OXFORD

Short- and long-term neuroplasticity interact during the perceptual learning of concurrent speech

Jessica MacLean [ID][1,2], Jack Stirn[1], Alexandria Sisson[1], Gavin M. Bidelman [ID][1,2,3,*]

[1]Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA,
[2]Program in Neuroscience, Indiana University, Bloomington, IN, USA,
[3]Cognitive Science Program, Indiana University, Bloomington, IN, USA
*Corresponding author: Department of Speech, Language and Hearing Sciences, 2631 East Discovery Parkway, Bloomington, IN 47408, United States.
Email: gbidel@indiana.edu

Plasticity from auditory experience shapes the brain's encoding and perception of sound. However, whether such long-term plasticity alters the trajectory of short-term plasticity during speech processing has yet to be investigated. Here, we explored the neural mechanisms and interplay between short- and long-term neuroplasticity for rapid auditory perceptual learning of concurrent speech sounds in young, normal-hearing musicians and nonmusicians. Participants learned to identify double-vowel mixtures during ∼ 45 min training sessions recorded simultaneously with high-density electroencephalography (EEG). We analyzed frequency-following responses (FFRs) and event-related potentials (ERPs) to investigate neural correlates of learning at subcortical and cortical levels, respectively. Although both groups showed rapid perceptual learning, musicians showed faster behavioral decisions than nonmusicians overall. Learning-related changes were not apparent in brainstem FFRs. However, plasticity was highly evident in cortex, where ERPs revealed unique hemispheric asymmetries between groups suggestive of different neural strategies (musicians: right hemisphere bias; nonmusicians: left hemisphere). Source reconstruction and the early (150–200 ms) time course of these effects localized learning-induced cortical plasticity to auditory-sensory brain areas. Our findings reinforce the domain-general benefits of musicianship but reveal that successful speech sound learning is driven by a critical interplay between long- and short-term mechanisms of auditory plasticity, which first emerge at a cortical level.

Key words: auditory perceptual learning; EEG; event-related brain potentials (ERP); frequency-following response (FFR); speech-in-noise perception.

Introduction

Experience on multiple timescales shapes our sensory systems. For example, studies have shown short-term changes in the selective tuning properties of cortical neurons that facilitate speech-in-noise (SIN) perception within a single training session (Ahveninen et al. 2011; Da Costa et al. 2013). Auditory perceptual learning studies also demonstrate rapid changes in behavior that track with subcortical (Carcagno and Plack 2011) and cortical (Ozaki et al. 2004) reorganization as indexed by scalp-recorded event-related brain potentials (ERPs). Similarly, neuroimaging studies show long-term listening experiences (e.g. specific language expertise, bilingualism) shape the brain's representations for speech sounds, selectively enhancing native compared to nonnative elements of a listener's lexicon (Kuhl et al. 1992; Krishnan et al. 2010; Jeng et al. 2011; Krishnan et al. 2011; Bidelman and Lee 2015). Yet, because they are studied in isolation, the way in which short-term and long-term plasticity interact within the auditory system remains poorly understood.

Musicianship offers one approach to investigate brain plasticity along multiple timescales (Kraus and Chandrasekaran 2010; Herholz and Zatorre 2012; Alain et al. 2014; Moreno and Bidelman 2014). For example, cross-sectional studies show that musicians demonstrate improved behavioral performance in difficult listening scenarios, such as SIN and other "cocktail party" environments (Parbery-Clark et al. 2009b; Coffey et al. 2017; Yoo and Bidelman 2019; Bidelman and Yoo 2020). On a smaller timescale, studies

have demonstrated behavioral and neural encoding improvements in older adults' speech and auditory processing following short-term music interventions (Alain et al. 2019; Dubinsky et al. 2019; Bidelman et al. 2022), providing more definitive evidence that music activities causally improve aspects of behavior. Improvements to SIN perception following music training have also been observed in clinical populations, such as individuals with hearing loss (Lo et al. 2020). One explanation for musicians' SIN benefit is the OPERA hypothesis, which posits that the precise auditory demands of music, when combined with repetition and high emotional reward, require attention-modulated engagement in overlapping networks for music and speech perception, conferring benefits to speech processing (Patel 2011, 2014). Collectively, these studies suggest that long-term musicianship (and related short-term interventions) bolster the brain's ability to precisely encode speech sounds, which may confer advantages not only to everyday speech processing but also novel sound acquisition via learning (e.g. second language development; Chobert and Besson 2013; Picciotti et al. 2018; Slevc and Miyake 2006).

Musicians' behavioral advantages for SIN processing are supported biologically through their stronger and faster sound-evoked neural responses at both brainstem and cortical levels of auditory processing (Shahin et al. 2003; Parbery-Clark et al. 2009a; Bidelman and Krishnan 2010; Bidelman et al. 2014; Puschmann et al. 2018). In particular, a large body of research has shown enhanced subcortical responses to sound in musicians, as

measured via frequency-following responses (FFRs) (Musacchia et al. 2007; Carcagno and Plack 2011; Strait et al. 2012; Weiss and Bidelman 2015). At the cortical level, musicians show enhanced responses in the timeframe of the P2 wave of the ERP, an early (150–200 ms) positive component reflecting perceptual auditory object formation, speech identification, and concurrent speech segregation abilities (Shahin et al. 2003; Bidelman et al. 2013; Leung et al. 2013; Ross et al. 2013; Bidelman and Yellamsetty 2017). Presumably, such coordination in neural encoding across different levels of the auditory system might account for musicians' superior ability to cope with real-world speech listening scenarios, including parsing target speech from background noise or competing talkers (Yoo and Bidelman 2019; Bidelman and Yoo 2020; Brown and Bidelman 2022).

One way to assess real-world SIN listening skills (and neuroplasticity therein) is via concurrent vowel identification tasks (Assmann and Summerfield 1989, 1990; Alain et al. 2007; Bidelman and Yellamsetty 2017). In these paradigms, listeners hear two simultaneous vowels and are asked to correctly identify both tokens as quickly and accurately as possible. Behaviorally, speech identification accuracy improves with increasing pitch differences between vowels for fundamental frequency (F0) separations from 0 to about 4 semitones (Arehart et al. 1997; Chintanpalli and Heinz 2013; Chintanpalli et al. 2016). Critically, task success requires multiple processes: listeners must first segregate and then identify both elements of the speech mixture. The segregation of complex auditory mixtures is thought to reflect a complex, distributed neural network involving both subcortical and cortical brain regions (Palmer 1990; Sinex et al. 2002; Dyson and Alain 2004; Alain et al. 2005; Bidelman and Alain 2015a). Moreover, when the task is conducted without feedback, the learning (and related short-term plasticity) is implicit and based on exposure. As such, double-vowel identification provides an ideal avenue for studying possible differential plasticity across the auditory system as it relates to learning and improving complex listening skills.

Using double vowel tasks, multiple studies have shown that short-term auditory perceptual learning (in nonmusicians) results in enhancements in the auditory cortical ERPs (Atienza and Cantero 2001; Reinke et al. 2003; Alain et al. 2007; Alain et al. 2015). The early timing of these neural changes (~100–250 ms) suggests that learning induces reorganization of the sensory-receptive fields of auditory cortical neurons rather than procedural learning alone (Fritz et al. 2003; Alain et al. 2007). Similar short-term plasticity has been observed at a subcortical level in brainstem FFRs—though for isolated rather than double vowel speech sound training (Reetzke et al. 2018). However, training sessions in most of these studies took place over multiple days or even weeks, which probably involved longer term mechanisms of learning (e.g. memory/sleep consolidation, overlearning) rather than sensory plasticity, per se (but see Alain et al. 2007). It is also unclear from the extant literature whether learning (i) engenders similar magnitudes of plasticity in auditory brainstem and cortex, and (ii) whether it proceeds in a bottom-up (e.g. brainstem→cortex) or top-down (e.g. cortex→brainstem) guided manner (cf. Ahissar and Hochstein 2004; Reetzke et al. 2018).

Here, we aimed to extend prior studies on the neural mechanisms of auditory perceptual learning by investigating the interplay between short- and long-term plasticity on concurrent speech processing. We build upon prior work by utilizing a concurrent double-vowel learning paradigm previously shown to induce rapid cortical plasticity in the ERPs (Alain et al. 2007). We measured simultaneous behavioral and multichannel

EEG responses while listeners completed ~45 min of training to assess short-term perceptual learning effects. The double-vowel paradigm provided an ideal test bed to assess changes in ecological speech listening skills. Our EEG approach also included the tandem recording of both brainstem FFRs with cortical ERPs to assess potential differential neuroplasticity at sub- vs. neocortical levels of the auditory-speech hierarchy. Cross-sectional comparisons between trained musician and nonmusician listeners allowed us to further assess the putative impact of long-term auditory plasticity on the trajectory of rapid perceptual learning. Our findings demonstrate that brainstem and cortical levels of auditory processing are subject to different time courses of plasticity, revealing a critical interaction between long- and short-term neural mechanisms in the context of speech-sound learning.

Materials and methods
Participants

Twenty-seven young adults (ages 18–34; mean $\pm$ SD: 23.68 ± 4.22; 13 female) participated in this study. This sample size was determined a priori to align with comparable studies investigating brainstem FFRs and cortical ERPs in musicians (Parbery-Clark et al. 2009a; Bidelman et al. 2014; Bidelman and Alain 2015b; Coffey et al. 2016) and short-term plasticity following rapid auditory perceptual learning (Alain et al. 2007; Carcagno and Plack 2011; Mankel et al. 2022). All participants had normal hearing thresholds bilaterally (pure tone average < 25 dB HL) at octave frequencies between 250 and 8,000 Hz, were fluent in American English, and reported no history of neurologic or psychiatric disorders. Participants gave written, informed consent in accordance with a protocol approved by the Indiana University Institutional Review Board.

Participants were recruited and separated into two groups based on their amount of music training. Musicians (M; $n = 13$) had ≥ 10 years of formal music training starting at or before age 12 (Wong et al. 2007). Nonmusicians (NM; $n = 14$) had ≤ 5 years of lifetime music training. Groups did not differ on age ($t(25) = 1.58$; $P = 0.413$; M: $M = 24.85$, $SD = 3.41$; NM: $M = 22.36$, $SD = 4.70$), cognitive ability as assessed through the Montreal Cognitive Assessment (Nasreddine et al. 2005) ($t(25) = 1.78$; $P = 0.088$; M: $M = 28.85$, $SD = 1.14$; NM: $M = 27.79$, $SD = 1.85$), self-reported bilingualism, ($\chi^2(1, n = 27) = 0.022$, $P = 0.883$; M: 5 bilinguals; NM: 5 bilinguals), sex balance ($\chi^2(1, n = 27) = 1.78$, $P = 0.182$; M: 6 females; NM: 10 females), nor handedness as assessed via the Edinburgh Handedness Inventory ($t(25) = -0.615$; $P = 0.544$; M: $M = 78.24$, $SD = 22.02$; NM: $M = 83.67$, $SD = 23.74$) (Oldfield 1971). As a confirmation of our group separation, musicians expectedly had ~14 more years of music training than their nonmusician peers (M: 16.1 ± 4.3 years; NM: 2.4 ± 1.7 years; $t(25) = 10.93$; $P < 0.001$).

Double-vowel stimuli and task

Concurrent vowel stimuli were modeled after previous studies (Assmann and Summerfield 1989, 1990; Alain et al. 2007; Bidelman and Yellamsetty 2017). Stimuli consisted of synthesized, steady-state vowels (/a/, /e/, and /i/), which were presented as simultaneous pairs in 3 unique combinations (i.e. /a/ + /e/; /e/ + /i/; /a/ + /i/). Vowels were never paired with themselves. Stimuli were created with a Klatt-based synthesizer (Klatt 1980) coded in MATLAB (v 2021; The MathWorks, Inc., Natick, MA). Each vowel was 100 ms in duration with 10-ms cos^2 onset/offset ramping to prevent spectral splatter. The fundamental frequency (F0) between vowels was set at 4 semitones (150 and 190 Hz),

which promotes segregation for most listeners (Assmann and Summerfield 1990; Bidelman and Yellamsetty 2017). Importantly, the high F0s of both speech tokens were well above the phase-locking limit of cortical neurons and thus ensured FFRs were of a subcortical origin (Joris et al. 2004; Brugge et al. 2009; Bidelman 2018; Gorina-Careta et al. 2021). F0 and the first two formant frequencies ($F1_{a,e,i}$ = 787, 583, 300 Hz; $F2_{a,e,i}$ = 1,307, 1,753, 2,805 Hz) remained constant for the duration of the token.

The speech sounds were presented in rarefaction phase through a TDT RZ6 interface (Tucker-Davis Technologies, Alachua, FL) controlled via MATLAB. Stimuli were presented binaurally at 79 dB SPL through electromagnetically shielded (Campbell et al. 2012; Price and Bidelman 2021) ER-2 insert earphones (Etymotic Research, Elk Grove, IL). Prior to EEG testing, we required all participants to identify single vowels with 100% accuracy. This ensured subsequent learning would be based on improvements in *concurrent* speech identification rather than isolated sound labeling, per se.

We used a clustered stimulus paradigm (Bidelman 2015b) employing interspersed fast and slow interstimulus intervals (ISIs) to simultaneously collect brainstem FFRs and cortical ERPs during the active perceptual task. Each trial consisted of 1 of the 3 vowel combinations. During a trial, 20 repetitions of the vowel pair were presented with a fast ISI of 10 ms to elicit the FFR. The ISI was then slowed to 1,100 ms, and a single stimulus was presented to evoke the ERP. Participants then identified both vowels through keyboard responses following the isolated vowel pair. The next trial began after the participants' response and 250 ms of silence. Participants were asked to identify both vowels as quickly and accurately as possible (no feedback was provided) by sequentially selecting two keys on the keyboard from a closed set of options (labeled "ah," "eh," or "ee") after each stimulus trial. Double vowel pairs were randomized in order. This identical task was repeated over 4 learning blocks. In total, each block generated 3,000 FFR trials and 150 ERP trials. Each block took 10–15 min to complete. Participants were offered a short (2–3 min) break after each block to avoid fatigue. Reaction times were recorded for each trial. RTs were calculated using a trimmed mean (250 to 6,000 ms) applied to all trials to exclude improbably short or long responses (e.g. fast guesses, lapes of attention; Bidelman and Walker 2017).

Electrophysiological recording and preprocessing

We used Curry 9 (Compumedics Neuroscan, Charlotte, NC) and BESA Research 7.1 (BESA, GmbH) to record and preprocess the continuous EEG data. Continuous EEGs were acquired from 64-channel Ag/AgCl electrodes positioned at 10–10 scalp locations (Oostenveld and Praamstra 2001). Recordings were digitized at 5 kHz using Neuroscan Synamps RT amplifiers. Data were referenced to an electrode placed 1 cm behind Cz during online recording. Data were re-referenced to linked mastoids (FFR) or common average reference (ERP) for subsequent analysis. Impedances were kept below 25 kΩ. Single electrodes were also placed on the outer canthi of the eyes and superior and inferior orbit to capture ocular movements. Eyeblinks were corrected using principal component analysis (Wallstrom et al. 2004). Responses were collapsed across vowel pairs to obtain an adequate number of trials for FFR/ERP analysis and reduce the dimensionality of the data (Alain et al. 2007; Bidelman and Yellamsetty 2017; Yellamsetty and Bidelman 2018; Price et al. 2019). Responses exceeding 150 μV were rejected as further artifacts. We then separately bandpass filtered the full-band responses from 120 to 1,500 Hz and 1 to 30 Hz (zero-phase Butterworth filters; slope = 48 dB/octave) to isolate FFRs and ERPs, respectively (Musacchia et al. 2008; Bidelman et al. 2013;

Price and Bidelman 2021). Data were then epoched (FFR: 0–105 ms; ERP: −200—1,000 ms), prestimulus baselined, and ensemble averaged to obtain speech-evoked FFR and ERP potentials for each stimulus.

Frequency-following response analysis

Brainstem responses were analyzed at the Cz electrode, where speech-FFRs are optimally recorded at the scalp (Bidelman 2015a). From FFR waveforms, we computed the Fast Fourier Transform for each block to analyze responses in the spectral domain. We measured the magnitude of the F0 response to both vowels, corresponding to the lower and higher vowel's voice pitch (i.e. 140–160 Hz and 180–200 Hz). Prior literature has shown perceptual effects in speech-FFR are largely captured at the response F0 (Price and Bidelman 2021; Carter and Bidelman 2023; Rizzi and Bidelman 2023). The ±20 Hz search window was guided by the F0 of the evoking double-vowel stimulus. Peak F0 amplitudes were averaged across vowels to derive a singular measure of FFR strength for each training block.

Event-related potential analysis

ERPs were analyzed at both the sensor and source level (Bidelman and Howell 2016). As with FFRs, scalp-level ERPs were quantified at the Cz electrode. To analyze the data at the source-level, we transformed each listener's scalp potentials into source space using BESA's Auditory Evoked Potential (AEP) virtual source montage (Scherg et al. 2002; Bidelman et al. 2018a; Mankel et al. 2020). This applied a spatial filter to all electrodes that calculated their weighted contribution to the scalp recordings. We used a 4-shell spherical volume conductor head model (Sarvas 1987; Berg and Scherg 1994) with relative conductivities (1/Ωm) of 0.33, 0.33, 0.0042, and 1 for the head, scalp, skull, and cerebrospinal fluid, respectively, and compartment sizes of 85 mm (radius), 6 mm (thickness), 7 mm (thickness), and 1 mm (thickness) (Picton et al. 1999; Herdman et al. 2002). The AEP model includes 11 regional dipoles distributed across the brain including bilateral auditory cortex (AC) [Talairach coordinates (x,y,z; in mm): $left$ = (−37, −18, 17) and $right$ = (37, −18, 17)]. Regional sources consist of 3 dipoles describing current flow (units nAm) in each cardinal plane. We extracted the time courses of the tangential component for left and right AC sources as this orientation captures the majority of variance describing the auditory cortical ERPs (Picton et al. 1999). This approach allowed us to reduce each listener's 64 channel ERP data to 2 source channels describing neuronal activity localized to the left (LH) and right hemisphere (RH) AC (Price et al. 2019; Mankel et al. 2020; Momtaz et al. 2021).

From sensor and source waveforms for each block, we measured the amplitude and latency of the P2 deflection between 130 and 170 ms. The analysis window was guided by visual inspection of the grand averaged data. We focus on the P2, as we have previously shown that this neural index is sensitive to long-term plasticity of musicianship (Bidelman et al. 2014; Mankel et al. 2020), success in double-vowel identification (Bidelman and Yellamsetty 2017), and tracks with perceptual learning during auditory categorization tasks (Mankel et al. 2022; see also Alain et al. 2007).

Statistical analysis

Unless otherwise noted, we analyzed the dependent variables using mixed-model ANOVAs in R (version 4.2.2) (R Core Team, 2020) and the *lme4* package (Bates et al. 2015). Fixed effects were block (4 levels; 1–4) and group (2 levels; musicians vs. nonmusicians). Subjects served as a random effect. Multiple comparisons
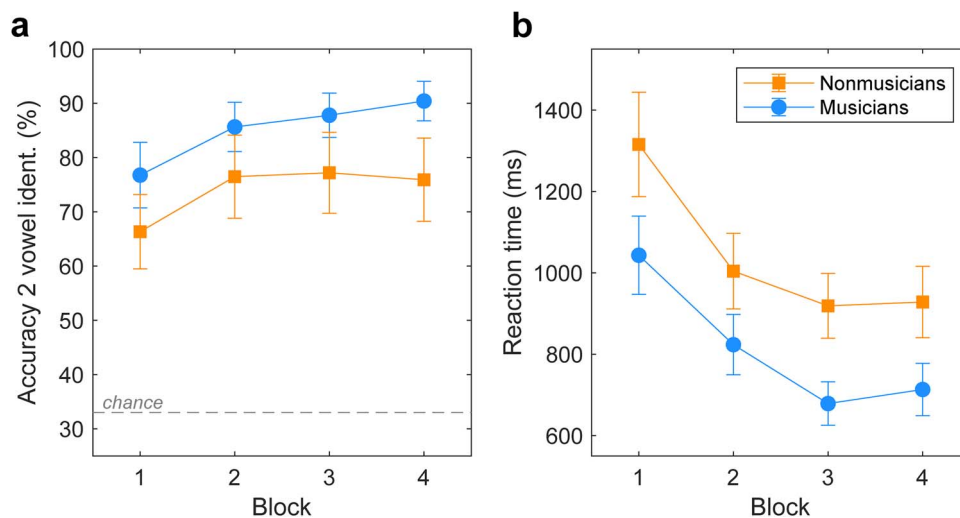
Fig. 1. Behavioral performance across blocks show rapid perceptual learning of double-vowel speech stimuli. a) Identification accuracy improved across blocks (each 10–15 min) for both groups. Chance performance on the task was 33%. b) Reaction times improved across blocks for both groups but musicians were faster overall. Error bars = ± 1.S.E.M.

were corrected via Tukey–Kramer adjustments. Effect sizes are reported as partial eta squared (η_p^2) and degrees of freedom (*d.f.*) using Satterthwaite's method. Percent correct data were arcsine-transformed to improve homogeneity of variance assumptions necessary for parametric ANOVA (Studebaker 1985). Reaction times were log-transformed for statistical analyses, though we note that untransformed RTs produced equivalent results. A priori significance level was set at $\alpha = 0.05$.

We used repeated measure correlations (rmCorr, version 0.6.0) (Bakdash and Marusich 2017) to assess brain–behavior relationships within each group. Unlike conventional correlations, rmCorrs account for nonindependence among observations (here, blocks within subjects) and measures within-subject correlations by evaluating the common intra-individual association between measures.

Results
Behavioral data
Figure 1 shows behavioral results across training blocks for both groups. RTs and identification accuracy were highly negatively correlated [$r = -0.50$, $P < 0.001$], indicating a typical time-accuracy tradeoff in concurrent vowel identification with improvements in learning (Bidelman and Yellamsetty 2017; Yellamsetty and Bidelman 2019). An ANOVA conducted on double-vowel identification accuracy (Fig. 1a) revealed a sole main effect of block [$F(3, 75) = 12.13$, $P < 0.001$, $\eta_p^2 = 0.33$]; the group [$F(1, 25) = 1.88$, $P = 0.18$, $\eta_p^2 = 0.07$] and block × group interaction effects [$F(3, 75) = 0.60$, $P = 0.61$, $\eta_p^2 = 0.02$] were not significant. The block effect was attributed to a steady increase in performance with training for both groups [linear contrast: M: $t(75) = 4.34$, $P < 0.001$; NM: $t(75) = 3.01$, $P = 0.0035$].

In contrast, reaction times (Fig. 1b) were modulated by main effects of both block [$F(3, 75) = 13.96$, $P < 0.001$, $\eta_p^2 = 0.36$] and group [$F(1, 25) = 8.93$, $P = 0.0062$, $\eta_p^2 = 0.26$] [block × group interaction: $F(3, 75) = 0.39$, $P = 0.76$, $\eta_p^2 = 0.02$]. The block effect was attributed to more rapid decision speeds at the end compared to the beginning of training [M: $t(75) = -4.34$, $P < 0.001$; NM: $t(75) = -3.75$, $P < 0.001$]. The main effect of group indicates that musicians showed faster RTs across the board. This was

further confirmed via correlational analyses, which showed that listeners' degree of musical training negatively predicted RTs; i.e. more highly trained individuals achieved faster decision speeds [$r = -0.26$, $P = 0.008$] (see Supplemental Information, Fig. S1). These data suggest that, behaviorally, all participants improved in speed and accuracy with training, but musicians had overall faster response times to speech than nonmusicians.

Subcortical responses
Grand average FFR time waveforms and spectra are shown for each group and training block in Fig. 2. FFRs showed phase-locked energy corresponding to the periodicities of both vowel stimuli. Responses showed energy at both F0s (i.e. 150 and 190 Hz) and their integer-related multiples up to the frequency ceiling of phase locking in the midbrain (~1,100 Hz) (Liu et al. 2006; Bidelman and Powers 2018). However, an ANOVA on FFR F0 amplitudes failed to reveal effects of block [$F(3, 74) = 0.53$, $P = 0.66$; $\eta_p^2 = 0.02$], group [$F(1, 25) = 0.42$, $P = 0.52$; $\eta_p^2 = 0.02$], nor their interaction [$F(3, 75) = 0.52$, $P = 0.67$, $\eta_p^2 = 0.02$]. These data suggest that brainstem speech representations were not modulated by either long- or short-term plasticity during rapid auditory perceptual learning.

Cortical responses
We analyzed cortical responses at the scalp and source levels. Figure 3a depicts scalp ERP responses at the electrode level (Cz) across the 4 blocks. Note prominent modulations in the P2 wave (~150 ms) across training blocks. This deflection was maximal near the scalp vertex and inverted at the mastoids, consistent with generators in the supratemporal plane (Picton et al. 1999; Alain et al. 2007). For P2 latencies, an ANOVA did not reveal any effects of group [$F(1, 25.1) = 0.23$, $P = 0.64$; $\eta_p^2 < 0.01$], block [$F(3, 74.2) = 2.59$, $P = 0.06$; $\eta_p^2 = 0.09$], nor group × block interaction [$F(3, 74.2) = 0.15$, $P = 0.93$; $\eta_p^2 < 0.01$]. In contrast, P2 amplitudes were strongly modulated by both factors [block × group interaction: $F(3, 74.2) = 3.88$, $P = 0.012$; $\eta_p^2 = 0.14$] (Fig. 3b) [group effect: $F(1, 25.1) = 2.15$, $P = 0.15$; $\eta_p^2 = 0.08$; block effect: $F(3, 74.2) = 0.71$, $P = 0.55$; $\eta_p^2 = 0.03$]. This interaction was attributed to Ms showing increased gains in P2 amplitude across blocks, whereas NMs' responses were invariant to learning [linear contrast: M: $t(74) = 2.70$, $P = 0.009$; NM: $t(74) = -1.59$, $P = 0.12$]. Collectively, these data suggest a
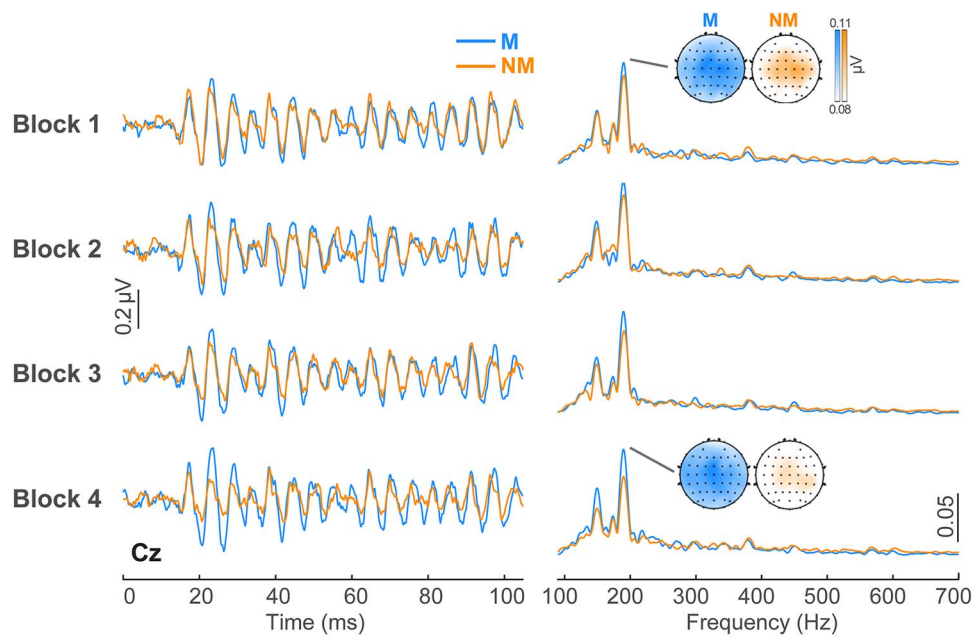
Fig. 2. Subcortical FFR responses are not sensitive to rapid perceptual learning. FFR waveforms are shown across the 4 blocks for both groups in the time (left) and frequency (right) domains. Heatmaps show response strength at 190 Hz for blocks 1 and 4, respectively. No group or block effects were observed in FFRs.

differential pattern in the strength of cortical speech encoding between groups (when measured at the scalp), with stronger training-related changes in neural processing among musically trained listeners. However, the volume-conducted nature of EEG does not allow us to adjudicate the intracranial generators underlying these effects in scalp data alone. Consequently, we used source analysis to determine if these neuroplastic changes observable in the sensor data were attributed to rapid changes within the auditory cortices themselves.

Figure 4 depicts grand average source responses for each group and a subset of blocks. Distributed imaging using Cortical Low resolution electromagnetic tomography Analysis Recursively Applied (CLARA; BESA v7.1) at the peak latency of the P2 (130–170 ms) (Iordanov et al. 2014) localized activity to bilateral AC (Fig. 4a). Source time courses from LH and RH are shown in Fig. 4b. We found training-related gains in P2 magnitude and latency that varied between groups and cerebral hemisphere (Fig. 5). Although LH source magnitudes were invariant [group main effect: $F(1, 25.07) = 0.025$, $P = 0.88$; $\eta_p^2 < 0.01$; block main effect: $F(3, 74.2) = 1.42$, $P = 0.24$; $\eta_p^2 = 0.05$; group × block: $F(3, 74.24) = 0.68$, $P = 0.57$; $\eta_p^2 = 0.03$] (Fig. 5a), LH P2 latencies varied strongly with both training block [$F(3, 73.5) = 5.28$, $P = 0.002$; $\eta_p^2 = 0.18$] and group membership [$F(1, 24) = 10.58$, $P = 0.003$; $\eta_p^2 = 0.31$] [group × block: $F(3, 73.5) = 1.64$, $P = 0.19$, $\eta_p^2 = 0.06$]. The block effect was attributed to responses becoming progressively later with training (Fig. 5b), whereas the group effect was due to NMs having earlier (~8 ms) responses compared to Ms overall.

In contrast, RH P2 magnitudes showed a critical block × group interaction [$F(3, 74.1) = 2.99$, $P = 0.036$; $\eta_p^2 = 0.11$] [block main effect: $F(3, 74.1) = 0.70$, $P = 0.55$; $\eta_p^2 = 0.03$; group main effect: $F(1, 24.9) = 0.29$, $P = 0.59$; $\eta_p^2 = 0.01$]. Whereas NMs' RH response strength remained static across blocks [$t(74) = 0.57$, $P = 0.57$], Ms' responses began more robust and became weaker with training to eventually converge with those of NMs [linear contrast: $t(74.1) = -2.35$, $P = 0.022$] (Fig. 5c). RH latencies revealed a sole main effect of group [$F(1, 25.2) = 5.52$, $P = 0.027$; $\eta_p^2 = 0.18$; block main effect: $F(3, 74.5) = 0.56$, $P = 0.64$; $\eta_p^2 = 0.02$; group × block: $F(3, 74.5) = 0.52$, $P = 0.66$; $\eta_p^2 = 0.02$] that was again attributed to

faster responses in NMs across the board (Fig. 5d). Collectively, these source data argue that the AC is sensitive to both perceptual learning for speech and prior listening experience (i.e. musicianship), which manifest in different lateralized effects in the LH vs. RH. In sum, we find that with speech-sound learning, auditory cortical processing (a) becomes prolonged in LH (though not made stronger/weaker) for all listeners; (b) is slightly delayed in Ms overall; and (c) is modulated in strength in RH among musicians.

Brain–behavior relationships

To assess correspondences between the neural and behavioral data, we performed repeated measure correlations (rmCorr) between RTs and source ERP amplitudes/latencies separately for Ms and NMs (Fig. 6). We selected these measures given their sensitivity to group differences in our main analysis. We used rmCorr to account for the within-subject correlations stemming from the repeated testing across training blocks. Interestingly, we observed different trends between groups. In the LH, RT was negatively correlated with P2 latency for NMs [$r = -0.38$, $P = 0.011$], but not for Ms [$r = -0.08$, $P = 0.61$]. In the RH, RT was negatively correlated with P2 latency for Ms [$r = -0.36$, $P = 0.025$] but not NMs [$r = 0.19$, $P = 0.21$]. No significant correlations were observed for RH P2 magnitude for either group [M: $r = 0.26$, $P = 0.11$; NM: $r = -0.18$, $P = 0.26$] (data not shown).

Discussion

By simultaneously measuring behavior and EEG during a double-vowel learning task in musicians and nonmusicians, our data reveal three primary findings: (i) although both groups successfully learned to segregate speech mixtures, musicians were overall faster in concurrent speech identification than nonmusicians; (ii) short-term plasticity related to auditory perceptual learning for speech was not observed at the subcortical level; (iii) plasticity was highly evident at the cortical level where ERP responses revealed unique hemispheric asymmetries suggestive of different neural strategies between groups. Our findings demonstrate sub- vs. neocortical levels of auditory processing are subject to
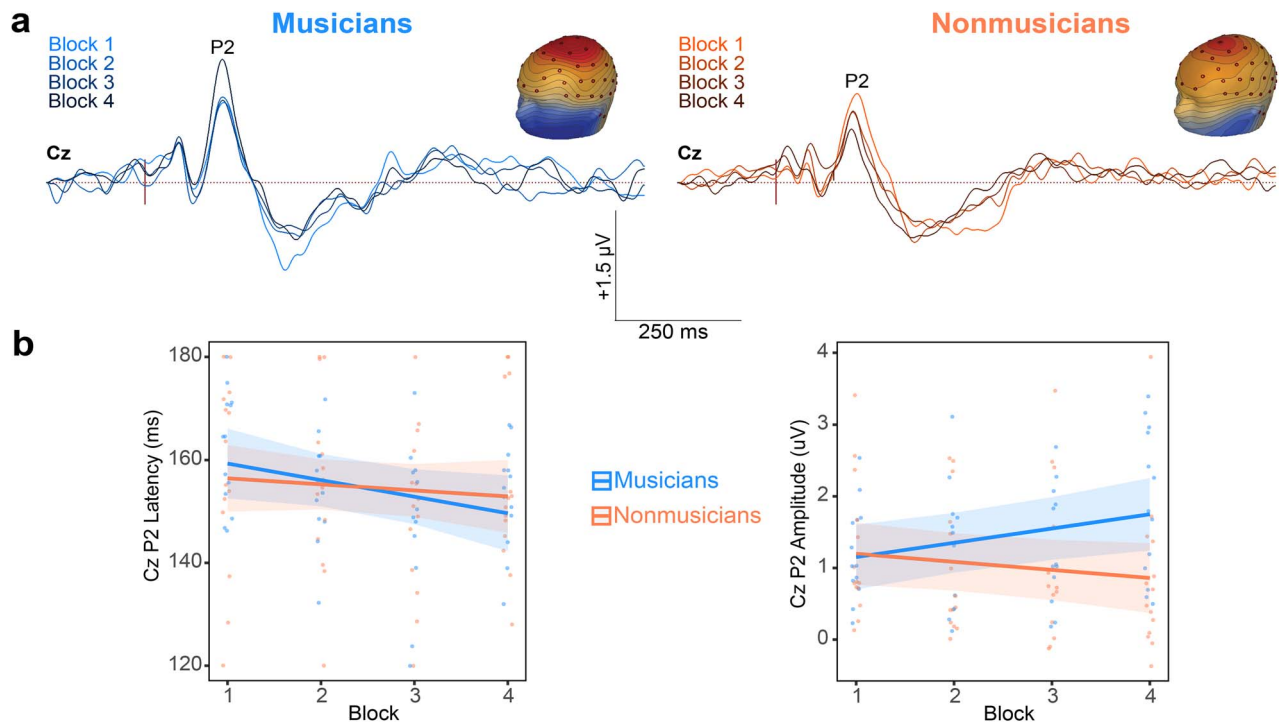
Fig. 3. Cortical ERP responses (electrode level) are sensitive to rapid perceptual learning. a) Grand average responses at Cz are displayed for both groups across the 4 blocks. Heatmaps show the scalp topography of P2 for block 4. b) P2 amplitude and latency across blocks per group. Musicians showed increasing response strength with training, whereas nonmusicians responses did not. Shading = ± 95% CI.
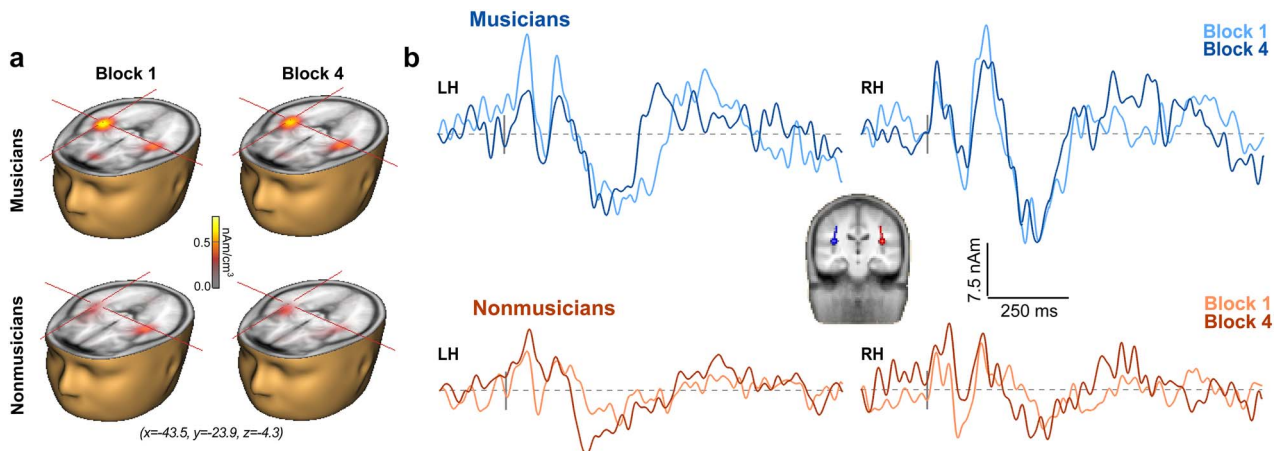

Fig. 4. Speech coding in AC varies differentially with perceptual learning in musicians vs. nonmusicians. a) Brain volumes show distributed source activation maps using CLARA imaging at the peak latency of the P2 (130–170 ms). Functional data are overlaid on the BESA template brain. The crosshair demarks a representative voxel in RH primary auditory cortex (PAC; Talairach coordinates). b) Source waveform time courses extracted from left (LH) and right (RH) hemisphere AC show different response trajectories as a result of learning between groups.

different time courses of plasticity and reveal a critical interaction between long- and short-term neural mechanisms with regard to speech-sound learning.

Short-term plasticity from perceptual learning differs with long-term auditory experience

Our behavioral data replicate findings of Alain et al. (2007) demonstrating ~45 min of training leads to rapid perceptual learning in deciphering speech mixtures. However, we extend these prior results by demonstrating faster overall reaction times for musicians relative to nonmusicians in double vowel identification across training blocks. This finding agrees with earlier data suggesting enhanced speech categorization in trained musicians (Elmer et al. 2012; Bidelman et al. 2014;

Bidelman and Walker 2019) and highly musical novices (Mankel et al. 2020). Previous evidence suggests that musicianship automatizes categorical perception earlier in the auditory-linguistic brain networks subserving sound-to-label formation (Bidelman and Walker 2019), resulting in domain-general benefits applicable to speech processing. The faster behavioral responses observed here for musicians also align with previous findings (Schneider et al. 2002; Chartrand and Belin 2006; Bidelman et al. 2014; Bidelman and Walker 2019) by suggesting that stronger category representation in early auditory cortical structures might dictate the speed and trajectory of how listeners acquire speech labels during novel learning. This notion is further bolstered by the correlations we find between listeners' RTs and source responses localized to primary AC.
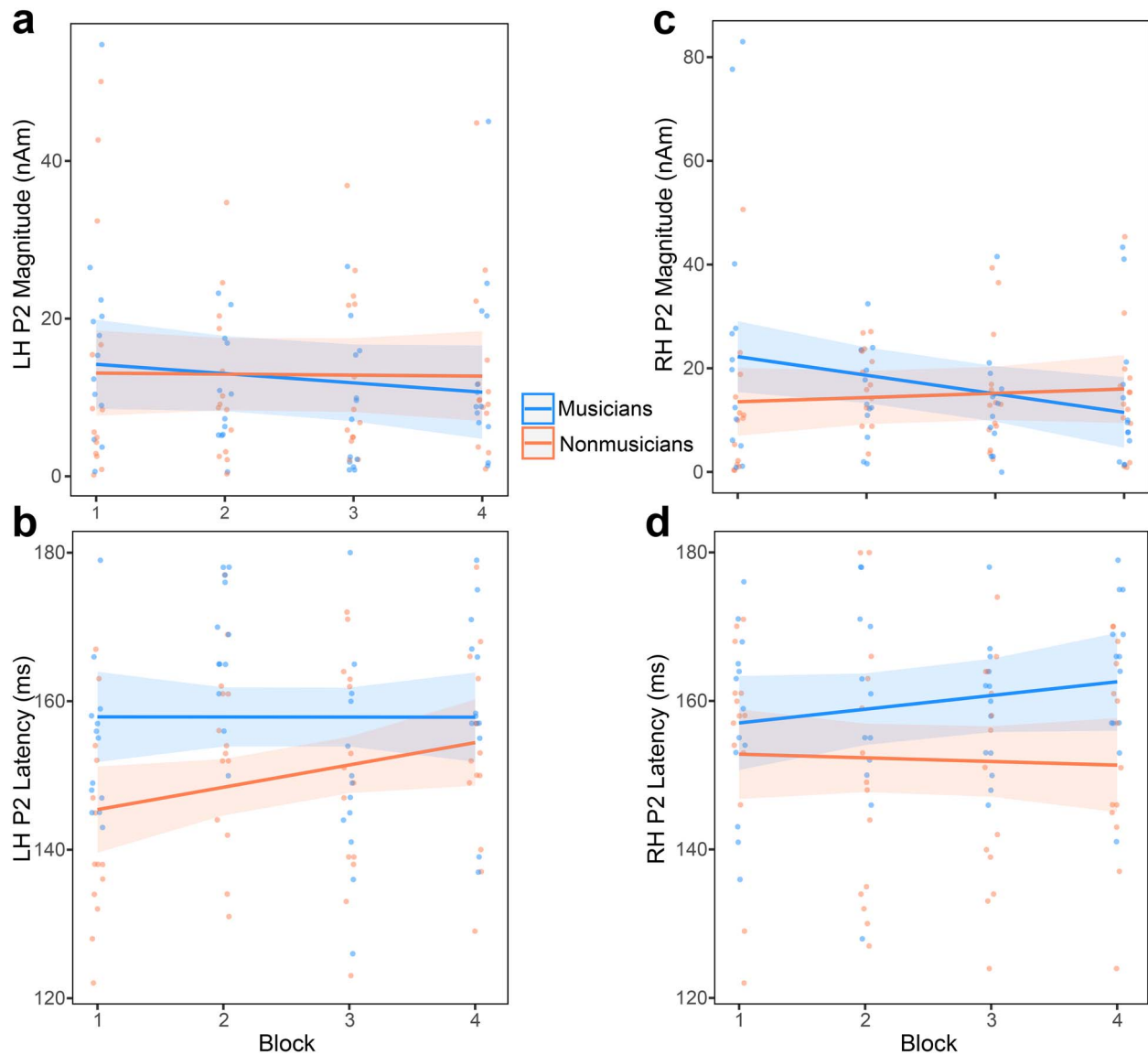
Fig. 5. Latency and amplitude of source responses vary with group and hemisphere. a) No effects were observed for LH P2 magnitude. b) LH responses increased in latency with training, and NM had earlier responses than M overall. c) RH magnitude remained static for NM, while M decreased across blocks. d) NMs displayed earlier RH responses than Ms shading = ± 95% CI.

Several previous studies have shown benefits of musicianship for "cocktail party" speech listening tasks (Bidelman and Krishnan 2010; Parbery-Clark et al. 2011; Swaminathan et al. 2015; Clayton et al. 2016; Coffey et al. 2017; Deroche et al. 2017; Du and Zatorre 2017; Mankel and Bidelman 2018; Torppa et al. 2018; Yoo and Bidelman 2019; Maillard et al. 2023). The double vowel identification task used here serves as an extension of cocktail-party listening as it requires accurate identification of concurrent speech tokens. We found superior RTs (but not accuracy) in concurrent speech identification among musicians. The fact that group benefits were limited to speed may be due to the relative simplicity of our task. While accuracy was not at ceiling, participants' performance was largely successful (>70%) even in the first training block. Concomitant neural changes notwithstanding, it is possible our task may not have "taxed" the perceptual system enough to elucidate strong behavioral differences beyond those seen in RT speeds. Nevertheless, our perceptual data alone support domain-general benefits of musicianship on the sound-to-label mapping process inherent to speech perception (Patel and Iversen 2007).

As in all studies investigating the effects of musicianship with cross-sectional designs, our data cannot definitively isolate the effects of music *training* from possible genetic or environmental predispositions, such as socioeconomic status, that might drive putative group differences; additionally, possible publication biases and flawed experimental designs add to the difficulty of determining specific music training effects in the extant literature (Sala and Gobet 2020; Neves et al. 2022; Schellenberg and Lima 2024). This makes it difficult to tease apart whether the observed group effects in musicians result from nature or nurture and whether they are due to neuroplastic effects or innate abilities (or combination of both; Mankel and Bidelman 2018). Indeed, some individuals without music training have high levels of musicality and improved speech/SIN processing ("musical sleepers"), whereas other individuals can have substantial music training but perform poorly on musical aptitude and speech tasks ("sleeping musicians") (Mankel and Bidelman 2018). Although our current data cannot resolve this issue, we demonstrate differences of behavioral and neural performance on SIN tasks
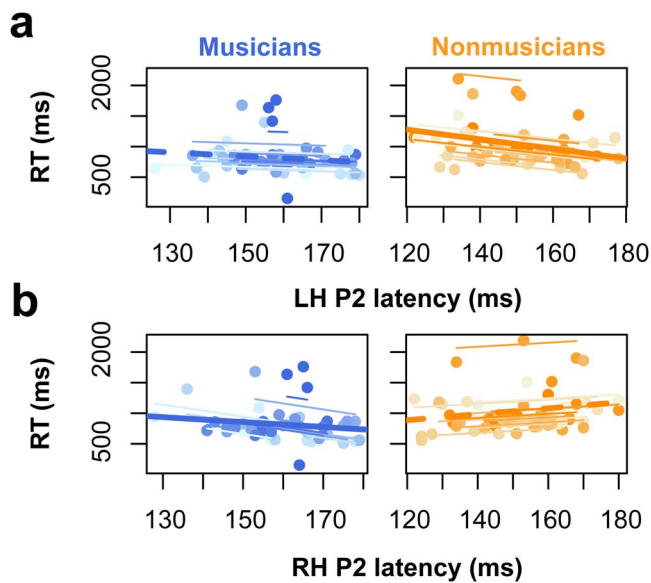
Fig. 6. Correlations between neural and behavioral measures. Repeated measures correlations (rmCorr) (Bakdash and Marusich 2017) show the within-subject relation between measures for each listener (single points and thin lines) and overall association for the aggregate sample (thick lines). a) RT was negatively correlated with LH P2 latency for NMs (but not for Ms). b) RT was negatively correlated with RH P2 latency for Ms (but not for NMs). Solid trend lines = significant correlations; dotted trend lines = *n.s.*

between musicians and nonmusicians, which differ significantly in the amount of music training but do not differ on other important demographic variables such as age, cognitive status, or language experience. Correlations between listeners' music engagement and improved speech behaviors also support this notion (Fig. S1). Still, future longitudinal studies are needed to further explore the interactions between short-term plasticity and long-term effects of music training on SIN perception. In this vein, randomized longitudinal studies have begun to demonstrate neural and behavioral domain-general benefits of music training for speech perception (Kraus et al. 2014; Slater et al. 2014; Slater et al. 2015), which support our cross-sectional findings herein.

Neural correlates of auditory perceptual learning are absent in subcortex

In stark contrast to the cortical ERPs (present study; Alain et al. 2007), 45 min of training failed to induce rapid plasticity in brainstem FFRs. The lack of training-related gains in the FFR is unlikely due to differences in the sensitivity of measurement or noise level across the two classes of response. First, FFRs were evoked by an order of magnitude more trials (FFR: several 1,000 vs. ERP: several 100) and so had considerably better signal to noise ratio than the ERPs. Moreover, although test–retest data show both cortical and brainstem evoked potentials are highly repeatable, FFR measures are considerably more stable and yield less inter- and intra-subject variability than their cortical counterparts (Bidelman et al. 2018b). Instead, our data suggest sensory enhancements in brainstem auditory processing are neither necessary nor sufficient to yield learning-related reorganization in the AC over the same (short) time course of training (present study; Alain et al. 2007). This notion aligns with recent studies suggesting cortical changes precede those in the brainstem (Reetzke et al. 2018; Skoe et al. 2021) by several days/weeks, as well as theoretical accounts that learning proceeds in a top-down guided manner (Ahissar and Hochstein 2004), with sensory

change in the brainstem only emerging at expert (or perhaps overlearned) stages of learning (Reetzke et al. 2018).

Our data also failed to reveal FFR differences between musicians and nonmusicians. This contrasts with several studies that have shown enhanced F0-pitch encoding in speech-FFRs of musicians (Musacchia et al. 2007; Bidelman et al. 2011a; Bidelman et al. 2011b; Bidelman et al. 2014; Coffey et al. 2016), though not always consistently (e.g. see Strait et al. 2012; Bidelman and Alain 2015b; Dawson et al. 2018). Indeed, at the behavioral level, musicians are not always better at exploiting F0 cues for voice segregation than their nonmusician peers (Deroche et al. 2017). The surprising lack of FFR group differences in the present data could be due to the relatively high F0 of our stimuli (> 150 Hz). The FFR contains multiple subcortical and cortical generators along the auditory pathway whose independent contributions vary in a stimulus-dependent matter (Tichko and Skoe 2017; Bidelman 2018; Coffey et al. 2019; Gorina-Careta et al. 2021). High F0 voice pitch stimuli (like those used here) minimize cortical contributions to the FFR (Bidelman 2018) and result in a dominantly brainstem-centric response that largely reflects exogenous processing of double-vowel stimuli (Yellamsetty and Bidelman 2019). Previous studies that have found musician encoding advantages and learning-related effects in the FFR have also used much lower F0s (~100 Hz) (e.g. Song et al. 2008; Carcagno and Plack 2011; Chandrasekaran et al. 2012; Reetzke et al. 2018), which may have reflected cortical rather than subcortical plasticity, per se (Coffey et al. 2019). It is also possible that auditory plasticity is stronger and emerges earlier at cortical relative to brainstem levels (e.g. Reetzke et al. 2018; Skoe et al. 2021; Bidelman et al. 2022; Lai et al. 2022) and varies in a stimulus-specific manner (Holmes et al. 2018). Nascent changes in the brainstem FFR might therefore require more protracted training than the short learning tasks used here. Indeed, lasting changes in the neural differentiation of speech, as indexed by the FFR, are observable no earlier than several days of training (Song et al. 2008; Reetzke et al. 2018) and even one year by some accounts (Kraus et al. 2014).

Additionally, prior studies demonstrating musician F0 benefits have exclusively used passive listening tasks. Attention varies with musicianship (Strait et al. 2010; Yoo and Bidelman 2019) and is known to enhance the speech FFR (Price and Bidelman 2021; Lai et al. 2022; Carter and Bidelman 2023). Consequently, experience-dependent effects of music on FFR strength might be more muted under states of active attentional engagement if nonmusicians deploy more attentional resources during speech processing. This notion is indeed supported by the longer RTs and more invariant cortical P2 across blocks we find in nonmusicians, which suggest our behavioral task was more demanding and/or recruited more attentional resources in this group. Even so, participants were able to accomplish our task early in the training regimen, and therefore change in sensory representation at a subcortical level was perhaps unnecessary for task success. Instead, we observe salient changes in RTs and ERP neural timing, suggesting that learning in our task probably reflected improved *access to*, rather than strength of, sensory representation at a cortical level of processing (Binder et al. 2004; Bidelman et al. 2014).

Neural correlates of perceptual learning are robust in cortex and reflect different neural strategies between musicians and nonmusicians

Both our sensor and source ERP data were consistent in showing stronger learning-related neural changes in musicians. In particular, hemispheric PAC activity suggested distinct neural strategies with learning based on prior auditory experience:

behavioral RTs correlated with RH P2 latency in musicians but LH P2 latency in nonmusicians. Given musicians' faster overall behavioral speed, the double-dissociation in hemispheric latencies between groups may point to superior "cue-weighting" by musicians towards pitch-related cues (Zatorre et al. 1992). Relationships between behavioral decision speeds and right hemispheric learning patterns for musicians could reflect their focus on pitch interval features (i.e. "musical" content) between vowels, whereas nonmusicians' left hemispheric pattern may reflect heavier reliance on linguistic information (Mankel et al. 2022). Though our task relies on segregation and identification of speech tokens, musicians' stronger learning-related changes in RH may indicate a focus on pitch rather than linguistic information of the speech stimuli (Alain et al. 2005), indicating distinct task strategies as a result of long-term experience in music.

A rightward biased mechanism in musicians is additionally supported by their decreased RH P2 magnitude with training. Declines in P2 strength with learning is consistent with other single-session, short-term learning experiments in which sensory-evoked neural responses become more efficient during active task engagement (Guenther et al. 2004; Alain et al. 2010; Ben-David et al. 2011; Pérez-Gay Juárez et al. 2019; Mankel et al. 2022). Our results converge with other studies showing a reduction (habituation) in cortical responses following rapid auditory perceptual learning especially in highly skilled listeners (e.g. musicians) (Seppänen et al. 2012). Stronger engagement of RH may also indicate increased attention to frequency-related information (Crowley and Colrain 2004). P2 latency, especially in RH, better differentiates phonetic speech categories in nonmusicians with higher musicality (Mankel et al. 2020), paralleling our findings here in trained musicians. Our results are also consistent with prior neuroimaging work suggesting musicians and nonmusicians process speech categories by differentially engaging different nodes of the auditory-linguistic network for otherwise identical perceptual tasks. For example, whereas musicians regulate speech coding to relatively early auditory areas (e.g. PAC), nonmusicians recruit additional downstream brain mechanisms (e.g. inferior frontal regions; Broca's area) to decode the same speech category labels (Bidelman and Walker 2019).

Somewhat surprisingly, we found learning-related changes in P2 timing were *negatively* correlated with behavioral RTs (in both groups though in opposite cerebral hemispheres). That is, later neural responses in PAC predicted faster speeds in double-vowel identification. The direction of this effect is not immediately apparent, as we would have expected faster RTs to correlate with earlier P2 responses. With respect to the direction of P2 modulation with learning, the literature has been somewhat equivocal. Different experiments have reported changes in evoked response in seemingly opposite directions (Tremblay et al. 2001; Atienza et al. 2002; Bosnyak et al. 2004; Sheehan et al. 2005; Zhang et al. 2005; Tong et al. 2009; Alain et al. 2010; Ben-David et al. 2011; Carcagno and Plack 2011; Ross et al. 2013; Wisniewski et al. 2020). It is possible that the counterintuitive earlier cortical responses—as we observe in nonmusicians—reflect increased arousal during task engagement. Indeed, RTs follow a U-shape with changes in arousal level such that they are fastest at intermediate levels and deteriorate (slow) in overly relaxed or tense states (Broadbent 1971; Welford 1980). Similarly, earlier P2 latency has been associated with more aroused and wakeful states of attention (Crowley and Colrain 2004). Consequently, it is possible NMs were more taxed during the rapid speech identification task leading to increased arousal that manifested in their longer behavioral RTs and counterintuitively earlier P2 responses.

Variations in arousal might also explain the negative RT-P2 relations we find in both groups. However, we also note the P2 itself reflects multiple sources with subcomponents in Heschl's gyrus, planum temporale, and surrounding auditory associations in both hemispheres (Steinmetzger and Rupp 2023). Thus, it is also possible the hemispheric differences we find in RT-P2 relations between groups reflect unique engagement of these multiple P2 generators that are not captured by our single dipole foci.

Interplay between short- and long-term plasticity in early auditory cortex

The P2 component of the ERPs occurs relatively early in the auditory cortical hierarchy (∼150 ms after stimulus onset). The learning-related changes seen in our data agree with previous studies showing associations between P2 and speech discrimination (Alain et al. 2010; Ben-David et al. 2011), sound object identification (Leung et al. 2013; Ross et al. 2013), and early speech category representation (Bidelman et al. 2013; Bidelman and Lee 2015; Alho et al. 2016; Bidelman and Walker 2019; Mankel et al. 2020). Interestingly, we show differential trajectories of neuroplasticity in sound encoding with rapid learning as a function of previous auditory experience. Musicians responded faster to vowel pairs and displayed greater learning-related changes compared to nonmusicians' at a cortical level. This suggests that the long-term auditory experience of musicianship might act as a catalyst for novel sound learning that would be highly relevant in other domains (e.g. second language learning; Slevc and Miyake 2006; Seppänen et al. 2012; Chobert and Besson 2013; Picciotti et al. 2018). Our results broadly align with the findings of Seppänen et al. (2012) who showed a similar reduction (habituation) in the attention-related P3b during auditory perceptual leaning in musicians, but not nonmusicians. We extend these findings by demonstrating similar short and long-term neuroplastic interaction in earlier auditory sensory processing indexed by the P2. We argue the early nature of these effects in waves that localize to auditory-perceptual areas (and well before motor responses) suggests the rapid plasticity we observe with auditory learning can be attributed to changes in sensory encoding rather than later procedural learning (Alain et al. 2007; Mankel et al. 2022). Indeed, neither task familiarity (i.e. procedural learning) nor stimulus repetition alone are sufficient to produce changes in the early cortical ERPs (Alain et al. 2007; Mankel et al. 2022).

Despite clear musician advantages at behavioral and cortical levels, our speech learning task was not sufficient to observe subcortical changes, which have been observed in longer training regimens over several sessions and days (Song et al. 2008; Carcagno and Plack 2011; Reetzke et al. 2018). Future studies using more difficult tasks and longer learning paradigms should be conducted to determine the dosage of training needed to induce learning-related plasticity at brainstem vs. cortical levels of the auditory pathway (cf. Reetzke et al. 2018), along with the effects of consolidation to learning gains (cf. Alain et al. 2015). More broadly, understanding the differential timelines in plasticity for speech coding resulting from musicianship would further support the use of music-based interventions to enhance speech and language outcomes.

Acknowledgments

Author contributions

Jessica MacLean (Design, Data Collection, Data Analysis, Writing), Jack Stirn (Data Collection, Data Analysis, Writing), Alexandria Sisson (Data Collection, Writing), and Gavin M. Bidelman (Design, Data Collection, Data Analysis, Writing).

Supplementary material

Supplementary material is available at *Cerebral Cortex* online.

Funding

References

Ahissar M, Hochstein S. The reverse hierarchy theory of visual perceptual learning. *Trend Cogn Sci*. 2004:8(10):457–464.

Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin F-H, Raij T, Sams M, Vasios CE, Belliveau JW. Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc Natl Acad Sci U S A*. 2011:108(10): 4182–4187.

Alain C, Reinke K, McDonald KL, Chau W, Tam F, Pacurar A, Graham S. Left thalamo-cortical network implicated in successful speech separation and identification. *Neuroimage*. 2005:26(2):592–599.

Alain C, Snyder JS, He Y, Reinke KS. Changes in auditory cortex parallel rapid perceptual learning. *Cereb Cortex*. 2007:17(5):1074–1084.

Alain C, Campeanu S, Tremblay K. Changes in sensory evoked responses coincide with rapid improvement in speech identification performance. *J Cogn Neurosci*. 2010:22(2):392–403.

Alain C, Zendel BR, Hutka S, Bidelman GM. Turning down the noise: the benefit of musical training on the aging auditory brain. *Hear Res*. 2014:308:162–173.

Alain C, Zhu KD, He Y, Ross B. Sleep-dependent neuroplastic changes during auditory perceptual learning. *Neurobiol Learn Mem*. 2015:118:133–142.

Alain C, Moussard A, Singer J, Lee Y, Bidelman GM, Moreno S. Music and visual art training modulate brain activity in older adults. *Front Neurosci*. 2019:13(182):1–15.

Alho J, Green BM, May PJC, Sams M, Tiitinen H, Rauschecker JP, Jääskeläinen IP. Early-latency categorical speech sound representations in the left inferior frontal gyrus. *Neuroimage*. 2016:129: 214–223.

Arehart KH, King CA, McLean-Mudgett KS. Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res*. 1997:40(6):1434–1444.

Assmann PF, Summerfield Q. Modeling the perception of concurrent vowels: vowels with the same fundamental frequency. *J Acoust Soc of Am*. 1989:85(1):327–338.

Assmann PF, Summerfield Q. Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *J Acoust Soc of Am*. 1990:88(2):680–697.

Atienza M, Cantero JL. Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Res*. 2001:901(1–2):151–160.

Atienza M, Cantero JL, Dominguez-Marin E. The time course of neural changes underlying auditory perceptual learning. *Learn Mem*. 2002:9(3):138–150.

Bakdash JZ, Marusich LR. Repeated measures correlation. *Front Psychol*. 2017:8:456–456.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015:67(1):1–48.

Ben-David BM, Campeanu S, Tremblay KL, Alain C. Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology*. 2011:48(6): 797–807.

Berg P, Scherg M. A fast method for forward computation of multiple-shell spherical head models. *Electroencephalogr Clin Neurophysiol*. 1994:90(1):58–64.

Bidelman GM. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient abr. *Hear Res*. 2015a:323: 68–80.

Bidelman GM. Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J Neurosci Methods*. 2015b:241:94–100.

Bidelman GM. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*. 2018:175:56–69.

Bidelman GM, Alain C. Hierarchical neurocomputations underlying concurrent sound segregation: connecting periphery to percept. *Neuropsychologia*. 2015a:68:38–50.

Bidelman GM, Alain C. Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci*. 2015b:35(2):1240–1249.

Bidelman GM, Howell M. Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *NeuroImage*. 2016:124(Pt A):581–590.

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

Bidelman GM, Lee C-C. Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage*. 2015:120:191–200.

Bidelman G, Powers L. Response properties of the human frequency-following response (FFR) to speech and non-speech sounds: level dependence, adaptation and phase-locking limits. *Int J Audiol*. 2018:57(9):665–672.

Bidelman GM, Walker B. Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *Eur J Neurosci*. 2017:45(5):690–699.

Bidelman GM, Walker BS. Plasticity in auditory categorization is supported by differential engagement of the auditory-linguistic network. *Neuroimage*. 2019:201(116022):1–10.

Bidelman GM, Yellamsetty A. Noise and pitch interact during the cortical segregation of concurrent speech. *Hear Res*. 2017:351: 34–44.

Bidelman GM, Yoo J. Musicians show improved speech segregation in competitive, multi-talker cocktail party scenarios. *Front Psychol*. 2020:11:1927.

Bidelman GM, Gandour JT, Krishnan A. Musicians demonstrate experience-dependent brainstem enhancement of musical scale features within continuously gliding pitch. *Neurosci Lett*. 2011a:503(3):203–207.

Bidelman GM, Krishnan A, Gandour JT. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci*. 2011b:33(3):530–538.

Bidelman GM, Moreno S, Alain C. Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*. 2013:79(1):201–212.

Bidelman GM, Weiss MW, Moreno S, Alain C. Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci*. 2014:40(4): 2662–2673.

Bidelman GM, Davis MK, Pridgen MH. Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear Res*. 2018a:367:149–160.

Bidelman GM, Pousson M, Dugas C, Fehrenbach A. Test-retest reliability of dual-recorded brainstem vs. cortical auditory evoked potentials to speech. *J Am Acad Audiol*. 2018b:29(02):164–174.

Bidelman GM, Chow R, Noly-Gandon A, Ryan JD, Bell KL, Rizzi R, Alain C. Transcranial direct current stimulation combined with listening to preferred music alters cortical speech processing in older adults. *Front Neurosci*. 2022:16:884130.

Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD. Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci*. 2004:7(3):295–301.

Bosnyak DJ, Eaton RA, Roberts LE. Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb Cortex*. 2004:14(10):1088–1099.

Broadbent DE. *Decision and stress*. London: Academic Press; 1971.

Brown JA, Bidelman GM. Familiarity of background music modulates the cortical tracking of target speech at the "cocktail party". *Brain Sci*. 2022:12(10):1320.

Brugge JF, Nourski KV, Oya H, Reale RA, Kawasaki H, Steinschneider M, Howard MA 3rd. Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J Neurophysiol*. 2009:102(4):2358–2374.

Campbell T, Kerlin JR, Bishop CW, Miller LM. Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography. *Ear Hear*. 2012:33(1):144–150.

Carcagno S, Plack CJ. Subcortical plasticity following perceptual learning in a pitch discrimination task. *J Assoc Res Otolaryngol*. 2011:12(1):89–100.

Carter JA, Bidelman GM. Perceptual warping exposes categorical representations for speech in human brainstem responses. *NeuroImage*. 2023:269:119899.

Chandrasekaran B, Kraus N, Wong PC. Human inferior colliculus activity relates to individual differences in spoken language learning. *J Neurophysiol*. 2012:107(5):1325–1336.

Chartrand JP, Belin P. Superior voice timbre processing in musicians. *Neurosci Lett*. 2006:405(3):164–167.

Chintanpalli A, Heinz MG. The use of confusion patterns to evaluate the neural basis for concurrent vowel identification. *J Acoust Soc of Am*. 2013:134(4):2988–3000.

Chintanpalli A, Ahlstrom JB, Dubno JR. Effects of age and hearing loss on concurrent vowel identification. *J Acoust Soc of Am*. 2016:140(6): 4142–4153.

Chobert J, Besson M. Musical expertise and second language learning. *Brain Sci*. 2013:3(2):923–940.

Clayton KK, Swaminathan J, Yazdanbakhsh A, Zuk J, Patel AD, Kidd G Jr. Executive function, visual attention and the cocktail party problem in musicians and non-musicians. *PLoS One*. 2016:11(7):e0157638.

Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun*. 2016:7(1):11070.

Coffey EBJ, Mogilever NB, Zatorre RJ. Speech-in-noise perception in musicians: a review. *Hear Res*. 2017:352:49–69.

Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N. Evolving perspectives on the sources of the frequency-following response. *Nat Commun*. 2019:10(1):5036.

Crowley KE, Colrain IM. A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clin Neurophysiol*. 2004:115(4):732–744.

Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J Neurosci*. 2013:33(5):1858–1863.

Dawson C, Tervaniemi M, Aalto D. Behavioral and subcortical signatures of musical expertise in Mandarin Chinese speakers. *PLoS One*. 2018:13(1):e0190793.

Deroche MLD, Limb CJ, Chatterjee M, Gracco VL. Similar abilities of musicians and non-musicians to segregate voices by fundamental frequency. *J Acoust Soc of Am*. 2017:142(4):1739–1755.

Du Y, Zatorre RJ. Musical training sharpens and bonds ears and tongue to hear speech better. *Proc Natl Acad Sci U S A*. 2017:114(51): 13579–13584.

Dubinsky E, Wood EA, Nespoli G, Russo FA. Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Front Neurosci*. 2019:13:13.

Dyson BJ, Alain C. Representation of concurrent acoustic objects in primary auditory cortex. *J Acoust Soc of Am*. 2004:115(1):280–288.

Elmer S, Meyer M, Jancke L. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. *Cereb Cortex*. 2012:22(3): 650–658.

Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*. 2003:6(11):1216–1223.

Gorina-Careta N, Kurkela JLO, Hämäläinen J, Astikainen P, Escera C. Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (MEG) study. *Neuroimage*. 2021:231:117866.

Guenther FH, Nieto-Castanon A, Ghosh SS, Tourville JA. Representation of sound categories in auditory cortical maps. *J Speech Lang Hear Res*. 2004:47(1):46–57.

Herdman AT, Lins O, van Roon P, Stapells DR, Scherg M, Picton T. Intracerebral sources of human auditory steady-state responses. *Brain Topogr*. 2002:15(2):69–86.

Herholz SC, Zatorre RJ. Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron*. 2012:76(3): 486–502.

Holmes E, Purcell DW, Carlyon RP, Gockel HE, Johnsrude IS. Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. *J Assoc Res Otolaryngol*. 2018:19(1):83–97.

Iordanov T, Hoechstetter K, Berg P, Paul-Jordanov I, Scherg M. *Clara: classical loreta analysis recursively applied*. Hamburg, Germany: Paper presented at OHBM; 2014.

Jeng FC, Hu J, Dickman B, Montgomery-Reagan K, Tong M, Wu G, Lin CD. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear Hear*. 2011:32(6):699–707.

Joris PX, Schreiner CE, Rees A. Neural processing of amplitude-modulated sounds. *Physiol Rev*. 2004:84(2):541–577.

Klatt DH. Software for a cascade/parallel formant synthesizer. *J Acoust Soc of Am*. 1980:67(3):971–995.

Kraus N, Chandrasekaran B. Music training for the development of auditory skills. *Nat Rev Neurosci*. 2010:11(8):599–605.

Kraus N, Slater J, Thompson EC, Hornickel J, Strait DL, Nicol T, White-Schwoch T. Music enrichment programs improve the neural encoding of speech in at-risk children. *J Neurosci*. 2014:34(36): 11913–11918.

Krishnan A, Gandour JT, Bidelman GM. The effects of tone language experience on pitch processing in the brainstem. *J Neurolinguistics*. 2010:23(1):81–95.

Krishnan A, Gandour JT, Ananthakrishnan S, Bidelman GM, Smalt CJ. Linguistic status of timbre influences pitch encoding in the brainstem. *Neuroreport*. 2011:22(16):801–803.

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992:255(5044):606–608.

Lai J, Price CN, Bidelman GM. Brainstem speech encoding is dynamically shaped online by fluctuations in cortical α state. *Neuroimage*. 2022:263:119627.

Leung AWS, He Y, Grady CL, Alain C. Age differences in the neuroelectric adaptation to meaningful sounds. *PLoS One*. 2013:8(7):e68892.

Liu LF, Palmer AR, Wallace MN. Phase-locked responses to pure tones in the inferior colliculus. *J Neurophysiol*. 2006:95(3):1926–1935.

Lo CY, Looi V, Thompson WF, McMahon CM. Music training for children with sensorineural hearing loss improves speech-in-noise perception. *J Speech Lang Hear Res*. 2020:63(6):1990–2015.

Maillard E, Joyal M, Murray MM, Tremblay P. Are musical activities associated with enhanced speech perception in noise in adults? A systematic review and meta-analysis. *Curr Res Neurobiol*. 2023:4:100083.

Mankel K, Bidelman GM. Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc Natl Acad Sci U S A*. 2018:115(51):13129–13134.

Mankel K, Barber J, Bidelman GM. Auditory categorical processing for speech is modulated by inherent musical listening skills. *Neuroreport*. 2020:31(2):162–166.

Mankel K, Shrestha U, Tipirneni-Sajja A, Bidelman GM. Functional plasticity coupled with structural predispositions in auditory cortex shape successful music category learning. *Front Neurosci*. 2022:16:16.

Momtaz S, Moncrieff D, Bidelman GM. Dichotic listening deficits in amblyaudia are characterized by aberrant neural oscillations in auditory cortex. *Clin Neurophysiol*. 2021:132(9):2152–2162.

Moreno S, Bidelman GM. Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear Res*. 2014:308:84–97.

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(40):15894–15898.

Musacchia G, Strait D, Kraus N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res*. 2008:241(1–2):34–42.

Nasreddine ZS, Phillips NA, Bedirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MOCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005:53(4):695–699.

Neves L, Correia AI, Castro SL, Martins D, Lima CF. Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence. *Neurosci Biobehav Rev*. 2022:140:104777.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971:9(1):97–113.

Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol*. 2001:112(4):713–719.

Ozaki I, Jin CY, Suzuki Y, Baba M, Matsunaga M, Hashimoto I. Rapid change of tonotopic maps in the human auditory cortex during pitch discrimination. *Clin Neurophysiol*. 2004:115(7):1592–1604.

Palmer AR. 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 of Am*. 1990:88(3):1412–1426.

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(45):14100–14107.

Parbery-Clark A, Skoe E, Lam C, Kraus N. Musician enhancement for speech-in-noise. *Ear Hear*. 2009b:30(6):653–661.

Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N. Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS One*. 2011:6(5):e18082.

Patel AD. Why would musical training benefit the neural encoding of speech? The opera hypothesis. *Front Psychol*. 2011:2:142.

Patel AD. Can nonlinguistic musical training change the way the brain processes speech? *The expanded opera hypothesis Hear Res*. 2014:308:98–108.

Patel AD, Iversen JR. The linguistic benefits of musical abilities. *Trends Cogn Sci*. 2007:11(9):369–372.

Pérez-Gay Juárez F, Sicotte T, Thériault C, Harnad S. Category learning can alter perception and its neural correlates. *PLoS One*. 2019:14(12):e0226000.

Picciotti PM, Bussu F, Calò L, Gallus R, Scarano E, di Cintio G, Cassarà F, D'Alatri L. Correlation between musical aptitude and learning foreign languages: an epidemiological study in secondary school Italian students. *Acta Otorhinolaryngol Ital*. 2018:38(1):51–55.

Picton TW, Alain C, Woods DL, John MS, Scherg M, Valdes-Sosa P, Bosch-Bayard J, Trujillo NJ. Intracerebral sources of human auditory-evoked potentials. *Audiol Neurootol*. 1999:4(2):64–79.

Price CN, Bidelman GM. Attention reinforces human corticofugal system to aid speech perception in noise. *Neuroimage*. 2021:235:118014.

Price CN, Alain C, Bidelman GM. Auditory-frontal channeling in α and β bands is altered by age-related hearing loss and relates to speech perception in noise. *Neuroscience*. 2019:423:18–28.

Puschmann S, Baillet S, Zatorre RJ. Musicians at the cocktail party: neural substrates of musical training during selective listening in multispeaker situations. *Cereb Cortex*. 2018:29(8): 3253–3265.

R-Core-Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. https://www.R-project.org/.

Reetzke R, Xie Z, Llanos F, Chandrasekaran B. Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Curr Biol*. 2018:28(9):1419–1427.e1414.

Reinke K, He Y, Wang C, Alain C. Perceptual learning modulates sensory evoked response during vowel segregation. *Cogn Brain Res*. 2003:17(3):781–791.

Rizzi R, Bidelman GM. Duplex perception reveals brainstem auditory representations are modulated by listeners' ongoing percept for speech. *Cereb Cortex*. 2023:33(18):10076–10086.

Ross B, Jamali S, Tremblay KL. Plasticity in neuromagnetic cortical responses suggests enhanced auditory object representation. *BMC Neurosci*. 2013:14(1):151.

Sala G, Gobet F. Cognitive and academic benefits of music training with children: a multilevel meta-analysis. *Mem Cogn*. 2020:48(8): 1429–1441.

Sarvas J. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. *Phys Med Biol*. 1987:32(1):11–22.

Schellenberg EG, Lima CF. Music training and nonmusical abilities. *Ann Rev Psychol*. 2024:75(1):1–42.

Scherg M, Ille N, Bornfleth H, Berg P. Advanced tools for digital EEG review: virtual source montages, whole-head mapping, correlation, and phase analysis. *J Clin Neurophysiol*. 2002:19(2):91–112.

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(7):688–694.

Seppänen M, Hamalainen J, Pesonen A-K, Tervaniemi M. Music training enhances rapid neural plasticity of N1 and P2 source activation for unattended sounds. *Front Hum Neurosci*. 2012:43(1):1–13.

Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE. Enhancement of neuroplastic P2 and N1C auditory evoked potentials in musicians. *J Neurosci*. 2003:23(13):5545–5552.

Sheehan KA, McArthur GM, Bishop DV. Is discrimination training necessary to cause changes in the P2 auditory event-related brain potential to speech sounds? *Brain Res Cogn Brain Res*. 2005:25(2):547–553.

Sinex DG, Sabes JH, Li H. Responses of inferior colliculus neurons to harmonic and mistuned complex tones. *Hear Res*. 2002:168(1–2):150–162.

Skoe E, Krizman J, Spitzer ER, Kraus N. Auditory cortical changes precede brainstem changes during rapid implicit learning: evidence from human EEG. *Front Neurosci*. 2021:15:718230.

Slater J, Strait DL, Skoe E, O'Connell S, Thompson E, Kraus N. Longitudinal effects of group music instruction on literacy skills in low-income children. *PLoS One*. 2014:9(11):e113383.

Slater J, Skoe E, Strait D, O'Connell S, Thompson E, Kraus N. Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav Brain Res*. 2015:291:244–252.

Slevc RL, Miyake A. Individual differences in second-language proficiency: does musical ability matter? *Psychol Sci*. 2006:17(8):675–681.

Song JH, Skoe E, Wong PC, Kraus N. Plasticity in the adult human auditory brainstem following short-term linguistic training. *J Cogn Neurosci*. 2008:20(10):1892–1902.

Steinmetzger K, Rupp A. The auditory P2 evoked by speech sounds consists of two separate subcomponents. bioRxiv. 2023. https://doi.org/10.1101/2023.06.30.547226.

Strait DL, Kraus N, Parbery-Clark A, Ashley R. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. *Hear Res*. 2010:261(1–2):22–29.

Strait DL, Parbery-Clark A, Hittner E, Kraus N. Musical training during early childhood enhances the neural encoding of speech in noise. *Brain Lang*. 2012:123(3):191–201.

Studebaker GA. A "rationalized" arcsine transform. *J Speech Lang Hear Res*. 1985:28(3):455–462.

Swaminathan J, Mason CR, Streeter TM, Best V, Kidd G Jr, Patel AD. Musical training, individual differences and the cocktail party problem. *Sci Rep*. 2015:5(1):11628.

Tichko P, Skoe E. Frequency-dependent fine structure in the frequency-following response: the byproduct of multiple generators. *Hear Res*. 2017:348:1–15.

Tong Y, Melara RD, Rao A. P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Res*. 2009:1297:80–88.

Torppa R, Faulkner A, Kujala T, Huotilainen M, Lipsanen J. Developmental links between speech perception in noise, singing, and cortical processing of music in children with cochlear implants. *Mus Per*. 2018:36(2):156–174.

Tremblay K, Kraus N, McGee T, Ponton C, Otis B. Central auditory plasticity: changes in the N1-P2 complex after speech-sound training. *Ear Hear*. 2001:22(2):79–90.

Wallstrom GL, Kass RE, Miller A, Cohn JF, Fox NA. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *Int J Psychophysiol*. 2004:53(2):105–119.

Weiss MW, Bidelman GM. Listening to the brainstem: musicianship enhances intelligibility of subcortical representations for speech. *J Neurosci*. 2015:35(4):1687–1691.

Welford AT. Choice reaction time: Basic concepts. In:Welford AT, editors. *Reaction times*. New York: Academic Press; 1980. pp. 73–128.

Wisniewski MG, Ball NJ, Zakrzewski AC, Iyer N, Thompson ER, Spencer N. Auditory detection learning is accompanied by plasticity in the auditory evoked potential. *Neurosci Lett*. 2020:721:134781.

Wong PC, Skoe E, Russo NM, Dees T, Kraus N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci*. 2007:10(4):420–422.

Yellamsetty A, Bidelman GM. Low- and high-frequency cortical brain oscillations reflect dissociable mechanisms of concurrent speech segregation in noise. *Hear Res*. 2018:361:92–102.

Yellamsetty A, Bidelman GM. Brainstem correlates of concurrent speech identification in adverse listening conditions. *Brain Res*. 2019:1714:182–192.

Yoo J, Bidelman GM. Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hear Res*. 2019:377:189–195.

Zatorre R, Evans A, Meyer E, Gjedde A. Lateralization of phonetic and pitch discrimination in speech processing. *Science*. 1992:256(5058):846–849.

Zhang Y, Kuhl PK, Imada T, Kotani M, Tohkura Y. Effects of language experience: neural commitment to language-specific auditory patterns. *Neuroimage*. 2005:26(3):703–720.