

Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound

Alexandra Parbery-Clark,^{1,2} Erika Skoe,^{1,2} and Nina Kraus^{1,2,3,4}

¹Auditory Neuroscience Laboratory, Departments of ²Communication Sciences, ³Neurobiology and Physiology, and ⁴Otolaryngology, Northwestern University, Evanston, Illinois 60208

Musicians have lifelong experience parsing melodies from background harmonies, which can be considered a process analogous to speech perception in noise. To investigate the effect of musical experience on the neural representation of speech-in-noise, we compared subcortical neurophysiological responses to speech in quiet and noise in a group of highly trained musicians and nonmusician controls. Musicians were found to have a more robust subcortical representation of the acoustic stimulus in the presence of noise. Specifically, musicians demonstrated faster neural timing, enhanced representation of speech harmonics, and less degraded response morphology in noise. Neural measures were associated with better behavioral performance on the Hearing in Noise Test (HINT) for which musicians outperformed the nonmusician controls. These findings suggest that musical experience limits the negative effects of competing background noise, thereby providing the first biological evidence for musicians' perceptual advantage for speech-in-noise.

Introduction

Musical performance is one of the most complex and cognitively demanding tasks that humans undertake (Parsons et al., 2005). By the age of 21, professional musicians have spent ~10,000 h practicing their instruments (Ericsson et al., 1993). This long-term sensory exposure may account for their enhanced auditory perceptual skills (Micheyl et al., 2006; Rammsayer and Altenmüller, 2006) as well as functional and structural adaptations seen at subcortical and cortical levels for speech and music (Pantev et al., 2003; Peretz and Zatorre, 2003; Trainor et al., 2003; Shahin et al., 2004; Besson et al., 2007; Musacchia et al., 2007; Wong et al., 2007; Lee et al., 2009; Strait et al., 2009). One critical aspect of musicianship is the ability to parse concurrently presented instruments or voices. Given this, we hypothesized that a musician's lifelong experience with musical stream segregation would transfer to its linguistic homolog, speech-in-noise (SIN) perception. To test this, we recorded speech-evoked auditory brainstem responses (ABRs) in both quiet and noise and tested SIN perception in a group of musicians and nonmusicians.

Speech perception in noise is a complex task requiring the segregation of the target signal from competing background noise. This task is further complicated by the degradation of the acoustic signal, with noise particularly disrupting the perception of fast spectrotemporal features (e.g., stop consonants) (Brandt and Rosen, 1980). Whereas children with language-based learn-

ing disabilities (Bradlow et al., 2003; Ziegler et al., 2005) and hearing-impaired adults (Gordon-Salant and Fitzgibbons, 1995) are especially susceptible to the negative effects of background noise, musicians are less affected and demonstrate better performance for SIN when compared with nonmusicians (Parbery-Clark et al., 2009).

Recent work points to a relationship between brainstem timing and SIN perception (Hornickel et al., 2009). The ABR, which reflects the activity of subcortical nuclei (Jewett et al., 1970; Lev and Sohmer, 1972; Smith et al., 1975; Chandrasekaran and Kraus, 2009), is widely used to assess the integrity of auditory function (Hall, 1992). The speech-evoked ABR represents the neural encoding of stimulus features with considerable fidelity (Kraus and Nicol, 2005). Nonetheless, the ABR is not hardwired; rather, it is experience dependent and varies with musical and linguistic experience (Krishnan et al., 2005; Song et al., 2008; for review, see Tzounopoulos and Kraus, 2009). Compared with nonmusicians, musicians exhibit enhanced subcortical encoding of sounds with both faster responses and greater frequency encoding. These enhancements are not simple gain effects. Rather, musical experience selectively strengthens the underlying neural representation of sounds reflecting the interaction between cognitive and sensory factors (Kraus et al., 2009), with musicians demonstrating better encoding of complex stimuli (Wong et al., 2007; Strait et al., 2009) as well as behaviorally relevant acoustic features (Lee et al., 2009). We hypothesized that, despite the well documented disruptive effects of noise (Don and Eggermont, 1978; Cunningham et al., 2001; Russo et al., 2004), musicians have enhanced encoding of the noise-vulnerable temporal stimulus events (onset and consonant–vowel formant transition) and increased neural synchrony in the presence of background noise resulting in a more precise temporal and spectral representation of the signal.

Received July 8, 2009; revised Aug. 29, 2009; accepted Oct. 5, 2009.

This work was supported by National Science Foundation Grant 0842376. We thank everyone who participated in this study. We also thank Dr. Samira Anderson and Carrie Lam for their help with peak picking and Dr. Beverly Wright, Dr. Frederic Marmel, Dr. Samira Anderson, Trent Nicol, and Judy Song for suggestions made on a previous version of this manuscript.

Correspondence should be addressed to Dr. Nina Kraus, Department of Communication Sciences, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: nkraus@northwestern.edu.

DOI:10.1523/JNEUROSCI.3256-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2914100-08\$15.00/0

Materials and Methods

Sixteen musicians (10 females) and 15 nonmusicians (9 females) participated in this study. Participants' ages ranged from 19 to 30 years (mean age, 23 ± 3 years). Participants categorized as musicians started instrumental training before the age of 7 and practiced consistently for at least 10 years before enrolling in the study. Nonmusicians were required to have had <3 years of musical training, which must have occurred >7 years before their participation in the study. All participants were right-handed, had normal hearing thresholds from 125 to 8000 Hz ≤ 20 dB, no conductive hearing loss, and normal ABRs to a click and speech syllable as measured by BioMARK (Biological Marker of Auditory Processing) (Natus Medical). No participant reported any cognitive or neurological deficits.

Stimuli

The speech syllable /da/ was a 170 ms six-formant speech sound synthesized using a Klatt synthesizer (Klatt, 1980) at a 20 kHz sampling rate. Except for the initial 5 ms stop burst, this syllable is voiced throughout with a steady fundamental frequency ($f_0 = 100$ Hz). This consonant-vowel syllable is characterized by a 50 ms formant transition (transition between /d/ and /a/) followed by a 120 ms steady-state (unchanging formants) portion corresponding to /a/ (see Fig. 2A). During the formant transition period, the first formant rises linearly from 400 to 720 Hz, and the second and third formants fall linearly from 1700 to 1240 Hz and 2580 to 2500 Hz, respectively. The fourth, fifth, and sixth formants remain constant at 3330, 3750, and 4900 Hz for the entire syllable.

The background noise consisted of multitalker babble created by the superimposition of grammatically correct but semantically anomalous sentences spoken by six different speakers (two males and four females). These sentences were recorded for a previous experiment and the specific recording parameters can be found in Smiljanic and Bradlow (2005). The noise file was 45 s in duration.

Procedure

The speech syllable /da/ was presented in alternating polarities at 80 dB sound pressure level (SPL) binaurally with an interstimulus interval of 83 ms (Neuro Scan Stim 2; Compumedics) through insert ear phones (ER-3; Etymotic Research). In the noise condition, both the /da/ and the multitalker babble were presented simultaneously to both ears. The /da/ was presented at a +10 signal-to-noise ratio over the background babble, which was looped for the duration of the condition. The responses to two background conditions, quiet and noise, were collected using NeuroScan Acquire 4.3 recording system (Compumedics) with four Ag-AgCl scalp electrodes. Responses were differentially recorded at a 20 kHz sampling rate with a vertical montage (Cz active, forehead ground, and linked-earlobe references), an optimal montage for recording brainstem activity (Galbraith et al., 1995; Chandrasekaran and Kraus, 2009). Contact impedance was $2 \text{ k}\Omega$ or less between electrodes. Six thousand artifact-free sweeps were recorded for each condition, with each condition lasting between 23 and 25 min. Participants watched a silent, captioned movie of their choice to facilitate a wakeful, yet still state for the recording session.

To limit the inclusion of low-frequency cortical activity, brainstem responses were off-line bandpass filtered from 70 to 2000 Hz (12 dB/octave, zero phase-shift) using NeuroScan Edit 4.3. The filtered recordings were then epoched using a -40 to 213 ms time window with the stimulus onset occurring at 0 ms. Any sweep with activity greater than $\pm 35 \mu\text{V}$ was considered artifact and rejected. The responses to the two polarities were added together to minimize the presence of the cochlear microphonic and stimulus artifact on the neural response (Gorga et al., 1985; Aiken and Picton, 2008). Last, responses were amplitude-baselined to the prestimulus period.

Analysis

Brainstem response

Timing and amplitude of onset and transition peaks. The neural response to the onset of sound ("onset peak") and the formant transition ("transition peak") are represented by large positive peaks occurring between

9–11 and 43–45 ms poststimulus onset (0 ms), respectively. These peaks were independently identified using NeuroScan Edit 4.3 (Compumedics) by the primary author and a second peak picker, who was blind to the participants' group. In the case of disagreement with peak identification, the advice of a third peak picker was sought. All participants had distinct onset peaks in the quiet condition, but three participants (two nonmusicians and one musician) had nonobservable onset peaks in the noise condition. Statistical analyses for onset peak latency and amplitude only included those participants who had clearly discriminable peaks in both quiet and noise ($n = 28$). The transition peak was the most reliable peak in the transition response for both the quiet and noise condition and was clearly identifiable in all participants ($n = 31$).

Quiet-to-noise correlation. To measure the effect of noise on the response morphology, the degree of correlation between each participant's response in quiet and in noise was calculated. Correlation coefficients were calculated by shifting, in the time domain, the response waveform in noise relative to the response waveform in quiet until a maximum correlation was found. This calculation resulted in a Pearson's r value, with smaller values indicating a more degraded response. Because the presence of noise typically causes the response to be delayed, the response in noise was shifted in time by up to 2 ms, and the maximum correlation over the 0–2 ms shift was recorded. The response time region used for this analysis was from 5 to 180 ms, which encompassed the complete neural response (onset, transition, and steady state).

Stimulus-to-response correlations. To gauge the effect of noise on the neural response to the steady-state vowel, the stimulus and response waveforms were compared via cross-correlation. The degree of similarity was calculated by shifting the stimulus waveform in time by 8–12 ms relative to the response, until a maximum correlation was found between the stimulus and the region of the response corresponding to the vowel. This time lag (8–12 ms) was chosen because it encompassed the stimulus transmission delay (from the ER-3 transducer and ear insert ~ 1.1 ms) and the neural lag between the cochlea and the rostral brainstem. This calculation resulted in a Pearson's r value for both the quiet and noise conditions.

Harmonic representation. To assess the impact of background noise on the neural encoding of the stimulus spectrum, a fast Fourier transform was performed on the steady-state portion of the response (60–180 ms). From the resulting amplitude spectrum, average spectral amplitudes of specific frequency bins were calculated. Each bin was 60 Hz wide and centered on the stimulus f_0 (100 Hz) and the subsequent harmonics H_2 – H_{10} (200–1000 Hz; whole-integer multiples of the f_0). To create a composite score representing the strength of the overall harmonic encoding, the average amplitudes of the H_2 to H_{10} bins were summed.

SIN perceptual measures

The behavioral data used for this study are reported in Parbery-Clark et al. (2009) and are used here for correlative purposes with the brainstem measures. Two commonly used clinical tests for speech-in-noise were administered: Hearing in Noise Test (HINT) and Quick Speech-in-Noise Test (QuickSIN).

Hearing in Noise Test. Hearing in Noise Test (HINT; Bio-logic Systems) (Nilsson et al., 1994) is an adaptive test of speech recognition that measures speech perception ability in speech-shaped white noise. The full test administration protocol is described by Parbery-Clark et al. (2009). For the purpose of this study, we restricted our analyses to only include the condition in which the speech and the noise originated from the same location because it most closely mirrored the stimulus presentation setup for the electrophysiological recordings. During HINT, participants repeated short semantically and syntactically simple sentences spoken by a male (e.g., "She stood near the window") presented in a speech-shaped noise background. Participants sat 1 m in front of the speaker from which the target sentences and the background noise were delivered. The noise presentation level was fixed at 65 dB SPL with the target sentence intensity level increasing or decreasing depending on performance. A final threshold signal-to-noise ratio (SNR)—defined as the difference in decibels between the speech and noise presentation levels for which 50% of

sentences are correctly repeated—was calculated with a negative threshold SNR indicating better performance on the task.

Quick Speech-in-Noise Test. Quick Speech-in-Noise Test (QuickSIN; Etymotic Research) (Killion et al., 2004) is a nonadaptive test of speech perception in four-talker babble that is presented binaurally through insert earphones (ER-2; Etymotic Research). Sentences were presented at 70 dB SPL, with the first sentence starting at a SNR of 25 dB and with each subsequent sentence being presented with a 5 dB SNR reduction down to 0 dB SNR. The sentences, which are spoken by a female, are syntactically correct yet have minimal semantic or contextual cues (Wilson et al., 2007). Participants repeated each sentence (e.g., “The square peg will settle in the round hole”), and their SNR score was based on the number of correctly repeated key words (underlined). For each participant, four lists were selected, with each list consisting of six sentences with five target words per sentence. Each participant’s final score, termed “SNR loss,” was calculated as the average score across each of the four administered lists. A more negative SNR loss is indicative of better performance on the task. For more information about SNR loss and its calculation, see the study by Killion et al. (2004).

Analytical and statistical methods

Correlations [quiet-to-noise and stimulus-to-response (SR)] and fast Fourier transforms were conducted with Matlab 7.5.0 routines (Mathworks).

All statistical analyses were conducted with SPSS. For all between- and within-group comparisons, a mixed-model repeated-measures ANOVA was conducted, with the subsequent planned *post hoc* tests when appropriate. Assumptions of normality, linearity, outliers, and multicollinearity were met for all analyses. In the case of group comparisons on a single variable, one-way ANOVAs were used. To investigate the relationship between frequency encoding and the stimulus-to-response correlation, a series of Pearson’s *r* correlations using all subjects, regardless of group, were used. Bonferroni’s corrections were applied when required.

Results

Musicians exhibited more robust speech-evoked auditory brainstem responses in background noise (Fig. 1). Musicians had earlier response onset timing, as well as greater phase-locking to the temporal waveform and stimulus harmonics, than nonmusicians. We also found that earlier response timing and more robust brainstem responses to speech in background noise were both related to better speech-in-noise perception as measured through HINT.

Brainstem response: quiet-to-noise correlations

For musicians, there was a greater degree of similarity between their brainstem responses to noise compared with quiet (one-way ANOVA: $F_{(1,30)} = 6.082, p = 0.02$; musicians: mean = 0.79, $\sigma = 0.07$; nonmusicians: mean = 0.7, $\sigma = 0.14$). This suggests that the addition of background noise does not degrade the musician brainstem response to speech, relative to their response in quiet, to the same degree as in nonmusicians.

Brainstem response: timing of onset and transition peaks

Musical experience was found to limit the degradative effect of background noise on the peaks in the brainstem response corre-

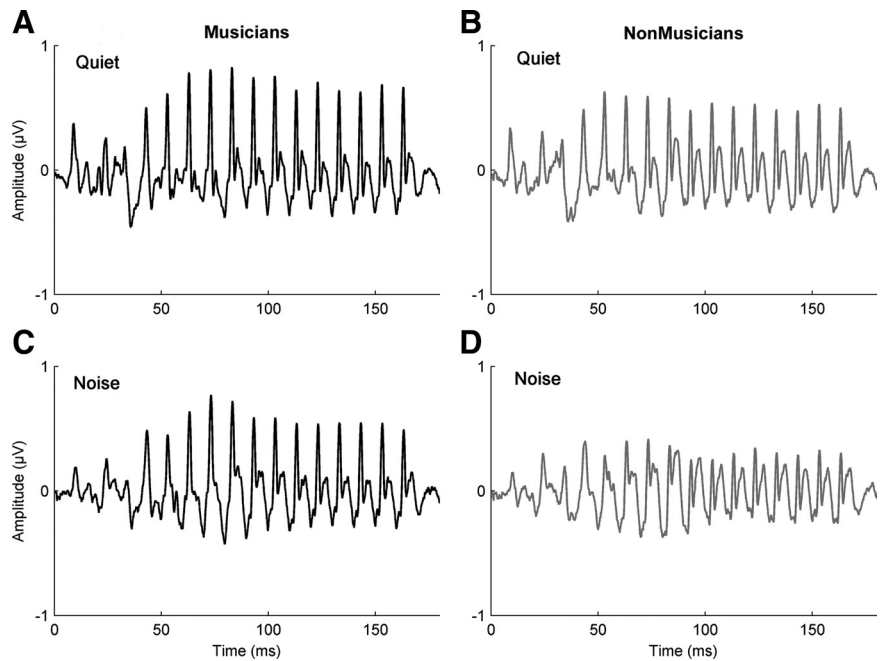


Figure 1. Auditory brainstem responses for musician and nonmusician groups. **A–D**, The speech-evoked auditory brainstem response in quiet [musicians (**A**) and nonmusicians (**B**)] and in background noise [musicians (**C**) and nonmusicians (**D**)]. Musicians demonstrated a greater degree of similarity between their responses in quiet and noise, suggesting that their responses are less degraded by the addition of background noise, unlike nonmusicians. A group comparison revealed a significant difference with musicians having a greater quiet-to-noise correlation than nonmusicians ($F_{(1,30)} = 6.082; p = 0.02$). The musicians are plotted in black and the nonmusicians in gray.

sponding to important temporal events in the stimulus. Typically, the addition of background noise delays the timing of the brainstem response, yet musicians exhibited smaller delays in timing than nonmusicians in noise. To investigate the effect of noise on the timing of the brainstem response to speech, we looked at the latencies of the onset peak (response to the onset of the stimulus, 9–11 ms) and the transition peak (response to the formant transition, 43–45 ms) for both the quiet and noise conditions. A mixed-model repeated-measures ANOVA, with group (musician/nonmusician) and background condition (quiet/noise) as the independent variables, and onset peak latency as the dependent variable, was performed. There was a significant main effect for background ($F_{(1,26)} = 307.841, p < 0.0005$), with noise resulting in delayed onset peaks, and a trend for group ($F_{(1,26)} = 3.219, p = 0.084$) with the musicians having earlier latencies in noise. There was also a significant interaction between group and background ($F_{(1,26)} = 4.936, p = 0.035$). Independent-samples *t* tests for each background condition revealed that the two groups had equivalent onset peak latencies in quiet ($t_{(26)} = 0.001, p = 0.976$; musicians: mean = 8.98 ms, $\sigma = 0.38$; nonmusicians: mean = 8.98 ms, $\sigma = 0.21$) but that, in noise, nonmusicians had significantly earlier onset responses ($t_{(26)} = 14.889, p = 0.001$; musicians: mean = 9.99 ms, $\sigma = 0.15$; nonmusicians: mean = 10.24 ms, $\sigma = 0.13$) (Fig. 2*B*). A similar relationship was found for the transition peak with significant main effects for both background and group, with the peak latencies being later in noise than in quiet ($F_{(1,29)} = 34.173, p < 0.0005$) and musicians having earlier transition latencies ($F_{(1,29)} = 8.937, p = 0.006$). There was a significant interaction between group and background ($F_{(1,29)} = 4.57, p = 0.041$). Again, *post hoc* comparisons revealed that the groups were not significantly different in quiet ($t_{(29)} = 1.43, p = 0.242$; musicians: mean = 43.03 ms, $\sigma = 0.41$; nonmusicians: mean = 43.22 ms, $\sigma = 0.46$), but musicians had earlier responses

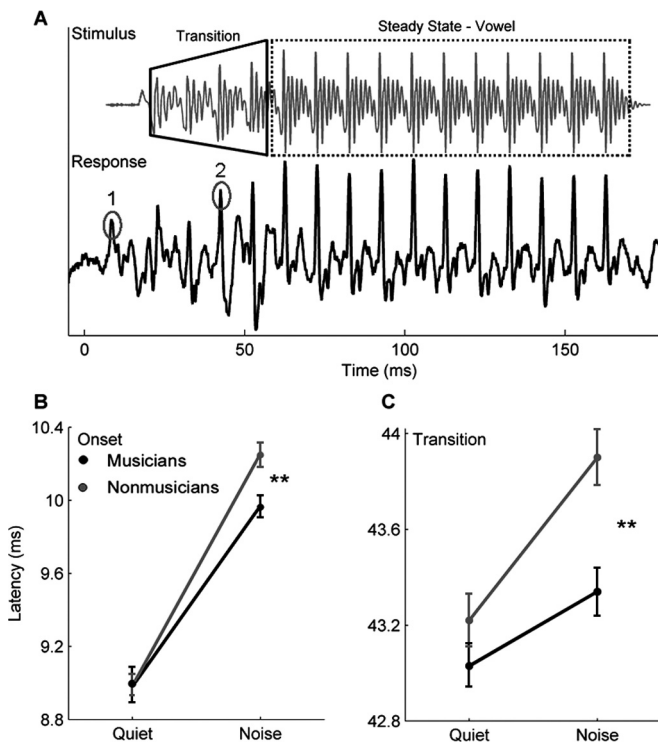


Figure 2. Group onset and transition differences. *A*, The stimulus /da/ and an individual (musician) response waveform. The stimulus waveform has been shifted forward in time (~8 ms) to align the stimulus and response onsets. The two major peaks corresponding to the onset (labeled 1) and the formant transition (labeled 2) of the stimulus are circled. *B*, *C*, Musicians (black) and nonmusicians (gray) have equivalent onset (*B*) and transition (*C*) peak latencies in the quiet condition. However, although both groups show an increase in onset and transition peak latency in the noise condition, the musicians are less affected. $**p < 0.01$. Error bars represent ± 1 SE.

in the presence of noise ($t_{(29)} = 15.2$, $p = 0.001$; musicians: mean = 43.34 ms, $\sigma = 0.35$; nonmusicians: mean = 43.90 ms, $\sigma = 0.43$) (Fig. 2*C*). Thus, for both groups, the addition of noise resulted in a delay in brainstem timing, but the latency shifts were smaller for the musicians, suggesting that their responses were less susceptible to the degradative effects of the background noise.

Brainstem response: amplitude of onset and transition peaks

Amplitudes of onset responses are known to be variable (Starr and Don, 1988; Hood, 1998), and previous research has found that when a speech stimulus is presented in noise the onset response is greatly reduced or eliminated (Russo et al., 2004). Consistent with these findings, the addition of background noise significantly reduced the onset peak amplitude for both groups with a significant main effect for background ($F_{(1,26)} = 179.715$, $p < 0.0005$; quiet: musicians: mean = 0.448 μ V, $\sigma = 0.112$; nonmusicians: mean = 0.390 μ V, $\sigma = 0.117$; noise: musicians: mean = 0.219 μ V, $\sigma = 0.796$; nonmusicians: mean = 0.172 μ V, $\sigma = 0.749$), and in the case of three participants (two nonmusicians and one musician) the onset response was completely eliminated. There were no significant group differences ($F_{(1,26)} = 2.014$, $p = 0.168$) nor a significant interaction between group and background ($F_{(1,26)} = 0.002$, $p = 0.961$). Although the transition response was more robust to the effects of noise than the onset response, in that all participants had reliably distinguishable peaks, the addition of background noise significantly reduced the amplitude of the transition peak ($F_{(1,29)} = 6.068$, $p < 0.02$; quiet: musicians: mean = 0.668 μ V, $\sigma = 0.249$; nonmusicians: mean =

0.578 μ V, $\sigma = 0.242$; noise: musicians: mean = 0.601 μ V, $\sigma = 0.202$; nonmusicians: mean = 0.498 μ V, $\sigma = 0.211$). There was no effect of group ($F_{(1,29)} = 1.476$, $p = 0.234$) nor an interaction between group and background ($F_{(1,29)} = 0.120$, $p = 0.732$), suggesting that noise had a similar effect on the peak amplitude of both groups.

Brainstem response: stimulus-to-response correlations and harmonic encoding

The presence of noise had a smaller degradative effect on the response to the vowel in musicians than nonmusicians. To quantify the effect of noise on the steady-state portion of the response, the degree of similarity between the stimulus and the corresponding brainstem response was calculated (SR correlation) for the quiet and noise conditions. A mixed-model ANOVA showed significant main effects of noise ($F_{(1,29)} = 17.49$, $p < 0.005$) and of group ($F_{(1,29)} = 6.01$, $p = 0.02$) and a marginally significant interaction ($F_{(1,29)} = 4.08$, $p = 0.052$). Subsequent independent-samples *t* tests indicated that the two groups had equivalent SR correlations in quiet ($t_{(29)} = 1.543$, $p = 0.134$; musicians: mean = 0.32, $\sigma = 0.04$; nonmusicians: mean = 0.28, $\sigma = 0.07$) but the musicians had significantly better SR correlations than the nonmusicians in noise ($t_{(29)} = 2.836$, $p = 0.008$; musicians: mean = 0.29, $\sigma = 0.05$; nonmusicians: mean = 0.22, $\sigma = 0.08$) (Fig. 3*A*). This suggests that the introduction of noise resulted in less degradation of the musician's response.

In the presence of background noise, musicians also showed significantly greater encoding of the harmonics (H_2 – H_{10}) (Fig. 3*C*). This was determined by spectrally analyzing the response to the stimulus steady state, the same time period used for calculating the SR correlations. Again, a repeated-measures ANOVA showed a main effect of noise ($F_{(1,29)} = 25.293$, $p < 0.005$) and group ($F_{(1,29)} = 6.255$, $p = 0.018$), but no interaction ($F_{(1,29)} = 0.004$, $p = 0.949$). *Post hoc* comparisons indicated that, although there was a trend in quiet for the musicians to have larger harmonic amplitudes ($t_{(29)} = 1.961$, $p = 0.06$; musicians: mean = 1.049 μ V, $\sigma = 0.302$; nonmusicians: mean = 0.859 μ V, $\sigma = 0.227$) (Fig. 3*B*), there was a significant difference in noise ($t_{(29)} = 2.871$, $p = 0.008$; musicians: mean = 0.886 μ V, $\sigma = 0.192$; nonmusicians: mean = 0.692 μ V, $\sigma = 0.182$). Conversely, when considering the f_0 , a repeated-measures ANOVA found no main effects for either noise ($F_{(1,29)} = 0.555$, $p = 0.462$), group ($F_{(1,29)} = 0.009$, $p = 0.924$), or an interaction ($F_{(1,29)} = 0.015$, $p = 0.904$; quiet: musicians: mean = 0.399 μ V, $\sigma = 0.201$; nonmusicians: mean = 0.395 μ V, $\sigma = 0.185$; noise: musicians: mean = 0.418 μ V, $\sigma = 0.200$; nonmusicians: mean = 0.409 μ V, $\sigma = 0.168$).

To elucidate the relationship between the SR correlation and the neural representation of the frequency components, Pearson's *r* correlations were conducted. We found that the summed harmonic representation (H_2 – H_{10}) in quiet was positively correlated with the SR correlation in quiet ($r = 0.581$, $p = 0.001$) and the summed harmonic representation (H_2 – H_{10}) in noise was positively correlated with the SR correlation in noise ($r = 0.648$, $p < 0.005$) (Fig. 3*D,E*). These results demonstrate that a greater harmonic representation is associated with a higher degree of correlation between the stimulus and the response. In quiet, the SR correlation of both groups was similar, and although the musicians tended to have greater harmonic representation, there were no significant group difference on either measure. However, in noise, musicians had greater harmonic encoding and also greater SR correlations. Therefore, it appears that musicians are better able to represent the stimulus harmonics in noise than nonmusicians and that this enhanced spectral representation contributes to their higher SR correlation. No group differences

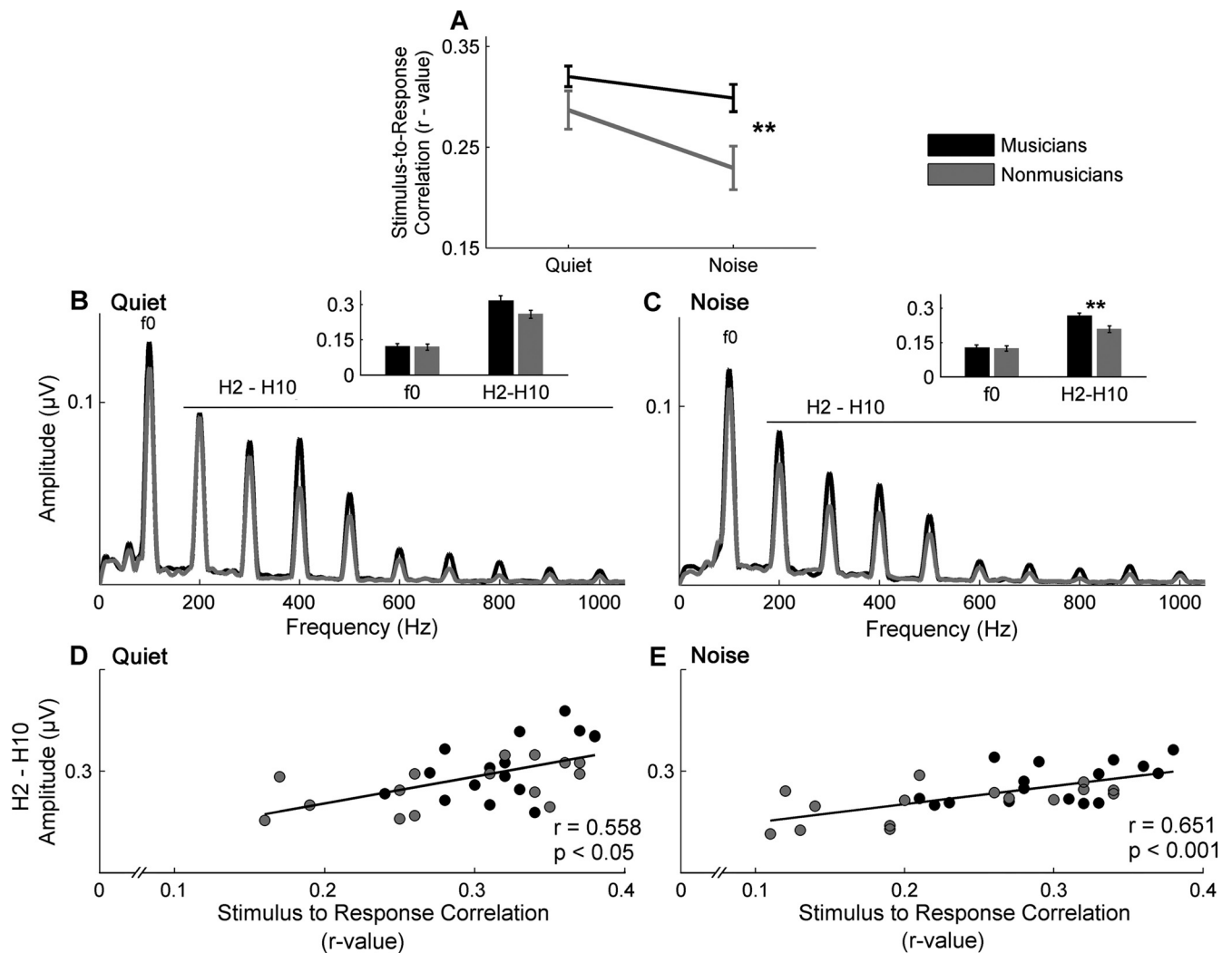


Figure 3. Stimulus-to-response correlation and harmonic representation. **A**, Musicians (black) show greater stimulus-to-response correlation in the presence of background noise, suggesting their brainstem responses are more resistant to the degradative effects of noise. **B, C**, The results of a fast Fourier transform show that, although musicians have equivalent encoding of the f_0 in both quiet (**B**) and noise (**C**), they demonstrate greater encoding of the harmonics (whole-integer multiples of the f_0). The inset bar graphs represent the f_0 and the summed representation of H_2-H_{10} in both quiet and noise, respectively. **D, E**, Finally, a greater summed representation of H_2-H_{10} was related to a higher stimulus-to-response correlation for both quiet (**D**) and noise (**E**) conditions. In quiet, the groups are relatively equivalent on these measures; however, in noise, there is a clear separation, with the musicians having both greater harmonic amplitudes and also higher stimulus-to-response correlations. $**p < 0.01$. Error bars represent ± 1 SE.

(one-way ANOVA) were found for the representation of the f_0 in either the quiet ($F_{(1,29)} = 0.511, p = 0.481$) or the noise condition ($F_{(1,29)} = 0.181, p = 0.673$) nor was there a relationship between f_0 and SR correlations (quiet: $r = 0.288, p = 0.116$; noise: $r = 0.275, p < 0.135$).

Brainstem-behavioral relationship

The brainstem measures in quiet (onset peak, transition peak, SR correlation) were not related to either behavioral test of speech-in-noise perception (HINT nor QuickSIN). For the brainstem responses in background noise, better HINT scores were related to earlier peak latencies for the onset and transition peaks (onset: $r = 0.551, p = 0.002$; transition: $r = 0.481, p = 0.006$) (Fig. 4A,B, respectively). In a similar vein, a greater SR correlation corresponded to better HINT scores ($r = -0.445, p = 0.01$) (Fig. 4C). These relationships suggest that better behavioral speech perception in noise, as measured by a lower HINT score, is associated with greater precision of brainstem timing in the presence of background noise (i.e., earlier peaks and higher SR correlations) (Fig. 4). QuickSIN, which

was previously found to be related to working memory and frequency discrimination ability (Parbery-Clark et al., 2009), showed no relationship with peak timing measures or the SR correlation for either the quiet or noise condition (all $p > 0.1$). Last, neither the representation of the f_0 nor the harmonics were related to performance on either speech-in-noise test (all $p > 0.1$).

Discussion

The present data show that, in background noise, musicians demonstrate earlier onset and transition response timing, better stimulus-to-response and quiet-to-noise correlations, and greater neural representation of the stimulus harmonics than nonmusicians. Earlier response timing as well as a better SR correlation in the noise condition were associated with better speech perception in noise as measured by HINT but not QuickSIN. Together, musical experience results in more robust subcortical representation of speech in the presence of background noise, which may contribute to the musician behavioral advantage for speech-in-noise perception.

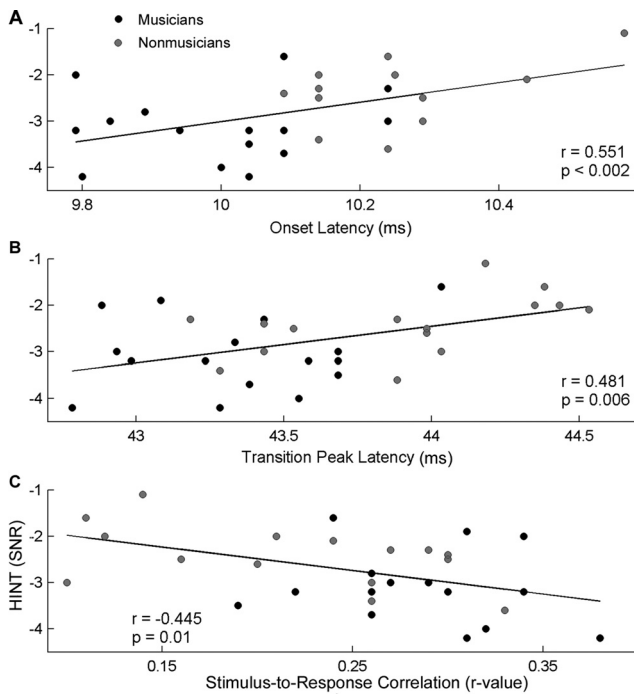


Figure 4. Relationship between speech perception in noise (HINT) and neurophysiological measures. **A, B,** Earlier onset (**A**) and transition (**B**) peak latencies are associated with better HINT scores. **C,** Likewise, a higher stimulus-to-response correlation between the eliciting speech stimulus /da/ and the brainstem response to this sound in the presence of background noise was found to correspond to a better HINT score. A more negative HINT SNR score is indicative of better performance. The black circles represent musicians, and the gray circles represent nonmusicians.

The subcortical representation of important stimulus temporal features was equivalent for musicians and nonmusicians in quiet, but the musicians' responses were less degraded by the background noise. The well documented increase in onset timing in background noise (Don and Eggermont, 1978; Burkard and Hecox, 1983a,b, 1987; Cunningham et al., 2001; Wible et al., 2005) was significantly smaller in the musicians. Previous work found no change in onset response timing with short-term auditory training (Hayes et al., 2003; Russo et al., 2005). In light of their results, Russo et al. (2005) postulated that the onset response, which originates from the primary afferent volley, may be more resistant to training effects. However, other studies, including ours, found earlier onset timing in musicians (Musacchia et al., 2007; Strait et al., 2009). It is therefore possible that extended auditory training, such as that experienced by musicians, is required for experience-related modulation of onset response timing.

Musicians also exhibited more robust responses to the steady-state portion of the stimulus in the presence of background noise. By calculating the degree of similarity between the stimulus waveform and the subcortical representation of the speech sound, we found that musicians had higher SR correlations in noise than nonmusicians. A greater SR correlation is indicative of more precise neural transcription of stimulus features. One possible explanation for this musician enhancement in noise may be based on the Hebbian principle, which posits that the associations between neurons that are simultaneously active are strengthened and those that are not are subsequently weakened (Hebb, 1949). Given the present results, we can speculate that extensive musical training may lead to greater neural coherence. This strengthening of the underlying neural circuitry would lead to a better bottom-up, feedforward representation of the signal. We can also inter-

pret these data within the framework of corticofugal modulation in which cortical processes shape the afferent auditory encoding via top-down processes. It is well documented that the auditory cortex sharpens the subcortical sensory representations of sounds through the enhancement of the target signal and the suppression of irrelevant competing background noise via the efferent system (Suga et al., 1997; Zhang et al., 1997; Luo et al., 2008). The musician's use of fine-grained acoustic information and lifelong experience with parsing simultaneously occurring melodic lines may refine the neural code in a top-down manner such that relevant acoustic features are enhanced early in the sensory system. This enhanced encoding improves the subcortical signal quality, resulting in a more robust representation of the target acoustic signal in noise. Although our data and experimental paradigm cannot tease apart the specific contributions of top-down or bottom-up processing, they are not mutually exclusive explanations. In all likelihood, top-down and bottom-up processes are reciprocally interactive with both contributing to the subcortical changes observed with musical training.

Interestingly, the improved stimulus-to-response correlation in the noise condition was related to greater neural representation of the stimulus harmonics (H_2 – H_{10}) but not the fundamental frequency in noise. Musicians, through the course of their training, spend thousands of hours producing, manipulating, and attending to musical sounds that are spectrally rich. The spectral complexity of music is partially attributable to the presence and relative strength of harmonics as well as the change in harmonics over time. Harmonics, which also underlie the perception of timbre or "sound color," enable us to differentiate between two musical instruments producing the same note. Musicians have enhanced cortical responses to their primary instrument, suggesting that their listening and training experience modulates the neural responses to specific timbres (Pantev et al., 2001; Margulis et al., 2009). Likewise, musicians demonstrate greater sensitivity to timbral differences and harmonic changes within a complex tone (Koelsch et al., 1999; Musacchia et al., 2008; Zedel and Alain, 2009). Within the realm of speech, timbral features provide important auditory cues for speaker and phonemic identification and contribute to auditory object formation (Griffiths and Warren, 2004; Shinn-Cunningham and Best, 2008). A potential benefit of heightened neural representation of timbral features would be the increased availability of harmonic cues, which can then be used to generate an accurate perceptual template of the target voice. An accurate template or perceptual anchor is considered a key element for improving signal perception (Best et al., 2008) and facilitates the segregation of the target voice from background noise (Mullennix et al., 1989; Ahissar, 2007). Zedel and Alain (2009) showed that musicians were more sensitive to subtle harmonic changes both behaviorally and cortically, which they interpret as a musician advantage for concurrent stream segregation—a skill considered important for speech perception in noise. In interpreting their results, Zedel and Alain (2009) postulate that the behavioral advantage and its corresponding cortical index may be attributable to a better representation of the stimulus at the level of the brainstem. Our findings, along with previous studies documenting enhanced subcortical representation of harmonics in musicians' responses, support this claim.

Limitations and future directions

These results provide biological evidence for the positive effect of lifelong musical training on speech-in-noise encoding. Nevertheless, we cannot determine the extent to which this enhancement

is mediated directly by musical training, group genetic differences, or a combination of the two. Longitudinal studies, akin to the large-scale design recently described by Forgeard et al. (2008) and Hyde et al. (2009), could not only elucidate the developmental time course and/or genetic disposition for the musician neural advantage for speech-in-noise, but may also help disentangle the relative influences of top-down and bottom-up processes on the neural encoding of speech-in-noise. Other important lines of research include the impact that the choice of musical instrument and musical genre, as well as extensive musical listening experience in the absence of active playing, have on the subcortical encoding of speech-in-noise.

Previous research has indicated that musical training may serve as a useful remediation strategy for children with language impairments (Overy et al., 2003; Besson et al., 2007; Jentschke et al., 2008; Jentschke and Koelsch, 2009). Our results imply that clinical populations known to have problems with speech perception in noise, such as children with language-based learning disabilities (e.g., poor readers) (Cunningham et al., 2001), may also benefit from musical training. More specifically, the subcortical deficits in sound processing seen in this population (e.g., timing and harmonics) (Wible et al., 2004; Banai et al., 2009; Hornickel et al., 2009) occur for the very elements that are enhanced in musicians. Moreover, for f_0 encoding, no group differences have been found between normal and learning-impaired children nor in the present study between musicians and nonmusicians; this is consistent with the previously described dissociation between the neural encoding of f_0 and the faster elements of speech (e.g., timing and harmonics) (Fant, 1960; Kraus and Nicol, 2005). By studying an expert population, we can investigate which factors contribute to an enhanced ability for speech perception in noise, providing future avenues for the investigation of speech perception deficits in noise as experienced by older adults and hearing-impaired and language-impaired children. By providing an objective biological index of speech perception in noise, brainstem activity may be a useful measure for evaluating the effectiveness of SIN-based auditory training programs.

Conclusion

Overall, our results offer evidence of musical expertise contributing to an enhanced subcortical representation of speech sounds in noise. Musicians had more robust temporal and spectral encoding of the eliciting speech stimulus, thus offsetting the deleterious effects of background noise. Faster neural timing and enhanced harmonic encoding in musicians suggests that musical experience confers an advantage resulting in more precise neural synchrony in the auditory system. These findings provide a biological explanation for musicians' perceptual enhancement for speech-in-noise.

References

- Ahissar M (2007) Dyslexia and the anchoring-deficit hypothesis. *Trends Cogn Sci* 11:458–465.
- Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. *Hear Res* 245:35–47.
- Banai K, Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009) Reading and subcortical auditory function. *Cereb Cortex* 19:2699–2707.
- Besson M, Schön D, Moreno S, Santos A, Magne C (2007) Influence of musical expertise and musical training on pitch processing in music and language. *Restor Neurol Neurosci* 25:399–410.
- Best V, Ozmeral EJ, Kopco N, Shinn-Cunningham BG (2008) Object continuity enhances selective auditory attention. *Proc Natl Acad Sci U S A* 105:13174–13178.
- Bradlow AR, Kraus N, Hayes E (2003) Speaking clearly for children with learning disabilities: sentence perception in noise. *J Speech Lang Hear Res* 46:80–97.
- Brandt J, Rosen JJ (1980) Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants. *Brain Lang* 9:324–337.
- Burkard R, Hecox K (1983a) The effect of broadband noise on the human brainstem auditory evoked response. I. Rate and intensity effects. *J Acoust Soc Am* 74:1204–1213.
- Burkard R, Hecox K (1983b) The effect of broadband noise on the human brainstem auditory evoked response. II. Frequency specificity. *J Acoust Soc Am* 74:1214–1223.
- Burkard R, Hecox KE (1987) The effect of broadband noise on the human brain-stem auditory evoked response. III. Anatomic locus. *J Acoust Soc Am* 81:1050–1063.
- Chandrasekaran B, Kraus N (2009) The scalp-recorded brainstem response to speech: neural origins. *Psychophysiology*. Advance online publication. Retrieved October 15, 2009. doi:10.1111/j.1469-8986.2009.00928.x.
- Cunningham J, Nicol T, Zecker SG, Bradlow A, Kraus N (2001) Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clin Neurophysiol* 112:758–767.
- Don M, Eggermont JJ (1978) Analysis of the click-evoked brainstem potentials in man using high-pass noise masking. *J Acoust Soc Am* 63:1084–1092.
- Ericsson KA, Krampe RT, Tesch-Römer C (1993) The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 100:363–406.
- Fant G (1960) *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
- Forgeard M, Winner E, Norton A, Schlaug G (2008) Practicing a musical instrument in childhood is associated with enhanced verbal ability and nonverbal reasoning. *PLoS One* 3:e3566.
- Galbraith GC, Arbagey PW, Branski R, Comerci N, Rector PM (1995) Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6:2363–2367.
- Gordon-Salant S, Fitzgibbons PJ (1995) Recognition of multiply degraded speech by young and elderly listeners. *J Speech Hear Res* 38:1150–1156.
- 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.
- Griffiths TD, Warren JD (2004) What is an auditory object? *Nat Rev Neurosci* 5:887–892.
- Hall JW (1992) *Handbook of auditory evoked responses*. Boston: Allyn and Bacon.
- Hayes EA, Warrier CM, Nicol TG, Zecker SG, Kraus N (2003) Neural plasticity following auditory training in children with learning problems. *Clin Neurophysiol* 114:673–684.
- Hebb DO (1949) *The organization of behavior*. New York: Wiley.
- Hood L (1998) *Clinical applications of the auditory brainstem response*. San Diego: Singular.
- Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N (2009) Subcortical differentiation of voiced stop consonants: relationships to reading and speech in noise perception. *Proc Natl Acad Sci U S A* 106:13022–13027.
- Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, Schlaug G (2009) Musical training shapes structural brain development. *J Neurosci* 29:3019–3025.
- Jentschke S, Koelsch S (2009) Musical training modulates the development of syntax processing in children. *Neuroimage* 47:735–744.
- Jentschke S, Koelsch S, Sallat S, Friederici AD (2008) Children with specific language impairment also show impairment of music-syntactic processing. *J Cogn Neurosci* 20:1940–1951.
- Jewett DL, Romano MN, Williston JS (1970) Human auditory evoked potentials: possible brain stem components detected on the scalp. *Science* 167:1517–1518.
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116:2395–2405.
- Klatt D (1980) Software for a Cascade/Parallel Formant Synthesizer. *J Acoust Soc Am* 67:13–33.
- Koelsch S, Schröger E, Tervaniemi M (1999) Superior attentive and pre-attentive auditory processing in musicians. *Neuroreport* 10:1309–1313.

- Kraus N, Nicol T (2005) Brainstem origins for cortical “what” and “where” pathways in the auditory system. *Trends Neurosci* 28:176–181.
- Kraus N, Skoe 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* 1169:543–557.
- 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.
- Lee KM, Skoe E, Kraus N, Ashley R (2009) Selective subcortical enhancement of musical intervals in musicians. *J Neurosci* 29:5832–5840.
- Lev A, Sohmer H (1972) Sources of averaged neural responses recorded in animal and human subjects during cochlear audiometry (electrocochleogram). *Arch Klin Exp Ohren Nasen Kehlkopfheilkd* 201:79–90.
- 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.
- Micheyl C, Delhommeau K, Perrot X, Oxenham AJ (2006) Influence of musical and psychoacoustical training on pitch discrimination. *Hear Res* 219:36–47.
- Mullennix JW, Pisoni DB, Martin CS (1989) Some effects of talker variability on spoken word recognition. *J Acoust Soc Am* 85:365–378.
- 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. *Hear Res* 241:34–42.
- Nilsson M, Soli SD, Sullivan JA (1994) Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95:1085–1099.
- Overy K, Nicolson RI, Fawcett AJ, Clarke EF (2003) Dyslexia and music: measuring musical timing skills. *Dyslexia* 9:18–36.
- Pantev C, Roberts LE, Schulz M, Engelen A, Ross B (2001) Timbre-specific enhancement of auditory cortical representations in musicians. *Neuroreport* 12:169–174.
- Pantev C, Ross B, Fujioka T, Trainor LJ, Schulte M, Schulz M (2003) Music and learning-induced cortical plasticity. *Ann N Y Acad Sci* 999:438–450.
- Parbery-Clark A, Skoe E, Lam C, Kraus N (2009) Musician enhancement for speech-in-noise. *Ear Hear* 30:653–661.
- Parsons LM, Sergent J, Hodges DA, Fox PT (2005) The brain basis of piano performance. *Neuropsychologia* 43:199–215.
- Peretz I, Zatorre RJ (2003) *The cognitive neuroscience of music*. Oxford: Oxford UP.
- Rammsayer T, Altenmüller E (2006) Temporal information processing in musicians and nonmusicians. *Music Percept* 24:37–48.
- Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. *Clin Neurophysiol* 115:2021–2030.
- Russo NM, Nicol TG, Zecker SG, Hayes EA, Kraus N (2005) Auditory training improves neural timing in the human brainstem. *Behav Brain Res* 156:95–103.
- Shahin A, Roberts LE, Trainor LJ (2004) Enhancement of auditory cortical development by musical experience in children. *Neuroreport* 15:1917–1921.
- Shinn-Cunningham BG, Best V (2008) Selective attention in normal and impaired hearing. *Trends Amplif* 12:283–299.
- Smiljanic R, Bradlow AR (2005) Production and perception of clear speech in Croatian and English. *J Acoust Soc Am* 118:1677–1688.
- 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.
- Song JH, Skoe E, Wong PC, Kraus N (2008) Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci* 20:1892–1902.
- Starr A, Don M (1988) Brain potentials evoked by acoustic stimuli. In: *Human event-related potentials. Handbook of electroencephalography and clinical neurophysiology, Vol 3* (Picton T, ed), pp 97–158. New York: Elsevier.
- Strait DL, Kraus N, Skoe E, Ashley R (2009) Musical experience influences subcortical processing of emotionally-salient vocal sounds. *Eur J Neurosci* 29:661–668.
- Suga N, Zhang Y, Yan J (1997) Sharpening of frequency tuning by inhibition in the thalamic auditory nucleus of the mustached bat. *J Neurophysiol* 77:2098–2114.
- 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.
- Tzounopoulos T, Kraus N (2009) Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62:463–469.
- Wible B, Nicol T, Kraus N (2004) Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems. *Biol Psychol* 67:299–317.
- Wible B, Nicol T, Kraus N (2005) Correlation between brainstem and cortical auditory processes in normal and language-impaired children. *Brain* 128:417–423.
- Wilson RH, McArdle RA, Smith SL (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856.
- 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.
- Zendel BR, Alain C (2009) Concurrent sound segregation is enhanced in musicians. *J Cogn Neurosci* 21:1488–1498.
- Zhang Y, Suga N, Yan J (1997) Corticofugal modulation of frequency processing in bat auditory system. *Nature* 387:900–903.
- Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C (2005) Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci U S A* 102:14110–14115.