

available at www.sciencedirect.comwww.elsevier.com/locate/brainres

**BRAIN
RESEARCH**

Research Report
Effects of reverberation on brainstem representation of speech in musicians and non-musicians
Gavin M. Bidelman*, Ananthanarayan Krishnan

 Department of Speech Language Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

ARTICLE INFO
Article history:

Accepted 29 July 2010

Available online 5 August 2010

Keywords:

Frequency-following response (FFR)

Speech perception

Hearing in noise

Auditory system

Electroencephalography (EEG)

Frequency difference limen

ABSTRACT

Perceptual and neurophysiological enhancements in linguistic processing in musicians suggest that domain specific experience may enhance neural resources recruited for language specific behaviors. In everyday situations, listeners are faced with extracting speech signals in degraded listening conditions. Here, we examine whether musical training provides resilience to the degradative effects of reverberation on subcortical representations of pitch and formant-related harmonic information of speech. Brainstem frequency-following responses (FFRs) were recorded from musicians and non-musician controls in response to the vowel /i/ in four different levels of reverberation and analyzed based on their spectro-temporal composition. For both groups, reverberation had little effect on the neural encoding of pitch but significantly degraded neural encoding of formant-related harmonics (i.e., vowel quality) suggesting a differential impact on the source-filter components of speech. However, in quiet and across nearly all reverberation conditions, musicians showed more robust responses than non-musicians. Neurophysiologic results were confirmed behaviorally by comparing brainstem spectral magnitudes with perceptual measures of fundamental (F0) and first formant (F1) frequency difference limens (DLs). For both types of discrimination, musicians obtained DLs which were 2–4 times better than non-musicians. Results suggest that musicians' enhanced neural encoding of acoustic features, an experience-dependent effect, is more resistant to reverberation degradation which may explain their enhanced perceptual ability on behaviorally relevant speech and/or music tasks in adverse listening conditions.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Human communication almost always requires the ability to extract speech sounds from competing background interference. In reverberant settings, sound waves are reflected (from

walls, floor, and ceiling) in an exaggerated manner resulting in a temporal overlap of the incident and reflected wavefronts. Because the reflected sounds travel along a longer path than the incident sound they arrive at the listener's ear delayed relative to the original, resulting in a distorted, noisier version

* Corresponding author. Purdue University, Department of Speech Language Hearing Sciences, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA. Fax: +1 765 494 0771.

E-mail addresses: gbidelma@purdue.edu (G.M. Bidelman), rkrish@purdue.edu (A. Krishnan).

Abbreviations: ACF, autocorrelation function; ACG, autocorrelogram; AFC, alternative forced choice task; DL, difference limen; F0, fundamental frequency; FFR, frequency-following response; IC, inferior colliculus; MOC, medial olivocochlear; OAE, otoacoustic emission

of the intended message. As such, reverberation can have deleterious effects on an individual's ability to identify and discriminate critical information in the speech stream (Culling et al., 2003; Gelfand and Silman, 1979; Nabelek and Robinson, 1982; Yang and Bradley, 2009).

It is important to note that the effects of reverberation on speech acoustics are fundamentally different than those of simple additive noise. Though both hinder speech intelligibility (Nabelek and Dagenais, 1986), noise effects occur with the addition of a masking signal to target speech whereas the effects of reverberation arise primarily from acoustic distortions to the target itself (for synergistic effects, see George et al., 2008). The deleterious effects of reverberation on speech intelligibility, as compared to quiet, can be ascribed to consequences of both "overlap-" (i.e., forward) and "self-masking" (Nabelek et al., 1989). As segments of the speech signal reflect in a reverberant space they act as forward maskers, overlapping subsequent syllables and inhibiting their discriminability. In addition, the concurrent reflections with incident sound dramatically change the dynamics of speech by blurring waveform fine-structure. When acting on a time-varying signal, this "temporal smearing" tends to transfer spectral features of the signal from one time epoch into later ones, inducing smearing effects in the spectrogram (Wang and Brown, 2006). As a consequence, this internal temporal smearing distorts the energy within each phoneme such that it can effectively act as its own (i.e., "self") masker. With such distortions, normal hearing listeners have difficulty identifying and discriminating consonantal features (Gelfand and Silman, 1979; Nabelek et al., 1989), vowels (Drgas and Blaszk, 2009; Nabelek and Letowski, 1988), and time-varying formant cues (Nabelek and Dagenais, 1986) in reverberant listening conditions. It should be noted that these confusions are further exacerbated with hearing impairment (Nabelek and Letowski, 1985; Nabelek and Dagenais, 1986; Nabelek, 1988).

Recent studies have shown that musical experience improves basic auditory acuity in both time and frequency as musicians are superior to non-musicians at detecting rhythmic irregularities and fine-grained manipulations in pitch both behaviorally (Jones and Yee, 1997; Kishon-Rabin et al., 2001; Micheyl et al., 2006; Rammsayer and Altenmuller, 2006; Spiegel and Watson, 1984) and neurophysiologically (Brattico et al., 2009; Crummer et al., 1994; Koelsch et al., 1999; Russeler et al., 2001; Tervaniemi et al., 2005). In addition, through their superior analytic listening, musicians parse and segregate competing signals in complex auditory scenes more effectively than non-musicians (Munte et al., 2001; Nager et al., 2003; Oxenham et al., 2003; van Zuijen et al., 2004; Zendel and Alain, 2009). Intriguingly, their domain specific music experience also influences faculties necessary for language. Musicians' perceptual enhancements improve language specific abilities including phonological processing (Anvari et al., 2002; Slevc and Miyake, 2006) and verbal memory (Chan et al., 1998; Franklin et al., 2008). Indeed, relative to non-musicians, English speaking musicians show better performance in the identification of lexical tones (Lee and Hung, 2008) and are more sensitive at detecting timbral changes in speech and music (Chartrand and Belin, 2006). These perceptual advantages are corroborated by electrophysiological evidence which demonstrates that both cortical (Chandrasekaran et al., 2009;

Moreno and Besson, 2005; Pantev et al., 2001; Schon et al., 2004) and subcortical (Bidelman et al., 2009; Musacchia et al., 2008; Wong et al., 2007) brain circuitry tuned by long-term music training facilitates the encoding of speech related signals. Taken together, these studies indicate that a musician's years of active engagement with complex auditory objects sharpens critical listening skills and furthermore, that these benefits assist and interact with brain processes recruited during linguistic tasks (Bidelman et al., 2009; Patel, 2008; Slevc et al., 2009).

Though musical training can improve speech related behaviors in dry or quiet listening environments, the question remains whether this enhancement provides increased resistance to signal degradation in adverse listening conditions (e.g., reverberation) as reflected in both perceptual and electrophysiological measures. For the click-evoked auditory brainstem response (ABR), prolonged latencies and reduced wave V amplitudes provide neural indices of the effects of additive noise to non-speech stimuli (Burkard and Hecox, 1983a,b; Don and Eggermont, 1978; Krishnan and Plack, 2009). Recently, Parbery-Clark et al. (2009a) have demonstrated that the frequency-following response (FFR) in musicians is both more robust and occurs with faster onset latency than in non-musician controls when speech stimuli are presented in the presence of noise interference. In addition, behavioral tests showed that musicians are more resilient to the deleterious effects of background noise obtaining better performance at lower signal-to-noise ratios (i.e., more challenging listening conditions) on clinical measures of speech-in-noise perception (Parbery-Clark et al., 2009b). These results suggest that musicians' strengthened sensory-level neural encoding of acoustic features is more resistant to noise degradation which in turn may contribute to their enhanced perceptual ability. However, to date, there are no published reports on the effects of reverberation on brainstem encoding of complex speech sounds.

As a window into the early, subcortical stages of speech processing we employ the electrophysiologic frequency-following response (FFR). The scalp recorded FFR reflects sustained phase-locked activity in a population of neural elements within the rostral brainstem and is characterized by a periodic waveform which follows the individual cycles of the stimulus (Chandrasekaran and Kraus, 2010; Krishnan, 2007). The FFR has provided a number of insights into the subcortical processing of ecologically relevant stimuli including speech (Krishnan and Gandour, 2009) and music (Bidelman and Krishnan, 2009; Kraus et al., 2009). Furthermore, the FFR has revealed that subcortical experience-dependent plasticity enhances neural representation of pitch in life-long speakers of a tone language (Krishnan et al., 2005, 2009b, 2010b) and individuals with extensive music experience (Bidelman et al., 2009; Lee et al., 2009; Wong et al., 2007). Here, we compare the spectral properties of musicians' and non-musicians' FFRs in response to vowel tokens presented in various amounts of reverberation (Fig. 1). The present framework allows us to assess whether or not there is a musician advantage to speech encoding in adverse listening conditions. We then compare properties of brainstem encoding to behavioral measures of pitch and first formant discrimination in reverberation to evaluate the role of subcortical sensory-level processing on perceptual measures related to speech intelligibility.

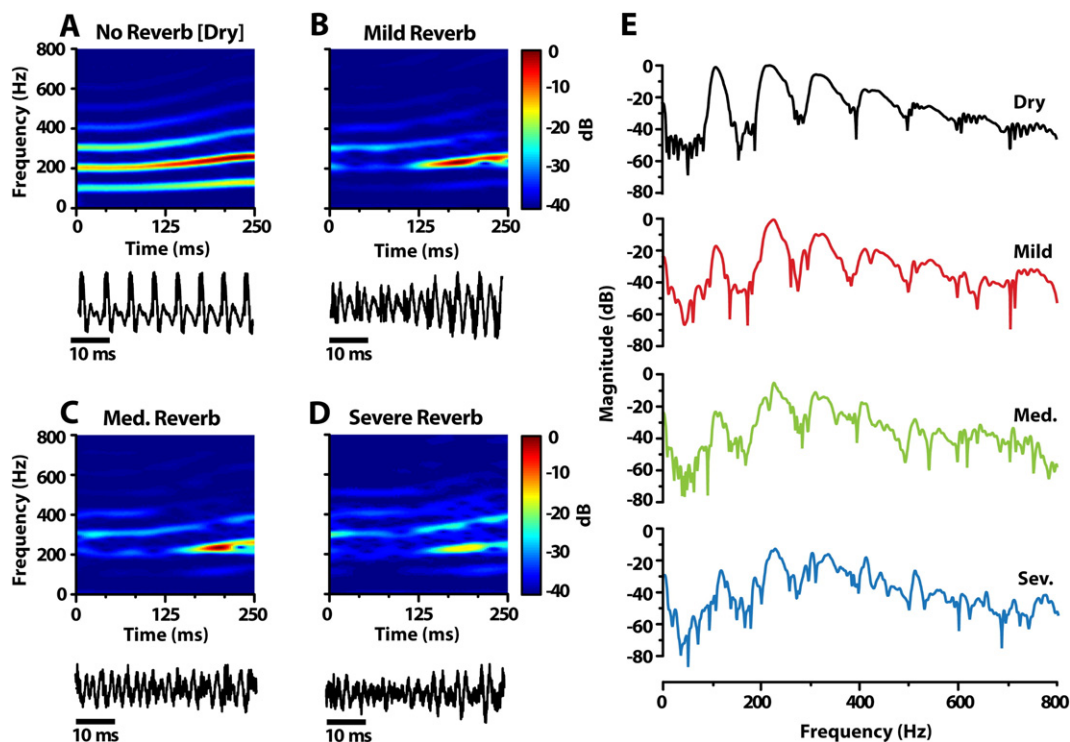


Fig. 1 – Reverberation stimuli and their acoustic characteristics. (A–D) Spectrograms for the vowel /i/ with time-varying fundamental frequency and formants $F1=300$; $F2=2500$; $F3=3500$; $F4=4530$ Hz. The addition of reverberation was accomplished by time domain convolution of the original vowel with room impulse responses recorded in a corridor at a distance of either 0.63 m (mild reverb; Reverberation Time (RT_{60}) ≈ 0.7 s), 1.25 m (medium reverb; $RT_{60}\approx 0.8$ s), or 5 m (severe reverb; $RT_{60}\approx 0.9$ s) from the sound source (Sayles and Winter, 2008; Watkins, 2005). With increasing reverberation, spectral components (top row) and the time waveform (bottom row) are “smeared”, resulting in a blurring of the fine time-frequency representation of the vowel (compare “dry” to “severe” condition). (E) FFT spectra over the duration of the stimuli as a function of reverberation strength. Smearing of high frequency spectral components (induced by blurring in the waveform’s temporal fine-structure) is evident with increasing levels of reverberation. However, note the relative resilience of the fundamental frequency ($F0$: 103–130 Hz) with increasing reverb.

2. Results

2.1. FFR temporal and spectral composition

Grand averaged FFR waveforms are shown for musicians and non-musicians per reverberation condition in Fig. 2. Musicians’ responses are more robust than those of non-musicians in clean (i.e., no reverb) and all but the most severe level of reverberation. The clearer, more salient response periodicity in musicians indicates enhanced phase-locked activity to speech relevant components (e.g., voice $F0$ and $F1$ harmonics) not only in clean but also adverse listening environments.

Grand averaged autocorrelograms derived from individual FFR waveforms in response to the vowel /i/ in three different levels of reverberation are shown in Fig. 3A (mild condition not shown). In all conditions, musicians show stronger bands of phase-locked activity at the fundamental frequency ($F0$) than their non-musician counterparts. To quantify the effects of reverberation on temporal pitch-relevant information present in the neural responses, we normalized each time-lag in the autocorrelogram by the corresponding stimulus period. This normalization essentially removes the time-varying aspect of

the signal by aligning each ACF slice according to the fundamental period (Sayles and Winter, 2008). Averaging slices over time yields a single summary ACF describing the degree of phase-locked activity present over the duration of the FFR response. Time-averaged ACFs are shown for musicians and non-musicians as a function of stimulus reverberation in Fig. 3B. For both groups, the neural representation of $F0$ remains fairly intact with the addition of reverberant energy (i.e., magnitude of summary ACF energy at fundamental period is invariant to increasing reverberation, Fig. 3B). Only with severe reverberation does the response magnitude seem to diminish. Yet, larger magnitudes in musicians across conditions illustrate their enhanced neural representation for pitch-relevant information in both quiet and reverberant listening conditions.

Grand averaged spectrograms and FFTs derived from individual FFRs (averaged over the duration of the response) are shown in Figs. 3C and D, respectively. Consistent with temporal measures of $F0$ (Figs. 3A and B), these spectral measures show robust encoding at $F0$ in both groups with little or no change in magnitude with increasing amounts of reverberation, except for a reduction in magnitude in the severe reverberant condition. The effects of added reverberation on

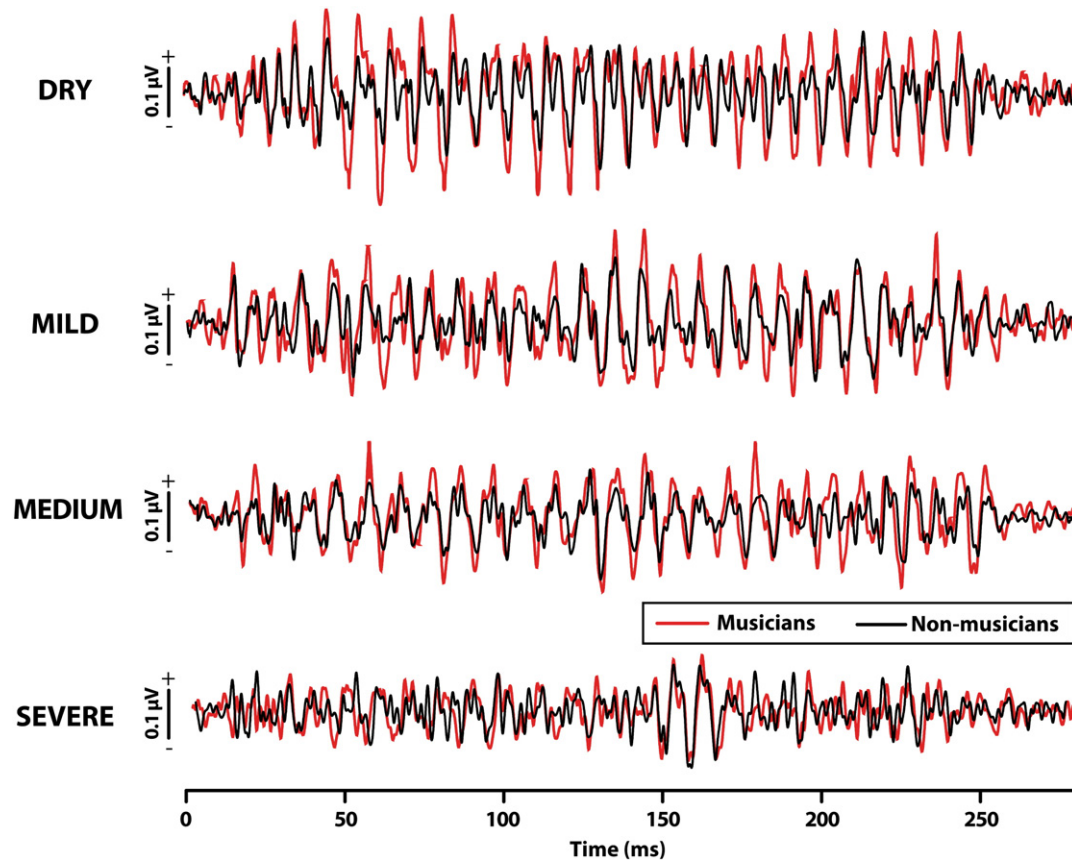


Fig. 2 – Grand average FFR waveforms in response to the vowel /i/ with increasing levels of reverberation. Musicians (red) have more robust brainstem responses than non-musicians (black) in clean (i.e., no reverb) and all but the most severe reverberation condition. The clearer, more salient response periodicity in musicians indicates enhanced phase-locked activity to speech relevant spectral components not only in clean but also adverse listening environments. FFR, frequency-following response.

first formant-related harmonics (mean spectral magnitude of response harmonics 2–4) are much more pronounced. Both groups show weaker, more diffuse encoding of higher harmonics relative to their responses in the dry condition (i.e., Fig. 3C, distorted energy in the spectrograms above about 200 Hz). Yet, compared to non-musicians, musicians still show larger magnitudes at F0 and its harmonics, indicating greater encoding of the “pitch” and formant-related spectral aspects of speech even in the presence of reverberation (Fig. 3D).

2.2. Brainstem magnitudes of F0 and formant-related harmonics

FFR encoding of F0 is shown in Fig. 4A. A two-way mixed-model ANOVA with group (2 levels; musicians and non-musicians), the between-subjects factor, and reverberation (4 levels; dry, mild, medium, and severe), the within-subject factor, was conducted on F0 magnitudes in order to evaluate the effects of long-term musical experience and reverberation on brainstem encoding of F0 (note: a similar ANOVA model was used on all other dependent variables). Results revealed significant main effects of group [$F_{1,18}=12.57$, $p=0.0023$] and reverb [$F_{3,54}=9.37$, $p<0.0001$]. The interaction failed to reach significance [$F_{3,54}=1.79$, $p=0.16$], indicating that musicians’ F0

magnitudes were greater than those of the non-musicians across all levels of reverberation. For both groups, posthoc Tukey–Kramer adjusted multiple comparisons ($\alpha=0.05$) revealed that F0 encoding was affected only by the strongest level of reverberation (i.e., dry=mild=medium>severe). Together, these results imply that reverberation had rather minimal effect on brainstem representation of F0 for both groups and that musicians consistently encoded F0 better than non-musicians.

Formant-related encoding, as measured by the mean spectral magnitude of response to harmonics 2–4, is shown in Fig. 4B. The mixed-model ANOVA revealed significant main effects of group [$F_{1,18}=24.29$, $p=0.0001$] and reverb [$F_{3,54}=20.57$, $p<0.0001$] on the representation of these harmonics. The group×reverb interaction was not significant [$F_{3,54}=0.51$, $p=0.68$], indicating that musicians’ encoding of formant-related harmonics was greater than non-musicians across the board. In contrast to F0 encoding, Tukey–Kramer multiple comparisons revealed that F1 harmonics were affected beginning as early as the mild reverberation (i.e., dry>mild) and that all higher levels of reverberation subsequently produced poorer encoding. Thus, although reverberation significantly degraded FFR encoding of F1 related harmonics in both groups, the amount of degradation was appreciably smaller in the musicians.

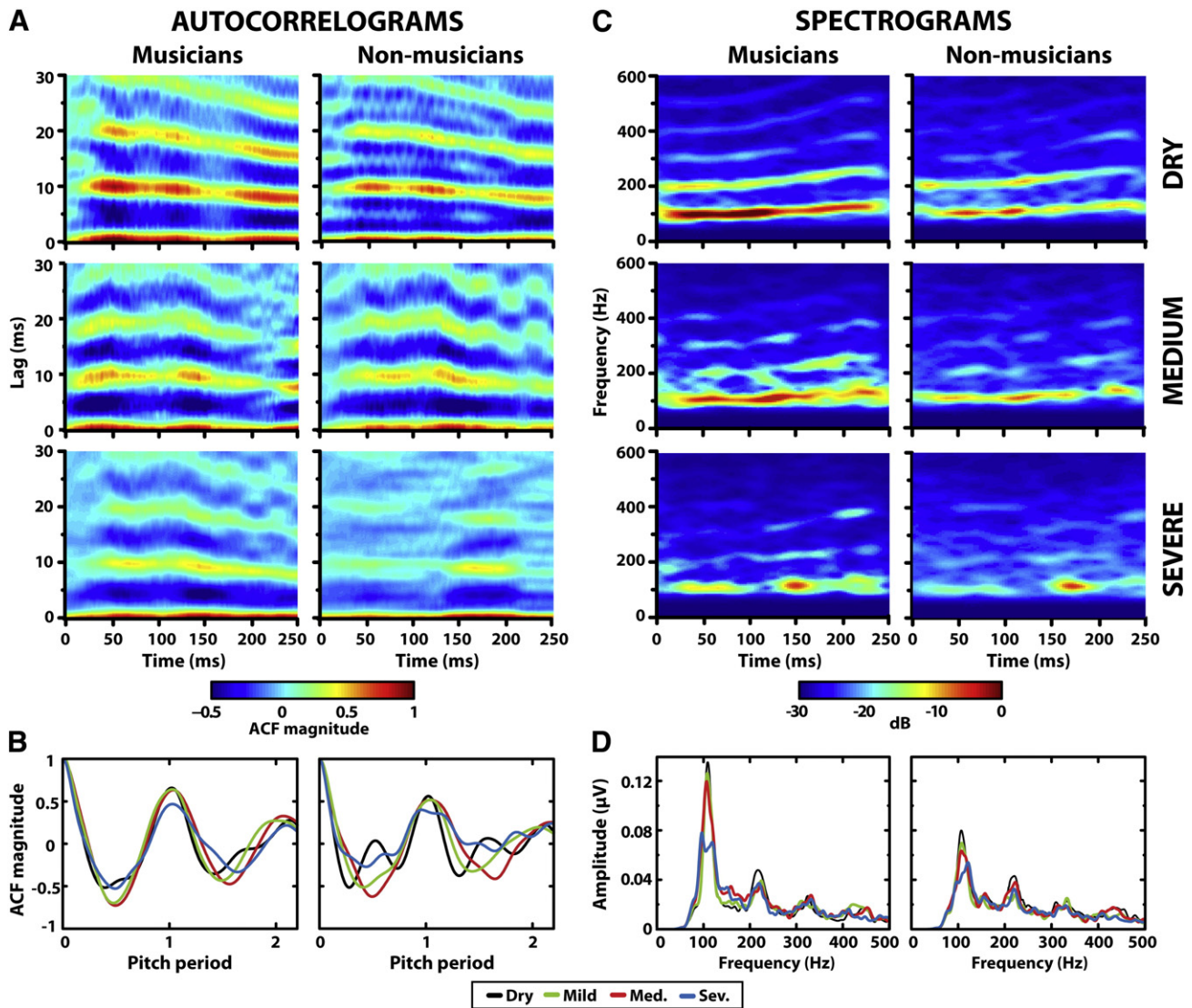


Fig. 3 – Autocorrelograms (A), time-averaged ACFs (B), spectrograms (C), and time-averaged FFTs (D) derived from FFR waveforms in response to the vowel /i/ in various amounts of reverberation. As indexed by the invariance of ACF magnitude at the fundamental period (i.e., F0 period = 1), increasing levels of reverberation has little effect on the neural encoding of pitch-relevant information for both groups (A–B). Yet, larger response magnitudes in musicians across conditions illustrate their enhanced neural representation for F0 in both dry and reverberant listening conditions (B). The effect of reverberation on FFR encoding of the formant-related harmonics is much more pronounced (C–D). As with the temporal measures, in both groups, the representation of F0 (100–130 Hz) remains more intact across conditions than higher harmonics (>200 Hz) which are smeared and intermittently lost in more severe amounts of reverberation (e.g., compare strength of F0 to strength of harmonics across conditions). However, as with F0, musicians demonstrate more resilience to the negative effects of reverberation than non-musicians showing more robust representation for F1 harmonics even in the most degraded conditions. ACF, autocorrelation function; FFR, frequency-following response; F0, fundamental frequency; F1, first formant.

2.3. Dry-to-reverberant response correlations

Dry-to-reverberant response correlations are shown per group in Fig. 5. A mixed-model ANOVA revealed significant main effects of group [$F_{1,18}=22.82, p=0.0002$] and reverb [$F_{2,36}=7.60, p=0.0018$] as well as their interaction [$F_{2,36}=3.92, p=0.0288$]. Multiple comparisons revealed that beginning with medium reverberation, musicians showed higher correlation between their reverberant and dry (i.e., no reverb) responses than non-musicians. Within the musician group, dry-to-reverberant response correlations

were unchanged with increasing reverberation strength. In contrast, non-musicians' reverberant responses became increasingly uncorrelated relative to their dry response with growing reverberation strength. These results indicate that the morphology of a musician's response in reverberant degraded listening conditions remains similar to their response morphology in dry (i.e., no reverb) conditions. The differential effect between groups suggests that reverberation does not degrade the musician brainstem response to speech, relative to their response in dry, to the same degree as in non-musicians.

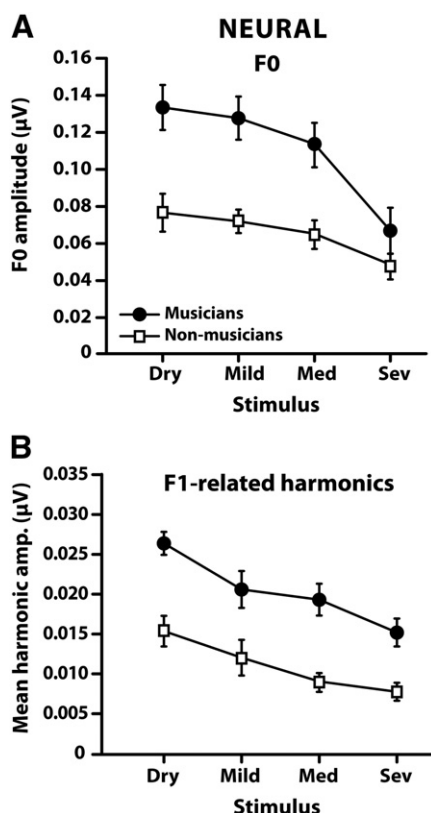


Fig. 4 – Neural encoding of F0 (A) and F1 harmonics (B) related information derived from brainstem FFRs. F0 encoding is defined as the spectral magnitude of the fundamental frequency and F1 harmonics as the mean spectral magnitude of harmonics 2–4. Regardless of group membership, reverberation negatively affects brainstem responses as evident in the roll-off of magnitudes with increasing levels of reverb. For both groups, reverberation has a rather minimal effect on the neural encoding of the vowel's F0 (A) but a pronounced effect on the encoding F1 harmonics (B). Yet, musicians, relative to non-musicians, show superior encoding for both F0 and F1 harmonics in dry as well as reverberant conditions. Note the difference in scales between the ordinates of A and B. Error bars indicate ± 1 SE. F0, fundamental frequency; F1, first formant.

2.4. Behavioral F0 and F1 frequency difference limens

Group fundamental frequency (F0 DLs) and first formant frequency (F1 DLs) difference limens are shown in Figs. 6A and B, respectively. Musicians obtained voice F0 DLs approximately four times smaller than non-musicians, consistent with previous reports examining pitch discrimination in musicians and non-musicians (Kishon-Rabin et al., 2001; Michey et al., 2006; Strait et al., 2010). A mixed-model ANOVA revealed significant main effects of group [$F_{1,6}=154.49$, $p<0.0001$] but not reverb [$F_{3,18}=0.59$, $p=0.631$]. These results indicate that although musicians were always superior to non-musicians at detecting changes in voice pitch (approximately 4 times better), reverberation itself did not hinder either group's overall performance. That is, within group, subjects performed equally well at detecting F0 differences in the mild,

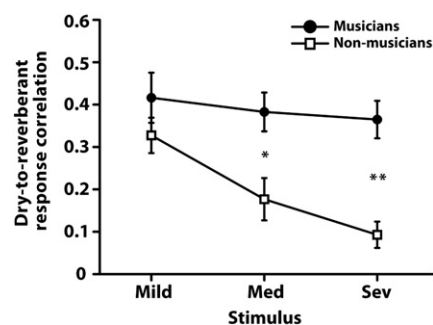


Fig. 5 – Dry-to-reverberant response correlations. The degree of correlation (Pearson's r) between reverberant (mild, medium, and severe) and dry responses was computed for each participant. Only the steady-state portions of the FFRs were considered (15–250 ms). Higher correlation values indicate greater correspondence between neural responses in the absence and presence of reverberation. Across increasing levels of reverberation, musicians show higher correlation between their reverberant and dry (i.e., no reverberation) responses than non-musician participants. A significant group \times stimulus interaction [$F_{2,36}=3.92$, $p=0.028$] indicates that musicians' FFRs are more resistant to the negative effects of reverberation than non-musicians, especially in medium and severe conditions. Error bars represent ± 1 SE. Group difference: * $p<0.01$ and ** $p<0.001$.

medium, and severe reverberation conditions as in the dry condition (i.e., no reverb). These results corroborate other studies which have shown that reverberation has little or no effect on F0 discrimination (Qin and Oxenham, 2005).

For first formant discrimination, F1 DLs were roughly an order of magnitude larger than F0 DLs. In other words, participants were much poorer at detecting changes in the vowel's formant structure than they were at changes in pitch. A mixed-model ANOVA revealed significant main effects of group [$F_{1,6}=30.28$, $p=0.0015$] indicating that musicians obtained better F1 DLs than their non-musician counterparts across the board (i.e., F1 DLs were approximately half as large in musicians). The main effect of reverb was also significant [$F_{3,18}=15.16$, $p<0.001$]. The group \times reverb interaction was not [$F_{3,18}=0.85$, $p=0.4827$]. Within group, Tukey-Kramer multiple comparisons revealed that all participants performed significantly better (i.e., smaller F1 DLs) in dry and mild reverberation than in the medium and severe cases. In other words, subjects had more difficulty detecting changes in the vowel's first formant frequency when degraded by reverberation. These results are consistent with the reduced ability of listeners to discriminate changes in complex timbre (Emiroglu and Kollmeier, 2008) and formant structure (Liu and Kewley-Port, 2004) in noisy listening conditions. Taken together, our results indicate that although reverberation hindered performance (i.e., higher F1 DLs) in both groups, musicians were superior to non-musicians at detecting changes in formant structure.

3. Discussion

The results of this study relate to three main observations: (1) increasing levels of reverberation degrades brainstem

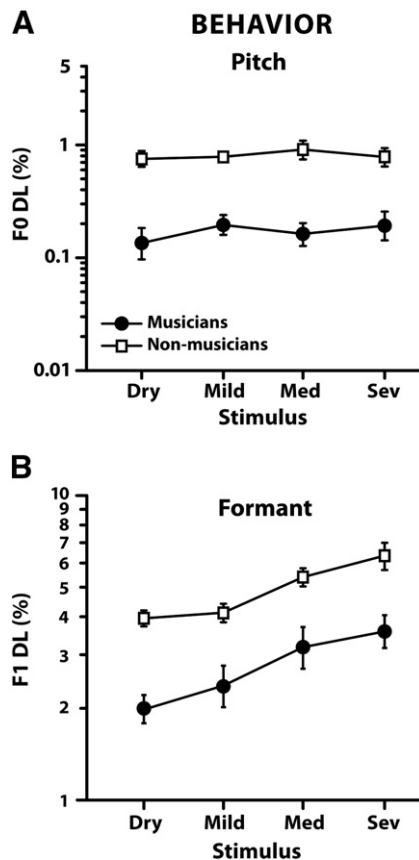


Fig. 6 – Mean behavioral fundamental (F0 DL) and first formant (F1 DL) frequency difference limens as a function of reverberation. F0 and F1 DLs were obtained separately by manipulating a vowel token’s pitch or first formant, respectively. (A) Musicians show superior ability at detecting changes in voice pitch obtaining F0 DLs approximately 4 times smaller than non-musicians. For both groups however, reverberation has little effect on F0 DLs indicating that it does not inhibit voice pitch discrimination. (B) Musicians detect changes in voice quality related to vowel’s first formant frequency better than non-musicians. However, in contrast to pitch, F1 DLs for both groups become poorer with increasing levels of reverberation suggesting that formant frequency discrimination is more difficult in challenging listening conditions. Error bars indicate ± 1 SE.

encoding of formant-related harmonics but not F0 which appears to be relatively more resistant to reverberation; (2) neural encoding of F1 harmonics and F0 is more robust in musicians as compared to non-musicians in both dry and reverberant listening conditions indicating that long-term musical experience enhances the subcortical representation of speech regardless of listening environment; (3) these electrophysiological effects correspond well with behavioral measures of vowel pitch and first formant discrimination (i.e., F0 and F1 DLs), that is, more robust brainstem representation corresponds with better ability to detect changes in the spectral properties of speech.

3.1. Neural basis for the differential effects of reverberation on F0 and formant-related encoding

Several studies have proposed that the pitch of complex sounds may be extracted by combining neural phase-locking to the temporal fine-structure for resolved harmonics and phase-locking to the temporal envelope modulation resulting from interaction of several unresolved harmonics (Cariani and Delgutte, 1996a,b; Meddis and O’Mard, 1997; Sayles and Winter, 2008). This latter pitch-relevant cue has been shown to degrade markedly with increasing reverberation, presumably due to the breakdown of temporal envelope modulation caused by randomization of phase relationships between unresolved harmonics (Sayles and Winter, 2008). This means the pitch of complex sounds containing only unresolved harmonics will be severely degraded by reverberation. In contrast, for those containing both resolved and unresolved components, pitch encoding in the presence of reverberation must rely solely on the temporally smeared fine-structure information in resolved regions (Sayles and Winter, 2008). The observation of little or no change in FFR phase-locking to F0 with increasing reverberation (Fig. 4, see also supplementary data, Fig. S1) is consistent with the lack of degradation in low frequency phase-locked units in the cochlear nucleus encoding spectrally resolved pitch (Sayles and Winter, 2008; Stasiak et al., 2010). Sayles and Winter (2008) suggest that both the more robust neural phase-locking in the low frequency channels in general and the more salient responses to resolved components increase their resistance to temporal smearing resulting from reverberation. It is also plausible that the slower rate of F0 change in our stimuli reduced the smearing effects of reverberation on the neural encoding of F0. However, it is not entirely clear why the encoding of the resolved first formant (F1) related harmonics in our FFR data showed greater degradation with increased reverberation. It is possible that this differential effect of reverberation on encoding may be due to the relatively greater spectro-temporal smearing of the formant-related higher frequency harmonics in our stimuli than the F0 component (see Fig. S1; Macdonald et al., 2010; Nabelek et al., 1989).

3.2. Relationship to psychophysical data

Psychophysical studies have shown that reverberation can have dramatically detrimental effects on speech intelligibility (Nabelek et al., 1989; Plomp, 1976) and on the ability to discriminate one voice from another (Culling et al., 2003; Darwin and Hukin, 2000) in normal hearing individuals. That it is the reduction in the effectiveness of the envelope modulation as a cue for extraction of pitch of complex sounds is suggested by the inability of normal hearing individuals to distinguish high-numbered (unresolved) harmonics from band-filtered noise in the presence of reverberation (Sayles and Winter, 2008). In contrast, the ability to discriminate a complex consisting of low-numbered (resolved) harmonics from noise is unaffected by reverberation. Our psychophysical data also show little or no effect of reverberation on the ability to discriminate F0 (pitch-relevant information) and is consistent with the observation of Sayles and Winter (2008) for their resolved stimuli (see also Qin and Oxenham, 2005). The

complementary results between brain and behavioral measures indicates that the temporal encoding of pitch-relevant information for complex sounds in reverberation is mediated by preservation of the degraded fine-structure in the low frequency channels as the temporal envelope from high frequency channels is severely degraded.

Paralleling our electrophysiological results, our psychophysical data showed a reduction in the ability to discriminate F1 related harmonics with increasing reverberation. Similar results have been reported for interfering noise which increases a listener's just noticeable difference (JND) for timbral aspects of musical instruments (Emiroglu and Kollmeier, 2008) and elevates vowel formant frequency DLs (Liu and Kewley-Port, 2004). Formants are relatively distinct in the speech signal and do not have the same redundancies as harmonics contributing to pitch, i.e., formants do not have harmonically related counterparts whereas integer related harmonics all reinforce a common F0. Given their uniqueness, it is likely that the voice quality characteristics of speech, including formants, are more susceptible to the effects of reverberation than pitch. In addition, the reduced performance in F1 discrimination (but not F0) may be due to the greater smearing of the higher frequency F1 related harmonics in the stimulus compared to the smearing of F0 (Fig. S1).

3.3. Neural basis for musician advantage to hearing in reverberation

Our findings provide further evidence that experience-dependent plasticity is induced by long-term experience with complex auditory signals. Consistent with previous psychophysical reports (Kishon-Rabin et al., 2001; Micheyl et al., 2006; Pitt, 1994), we observed that musicians' voice pitch (F0) and first formant (F1) discrimination performance was 2–4 times better (i.e., smaller DLs) than that of non-musicians (Fig. 6). Complementary results were seen in brainstem responses where FFR F0 and F1 harmonic magnitudes were nearly twice as large in musicians as non-musicians (Fig. 4). Therefore, musicians seem to have an advantage, and resilience (e.g., Fig. 5), in both their neural encoding and behavioral judgments of reverberant degraded speech. These data provide further evidence that the benefits of extensive musical training are not restricted solely to the domain of music. Indeed, other recent reports show that active engagement with music starting early in life transfers to functional benefits in language processing at both cortical (Chandrasekaran et al., 2009; Schon et al., 2004; Slevc et al., 2009) and subcortical (Bidelman et al., 2009; Musacchia et al., 2008; Wong et al., 2007) levels. Our data also converge and extend recent reports associating these enhancements with offsetting the deleterious effects of noise babble (Parbery-Clark et al., 2009a) and temporal degradation (Krishnan et al., 2010a) on the subcortical representation of speech relevant signals.

From a neurophysiologic perspective, the optimal encoding we find in musicians reflects enhancement in phase-locked activity within the rostral brainstem. IC architecture (Braun, 1999, 2000; Schreiner and Langner, 1997) and its response properties (Braun, 2000; Langner, 1981, 1997) provide optimal hardware in the midbrain for extracting spectrally complex information. The enhancements we observe in musicians

likely represent a strengthening of this subcortical circuitry developed from the many hours of repeated exposure to dynamic spectro-temporal properties found in music. These experience-driven enhancements may effectively increase the system's gain and therefore the ability to represent important characteristics of the auditory signal (Pantev et al., 2001; Pitt, 1994; Rammsayer and Altenmuller, 2006). Higher gain in sensory-level encoding may ultimately improve a musician's hearing in degraded listening conditions (e.g., noise or reverberation) as speech components would be extracted with higher neural signal-to-noise ratios (SNRs). A more favorable SNR may be one explanation for musicians' neurophysiological and perceptual advantages we observe here for speech-in-reverberation. Although noise masking and reverberation are fundamentally different interferences, they represent realistic forms of signal degradation to which listeners might be subjected. Yet, regardless of the nature of interference (noise or reverberation), musicians appear to show more resistance to these degradative effects (current study; Parbery-Clark et al., 2009a,b).

Overwhelming empirical evidence localizes the primary source generator of the FFR to the inferior colliculus (IC) of the midbrain (Akhoun et al., 2008; Galbraith et al., 2000; Smith et al., 1975; Sohmer and Pratt, 1977). However, because it is a far field evoked potential, the neurophysiologic activity recorded by surface electrodes may reflect the aggregate of both afferent and efferent auditory processing. Therefore, the experience-dependent brainstem enhancements we observe in musicians may reflect plasticity local to the IC (Bajo et al., 2010; Gao and Suga, 1998; Krishnan and Gandour, 2009; Yan et al., 2005), the influence of top-down corticofugal modulation (Kraus et al., 2009; Luo et al., 2008; Perrot et al., 2006; Tzounopoulos and Kraus, 2009), or more probably, a combination of these two mechanisms working in tandem (Xiong et al., 2009). Though none of these possibilities can be directly confirmed noninvasively in humans, indirect empirical evidence from the otoacoustic emissions (OAEs) literature posits that a musician's speech-in-noise encoding benefits may begin in auditory stages peripheral to the IC.

The cochlea itself receives efferent feedback from the medial olivocochlear (MOC) bundle, a fiber track originating in the superior olivary complex within the caudal brainstem and terminating on the outer hair cells in the Organ of Corti. As measured by contralateral suppression of OAEs, converging studies suggest that the MOC feedback loop is stronger and more active in musicians than non-musicians (Brashears et al., 2003; Micheyl et al., 1995, 1997; Perrot et al., 1999). Indeed, the MOC efferents have been implicated in hearing in adverse listening conditions (Guinan, 2006) and play an antimasking role during speech perception in noise (Giraud et al., 1997). Therefore, it is plausible that a musician's advantage for hearing in degraded listening conditions (e.g., reverberation) may result from enhancements in SNR beginning as early as the level of the cochlea (Micheyl and Collet, 1996). Although the FFR primarily reflects retrocochlear neural activity generated in the rostral brainstem, we cannot rule out the possibility that some amount of signal enhancement has already occurred in auditory centers preceding the IC. As such, the speech-in-reverberation encoding benefits we observe in musicians may reflect the output from these lower level structures (i.e., cochlea or caudal brainstem

nuclei) which is further tuned or at least maintained in midbrain responses upstream.

3.4. Reverberation differentially affects source-filter characteristics of speech

The source-filter theory of speech production postulates that speech acoustics result from the glottal source being filtered by the vocal tract's articulators. The fundamental frequency at which the vocal folds vibrate determines the pitch of a talker independently from the configuration of the vocal tract and oral cavity, which determine formant structure (i.e., voice quality) (Fant, 1960). Together, these two cues provide adequate information for identifying who is speaking (e.g., male vs. female talker) and what is being said (e.g., /a/ vs. /i/ vowel) (Assmann and Summerfield, 1989).

Recently, Kraus and colleagues have proposed a data-driven theoretical framework suggesting that the parallel source-filter streams found at a cortical level may emerge in subcortical sensory-level processing (Kraus and Nicol, 2005). FFRs convey, with high fidelity, acoustic structure related to both speaker identity (e.g., voice pitch) and the spoken message (i.e., formant information) and although these separate acoustic features occur simultaneously in the speech signal, they can be disentangled in brainstem response components. As such, Kraus and colleagues propose that brainstem encoding of speech may be a fundamental precursor to, or at least an influence on, the 'source-filter' (cf. pitch and formant structure) dichotomy found in speech perception.

Our results support the disassociation between pitch and formant structure put forth in this theoretical model. We found that FFR spectral (Figs. 3D and 4A) and temporal (Figs. 3A–B) measures of F0 were relatively unchanged with increasing levels of reverberation in contrast to formant-related harmonic magnitudes, which decreased monotonically with increasing reverberation (Fig. 4B). Complementary results were seen in the behavioral data where first formant, but not voice pitch discrimination, was affected by reverberation (Fig. 6). These results suggest that reverberation has a differential effect on the neural encoding of 'source-filter' components of speech. Similar differential effects on voice F0 and harmonics (i.e., source-filter) have been found in brainstem responses to noise degraded vowels. That is, source related response components (i.e., F0) are relatively immune to additive background noise whereas filter related components (i.e., formant structure/upper harmonics) become degraded (Cunningham et al., 2001; Russo et al., 2004; Wible et al., 2005). The similar differential effect on F0 and formant harmonics we observe here indicates that only the neural representation of 'filter' related components are appreciably degraded with reverberation with far less change in the neural representation of 'source' relevant information.

4. Conclusions

By comparing brainstem responses in musicians and non-musicians we found that musicians show more robust subcortical representation for speech in quiet and in the

presence of reverberation. Musicians have more robust encoding of the fine spectro-temporal aspects of speech which are cue talker identity (i.e., voice pitch) and speaker content (i.e., formant structure). We found that brainstem responses correspond well with perceptual measures of voice pitch and formant discrimination suggesting that that a listener's ability to extract speech in adverse listening conditions may, in part, depend on how well acoustic features are encoded at the level of the auditory brainstem. Reverberation hinders brainstem and perceptual measures of formant information but not voice pitch related properties of speech, F1 and F0 respectively. This differential effect may be one reason, for instance, why reverberation is tolerable (and often desirable) in concert music halls (Backus, 1977; Lifshitz, 1925) where pitch dominates the signal, but not in classrooms (Yang and Bradley, 2009), where target acoustics are geared toward speech intelligibility. These results provide further evidence that long-term musical training positively influences language specific behaviors.

5. Experimental procedures

5.1. Participants

Ten adult English speaking musicians and ten adult English speaking non-musicians were recruited from the Purdue University student body to participate in the experiment. Participation enrollment was based on the following stringent prescreening criteria (Bidelman et al., 2009; Krishnan et al., 2009a, 2010a): 20–30 years of age, right handed, normal hearing by self report, no familiarity with a tone language, no previous history of psychiatric or neurological illnesses, and at least eight years of continuous training on a musical instrument (for musicians) or less than three (for non-musicians). Upon passing these prescreening criteria, a series of questionnaires was used to assess the biographical and music history of all participants. Based on self report, musicians were amateur instrumentalists who had received at least 10 years of continuous instruction on their principal instrument ($\mu \pm \sigma$; 12.0 ± 1.8 years), beginning at or before the age of 10 (7.8 ± 1.6 years). Each had formal private or group lessons within the past 5 years and currently played his/her instrument(s). Non-musicians had no more than 3 years of formal music training (0.85 ± 1.0 years) on any combination of instruments in addition to not having received music instruction within the past 5 years (Table 1). Individuals with 3–10 years of musical training were not recruited to ensure proper split between musician and non-musician groups. There were no significant differences between the musician (4 female and 6 male) and non-musician (6 female and 4 male) groups in gender distribution ($p=0.65$, Fisher's exact test). All were strongly right handed (>81%) as measured by the Edinburgh Handedness inventory (Oldfield, 1971) and exhibited normal hearing sensitivity at octave frequencies between 500 and 4000 Hz. In addition, participants reported no previous history of neurological or psychiatric illnesses. The two groups were also closely matched in age (M: 22.4 ± 2.0 years, NM: 22.0 ± 2.7 years; $t_{18}=0.38$, $p=0.71$) and years of formal education (M: 16.8 ± 2.4 years, NM: 16.2 ± 2.5 years; $t_{18}=0.55$, $p=0.54$). All participants were students enrolled at Purdue University at the time of their participation. All were

Table 1 – Musical background of participants.

Participant	Instrument(s)	Years of music training	Age of onset
Musicians			
M1	Piano	12	8
M2	Saxophone/piano	13	8
M3	Piano/saxophone	11	8
M4	Piano/viola	14	4
M5	Trumpet	12	9
M6	Piano/clarinet	10	8
M7	Piano/guitar	10	9
M8	Piano	10	7
M9	Piano/flute/trumpet	15	7
M10	Saxophone/clarinet	13	10
Mean (SD)		12 (1.8)	7.8 (1.6)
Non-musicians			
NM1	Clarinet	1	12
NM2	Piano	1	9
NM3	Piano	2	10
NM4	Saxophone	3	11
NM5	Trumpet	1	11
NM6	Guitar	0.5	15
NM7	–	0	–
NM8	–	0	–
NM9	–	0	–
NM10	–	0	–
Mean (SD)		0.85 (1.0)	11.3 (2.06) ^a

^a Age of onset statistics for non-musicians was computed from the six participants with minimal musical training.

paid for their time and gave informed consent in compliance with a protocol approved by the Institutional Review Board of Purdue University.

5.2. Stimuli

A synthetic version of the vowel /i/ was generated using the Klatt cascade formant synthesizer (Klatt, 1980; Klatt and Klatt, 1990) as implemented in Praat (Boersma and Weenink, 2009). The fundamental frequency (F0) contour of the vowel was modeled after a rising time-varying linguistic F0 contour, traversing from 103 Hz to 130 Hz over the 250 ms duration of the stimulus (for further details see Bidelman et al., 2009; Krishnan et al., 2009b). Time-varying F0 was used because reverberation has a larger effect on dynamic rather than steady-state properties of complex sound (Sayles and Winter, 2008). Vowel formant frequencies were steady-state and held constant (in Hz): F1=300; F2=2500; F3=3500; F4=4530 Hz. The addition of reverberation was accomplished by time domain convolution of the original vowel with room impulse responses recorded in a corridor at a distance of either 0.63 m (mild reverb; Reverberation Time (RT₆₀)≈0.7 s), 1.25 m (medium reverb; RT₆₀≈0.8 s), or 5 m (severe reverb; RT₆₀≈0.9 s) from the sound source (Sayles and Winter, 2008; Watkins, 2005). In addition to the three reverberant conditions a “dry” vowel was used in which the original stimulus was unaltered. Following the convolution, the “tails” introduced by the reverberation were removed, the waveforms time normalized to 250 ms, and amplitude normalized to an RMS of 80 dB SPL. Normalization ensured that all stimuli, regardless of reverberation condition, were presented at equal intensities thus ruling out overall amplitude as a potential factor (Sayles and Winter, 2008).

For details of the room impulse responses and their recording, the reader is referred to Watkins (2005). Important features of the reverberation stimuli are presented in Fig. 1; (A–D) spectrograms and 70 ms portion of the time waveforms extracted from 80 to 150 ms and (E) FFTs computed over the duration of the response. Notice the more pronounced spectral smearing (induced by blurring in the temporal fine-structure of the waveform) and broadening of higher frequency components in the spectrograms and FFTs with increasing reverberation strength (see also Fig. S1). Conversely, note the relative resilience of the fundamental frequency (F0: 103–130 Hz) to these negative effects.

5.3. FFR data acquisition

FFR recording protocol was similar to previous reports from our laboratory (Bidelman and Krishnan, 2009; Krishnan et al., 2009a). Participants reclined comfortably in an acoustically and electrically shielded booth to facilitate recording of brainstem FFRs. They were instructed to relax and refrain from extraneous body movement to minimize myogenic artifacts. Subjects were allowed to sleep through the duration of the FFR experiment. FFRs were recorded from each participant in response to monaural stimulation of the right ear at an intensity of 80 dB SPL through a magnetically shielded insert earphone (Etymotic ER-3A). Each stimulus was presented using rarefaction polarity and a repetition rate of 2.76/s. The presentation order of the stimuli was randomized both within and across participants. Control of the experimental protocol was accomplished by a signal generation and data acquisition system (Tucker-Davis Technologies, System III) using a sampling rate of 24,414 Hz.

FFRs were recorded differentially between a non-inverting (+) electrode placed on the midline of the forehead at the hairline (Fz) and an inverting (–) reference electrode placed on the right mastoid (A2). Another electrode placed on the mid-forehead (Fpz) served as the common ground. Inter-electrode impedances were maintained at or below 1 kΩ. The EEG inputs were amplified by 200,000 and band-pass filtered from 70 to 3000 Hz (6 dB/octave roll-off). Sweeps containing activity exceeding ±30 μV were rejected as artifacts. In total, each response waveform represents the average of 3000 artifact free trials over a 280 ms acquisition window.

5.4. FFR data analysis

5.4.1. Temporal analysis

Short-term autocorrelation functions (ACFs) (i.e., running autocorrelograms) were computed from individual FFRs to index variation in neural periodicities over the duration of the responses. The autocorrelogram (ACG) represents the short term autocorrelation function of windowed frames of a compound signal, i.e., $ACG(\tau, t) = X(t) \times X(t-\tau)$ for each time t and time-lag τ . It is a three dimensional plot quantifying the variations in periodicity and “neural pitch strength” (i.e., degree of phase-locking) as a function of time. The horizontal axis represents the time at which single ACF “slices” are computed while the vertical axis represents their corresponding time-lags, i.e., periods. The intensity of each point in the image represents the instantaneous ACF magnitude computed at a given time within the response. Mathematically, the running autocorrelogram is the time domain analog to the frequency-domain

spectrogram. In terms of neurophysiology, it represents the running distribution of all-order interspike intervals present in the population neural activity (Cariani and Delgutte, 1996a; Sayles and Winter, 2008).

5.4.2. Spectral analysis

Fundamental frequency (F0) information provides pitch cues used for speaker identification (e.g., male vs. female talker) while harmonic structure and spectral envelope provide formant cues (e.g., vowel quality) used to decipher what is actually being said (Assmann and Summerfield, 1989, 1990; Palmer, 1990). Importantly, formant structure and voice fundamental frequency can be considered independent features in speech acoustics. Therefore, we chose to treat F0 and F1 related harmonic encoding of the FFR as distinct elements and analyze them separately.

The spectral composition of each FFR response waveform was quantified by measuring the magnitudes of the first four harmonics (H1=F0, H2, H3, and H4=F1 related harmonics). Individual frequency spectra were computed over the duration of each FFR by taking the FFT (5 Hz resolution) of a time-windowed version of the temporal waveform (Gaussian window). For each subject and condition, the magnitudes of F0 (i.e., H1) and its harmonics (H2–H4) were measured as the peaks in the FFT, relative to the noise floor, which fell in the same frequency range as those of the input stimulus—F0: 100–135 Hz, H2: 200–270 Hz, H3: 300–400 Hz, H4: 405–540 Hz (see stimulus Fig. 1). For the purpose of subsequent discussion and analysis, strength of pitch encoding is defined as the magnitude of F0 and the strength of formant-related encoding as the mean spectral magnitude of harmonics H2–H4 (cf. Banai et al., 2009; Strait et al., 2009). Low-order harmonics (i.e., H2–H4) also contribute to pitch as do all harmonically related spectral components. However, H2–H4 are proximal to the first formant in our stimuli (Fig. 1), and therefore, also provide cues for the quality of the vowel given their contribution to the formant structure. While we realize that this dichotomy is an overly simplistic view of speech processing, perceptual (Assmann and Summerfield, 1989, 1990) and electrophysiological (Kraus and Nicol, 2005) evidence suggest that these two parameters provide adequate information for speaker segregation (i.e., identifying who is talking) and verbal identification (i.e., determining what is being said), respectively. In addition to measuring harmonic peaks, narrowband spectrograms were computed from FFRs per condition to assess how spectral content changes with time. All data analyses were performed using custom routines coded in MATLAB® 7.9 (The MathWorks, Inc., Natick, MA).

5.4.3. Dry-to-reverberant response correlations

To measure the effects of reverberation on FFR morphology, the degree of correlation between reverberant and dry responses was computed for each participant. Correlations coefficients were computed by first temporally aligning the reverberant and dry response waveforms by maximizing their cross-correlation (e.g., Galbraith and Brown, 1990; Parbery-Clark et al., 2009a). Only the steady-state portions of the FFRs were considered in this analysis (15–250 ms). The dry-to-reverberant correlation was then computed as the correlation (Pearson's r) between the two response waveforms. Three such correlations were performed for each participant (i.e.,

dry-to-mild, dry-to-medium, and dry-to-severe). Higher values of this metric indicate greater correspondence between neural responses in the absence and presence of reverberation.

5.5. Behavioral voice pitch and formant discrimination task

Four musicians and four non-musicians who also took part in the FFR experiment participated in the perceptual task. Behavioral fundamental (F0 DL) and formant frequency difference limens (F1 DL) were estimated for each reverberation condition using a measure of individual frequency discrimination thresholds (Bernstein and Oxenham, 2003; Carlyon and Shackleton, 1994; Hall and Plack, 2009; Houtsma and Smurzynski, 1990; Krishnan et al., 2010). F0 DLs and F1 DLs were obtained in separate test sessions. For each type of DL, testing consisted of four blocks (one for each reverb condition; dry, mild, medium, and severe) in which participants performed a three-alternative forced choice task (3-AFC). For a given trial within a condition, they heard three sequential intervals, two containing a reference stimulus and one containing a comparison, assigned randomly. For F0 DL measurements, the reference /i/ vowel had fixed formant frequencies (F1=300; F2=2500; F3=3500; F4=4530 Hz) and a F0 frequency of 115 Hz. The comparison stimulus was exactly the same only F0 was varied adaptively based on the subject's response. In contrast, when measuring F1 DLs, the reference and comparison /i/ vowels contained the same F0 frequency (115 Hz), as well as second, third, and fourth formants (F2=2500; F3=3500; F4=4530 Hz). Only F1 differed between the standard and comparison. For both types of DLs the procedure was identical. The subject's task was to identify the interval in which the vowel sounded different. Following a brief training run, discrimination thresholds were measured using a two-down, one-up adaptive paradigm which tracks the 71% correct point on the psychometric function (Levitt, 1971). Following two consecutive correct responses, the frequency difference was decreased for the subsequent trial, and increased following a single incorrect response. Frequency difference between reference and comparison intervals was varied using a geometric step size of $\sqrt{2}$ between response reversals. For each reverberation condition, 16 reversals were measured and the geometric mean of the last 12 taken as the individual's frequency difference limen, that is, the minimum frequency difference needed to detect a change in the vowel's voice pitch (for F0 DL) or first formant (for F1 DL), respectively.

Supplementary materials related to this article can be found online at [doi:10.1016/j.brainres.2010.07.100](https://doi.org/10.1016/j.brainres.2010.07.100).

Acknowledgments

This research was supported by NIH R01 DC008549 (A.K.) and T32 DC 00030 NIDCD predoctoral traineeship (G.B.). The authors wish to thank Dr. Tony J. Watkins (University of Reading) for providing the room impulse responses used to create the reverberant stimuli and Dr. Chris Plack (University of Manchester) for supplying the MATLAB code for the behavioral experiments.

REFERENCES

- Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clin. Neurophysiol.* 119, 922–933.
- Anvari, S.H., Trainor, L.J., Woodside, J., Levy, B.A., 2002. Relations among musical skills, phonological processing and early reading ability in preschool children. *J. Exp. Child Psychol.* 83, 111–130.
- Assmann, P.F., Summerfield, Q., 1989. Modeling the perception of concurrent vowels: vowels with the same fundamental frequency. *J. Acoust. Soc. Am.* 85, 327–338.
- Assmann, P.F., Summerfield, Q., 1990. Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J. Acoust. Soc. Am.* 88, 680–697.
- Backus, J., 1977. *The Acoustical Foundations of Music*. Norton, New York.
- Bajo, V.M., Nodal, F.R., Moore, D.R., King, A.J., 2010. The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260.
- Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., Kraus, N., 2009. Reading and subcortical auditory function. *Cereb. Cortex* 19, 2699–2707.
- Bernstein, J.G., Oxenham, A.J., 2003. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* 113, 3323–3334.
- Bidelman, G.M., Krishnan, A., 2009. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. *J. Neurosci.* 29, 13165–13171.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2009. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* doi:10.1162/jocn.2009.21362.
- Boersma P, Weenink D, 2009. Praat: doing phonetics by computer (Version 5.1.17) [Computer program]. Retrieved September 1, 2009, from <http://www.praat.org>.
- Brashears, S.M., Morlet, T.G., Berlin, C.I., Hood, L.J., 2003. Olivocochlear efferent suppression in classical musicians. *J. Am. Acad. Audiol.* 14, 314–324.
- Brattico, E., Pallesen, K.J., Varyagina, O., Bailey, C., Anourova, I., Jarvenpaa, M., Eerola, T., Tervaniemi, M., 2009. Neural discrimination of nonprototypical chords in music experts and laymen: an MEG study. *J. Cogn. Neurosci.* 21, 2230–2244.
- Braun, M., 1999. Auditory midbrain laminar structure appears adapted to f0 extraction: further evidence and implications of the double critical bandwidth. *Hear. Res.* 129, 71–82.
- Braun, M., 2000. Inferior colliculus as candidate for pitch extraction: multiple support from statistics of bilateral spontaneous otoacoustic emissions. *Hear. Res.* 145, 130–140.
- 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.
- Cariani, P.A., Delgutte, B., 1996a. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716.
- Cariani, P.A., Delgutte, B., 1996b. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J. Neurophysiol.* 76, 1717–1734.
- Carlyon, R.P., Shackleton, T.M., 1994. Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms. *J. Acoust. Soc. Am.* 95, 3541–3554.
- Chan, A.S., Ho, Y.C., Cheung, M.C., 1998. Music training improves verbal memory. *Nature* 396, 128.
- Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.
- Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2009. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain Lang.* 108, 1–9.
- Chartrand, J.P., Belin, P., 2006. Superior voice timbre processing in musicians. *Neurosci. Lett.* 405, 164–167.
- Crummer, G.C., Walton, J.P., Wayman, J.W., Hantz, E.C., Frisina, R.D., 1994. Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *J. Acoust. Soc. Am.* 95, 2720–2727.
- Culling, J.F., Hodder, K.I., Toh, C.Y., 2003. Effects of reverberation on perceptual segregation of competing voices. *J. Acoust. Soc. Am.* 114, 2871–2876.
- Cunningham, J., Nicol, T., Zecker, S.G., 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.
- Darwin, C.J., Hukin, R.W., 2000. Effects of reverberation on spatial, prosodic, and vocal-tract size cues to selective attention. *J. Acoust. Soc. Am.* 108, 335–342.
- Don, M., Eggermont, J.J., 1978. Analysis of the click-evoked brainstem potentials in man using high-pass noise masking. *J. Acoust. Soc. Am.* 63, 1084–1092.
- Drgas, S., Blaszak, M.A., 2009. Perceptual consequences of changes in vocoded speech parameters in various reverberation conditions. *J. Speech Lang. Hear. Res.* 52, 945–955.
- Emiroglu, S., Kollmeier, B., 2008. Timbre discrimination in normal-hearing and hearing-impaired listeners under different noise conditions. *Brain Res.* 1220, 199–207.
- Fant, G., 1960. *Acoustic Theory of Speech Production*. Mouton, The Hague.
- Franklin, M.S., Sledge Moore, K., Yip, C.-Y., Jonides, J., Rattray, K., Moher, J., 2008. The effects of musical training on verbal memory. *Psychol. Music.* 36, 353–365.
- Galbraith, G.C., Brown, W.S., 1990. Cross-correlation and latency compensation analysis of click-evoked and frequency-following brain-stem responses in man. *Electroencephalogr. Clin. Neurophysiol.* 77, 295–308.
- Galbraith, G., Threadgill, M., 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.
- Gao, E., Suga, N., 1998. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. *Proc. Natl Acad. Sci. USA* 95, 12663–12670.
- Gelfand, S.A., Silman, S., 1979. Effects of small room reverberation upon the recognition of some consonant features. *J. Acoust. Soc. Am.* 66, 22–29.
- George, E.L., Festen, J.M., Houtgast, T., 2008. The combined effects of reverberation and nonstationary noise on sentence intelligibility. *J. Acoust. Soc. Am.* 124, 1269–1277.
- Giraud, A.L., Garnier, S., Micheyl, C., Lina, G., Chays, A., Chery-Croze, S., 1997. Auditory efferents involved in speech-in-noise intelligibility. *NeuroReport* 8, 1779–1783.
- Guinan Jr., J.J., 2006. Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear.* 27, 589–607.
- Hall, D.A., Plack, C.J., 2009. Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576–585.
- Houtsma, A., Smurzynski, J., 1990. Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304–310.
- Jones, M.R., Yee, W., 1997. Sensitivity to time change: the role of context and skill. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 693–709.

- Kishon-Rabin, L., Amir, O., Vexler, Y., Zaltz, Y., 2001. Pitch discrimination: are professional musicians better than non-musicians? *J. Basic Clin. Physiol. Pharmacol.* 12, 125–143.
- Klatt, D.H., 1980. Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am.* 67, 971–995.
- Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820–857.
- Koelsch, S., Schroger, E., Tervaniemi, M., 1999. Superior 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. *Ann. NY Acad. Sci.* 1169, 543–557.
- Krishnan, A., 2007. Human frequency following response. In: Burkard, R.F., Don, M., Eggermont, J.J. (Eds.), *Auditory Evoked Potentials: Basic Principles and Clinical Application*. Lippincott Williams & Wilkins, Baltimore, pp. 313–335.
- Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain Lang.* 110, 135–148.
- Krishnan, A., Plack, C.J., 2009. Auditory brainstem correlates of basilar membrane nonlinearity in humans. *Audiol. Neurootol.* 14, 88–97.
- Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res. Cogn. Brain Res.* 25, 161–168.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009a. Experience-dependent neural representation of dynamic pitch in the brainstem. *NeuroReport* 20, 408–413.
- Krishnan, A., Swaminathan, J., Gandour, J.T., 2009b. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.* 21, 1092–1105.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010a. Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Res.* 1313, 124–133.
- Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010b. The effects of tone language experience on pitch processing in the brainstem. *J. Neuroling.* 23, 81–95.
- Krishnan, A., Bidelman, G.M., Gandour, J.T., 2010. Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hear. Res.* 268, 60–66.
- Langner, G., 1981. Neuronal mechanisms for pitch analysis in the time domain. *Exp. Brain Res.* 44, 450–454.
- Langner, G., 1997. Neural processing and representation of periodicity pitch. *Acta Otolaryngol. Suppl.* 532, 68–76.
- Lee, C.Y., Hung, T.H., 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. *J. Acoust. Soc. Am.* 124, 3235–3248.
- Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840.
- Levitt, H., 1971. Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477.
- Lifshitz, S., 1925. Optimum reverberation for an auditorium. *Phys. Rev.* 25, 391–394.
- Liu, C., Kewley-Port, D., 2004. Formant discrimination in noise for isolated vowels. *J. Acoust. Soc. Am.* 116, 3119–3129.
- Luo, F., Wang, Q., Kashani, A., Yan, J., 2008. Corticofugal modulation of initial sound processing in the brain. *J. Neurosci.* 28, 11615–11621.
- Macdonald, E.N., Pichora-Fuller, M.K., Schneider, B.A., 2010. Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. *Hear. Res.* 261, 63–66.
- Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820.
- Micheyl, C., Collet, L., 1996. Involvement of the olivocochlear bundle in the detection of tones in noise. *J. Acoust. Soc. Am.* 99, 1604–1610.
- Micheyl, C., Carbonnel, O., Collet, L., 1995. Medial olivocochlear system and loudness adaptation: differences between musicians and non-musicians. *Brain Cogn.* 29, 127–136.
- Micheyl, C., Khalifa, S., Perrot, X., Collet, L., 1997. Difference in cochlear efferent activity between musicians and non-musicians. *NeuroReport* 8, 1047–1050.
- Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J., 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47.
- Moreno, S., Besson, M., 2005. Influence of musical training on pitch processing: event-related brain potential studies of adults and children. *Ann. NY Acad. Sci.* 1060, 93–97.
- Munte, T.F., Kohlmetz, C., Nager, W., Altenmuller, E., 2001. Neuroperception. Superior auditory spatial tuning in conductors. *Nature* 409, 580.
- Musacchia, G., Strait, D., Kraus, N., 2008. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* 241, 34–42.
- Nabelek, A.K., 1988. Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. *J. Acoust. Soc. Am.* 84, 476–484.
- Nabelek, A.K., Dagenais, P.A., 1986. Vowel errors in noise and in reverberation by hearing-impaired listeners. *J. Acoust. Soc. Am.* 80, 741–748.
- Nabelek, A.K., Letowski, T.R., 1985. Vowel confusions of hearing-impaired listeners under reverberant and nonreverberant conditions. *J. Speech Hear. Disord.* 50, 126–131.
- Nabelek, A.K., Letowski, T.R., 1988. Similarities of vowels in nonreverberant and reverberant fields. *J. Acoust. Soc. Am.* 83, 1891–1899.
- Nabelek, A.K., Robinson, P.K., 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. *J. Acoust. Soc. Am.* 71, 1242–1248.
- Nabelek, A.K., Letowski, T.R., Tucker, F.M., 1989. Reverberant overlap- and self-masking in consonant identification. *J. Acoust. Soc. Am.* 86, 1259–1265.
- Nager, W., Kohlmetz, C., Altenmuller, E., Rodriguez-Fornells, A., Munte, T.F., 2003. The fate of sounds in conductors' brains: an ERP study. *Brain Res. Cogn. Brain Res.* 17, 83–93.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Oxenham, A.J., Fligor, B.J., Mason, C.R., Kidd Jr., G., 2003. Informational masking and musical training. *J. Acoust. Soc. Am.* 114, 1543–1549.
- Palmer, A.R., 1990. The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers. *J. Acoust. Soc. Am.* 88, 1412–1426.
- Pantev, C., Roberts, L.E., Schulz, M., Engelen, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport* 12, 169–174.
- Parbery-Clark, A., Skoe, E., Kraus, N., 2009a. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* 29, 14100–14107.
- Parbery-Clark, A., Skoe, E., Lam, C., Kraus, N., 2009b. Musician enhancement for speech-in-noise. *Ear Hear.* 30, 653–661.
- Patel, A.D., 2008. *Music, Language, and the Brain*. Oxford University Press, NY.
- Perrot, X., Micheyl, C., Khalifa, S., Collet, L., 1999. Stronger bilateral efferent influences on cochlear biomechanical activity in musicians than in non-musicians. *Neurosci. Lett.* 262, 167–170.

- Perrot, X., Ryvlin, P., Isnard, J., Guenot, M., Catenoix, H., Fischer, C., Manguiere, F., Collet, L., 2006. Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb. Cortex* 16, 941–948.
- Pitt, M.A., 1994. Perception of pitch and timbre by musically trained and untrained listeners. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 976–986.
- Plomp, R., 1976. Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acustica* 34, 200–211.
- Qin, M.K., Oxenham, A.J., 2005. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear.* 26, 451–460.
- Rammesayer, T., Altenmüller, E., 2006. Temporal information processing in musicians and nonmusicians. *Music Percept.* 24, 37–48.
- Russeler, J., Altenmüller, E., Nager, W., Kohlmetz, C., Munte, T.F., 2001. Event-related brain potentials to sound omissions differ in musicians and non-musicians. *Neurosci. Lett.* 308, 33–36.
- Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030.
- Sayles, M., Winter, I.M., 2008. Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58, 789–801.
- Schon, D., Magne, C., Besson, M., 2004. The music of speech: music training facilitates pitch processing in both music and language. *Psychophysiology* 41, 341–349.
- Schreiner, C.E., Langner, G., 1997. Laminar fine structure of frequency organization in auditory midbrain. *Nature* 388, 383–386.
- Slevc, R.L., Miyake, A., 2006. Individual differences in second-language proficiency: does musical ability matter? *Psychol. Sci.* 17, 675–681.
- Slevc, L.R., Rosenberg, J.C., Patel, A.D., 2009. Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon. Bull. Rev.* 16, 374–381.
- Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalogr. Clin. Neurophysiol.* 39, 465–472.
- Sohmer, H., Pratt, H., 1977. Identification and separation of acoustic frequency following responses (FFRs) in man. *Electroencephalogr. Clin. Neurophysiol.* 42 (4), 493–500.
- Spiegel, M.F., Watson, C.S., 1984. Performance on frequency-discrimination tasks by musicians and nonmusicians. *J. Acoust. Soc. Am.* 76, 1690–1695.
- Stasiak, A., Winter, I., Sayles, M., 2010. The effect of reverberation on the representation of single vowels, double vowels and consonant–vowel syllables by single units in the ventral cochlear nucleus. (Eds.). Poster Session Conducted at the 33rd MidWinter Meeting of the Association for Research in Otolaryngology. Abstract 231, retrieved from <http://www.aro.org/abstracts/abstracts.html>.
- Strait, D.L., Kraus, N., Skoe, E., Ashley, R., 2009. Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* 29, 661–668.
- Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear. Res.* 261, 22–29.
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., Schroger, E., 2005. Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp. Brain Res.* 161, 1–10.
- Tzounopoulos, T., Kraus, N., 2009. Learning to encode timing: mechanisms of plasticity in the auditory brainstem. *Neuron* 62, 463–469.
- van Zuijlen, T.L., Sussman, E., Winkler, I., Naatanen, R., Tervaniemi, M., 2004. Grouping of sequential sounds—an event-related potential study comparing musicians and nonmusicians. *J. Cogn. Neurosci.* 16, 331–338.
- Wang, D.L., Brown, G.J., 2006. *Computational Auditory Scene Analysis*. Wiley, Hoboken, NJ.
- Watkins, A.J., 2005. Perceptual compensation for effects of reverberation in speech identification. *J. Acoust. Soc. Am.* 118, 249–262.
- Wible, B., Nicol, T., Kraus, N., 2005. Encoding of complex sounds in an animal model: implications for understanding speech perception in humans. In: Heil, K.R.P., Budinger, E., Scheich, H. (Eds.), *Auditory Cortex: Towards a Synthesis of Human and Animal Research*. Lawrence Erlbaum Associates, Oxford, pp. 241–254.
- Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Xiong, Y., Zhang, Y., Yan, J., 2009. The neurobiology of sound-specific auditory plasticity: a core neural circuit. *Neurosci. Biobehav. Rev.* 33, 1178–1184.
- Yan, J., Zhang, Y., Ehret, G., 2005. Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *J. Neurophysiol.* 93, 71–83.
- Yang, W., Bradley, J.S., 2009. Effects of room acoustics on the intelligibility of speech in classrooms for young children. *J. Acoust. Soc. Am.* 125, 922–933.
- Zendel, B.R., Alain, C., 2009. Concurrent sound segregation is enhanced in musicians. *J. Cogn. Neurosci.* 21, 1488–1498.