ ,  $P < 0.006$ ). Bars show average latency results from NonMus and MusYrs groups, respectively. Error bars represent one standard error.

MusAge also demonstrated enhanced representations of spectral peaks  $H_2$  and  $H_3$  in response to the complex portion of the stimulus, compared with NonMus (Fig. 2C;  $H_2$ :  $F_{1,28} = 7.95$ ,  $P < 0.01$ ;  $H_3$ :  $F_{1,28} = 6.16$ ,  $P < 0.02$ ).

## Discussion

We suggest that musical experience has more pervasive domain-general effects on the auditory system than previously documented, resulting in fine neural tuning to acoustic features important for vocal communication. Our results thus provide initial biological evidence for enhanced perception of emotion in musicians, indicating involvement of subcortical mechanisms in the auditory processing of communicated states of emotion.

The interplay between response enhancement and economy may contribute to musicians' enhanced perceptual capabilities of emotional cues in speech (Thompson *et al.*, 2004; Dmitrieva *et al.*, 2006). Our observations of enhancements only to the most acoustically complex stimulus portion are consistent with Ahissar and Hochstein's Reverse Hierarchy Theory for learning in the visual domain. The Reverse Hierarchy Theory suggests that learning results from a task-dependent backwards search for higher SNRs and that such learning is particularly characteristic of highly trained populations (musicians in the auditory domain; Ahissar & Hochstein, 2004).

Musicians' responses to the more periodic section of the stimulus demonstrate smaller response magnitudes, as fewer resources are required. This observation reflects musicians' neural efficiency in the processing of simpler acoustic features, previously shown to indicate domain-general expertise (Haier *et al.*, 1992; Neubauer *et al.*, 1995; Jausovec & Jausovec, 2000a,b; Grabner *et al.*, 2003, 2006; Doppelmayr *et al.*, 2005). Such efficiency has previously been documented in musicians by Magne *et al.*, who investigated effects of musical training on children's abilities to detect pitch violations in speech (Moreno & Besson, 2005; Magne *et al.*, 2006). They found that musical training resulted in decreased response magnitudes of cortical event-related potentials between 200 and 1200 ms. Other studies have similarly documented selective response enhancements to more complex stimuli. For example, Wong *et al.* (2007) showed enhanced subcortical representations of pitch in musicians' responses only to the sound with the most complex pitch trajectory. Likewise, after auditory training, Song *et al.* (2008) showed subcortical enhancements only to the sound with greatest pitch deviations with no enhancements to the more familiar, less complex sounds. In combination with our findings, these results suggest that extensive auditory training contributes to functional economy of neural processing when pitch changes are minimal and other more complex features dominate.

In regards to response timing, MusYrs had faster responses than NonMus as early as the response onset. This indicates that long-term musical experience contributes to enhanced subcortical timing, a factor that also accounts for enhancements in frequency representations (encoded through cycle-to-cycle neural phase-locking to stimulus periodicity). Synchronous response timing is reflected in response amplitudes and peak latencies that result from stimulus-induced synchrony of neural onsets and is considered a measure of neural enhancement. A balance of subcortical response enhancement and economy may contribute to the enhanced perceptual capabilities of musicians of emotional cues in speech (Thompson *et al.*, 2004; Dmitrieva *et al.*, 2006).

### *Musical training histories affect subcortical auditory processing*

Our data suggest that musical training prior to the age of 7 years impacts subcortical representations of pitch and timbre, whereas

timing-related enhancements are affected by duration of musical practice. Compared with non-musicians, subjects who began training prior to the age of 7 years (MusAge) showed enhanced representations of frequencies important for the perception of pitch and timbre ( $F_0$ ,  $H_2$  and  $H_3$ ). The absence of similar results for the MusYrs grouping indicates an optimal period in the development of encoding strategies for these features. Deprivation studies provide evidence for optimal periods in the acquisition of tonotopic maps in the primary auditory cortex, and exposure to spectrally and temporally complex auditory input has proven necessary for the normal development of auditory-evoked responses (Fallon *et al.*, 2002; Weinberger, 2004; Trainor, 2005). Music serves as a spectrally and temporally complex auditory stimulus, demonstrating a broader spectral range and wider variation in tone durations than speech, with pitch and timbre changes fundamental to its perception (Saunders, 1996; Zatorre *et al.*, 2002). In light of the multitude of hours musicians spend manipulating it, music's acoustic complexity makes it a powerful tool for engendering neural plasticity during optimal periods of auditory development.

There does not seem to be a similar optimal period in the development of neural representations of timing. Because subjects in our MusYrs group had at least 10 years of consistent musical experience, however, we cannot rule out the possibility for optimal periods ending after pre-teen years. While MusYrs showed faster responses than NonMus, these timing-related effects do not depend on the age that musical training began. This discrepancy between timing- and frequency-related effects of musical training aligns with previously documented evidence of distinct subcortical encoding mechanisms for different features of acoustic stimuli (Kraus & Nicol, 2005). Furthermore, our observations suggest that mechanisms for timing-related plasticity are more malleable later in life than frequency-related mechanisms.

Our results help to tease apart whether neural plasticity in musicians is genetically determined or stem from musical training. While it could be argued that someone born with better auditory processing is more likely to gravitate toward musical practice, genetic influences cannot fully account for our results. By showing relationships between response enhancements, the number of years an individual has spent participating in musical training and the age at which musical training began, we provide evidence for musical experience as the driving force behind the response differences observed in musicians. Thus, we join others in showing that the number of years an individual has consistently practiced relates to the extent of neural timing-related enhancement (Musacchia *et al.*, 2007; Wong *et al.*, 2007). This finding aligns with behavioral observations (Yee *et al.*, 1994) as well as cortical evidence (Besson *et al.*, 1994; Gaser & Schlaug, 2003; Hutchinson *et al.*, 2003). Experience-driven plasticity is further supported by work showing selective enhancements of neural activity to musicians' own instruments (Pantev *et al.*, 2001; Margulis *et al.*, 2009). Although genetic determinants surely influence neural specialization in the auditory domain (Peretz *et al.*, 2007), our data indicate that experience-driven plasticity brings about domain-general enhancements as manifested in responses to speech (Musacchia *et al.*, 2007; Wong *et al.*, 2007) and para-speech emotional sounds.

### *Cortex-brainstem interactions*

In the unimpaired system, cortical functions like music and language appear to fundamentally shape early auditory processing (Krishnan *et al.*, 2005; Musacchia *et al.*, 2007; Wong *et al.*, 2007). This occurs not only with lifelong experience but also as a result of short-term auditory training (Russo *et al.*, 2005; Song *et al.*, 2008). These enhancements are likely mediated by top-down influences that operate via an extensive corticofugal circuitry of descending efferent fibers



that synapse at a wealth of points along the auditory pathway (Suga *et al.*, 2001). The disruption of top-down influences may account for the auditory brainstem-encoding deficiencies seen with language and learning impairment, such as reading disorders (Banai *et al.*, 2005) and autism (Russo *et al.*, 2008). By helping to reinforce top-down mechanisms, music may provide a rehabilitative benefit for these populations. Through studying subcortical responses of musicians, we gain better insight into the mechanisms underlying experience-related effects that could aid in the development of treatments for auditory brainstem deficiencies.

### Subcortical fast pathway for the auditory processing of emotion

The involvement of subcortical pathways in the auditory detection of emotional states has previously been proposed (Bigand *et al.*, 2005). Peretz *et al.* (1998) played short emotional music excerpts for a female amusic subject, I.R., and control subjects. I.R.'s amusia occurred in conjunction with bilateral cortical lesions and other cortical surgery, and is thus indicative of cortical malfunction. In light of deficiencies, her emotional detection capabilities remained surprisingly intact. Although unable to recognize melodies, rhythms or other acoustic features of music, she performed similarly to controls in attributing emotions to musical samples. Preservation of her ability to perceive emotional properties of music may suggest subcortical auditory pathway involvement in the processing of aurally expressed emotion.

Fast subcortical–limbic pathways for emotional processing have previously been established in the visual system (for a review, see Johnson, 2005), with the existence of analogous pathways in auditory processing indicated more recently (Marsh *et al.*, 2002; Brandao *et al.*, 2005; Macedo *et al.*, 2005). Our findings provide a significant advance toward understanding biological mechanisms underlying human perception of biological states by demonstrating subcortical involvement in the encoding of acoustic features foundational to the vocal communication of emotion.

### Conclusions

Our results provide evidence for a subcortical role in the processing of emotional cues by showing that auditory brainstem responses to emotionally salient vocal sounds are dynamic, shaped by life-long, multisensory experience with auditory signals. These responses are not hard-wired, but malleable with extensive auditory training. Specifically, we found that musical training engenders subcortical efficiency that is intricately connected with acoustic features integral to the communication of emotional states. These findings provide the first biological evidence for musicians' perceptual enhancements in detecting vocally expressed emotion.

### Acknowledgements

This research was funded in part by the National Science Foundation Grant 0544846 and a Graduate Research Grant from the Northwestern University Graduate School. The authors would like to thank Jennifer Krizman for her assistance in preparing the final proofs.

### Abbreviations

ABR, auditory brainstem response;  $F_0$ , fundamental frequency; FFT, fast Fourier Transform;  $H_2$ ,  $H_3$ , second and third spectral peaks; MusAge, musicians, grouped by age of onset of musical practice; MusYrs, musicians, grouped by years of musical practice; NonMus, non-musicians grouped according to either MusYrs or MusAge criteria, depending on grouping employed; RMA, rectified mean amplitude; SAM, self-assessment manikin; SNR, signal to noise ratio.

### References

- Ahissar, M. & Hochstein, S. (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.*, **8**, 457–464.
- Akhoun, I., Gallégo, S., Moulin, A., Ménard, M., Veuillet, E., Berger-Vachon, C., Collet, S. & 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**, 922–933.
- Banai, K., Nicol, T., Zecker, S.G. & Kraus, N. (2005) Brainstem timing: implications for cortical processing and literacy. *J. Neurosci.*, **25**, 9850–9857.
- Bangert, M. & Schlaug, G. (2006) Specialization of the specialized in features of external human brain morphology. *Eur. J. Neurosci.*, **24**, 1832–1834.
- Besson, M., Faita, F. & Requin, J. (1994) Brain waves associated with musical incongruities differ for musicians and non-musicians. *Neurosci. Lett.*, **168**, 101–105.
- Bigand, E., Filipic, S. & Lalitte, P. (2005) The time course of emotional responses to music. *Ann. N. Y. Acad. Sci.*, **1060**, 429–437.
- Boersma, P. & Weenink, D. (2008) *Praat: doing phonetics by computer (Version 5.0.46)* [Computer program].
- Bradley, M.M. & Lang, P.J. (1994) Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry*, **25**, 49.
- Bradley, M.M. & Lang, P.J. (2007) *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective Ratings of Sounds and Instruction Manual*. NIMH Center for the Study of Emotion and Attention, Gainesville.
- Brandao, M.L., Borelli, K.G., Nobre, M.J., Santos, J.M., Albrechet-Souza, L., Oliveira, A.R. & Chacon, R. (2005) Gabaergic regulation of the neural organization of fear in the midbrain tectum. *Neurosci. Biobehav. Rev.*, **29**, 1299–1311.
- Costa-Giomi, E. (2005) Does music instruction improve fine motor abilities? *Ann. N. Y. Acad. Sci.*, **1060**, 262–264.
- Dmitrieva, E.S., Gel'man, V.Y., Zaitseva, K.A. & Orlov, A.M. (2006) Ontogenetic features of the psychophysiological mechanisms of perception of the emotional component of speech in musically gifted children. *Neurosci. Behav. Physiol.*, **36**, 53–62.
- Doppelmayr, M., Klimesch, W., Hodlmoser, K., Sauseng, P. & Gruber, W. (2005) Intelligence related upper alpha desynchronization in a semantic memory task. *Brain Res. Bull.*, **66**, 171–177.
- Fallon, J.B., Irvine, D.R.F. & Shepherd, R.K. (2002) Cochlear implants and brain plasticity. *Br. Med. Bull.*, **63**, 183–193.
- Galbraith, G.C. (1994) Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr. Clin. Neurophysiol.*, **92**, 321–330.
- Galbraith, G.C., Threadgill, M.R., Hemsley, J., Salour, K., Songdej, N., Ton, J. & Cheung, L. (2000) Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci. Lett.*, **292**, 123–127.
- Gaser, C. & Schlaug, G. (2003) Brain structures differ between musicians and non-musicians. *J. Neurosci.*, **23**, 9240–9245.
- Grabner, R.H., Stern, E. & Neubauer, A.C. (2003) When intelligence loses its impact: neural efficiency during reasoning in a familiar area. *Int. J. Psychophysiol.*, **49**, 89–98.
- Grabner, R.H., Neubauer, A.C. & Stern, E. (2006) Superior performance and neural efficiency: the impact of intelligence and expertise. *Brain Res. Bull.*, **69**, 422–439.
- Grandjean, D., Bänziger, T. & Scherer, K.R. (2006) Intonation as an interface between language and affect. *Prog. Brain Res.*, **156**, 235–268.
- Haier, R.J., Siegel, B., Tang, C., Abel, L. & Buchsbaum, M.S. (1992) Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence*, **16**, 415–426.
- Hall, J.W. III (1979) Auditory brainstem frequency following responses to waveform envelope periodicity. *Science*, **205**, 1297–1299.
- Hutchinson, S., Lee, L.H.L., Gaab, N. & Schlaug, G. (2003) Cerebellar volume of musicians. *Cereb. Cortex*, **13**, 943–949.
- Jausovec, N. & Jausovec, K. (2000a) Differences in event-related and induced brain oscillations in the theta and alpha frequency bands related to human intelligence. *Neurosci. Lett.*, **293**, 191–194.
- Jausovec, N. & Jausovec, K. (2000b) Correlations between ERP parameters and intelligence: a reconsideration. *Biol. Psychol.*, **55**, 137–154.
- Johnson, M.H. (2005) Subcortical face processing [Review]. *Nat. Rev. Neurosci.*, **6**, 766–774.
- Juslin, P.N. & Laukka, P. (2003) Communication of emotions in vocal expression and music performance: different channels, same code? *Psychol. Bull.*, **129**, 770–814.
- Kraus, N. & Nicol, T. (2005) Brainstem origins for cortical “what” and “where” pathways in the auditory system. *Trends Neurosci.*, **28**, 176–181.

- Krishnan, A., Xu, Y., Gandour, J. & Cariani, P. (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Cogn. Brain Res.*, **25**, 161–168.
- Macedo, C.E., Cuadra, G., Molina, V. & Brandao, M.L. (2005) Aversive stimulation of the inferior colliculus changes dopamine and serotonin extracellular levels in the frontal cortex: modulation by the basolateral nucleus of amygdala. *Synapse*, **55**, 58–66.
- 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**, 199–211.
- Margulis, E.H., Uppunda, A.K., Mlsna, L.M., Parrish, T.B. & Wong, P.C.M. (2009) Selective Neurophysiologic Responses to Timbre in Musicians with Different Listening Biographies. *Hum. Brain Mapp.*, **30**, 267–275.
- Marsh, R.A., Fuzessary, Z.M., Grosse, C.D. & Wenstrup, J.J. (2002) Projection to the inferior colliculus from the basal nucleus of the amygdala. *J. Neurosci.*, **22**, 10449–10460.
- Moreno, S. & Besson, M. (2005) Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann. N. Y. Acad. Sci.*, **1050**, 93–96.
- Musacchia, G., Sams, M., Skoe, E. & Kraus, N. (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl Acad. Sci. USA*, **104**, 15894–15898.
- Musacchia, G., Strait, D. & Kraus, N. (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. *Hear. Res.*, **241**, 34–42.
- Nair, D., Large, W.E., Steinberg, F. & Kelso, J.A.S. (2002) Expressive timing and perception of emotion in music: an fMRI study. *Proceedings of the 7th International Conference on Music Perception and Cognition*, **627**, 627–630.
- Neubauer, A., Freudenthaler, H.H. & Pfuerscheller, G. (1995) Intelligence and spatiotemporal patterns of event-related desynchronization (ERD). *Intelligence*, **20**, 249–266.
- Pantev, C., Roberts, L.E., Schulz, M., Engelien, A. & Ross, B. (2001) Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport*, **12**, 169–174.
- Peretz, I., Gagnon, L. & Bouchard, P. (1998) Music and emotion: perceptual determinants, immediacy and isolation after brain damage. *Cognition*, **68**, 111–141.
- Peretz, I., Cummings, S. & Dube, M.P. (2007) The genetics of congenital amusia (tone-deafness): a family aggregation study. *Am. J. Hum. Genet.*, **81**, 582–588.
- Price, C., Thierry, G. & Griffiths, T. (2005) Speech-specific auditory processing: where is it? *Trends Cogn. Sci.*, **9**, 271–276.
- Ragert, P., Schmidt, A., Altenmüller, E. & Dinse, H.R. (2004) Superior tactile performance and learning in professional pianists: evidence for metaplasticity in musicians. *Eur. J. Neurosci.*, **19**, 473–478.
- Russo, N., Nicol, T., Zecker, S., Hayes, E. & Kraus, N. (2005) Auditory training improves neural timing in the human brainstem. *Behav. Brain Res.*, **156**, 95–103.
- Russo, N.M., Bradlow, A.R., Skoe, E., Trommer, B.L., Nicol, T., Zecker, S. & Kraus, N. (2008) Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clin. Neurophysiol.*, **119**, 1720–1731.
- Saunders, J. (1996) Real-time discrimination of broadcast speech/music. *Proc. Int. Conf. Acoust. Speech Signal Process.*, **2**, 993–996.
- Scherer, K.R. (2003) Vocal communication of emotion: a review of research paradigms. *Speech Commun.*, **40**, 227–256.
- Schlaug, G. (2001) The brain of musicians: a model for functional and structural plasticity. *Ann. N. Y. Acad. Sci.*, **930**, 281–299.
- Schlaug, G., Norton, A., Overy, K. & Winner, E. (2005) Effects of music training on the child's brain and cognitive development. *Ann. N. Y. Acad. Sci.*, **1060**, 219–230.
- Schneider, P., Scherg, M., Dosch, H.G., Specht, H.J., Gutschalk, A. & Rupp, A. (2002) Morphology of Heschl's Gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.*, **5**, 688–694.
- Shahin, A.J., Roberts, L.E., Pantev, C., Aziz, M. & Picton, T.W. (2007) Enhanced anterior-temporal processing for complex tones in musicians. *Clin. Neurophysiol.*, **118**, 209–220.
- Song, J.H., Skoe, E., Wong, P.C.M. & Kraus, N. (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *J. Cogn. Neurosci.*, **20**, 1892–1902.
- Suga, N., Gao, E., Zhang, Y., Xiaofeng, M. & Olsen, J.F. (2001) The corticofugal system for hearing: recent progress. *Proc. Natl Acad. Sci.*, **97**, 11807–11814.
- Teas, D.C. (1993) *Cochlear Microphonic Response to Low Frequency Noise*. N.T.I.S., Springfield, IL.
- Thompson, W.F., Schellenberg, E.G. & Husain, G. (2004) Decoding speech prosody: do music lessons help? *Emotion*, **4**, 46–64.
- Trainor, L.J. (2005) Are there critical periods for music development? *Dev. Psychobiol.*, **46**, 262–278.
- Van Lancker, D., Cornelius, C. & Kreiman, J. (1989) Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Dev. Neuropsychol.*, **5**, 207–226.
- Weinberger, N.M. (2004) Specific long-term memory traces in primary auditory cortex. *Nat. Rev. Neurosci.*, **5**, 279–290.
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T. & Kraus, N. (2007) Musical training shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.*, **10**, 420–422.
- Yee, W., Holleran, S. & Jones, M.R. (1994) Sensitivity to event timing in regular and irregular sequences: influences of musical skill. *Percept. Psychophys.*, **56**, 461–471.
- Zatorre, R.J., Belin, P. & Penhune, V.B. (2002) Structure and function of auditory cortex: music and speech. *Trends Cogn. Sci.*, **6**, 37–46.