Musicians have fine-tuned neural distinction of speech syllables

Alexandra Parbery-Clark[1,2], **Adam Tierney**[1,2], **Dana L. Strait**[1,3], and **Nina Kraus**[1,2,3,4,5,*]

[1]Auditory Neuroscience Laboratory, Northwestern University, Evanston, IL, USA

[2]Communication Sciences, Northwestern University, Evanston, IL, USA

[3]Institute for Neuroscience, Northwestern University, Evanston, IL, USA

[4]Neurobiology and Physiology, Northwestern University, Evanston, IL, USA

[5]Otolaryngology, Northwestern University, Evanston, IL, USA

Abstract

One of the benefits musicians derive from their training is an increased ability to detect small differences between sounds. Here, we asked whether musicians' experience discriminating sounds on the basis of small acoustic differences confers advantages in the subcortical differentiation of closely-related speech sounds (e.g., /ba/ and /ga/), distinguishable only by their harmonic spectra (i.e., their second formant trajectories). Although the second formant is particularly important for distinguishing stop consonants, auditory brainstem neurons do not phase-lock to its frequency range (above 1000 Hz). Instead, brainstem nuclei convert this high-frequency content into neural response timing differences. As such, speech tokens with higher formant frequencies elicit earlier brainstem responses than those with lower formant frequencies. By measuring the degree to which subcortical response timing differs to the speech syllables /ba/, /da/, and /ga/ in adult musicians and nonmusicians, we reveal that musicians demonstrate enhanced subcortical discrimination of closely related speech sounds. Furthermore, the extent of subcortical consonant discrimination correlates with speech-in-noise perception. Taken together, these findings show a musician enhancement for the neural processing of speech and reveal a biological mechanism contributing to musicians' enhanced speech perception.

Keywords

auditory; brainstem; musical training; speech discrimination; speech in noise

Introduction

Musicians develop a number of auditory skills, including the ability to track a single instrument embedded within a multitude of sounds. This skill relies on the perceptual separation of concurrent sounds that can overlap in pitch but differ in timbre. One contributor to timbre is a sound's spectral fine structure: the amplitudes of different harmonics and how they change over time (Krimpoff et al. 1994, Caclin et al. 2005, Caclin

*Correspondence: Nina Kraus, Ph.D., Auditory Neuroscience Laboratory, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA, Ph: 847-491-3181, nkraus@northwestern.edu.
APC and AT are co-first authors and contributed equally to this work.

et al. 2006, Kong et al. 2011). Consequently, it is not surprising that musicians, compared to nonmusicians, demonstrate greater perceptual acuity of rapid spectro-temporal changes (Gaab et al. 2005), harmonic differences (Musacchia et al. 2008; Zendel and Alain 2009), as well as greater neural representation of harmonics (Koelsch et al. 1999, Shahin et al. 2005, Musacchia et al. 2008, Zendel and Alain 2009, Strait et al. 2009, Parbery-Clark et al. 2009a, Lee et al. 2009).

Harmonic amplitudes are one of the most important sources of information distinguishing speech sounds. During speech production, individuals modify the filter characteristics of the vocal apparatus to enhance or attenuate specific frequencies, creating spectral peaks (i.e., formants) that help distinguish vowels and consonants. Given the similarity of the acoustic cues that characterize speech and music–that is, differences in the distribution of energy across the harmonic spectrum–musicians' extensive experience distinguishing musical sounds may provide advantages for processing speech (Tallal and Gaab 2006, Parbery-Clark et al. 2009a, Lee et al. 2009, Kraus and Chandrasekaran 2010, Besson et al. 2011, Patel 2011, Marie et al. 2011a, Marie et al. 2011b, Chobert et al. 2011, Strait and Kraus 2011, Shahin, 2011). This hypothesis is supported by evidence that musicians demonstrate enhanced auditory function throughout the auditory pathway. Specifically, musicians have enhanced neural responses to music (Fujioka et al. 2004, Shahin et al. 2005, Musacchia et al. 2007, Lee et al. 2009, Bidelman, 2010, Bidelman et al. 2011a, 2011b) as well as speech (Musacchia et al. 2007, Parbery-Clark et al. 2009, Bidelman et al 2009), including changes in pitch, duration, intensity and voice onset time (Schön et al. 2004, Magne et al. 2006, Moreno and Besson 2006, Marques et al. 2007, Tervaniemi et al. 2009, Chobert et al. 2011, Besson et al. 2011, Marie et al. 2011a, Marie et al. 2011b).

The auditory brainstem response (ABR) is an evoked response of central brainstem origin (Galbraith et al. 2000, Chandrasekaran and Kraus 2010, Warrier et al. 2011) that faithfully represents spectral components of incoming sounds up to ~1000 Hz through neural phase-locking to stimulus periodicity (Marsh 1974, Liu et al. 2006). The frequency range of the second formant, although particularly important for distinguishing consonants, falls above this phase-locking range for many speech tokens. The tonotopic organization of the auditory system overcomes this, as higher frequencies represented at the base of the cochlea result in earlier neural responses relative to lower frequencies (Gorga et al. 1988). Consequently, consonants with higher frequency second formants elicit earlier ABRs than sounds with lower frequency second formants (Johnson et al. 2008).

The second formant is the most difficult acoustic feature to detect when speech is presented in noise (Miller and Nicely, 1955). Accordingly, the robustness of its neural encoding contributes to the perception of speech in noise. Given that musicians outperform nonmusicians on SIN perception (Parbery-Clark et al. 2009b, Zendel and Alain 2011), we asked whether musicians demonstrate greater subcortical differentiation of speech syllables that are distinguishable only by their second formant trajectories. To this aim, we assessed ABRs to the speech syllables /da/, /ba/, and /ga/ and SIN discrimination in adult musicians and nonmusicians. Timing differences in ABRs can be objectively measured by calculating response phase differences (Skoe et al. 2011, Tierney et al. 2011) and are present in both scalp-recorded responses and near-field responses recorded from within the inferior colliculus (Warrier et al. 2011), suggesting that they reflect a fundamental characteristic of temporal processing. We calculated the phase differences between subcortical responses to these syllables. We expected musicians to demonstrate greater phase shifts between speech-evoked ABRs than nonmusicians, indicating more distinct neural representations of each sound. Given the importance of accurate neural representation for SIN perception, we expected the degree of phase shift to relate to SIN perception.

Materials and Methods

Fifty young adults (ages 18–32, mean=22.0, SD=3.54) participated in this study. Twenty-three musicians (15 female) all started musical training by the age of 7 (mean=5.14, SD=1.03) and had consistently practiced a minimum of three times a week (mean=15.9 years, SD=4.0). Twenty-seven (15 female) had received < 3 years of music training and were categorized as nonmusicians. Nineteen of these nonmusicians subjects had received no musical training at any point in their lives. Eight had less than three years (mean=2.1, SD=0.8) of musical training which started after the age of 9 (range 9–16, mean=11.28, SD=2.6). See Table 1 for musical practice histories. All musicians reported that their training included experience playing in ensembles. All participants were right-handed, had pure tone air conduction thresholds = 20 dB HL from 0.125–8 kHz and had normal click-evoked ABRs. The two groups did not differ in age ($F_{(1,49)}$=0.253, p=0.617), sex (X^2 $_{(1,N = 50)}$=0.483, p=0.569), hearing ($F_{(1,35)}$=0.902, p=0.565) or nonverbal IQ ($F_{(1, 49)}$=2.503, p=0.120) (Test of Nonverbal Intelligence; Brown et al. 1997).

Behavioral testing

Speech-in-noise perception was assessed using the Quick Speech-in-Noise Test (QuickSIN, Etymotic Research), a nonadaptive test that presents a target sentence embedded in four-talker babble through insert earphones (ER-2, Etymotic Research). Participants were tested on four six-sentence lists, with each sentence containing five key words. For each list, initial target sentences are presented at a signal-to-noise ratio (SNR) of +25 dB and five subsequent sentences were presented with a progressive 5 dB reduction in SNR per sentence, ending at 0 dB SNR. Recall of the sentences occurred immediately after each sentence was heard. The length of the sentences was 8.5 ± 1.3 words. These sentences included (target words underlined): "The square peg will settle in the round hole." and "The sense of smell is better than that of touch." The number of total target words correctly recalled was subtracted from 25.5 to calculate an SNR loss (Killion et al. 2004). Performance across all four lists was averaged; lower scores indicated better performance.

Stimuli

The three speech syllables /ga/, /da/, and /ba/ were constructed using a Klatt-based synthesizer. Each syllable is 170 ms in duration with an unchanging fundamental frequency (F_0=100 Hz). For the first 50 ms of all three syllables, consisting of the transition between the consonant stop burst and the vowel, the first and third harmonics change over time (F_1=400–720 Hz; F_3=2580–2500 Hz) whereas the fourth, fifth and sixth harmonics remain steady (F_4=3300 Hz; F_5=3750 Hz; F_6=4900 Hz). The syllables are distinguished by the trajectory of their second formants: /ga/ and /da/ fall (from 2480–1240 Hz and 1700–1240 Hz, respectively) while /ba/ rises (from 900–1240 Hz) (Fig.1 left; Fig 2). The three sounds are identical for the duration of the vowel /a/ (50–170ms). The three syllables /ga/, /da/, and /ba/ were presented pseudo-randomly within the context of five other syllables with a probability of occurrence of 12.5%. The other five speech sounds were also generated using a Klatt-based synthesizer and differed by formant structure (/du/), voice-onset time (/ta/), F_0 (/da/ with a dipping contour, /da/ with an F_0 of 250 Hz), and duration (163 ms /da/). Syllables were presented in a single block with an interstimulus interval of 83ms. The recording session lasted 35±2 min. Because we were interested in quantifying the effects of neural discrimination of speech sounds differing only in formant structure, only responses to /ga/, /da/, and /ba/ are assessed here. See Chandrasekaran et al. (2009b) for further descriptions of these other syllables.

Procedure

ABRs were differentially recorded at a 20 kHz sampling rate using Ag-AgCl electrodes in a vertical montage (Cz active, FPz ground and linked-earlobe references) in Neuroscan Acquire 4.3 (Compumedics, Inc., Charlotte, NC). Contact impedance was 2 kΩ or less across all electrodes. Stimuli were presented binaurally at 80 dB SPL with an 83-ms inter-stimulus interval (Scan 2, Compumedics, Inc.) through insert earphones (ER-3, Etymotic Research, Inc., Elk Grove Village, IL). The speech syllables were presented in alternating polarities, a technique commonly used in brainstem recordings to minimize the contribution of stimulus artifact and cochlear microphonic. Since the stimulus artifact and cochlear microphonic follow the phase of the stimulus, when the responses to alternating polarities are added together the artifacts are reduced, leaving the phase-invariant component of the response intact (see Skoe and Kraus 2010 for more details). To change a stimulus from one polarity to another, the stimulus waveform was inverted by 180 degrees. During the recording, subjects watched a silent, captioned movie of their choice to facilitate a restful state. Seven hundred artifact-free trials were collected to each stimulus.

ABRs were bandpass filtered offline from 70–2000 Hz (12 dB/octave roll-off) to maximize auditory brainstem contributions to the signal and to reduce the inclusion of low-frequency cortical activity (Akhoun et al. 2008). Trials with amplitudes exceeding ±35 μV were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period (−40 to 0 ms). All data processing was conducted using scripts generated in Matlab 2007b (The Mathworks, Natick, 7.5.0).

Phase Analysis

The cross-phaseogram was constructed according to Skoe et al. (2011). First, we divided the response into overlapping 40-ms windows starting at 20-ms before the onset of the stimulus, with each window separated from the next by 1-ms. The center time point of each window ranged from 0 to 170 ms, with a total of 170 windows analyzed. These windows were baseline-corrected using Matlab's detrend function and ramped using a Hanning window. Within each 40-ms window, we then applied the Matlab cross-power spectral density function to each pair of responses--for example, the response to /da/ and the response to /ba/--and converted the resulting spectral power estimates to phase angles.

Because the phase shifts between pairs of responses to consonants reported in Skoe et al. (2011) were largest from 400 to 720 Hz, we restricted our analyses to this frequency region. The cross-phaseogram was split into two time regions according to the acoustic characteristics of the stimuli: a time region corresponding to the dynamic formant transition (5–45 ms) and another corresponding to the sustained vowel (45–170 ms). Repeated-measures ANOVAs were run on the two time regions with group as the fixed factor (musician/nonmusician) and speech syllable pairing as between-subject factor (/ga/-/ba/, /da/-/ba/, and /ga/-/da/). Post-hoc ANOVAs compared the amount of phase shift in musicians and nonmusicians for each condition and Pearson correlations were conducted to explore the relationship between the amount of phase shift and QuickSIN performance. All statistics were conducted in Matlab 2007b; results reported here reflect two-tailed significance values.

Results

Musicians demonstrated greater neural differentiation of the three speech sounds, with musicians' average phase shifts during the transition period across conditions being larger than the average shift for nonmusicians (main effect of group, $F(1,48)=7.58$, $p=0.008$). Musicians demonstrated a greater phase shift than nonmusicians between 400–720 Hz and 5–45 ms for /ga/ versus /ba/ ($F(1,49)=12.282$, $p=0.001$) and for /da/ versus /ba/ (F(1,

49)=5.69, p=0.02) (Figs. 3 and 4, left). Musicians and nonmusicians did not, however, show different degrees of phase shift for the /ga/ versus /da/ stimulus pairing (F(1, 49)=0.443, p=0.51), likely due to the fact that the /ga/-/da/ syllable pair is the most acoustically similar of the stimulus pairings. A greater number of response trials might have increased our ability to detect such subtle differences between these two syllables, especially in nonmusicians. During the control time range (Figs. 3 and 4, right), which reflects the neural encoding of the vowel, there was no effect of group (F(1,49)=0.549, p=0.46), stimulus pairing (F(2,49)=0.471, p=0.626), or group x pairing interaction (F(2,60)=.252, p=0.778).

Collapsing across both groups did not reveal a tendency for phase shifts for the three stimulus pairings to differ in magnitude (F(2,96)=0.836, p=0.437), a result previously reported by Skoe et al. (2011), likely due to the fact that only 700 sweeps were collected in this study versus the 6000 collected for Skoe et al. (2011). The musicians and nonmusicians, however, significantly differed in the extent to which the three syllable pairings gave rise to unique phase shift signatures (significant group x syllable pairing interaction, F(2,96)=5.719, p=0.005). Specifically, degree of phase shift differed between the three syllable pairings in musician responses (F(2,21)=10.083, p=0.001) but not in nonmusician responses (F(2,25)=0.901, p=0.419).

In addition to having greater neural discrimination of these speech syllables, musicians outperformed nonmusicians on the QuickSIN test (F(1,49)=15.798, p<0.001). Speech-in-noise performance correlated with the extent of phase shifts across two out of the three stimulus pairings (/da/-/ba/: *r*=−0.372, p=0.01; /ga/-/ba/: *r*=−0.485, p=0.008; /ga/-/da/: *r*=−0.042, p=0.773). Within the musician group, the extent of musical training did not correlate with the amount of phase shift (/ga/-/ba/: r = 0.157, p = 0.474, /da/-/ba/: r = −0.254, p = 0.242; /ga/-/da/: r = 0.206, p = 0.347).

Discussion

Here, we reveal that musicians have more distinct subcortical representations of contrastive speech syllables than nonmusicians and the extent of subcortical speech sound differentiation correlates with speech perception in noise. These findings demonstrate that musicians possess a neural advantage for distinguishing speech sounds that may contribute to their enhanced speech-in-noise perception. These findings provide the first neural evidence that musicians possess an advantage for representing differences in formant frequencies.

Extensive research has documented enhancements in the processing of an acoustic feature with repetition in a task-relevant context. Indeed, interactive experiences with sound shape frequency mapping throughout the auditory system (Fritz et al. 2003, 2007, 2010), with task-relevant frequencies becoming more robustly represented in auditory cortex (Fritz et al. 2003). Repetitive stimulation of auditory cortical neurons responsive to a particular frequency similarly leads to an increase in the population of responding neurons located within the inferior colliculus and the cochlear nucleus that phase-lock to that frequency (Gao and Suga 1998, Yan and Suga 1998, Luo et al. 2008). Again, this effect is strengthened if the acoustic stimulation is behaviorally relevant (Gao and Suga 1998, 2000). These cortical modifications of subcortical processing are likely brought about via the corticofugal system —a series of downward-projecting neural pathways originating in the cortex and terminating at subcortical nuclei (Suga & Ma 2003) and ultimately the cochlea (Brown and Nuttall 1986). Attention and behavioral relevance can, therefore, lead to broader representation of a frequency throughout the auditory system, both cortically and subcortically.

Musicians spend many hours on musical tasks that require fine spectral resolution. For example, musicians learn to use spectral cues to discriminate instrumental timbres—a skill that is particularly useful for ensemble musicians who regularly encounter simultaneous instrumental sounds that overlap in pitch, loudness and duration but differ in timbre. Timbre is perceptually complex and relies in part upon spectro-temporal fine structure (i.e., relative amplitudes of specific harmonics) (Krimpoff et al. 1994, Caclin et al. 2005, Caclin et al. 2006, Kong et al. 2011). Thus, musicians' goal-oriented, directed attention to specific acoustic features, such as timbre, may, via the corticofugal system, enhance auditory processing of the rapidly-changing frequency components of music, speech and other behaviorally-relevant sounds. This is supported with previous research demonstrating that musicians have heightened sensitivity to small harmonic differences (Musacchia et al. 2008, Zendel and Alain 2009) as well as greater neural representation of harmonics (Koelsch et al. 1999, Shahin et al. 2005, Musacchia et al. 2008, Zendel and Alain 2009, Strait et al. 2009, Parbery-Clark et al. 2009a, Lee et al. 2009) than nonmusicians.

That musicians must learn to map slight changes in relative harmonic amplitudes to behavioral relevance, or meaning, may contribute to these auditory perceptual and neurophysiological enhancements (Kraus and Chandrasekaran 2010). Here, we propose that the fine-tuning of neural mechanisms for encoding high frequencies within a musical context leads to enhanced subcortical representation of the higher frequencies present in speech, improving musicians' ability to neurally process speech formants. This evidence can be interpreted alongside a wealth of research substantiating enhanced cortical anatomical structures (e.g., auditory areas) (Pantev et al. 1998, Zatorre, 1998, Schneider et al., 2002, Sluming et al. 2002, Gaser and Schlaug, 2003, Bermudez and Zatorre, 2005, Lappe et al., 2008, Hyde et al. 2009) and circuitry connecting areas that undergird language processing (e.g., the arcuate fasciculus; Wan and Schlaug 2010, Halwani et al. 2011) in musicians. This may account for their more robust neural speech-sound processing compared to nonmusicians (Schön et al. 2004, Magne et al. 2006, Moreno and Besson 2006, Marques et al. 2007, Tervaniemi et al. 2009, Chobert et al. 2011, Besson et al. 2011, Marie et al. 2011a, Marie et al. 2011b). Taken together, this work lends support to our interpretation that musical training is associated with auditory processing benefits that are not limited to the music domain. We further propose that the subcortical auditory enhancements observed in musicians, such as for speech-sound differentiation, may reflect their cortical distinctions by means of strengthened top-down neural pathways that make up the corticofugal system for hearing. Such pathways facilitate cortical influences on subcortical auditory function (Gao and Suga 1998, 2000, Yan and Suga 1998, Suga et al 2002, Luo et al. 2008) and have been implicated in auditory learning (Bajo et al. 2010). It is also possible that over the course of thousands of hours of practice distinguishing between sounds on the basis of their spectra a musician's auditory brainstem function is locally, pre-attentively modified. Top-down and local mechanisms likely work in concert to drive musicians' superior neural encoding of speech.

Although we interpret these results in the context of training-related enhancements in musicians compared to nonmusicians, our outcomes cannot discount the possibility that the musician advantage for speech processing reflects genetic differences that exist prior to training. Future work is needed to tease apart the relative contributions of experience and genetics to differences in the subcortical encoding of speech. Although we interpret our results as reflecting the influence of musical training on the neural processing of speech, we did not find a correlation between number of years of musical training and neural distinction of speech sounds. One possible explanation for this is that our musicians were highly trained and had extensive musical experience; the benefits accrued from musical practice may be most significant during the first few years of training. If so, we would expect a group of individuals with minimal to medium amounts of musical experience to show a significant

correlation between years of practice and speech sound distinction. An alternate interpretation is that our results reflect innate biological differences between musicians and nonmusicians; enhanced subcortical auditory processing may be one of the factors that leads individuals to persevere with musical training.

More precise neural distinction in musicians relates with speech-in-noise perception

Speech perception in real-world listening environments is a complex task, requiring the listener to track a target voice within competing background noise. This task is further complicated by the degradation of the acoustic signal by noise, which particularly disrupts the perception of fast spectro-temporal features of speech (Brandt and Rosen 1980). While hearing in noise is challenging for everyone, musicians are less affected by the presence of noise than nonmusicians (Parbery-Clark et al. 2009a, Parbery-Clark et al. 2009b, Parbery-Clark et al. 2011, Bidelman et al. 2011, Zendel and Alain 2011). Previous research has found that speech-in-noise perceptual ability relates to the neural differentiation of speech sounds in children (Hornickel et al. 2009). Here, we extend this finding and show that adult musicians with enhanced SIN perception demonstrate greater neural distinction of speech sounds, potentially providing a biological mechanism accounting for musicians' advantage for hearing in adverse listening conditions.

While our nonmusician group, all normal- to high-functioning adults, demonstrated a range of SIN perception, they did not show significant phase shifts. Given the correlation between phase shifts and SIN perception, it is possible that the representation of high-frequency differences via timing shifts facilitates, but is not strictly necessary for, syllable discrimination. Alternately, musicians may have such robust neural representation of sound that subcortical differentiation of speech sounds can be clearly seen in an average comprised of only 700 trials. It is possible that nonmusicians would also demonstrate differentiated neural responses but only with a greater number of trials, as reported previously in nonmusician children (Skoe et al. 2011).

Certain populations, such as older adults (Ohde and Abou-Khalil, 2001) and children with reading impairments (Tallal et al., 1981, Maassen et al. 2001, Serniclaes et al. 2001, Serniclaes and Sprenger-Charolles 2003, Hornickel et al. 2009), demonstrate decreased perceptual discrimination of contrasting speech sounds. This is may be due to the impaired processing of high-frequency spectrotemporal distinctions that occur over the sounds' first 40 ms (i.e., their formant transitions). SIN perception is particularly difficult for these populations (older adults: Gordon-Salant and Fitzgibbons, 1995; children with reading impairments: Bradlow and Kraus, 2003). Indeed, children with poor reading and/or SIN perception show reduced neural differentiation of the same stop consonants (/ba/, /da/, and /ga/) used in the present study, thus providing a neural index of these behavioral difficulties (Hornickel et al. 2009). Musicians' enhanced subcortical differentiation of contrasting speech syllables suggests that musical training may provide an effective rehabilitative approach for children who experience difficulties with reading and hearing in noise. Future work could directly test the impact of musical training on both subcortical representations of speech sounds and language skills by randomly assigning music lessons or an auditory-based control activity to individuals with reading and speech perception impairment. Given that effects of musical training on speech-syllable discrimination may stem from musicians' experience with high-frequency spectrotemporal features that distinguish instrumental timbres, a training regimen that emphasizes timbre perception—e.g., ensemble work—may lead to enhanced speech-sound discrimination and related improvements in reading and SIN perception.

Conclusion

Here, we demonstrate that musicians show greater neural distinction between speech syllables than nonmusicians and that the extent of this neural differentiation correlates with the ability to perceive speech in noise. This musician enhancement may stem from their extensive experience distinguishing closely related sounds on the basis of timbre. We suggest that musical training may improve speech perception and reading skills in learning impaired children.

Acknowledgments

References

Aiken S, Picton T. Envelope and spectral frequency-following responses to vowel sounds. Hear Res. 2008; 245:35–47. [PubMed: 18765275]

Akhoun I, Gallégo S, Moulin A, Ménard M, Veuillet E, Berger-Vachon C, Collet L, Thai-Van H. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. Clinical Neurophysiology. 2008; 119:922–933. [PubMed: 18291717]

Bajo V, Nodal F, Moore D, King A. The descending corticocollicular pathway mediates learning-induced auditory plasticity. Nat Neurosci. 2010; 13:253–263. [PubMed: 20037578]

Bermudez P, Zatorre RJ. Differences in gray matter between musicians and nonmusicians. Ann N Y Acad Sci. 2005; 1060:395–399. [PubMed: 16597791]

Besson M, Schon D, Moreno S, Santos A, Magne C. Influence of musical expertise and musical training on pitch processing in music and language. Restor Neurol Neuros. 2007; 25:399–410.

Besson M, Chobert J, Marie C. Transfer of training between music and speech: common processing, attention and memory. Front Psychol. 2011; 2:1–12. [PubMed: 21713130]

Bidelman GM, Krishnan A. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 2010; 1355:112–25. [PubMed: 20691672]

Bidelman GM, Krishnan A. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. Brain Res. 2010; 1355:112–125. [PubMed: 20691672]

Bidelman GM, Krishnan A, Gandour JT. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. Eur J Neurosci. 2011; 33:530–538. [PubMed: 21198980]

Bidelman GM, Gandour JT, Krishnan A. Musicians demonstrate experience-dependent brainstem enhancements of musical scale features within continuously gliding pitch. Neuroscience Letters. 2011; 503:203–207. [PubMed: 21906656]

Bradlow A, Kraus N. Speaking clearly for children with learning disabilities: sentence perception in noise. J Speech Lang Hear R. 2003; 46:80–97.

Brandt J, Rosen JJ. Auditory phonemic perception in dyslexia: categorical identification and discrimination of stop consonants. Brain Lang. 1980; 9:324–337. [PubMed: 7363076]

Brown MC, Nuttall AL. Efferent control of cochlear inner hair cell responses in the guinea-pig. J Physiol. 1984; 354:625–646. [PubMed: 6481647]

Brown, L.; Sherbenou, RJ.; Johnsen, SK. Test of Nonverbal Intelligence (*TONI-3*). Austin, TX: Pro-ed, Inc; 1997.

Caclin A, McAdams S, Smith B, Winsberg S. Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. J Acoust Soc Am. 2005; 118:471–482. [PubMed: 16119366]

Caclin A, Brattico E, Tervaniemi M, Naatanen R, Morlet D, Giard M, McAdams S. Separate neural processing of timbre dimensions in auditory sensory memory. J Cog Neurosci. 2006; 18:1959–1972.

Chan S, Ho Y, Cheung M. Music training improves verbal memory. Nature. 1998; 396:128. [PubMed: 9823892]

Chandrasekaran B, Krishnan A, Gandour J. Mismatch negativity to pitch contours is influenced by language experience. Brain Res. 2007; 1128:148–156. [PubMed: 17125749]

Chandrasekaran B, Krishnan A, Gandour J. Relative influence of musical and linguistic experience on early cortical processing of pitch contours. Brain Lang. 2009a; 108:1–9. [PubMed: 18343493]

Chandrasekaran B, Hornickel J, Skoe E, Nicol T, Kraus N. Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. Neuron. 2009b; 64:11–319.

Chandrasekaran B, Kraus N. The scalp-recorded brainstem response to speech: neural origins. Psychophysiology. 2010; 47:236–246. [PubMed: 19824950]

Chobert J, Marie C, Francois C, Schon D, Besson M. Enhanced passive and active processing of syllables in musician children. J Cog Neuro. 2011; 23:3874–3887.

Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci. 2003; 6:1216–1224. [PubMed: 14583754]

Fritz J, David S, Radtke-Schuller S, Yin P, Shamma S. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat Neurosci. 2010; 13:1011–1022. [PubMed: 20622871]

Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C. Musical training enchances automatic encoding of melodic contour and interval structure. J Cognitive Neuroscience. 2004; 16:1010–1021.

Gaab N, Tallal P, Lakshminarayanan K, Archie JJ, Glover GH, Gabriel JDE. Neural Correlates of Rapid Spectrotemporal Processing in Musicians and Nonmusicians. Ann NY Acad Sci: Neurosciences and Music. 2005; 1060:82–88.

Galbraith GC, Threadgill MR, Hemsley J, Salour K, Songdej N, Ton J, Cheung L. Putative measure of peripheral and brainstem frequency-following in humans. Neurosci Lett. 2000; 292:123–7. [PubMed: 10998564]

Gao E, Suga N. Experience-dependent corticofugal adjustment of midbrain frequency map in bat auditory system. Proc Natl Acad Sci. 1998; 95:12663–12670. [PubMed: 9770543]

Gao E, Suga N. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: role of the corticofugal system. Proc Natl Acad Sci. 2000; 97:8081–8086. [PubMed: 10884432]

Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. J Neurosci. 2003; 23:9240–9245. [PubMed: 14534258]

Gordon-Salant S, Fitzgibbons PJ. Recognition of multiply degraded speech by young and elderly listeners. J Speech Hear Res. 1995; 38:1150–1156. [PubMed: 8558883]

Gorga MP, Kaminski JR, Beauchaine KA, Jesteadt W. Auditory brainstem responses to tone bursts in normally hearing subjects. J Speech Hear Res. 1988; 31:87–97. [PubMed: 3352259]

Halwani GF, Loui P, Rüber T, Schlaug G. Effects of practice and experience on the arcuate fasciculus: comparing singers, instrumentalists, and non-musicians. Front Psychol. 2011; 2:113.10.3389/fpsyg.2011.00156 [PubMed: 21716636]

Hornickel J, Skoe E, Nicol T, Zecker S, Kraus N. Subcortical differentiation of voiced stop consonants: relationships to reading and speech in noise perception. Proc Natl Acad Sci. 2009; 106:13022–13027. [PubMed: 19617560]

Hurwitz I, Wolff P, Bortnick B, Kokas K. Nonmusical effects of the Kodaly music curriculum in primary grade children. J Learning Disabilities. 1975; 3:167–174.

Hyde KL, Lerch J, Norton A, Forgeard M, Winner E, Evans AC, Schlaug G. Musical training shapes structural brain development. J Neurosci. 2009; 29:3019–3025. [PubMed: 19279238]

Johnson KL, Nicol T, Kraus N. Developmental plasticity in the human auditory brainstem. J Neurosci. 2008; 28:4000–4007. [PubMed: 18400899]

Killion MC, Niquette Pa, Gudmundsen GI, Revit LJ, Banerjee S. 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. 2004; 116:2395–2405. [PubMed: 15532670]

Klatt D. Software for a cascade/parallel formant synthesizer. J Acoust Soc Am. 1980; 67:13–33.

Koelsch S, Schroger E, Tervaniemi M. Superior pre-attentive auditory processing in musicians. NeuroReport. 1999; 10:1309–1313. [PubMed: 10363945]

Kong Y, Mullangi A, Marozeau J, Epstein M. Temporal and spectral cues for musical timbre perception in electric hearing. J Speech Lang Hear. 2011; 54:981–994.

Kraus N, Chandrasekaran B. Music training for the development of auditory skills. Nat Rev Neurosci. 2010; 11:599–605. [PubMed: 20648064]

Krimphoff J, McAdams S, Winsberg S. Caractérisation du timbre des sons complexes. II: Analyses acoustiques et quantification psychophysique. J Phys-Paris. 1994; 4:625–628.

Lappe C, Herholz SC, Trainor LJ, Pantev C. Cortical plasticity induced by short-term unimodal and multimodal musical training. J Neurosci. 2008; 28:9632–9639. [PubMed: 18815249]

Lee KM, Skoe E, Kraus N, Ashley R. Selective subcortical enhancement of musical intervals in musicians. J Neurosci. 2009; 29:5832–5840. [PubMed: 19420250]

Liu L-F, Palmer AR, Wallace MN. Phase-locked responses to pure tones in the inferior colliculus. J Neurophysiol. 2006; 95:1926–1935. [PubMed: 16339005]

Luo F, Wang Q, Kashani A, Yan J. Corticofugal modulation of initial sound processing in the brain. J Neurosci. 2008; 28:11615–11621. [PubMed: 18987197]

Maassen B, Groenen P, Crul T, Assman-Hulsmans C, Gabreels F. Identification and discrimination of voicing and place-of-articulation in developmental dyslexia. Clinical Linguistics & Phonetics. 2001; 15:319–339.

Magne C, Schon D, Besson M. Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. J Cogn Neurosci. 2006; 18:199–211. [PubMed: 16494681]

Marie C, Kujala T, Besson M. Musical and linguistic expertise influence preattentive and attentive processing of non-speech sounds. Cortex. 2010; 48:447–457. [PubMed: 21189226]

Marie C, Delogu F, Lampis G, Olivetti Belardinelli M, Besson M. Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. J Cogn Neurosci. 2011a; 10:2701–15.

Marie C, Magne C, Besson M. Musicians and the metric structure of words. J Cogn Neurosci. 2011b; 23:294–305. [PubMed: 20044890]

Margulis E, Misna L, Uppunda A, Parrish T, Wong P. Selective neurophysiologic responses to music in instrumentalists with different listening biographies. Hum Brain Mapp. 2009; 30:267–275. [PubMed: 18072277]

Marques C, Moreno S, Castro S, Besson M. Musicians detect pitch violation in a foreign language better than non-musicians: behavioural and electrophysiological evidence. Journal of Cognitive Neuroscience. 2007; 19:1453–1463. [PubMed: 17714007]

Miller G, Nicely P. An analysis of perceptual confusions among some English consonants. J Acoust Soc Am. 1955; 27:338–352.

Moreno S, Besson M. Musical training and language-related brain electrical activity in children. Psychophysiology. 2006; 43:287–291. [PubMed: 16805867]

Musacchia G, Sams M, Skoe E, Kraus N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc Natl Acad Sci U S A. 2007; 104:15894–15898. [PubMed: 17898180]

Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and nonmusicians. Hear Res. 2008; 241:34–42. [PubMed: 18562137]

Naatanen R, Lehtokoski A, Lennes M, Cheour M, Houtilainen M, Ilvonen A, Vainio M, Alku P, Ilmoniemi R, Luuk A, Allik J, Sinkkonen J, Alho K. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature. 1997; 385:432–435. [PubMed: 9009189]

Nager W, Kohlmetz C, Altenmuller E, Rodriguez-Fornells A, Munte T. The fate of sounds in conductors' brains: an ERP study. Cog Brain Res. 2003; 17:83–93.

Nikjeh D, Lister J, Frisch S. Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. Ear Hear. 2009; 30:432–446. [PubMed: 19494778]

Ohde RN, Abou-Khalil R. Age differences for stop-consonant and vowel perception in adults. J Acoust Soc Am. 2001; 110:2156–2166. [PubMed: 11681392]

Pantev C, Oostenveld R, Engelien A, Ross B, Roberts LE, Hoke M. Increased auditory cortical representation in musicians. Nature. 1998; 392:811– 814. [PubMed: 9572139]

Pantev C, Roberts L, Schulz M, Engelien A, Ross B. Timbre-specific enhancement of auditory cortical representations in musicians. NeuroReport. 2001; 12:1–6. [PubMed: 11201065]

Parbery-Clark A, Skoe E, Kraus N. Musical experience limits the degradative effects of background noise on the neural processing of sound. J Neurosci. 2009a; 29:14100–14107. [PubMed: 19906958]

Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech in noise. Ear Hear. 2009b; 30:653–661. [PubMed: 19734788]

Patel A. Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front Psychology. 2011; 2:142.10.3389/fpsyg.2011.00142

Schneider P, Scherg M, Dosch HG, Specht HJ, Gutschalk A, Rupp A. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. Nat Neurosci. 2002; 5:688–694. [PubMed: 12068300]

Schlaug GL, Jäncke Y, Huang, Staiger JF, Steinmetz H. Increased corpus callosum size in musicians. Neuropsychologia. 1995; 33:1047–1054. [PubMed: 8524453]

Schon D, Magne C, Besson M. The music of speech: music training facilitates pitch processing in both music and language. Psychophysiology. 2004; 41:341–349. [PubMed: 15102118]

Serniclaes W, Sprenger-Charolles L. Categorical perception of speech sounds and dyslexia. Current Psychology Letters: Behaviour, Brain & Cognition. 2003; 10(1)

Serniclaes W, Sprenger-Charolles L, Carre R, Demonet JF. Perceptual discrimination of speech sounds in developmental dyslexia. Journal of Speech, Language and Hearing Research. 2001; 44:384.10.1044/1092-4388(2001/032)

Shahin AJ, Roberts LR, Chau W, Trainor LJ, Miller L. Musical training leads to the development of timbre-specific gamma band activity. NeuroImage. 2008; 41:113–122. [PubMed: 18375147]

Shahin A. Neurophysiological influence of musical training on speech perception. Front Psychology. 2011; 2:126.10.3389/fpsyg.2011.00126

Shahin A, Roberts L, Pantev C, Trainor L, Ross B. Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. NeuroReport. 2005; 16:1781–1785. [PubMed: 16237326]

Skoe E, Nicol T, Kraus N. Cross-phaseogram: Objective neural index of speech sound differentiation. J Neurosci Meth. 2011; 196:308–317.

Sluming V, Barrick T, Howard M, Cezayirli E, Mayes A, Roberts N. Voxel-based morphometry reveals increased gray matter density in Broca's area in male symphony orchestra musicians. Neuroimage. 2002; 17:1613–22. [PubMed: 12414299]

Strait DL, Kraus N. Playing music for a smarter ear: Cognitive, perceptual and neurobiological evidence. Music Perception. 2011; 29:133–146.

Strait DL, Skoe E, Kraus N, Ashley R. Musical experience promotes subcortical efficiency in processing emotional vocal sounds. Ann NY Acad Sci: Neurosciences and Music III. 2009; 1169:209–213.

Strait DL, Kraus N. Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. Front Psychol. 2011; 2:113. [PubMed: 21716636]

Suga N, Xiao Z, Ma X, Ji W. Plasticity and corticofugal modulation for hearing in adult animals. Neuron. 2002; 36:9–18. [PubMed: 12367501]

Tallal P, Stark R. Speech acoustic-cue discrimination abilities of normally developing and language-impaired children. J Acoust Soc Am. 1981; 69:568–575. [PubMed: 7462479]

Tallal P, Gaab N. Dynamic auditory processing, musical experience and language development. Trends Neurosci. 2006; 29:382–390. [PubMed: 16806512]

Tervaniemi M, Jacobsen T, Rottger S, Kujala T, Widmann A, Vainio M, Naatanen R, Scroger E. Selective tuning of cortical sound-feature processing by language experience. Eur J Neurosci. 2006a; 23:2538–2541. [PubMed: 16706861]

Tervaniemi M, Castaneda A, Knoll M, Uther M. Sound processing in amateur musicians and nonmusicians: event-related potential and behavioral indices. NeuroReport. 2006b; 17:1225–1228. [PubMed: 16837859]

Tervaniemi M, Kruck S, De Baene W, Schroger E, Alter K, Friderici A. Top-down modulation of auditory processing: effects of sound context, musical expertise and attentional focus. Eur J Neurosci. 2009; 30:1636–1642. [PubMed: 19821835]

Tierney A, Parbery-Clark A, Skoe E, Kraus N. Frequency-dependent effects of background noise on subcortical response timing. Hear Res. 201110.1016/j.heares.2011.08.014

Wan C, Schlaug G. Music making as a tool for prompting brain plasticity across the life span. Neuroscientist. 2010; 16:566–577. [PubMed: 20889966]

Warrier C, Nicol T, Abrams D, Kraus N. Inferior colliculus contributions to phase encoding of stop consonants in an animal model. Hear Res. 201110.1016/j.heares.2011.09.001

Winkler I, Kajula T, Tiitinen H, Sivonen P, Alku P, Lehtokoski A, Czigler I, Csepe V, Ilmoniemi R, Naatanen R. Brain responses reveal the learning of foreign language phonemes. Psychophysiology. 1999; 36:638–642. [PubMed: 10442032]

Wong P, Skoe E, Russo N, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat Neurosci. 2007; 10:420–422. [PubMed: 17351633]

Yan W, Suga N. Corticofugal modulation of the midbrain frequency map in the bat auditory system. Nat Neurosci. 1998; 1:54–58. [PubMed: 10195109]

Ylinen S, Shestakova A, Huotilainen M, Alku P, Naatanen R. Mismatch negativity (MMN) elicited by changes in phoneme length: a cross-linguistic study. Brain Res. 2006; 1072:175–185. [PubMed: 16426584]

Zatorre RJ. Functional specialization of human auditory cortex for musical processing. Brain. 1998; 121:1817–1818. [PubMed: 9798739]

Zendel B, Alain C. Concurrent sound segregation is enhanced in musicians. J Cog Neurosci. 2009; 21:1488–1498.

Zendel B, Alain C. Musicians experience less age-related decline in central auditory processing. Psychol Aging. 201110.1037/a0024816

Certain speech sounds (i.e. /ba/, /da/, and /ga/) are commonly confused

Musicians demonstrate greater neural distinction of these sounds

Greater neural distinction of these sounds relates to the ability to hear in background noise

Older adults and learning-impaired children have decreased discrimination of speech syllables

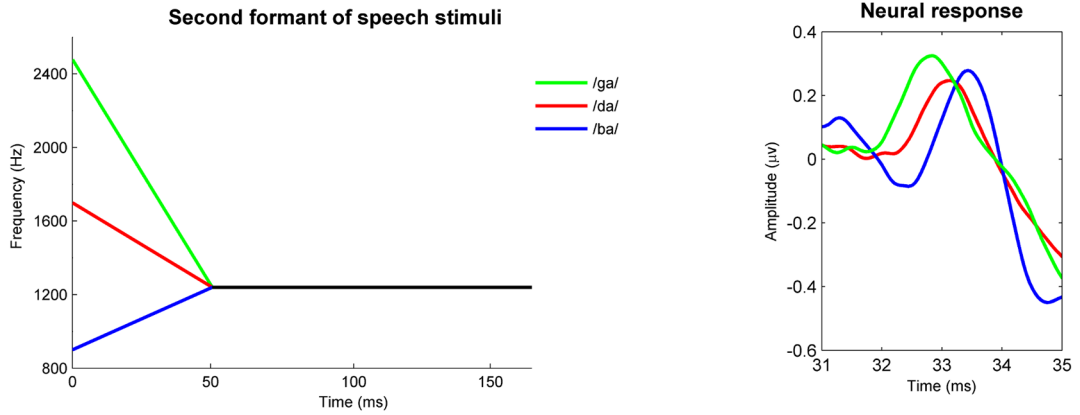Musical training may be an effective vehicle for auditory remediation and habilitation

Fig. 1. Second formant trajectories of the three speech stimuli (left)
The speech sounds /ga/ /da/ and /ba/ differ in their second formant trajectories during the formant transition period (5–50 ms) but are identical during the vowel /a/ (50–170ms). **Illustrative example of the neural timing differences elicited by the three stimuli in the formant transition region (right).** Consistent with the auditory system's tonotopic organization that engenders earlier neural response timing to higher frequencies relative to lower frequencies, the neural response timing to /ga/ is earlier relative to the responses to /da/ and /ba/ during the formant transition region.
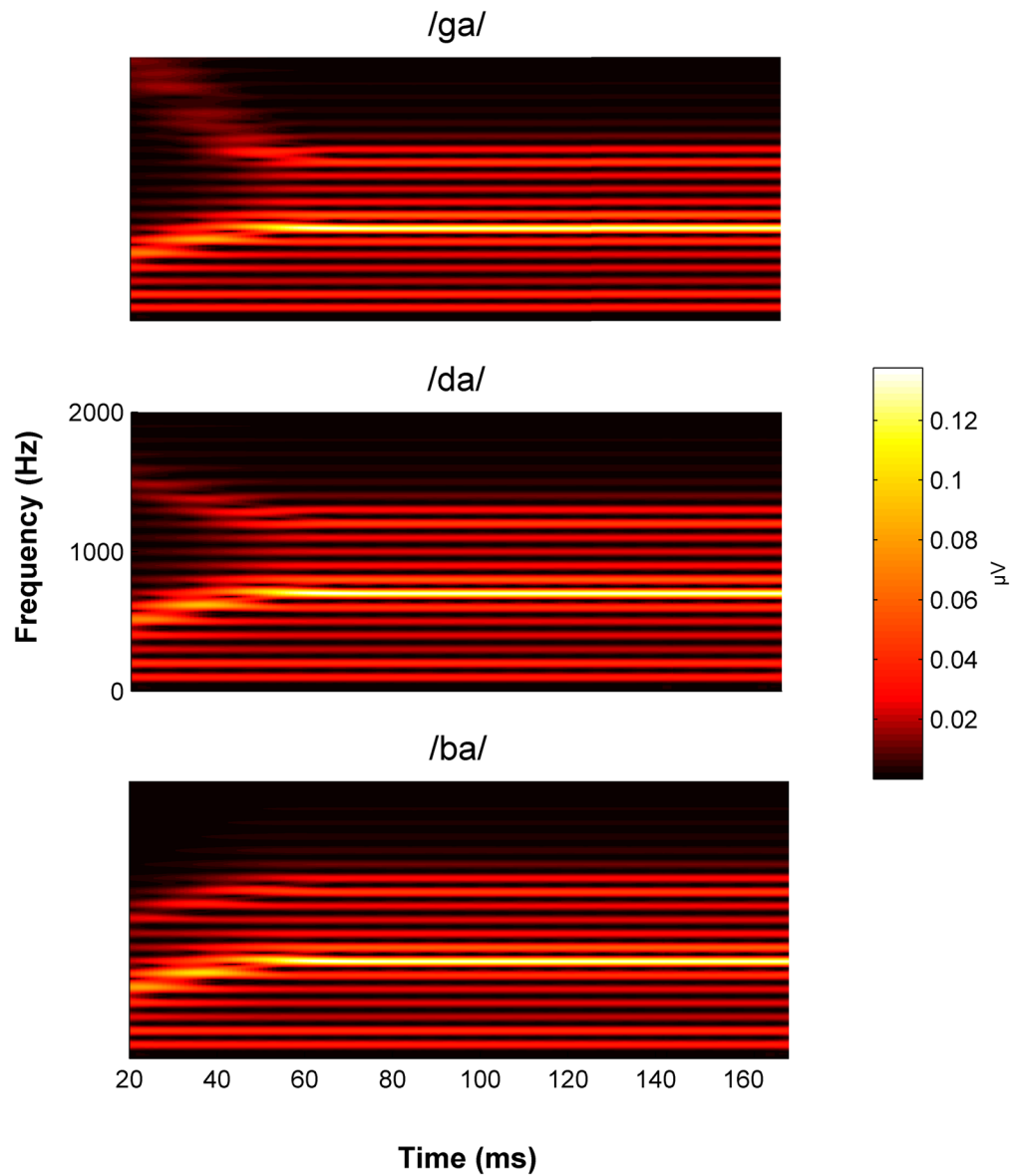
Fig 2. Spectrograms of the three speech stimuli
The spectrograms of the three speech stimuli /ba/ (top), /da/ (middle), and /ga/ (bottom) differ only in their second formant trajectory.
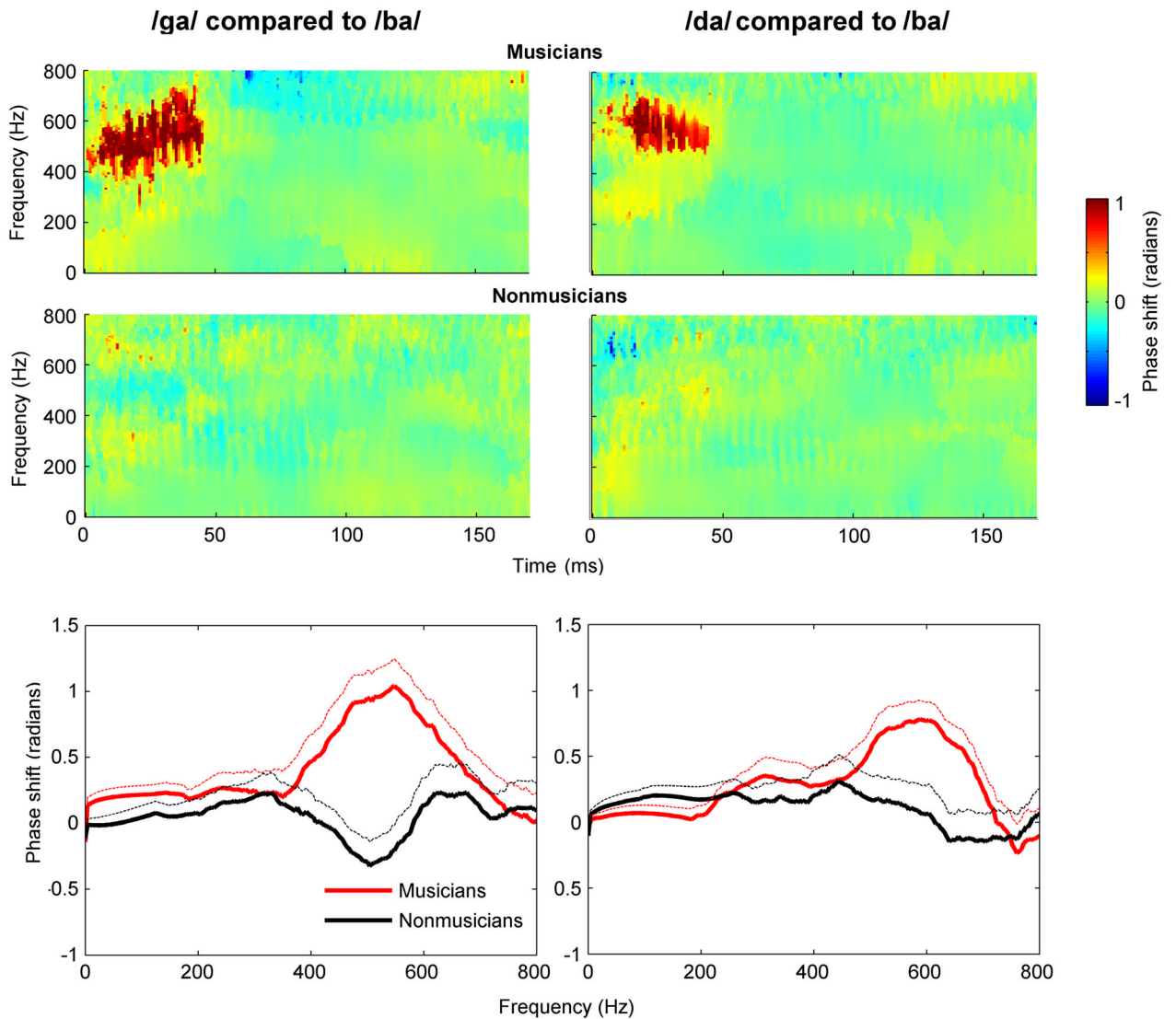
Fig. 3. Musicians show greater subcortical differentiation of speech sounds than non-musicians
The neural timing difference between the brainstem responses to the syllables /ga/ and /ba/ (left) and /da/ and /ba/ (right) can be measured as phase shifts. Warm colors indicate that the response to /ga/ led the response to /ba/; cool colors indicate the opposite. Musicians demonstrate clear phase differences in responses to these syllables during the transition region (top). These phase shifts are considerably weaker in nonmusicians (middle). The average phase shift and standard error over the transition portion plotted across frequency (bottom). Musicians have larger phase shifts during the formant transition than nonmusicians, with the greatest phase shifts occurring between 400–720 Hz.
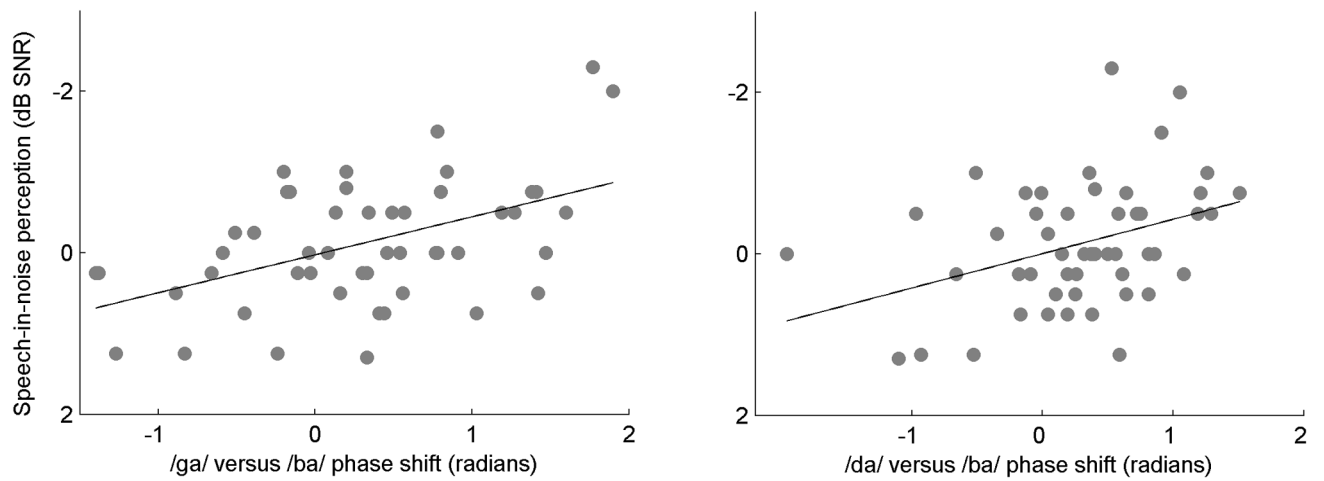
Fig. 4. Relationship between phase shift and speech-in-noise perception
The extent of the phase shifts between the responses to /ga/ and /ba/ (top) and /da/ and /ba/ (bottom) correlate with speech-in-noise perception (/ga/-/ba/: $r=-0.41$, p=0.021, /da/-/ba/: $r=-0.73$, p<0.0001). [Note axes have been flipped to show better SIN perception (more negative SNR) in the upward direction]

Table 1
Participants' musical practice history

Age at which musical training began, years of musical training and major instrument(s) are indicated for all participants with musical experience. Means for years of musical training and age at onset for the nonmusicians were calculated from the eight participants who had musical experience.

Musician	Age Onset (Years)	Years of training (Years)	Instrument
#1	5	16	Bassoon/piano
#2	4	26	Cello
#3	7	19	Cello
#4	5	13	Flute
#5	5	19	Piano
#6	6	17	Piano
#7	5	16	Piano
#8	6	26	Piano
#9	6	12	Piano
#10	6	15	Piano
#11	6	19	Piano/horn
#12	6	12	Piano/percussion
#13	5	18	Piano/percussion
#14	7	16	Piano/voice
#15	6	19	Piano/voice
#16	5	20	Piano/voice
#17	5	14	Piano/voice
#18	3	18	Violin
#19	5	14	Violin
#20	5	16	Violin
#21	6	15	Violin
#22	5	15	Violin
#23	5	23	Violin/piano
Mean	**5.4**	**17.3**	
Nonmusician			
#24	11	2	Flute
#25	9	2	Flute
#26	16	1	Guitar
#27	9	3	Guitar
#28	13	1	Piano
#29	12	2	Piano
#30	9	3	Piano/voice
#31	12	3	Voice
#32	0	0	N/A

Musician	Age Onset (Years)	Years of training (Years)	Instrument
#33	0	0	N/A
#34	0	0	N/A
#35	0	0	N/A
#36	0	0	N/A
#37	0	0	N/A
#38	0	0	N/A
#39	0	0	N/A
#40	0	0	N/A
#41	0	0	N/A
#42	0	0	N/A
#43	0	0	N/A
#44	0	0	N/A
#45	0	0	N/A
#46	0	0	N/A
#47	0	0	N/A
#48	0	0	N/A
#49	0	0	N/A
#50	0	0	N/A
Mean	**11.4**	**2.1**	