



# Inherent auditory skills rather than formal music training shape the neural encoding of speech

Kelsey Mankel<sup>a,b</sup> and Gavin M. Bidelman<sup>a,b,c,1</sup>

<sup>a</sup>School of Communication Sciences and Disorders, University of Memphis, Memphis, TN 38152; <sup>b</sup>Institute for Intelligent Systems, University of Memphis, Memphis, TN 38152; and <sup>c</sup>Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center, Memphis, TN 38163

Edited by Dale Purves, Duke University, Durham, NC, and approved November 6, 2018 (received for review July 9, 2018)

**Musical training is associated with a myriad of neuroplastic changes in the brain, including more robust and efficient neural processing of clean and degraded speech signals at brainstem and cortical levels. These assumptions stem largely from cross-sectional studies between musicians and nonmusicians which cannot address whether training itself is sufficient to induce physiological changes or whether preexisting superiority in auditory function before training predisposes individuals to pursue musical interests and appear to have similar neuroplastic benefits as musicians. Here, we recorded neuroelectric brain activity to clear and noise-degraded speech sounds in individuals without formal music training but who differed in their receptive musical perceptual abilities as assessed objectively via the Profile of Music Perception Skills. We found that listeners with naturally more adept listening skills (“musical sleepers”) had enhanced frequency-following responses to speech that were also more resilient to the detrimental effects of noise, consistent with the increased fidelity of speech encoding and speech-in-noise benefits observed previously in highly trained musicians. Further comparisons between these musical sleepers and actual trained musicians suggested that experience provides an additional boost to the neural encoding and perception of speech. Collectively, our findings suggest that the auditory neuroplasticity of music engagement likely involves a layering of both preexisting (nature) and experience-driven (nurture) factors in complex sound processing. In the absence of formal training, individuals with intrinsically proficient auditory systems can exhibit musician-like auditory function that can be further shaped in an experience-dependent manner.**

EEG | experience-dependent plasticity | auditory event-related brain potentials | frequency-following responses | nature vs. nurture

It is widely reported that musical training alters structural and functional properties of the human brain. Music-induced neuroplasticity has been observed at every level of the auditory system, arguably making musicians an ideal model to understand experience-dependent tuning of auditory system function (1–4). Most notably among their auditory-cognitive benefits, musicians are particularly advantaged in speech and language tasks including speech-in-noise (SIN) recognition (for a review see ref. 5). This suggests that musicianship may increase listening capacities and aid the deciphering of speech not only in ideal acoustic conditions but also in difficult acoustic environments (e.g., noisy “cocktail party” scenarios).

Electrophysiological recordings have been useful in demonstrating music-related neuroplasticity at different levels of the auditory neuroaxis. In particular, frequency-following responses (FFRs), predominantly reflecting phase-locked activity from the brainstem (6, 7) and, under some circumstances, from the cortex (7, 8), serve as a “neural fingerprint” of sound coding in the EEG (9). The strength with which speech-evoked FFRs capture voice pitch (i.e., fundamental frequency; F0) and harmonic timbre cues of complex signals is causally related to listeners’ perception of speech material (9). Interestingly, FFRs are augmented and shorter in latency in musicians than in nonmusicians, particularly for noise-degraded speech (10–12), providing a neural account of their enhanced SIN perception observed behaviorally. Similarly,

event-related potentials (ERPs) and fMRI show differential speech activity in musicians at cortical levels of the nervous system (13–16). Collectively, an overwhelming number of studies have implied that musical training shapes auditory brain function at multiple stages of subcortical and cortical processing and, in turn, bolsters the perceptual organization of speech.

Problematically, innate differences in auditory system function could easily masquerade as plasticity in cross-sectional studies on auditory learning (17) and music-related plasticity (18–20). This concern is reinforced by the fact that musical skills such as pitch and timing perception develop very early in infancy (i.e., 6 mo of age; ref. 21) and may even be linked to certain genetic markers (22–25). Unfortunately, the majority of studies linking musicianship to speech–language plasticity are cross-sectional and include only self-reports of musicians’ experience (e.g., refs. 10, 14, and 18). It remains possible that certain individuals have naturally enriched auditory systems that predispose them to pursue musical interests (e.g., ref. 26). Consequently, widely reported musician advantages in auditory perceptual tasks may be due to intrinsic differences unrelated to formal training.

Current conceptions of musicality define it as an innate, natural, and spontaneous development of widely shared traits, constrained by our cognitive abilities and biology, that underlie the capacity for music (27). While being “musical” no doubt encompasses more than simple hearing or perceptual abilities (e.g., instrumental production creativity), for this investigation we focus on the receptive aspects of musicality (i.e., auditory perceptual skills), following the long tradition of assessing formal music abilities through strictly perceptual measures (28–32). Among the normal

## Significance

**Musicianship is widely reported to enhance auditory brain processing related to speech–language function. Such benefits could reflect true experience-dependent plasticity, as often assumed, or innate, preexisting differences in auditory system function (i.e., nurture vs. nature). By recording speech-evoked electroencephalograms, we show that individuals who are naturally more adept in listening tasks but possess no formal music training (“musical sleepers”) have superior neural encoding of clear and noise-degraded speech, mirroring enhancements reported in trained musicians. Our findings provide clear evidence that certain individuals have musician-like auditory neurobiological function and temper assumptions that music-related neuroplasticity is solely experience-driven. These data reveal that formal music experience is neither necessary nor sufficient to enhance the brain’s neural encoding and perception of sound.**

Author contributions: K.M. and G.M.B. designed research; K.M. performed research; K.M. and G.M.B. analyzed data; and K.M. and G.M.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence should be addressed. Email: gmbldman@memphis.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811793115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1811793115/-DCSupplemental).

Published online December 3, 2018.

population, “musical sleepers” (i.e., nonmusicians with a high level of receptive musicality) are identified as individuals having naturally superior auditory and music listening skills but who lack formal musical training (30). In the nature-vs.-nurture debate of music and the brain, distinguishing between innate and experience-dependent effects is accomplished only through longitudinal training paradigms (which are costly and often impractical for assessing decades-long training effects) or utilizing objective measures of listening skills that can identify people with highly acute (i.e., musician-like) auditory abilities.

To this end, the aim of the present study was to determine if preexisting differences in auditory skills might account for at least some of the neural enhancements in speech processing as frequently reported in trained musicians. Our study was not intended to refute the possible connections between music training and enhanced linguistic brain function. Rather, we aimed to test the possibility that preexisting auditory skills might at least partially mediate neural enhancements in speech processing. Our design included neuroimaging (FFR and ERP) and behavioral measures to replicate the major experimental designs of previous work documenting neuroplasticity in speech and SIN processing among trained musicians. We hypothesized that musical sleepers would show enhanced neurophysiological encoding of normal and noise-degraded speech, consistent with widespread findings reported in trained musicians (10). Our findings demonstrate a critical but underappreciated role of preexisting auditory skills in the neural encoding of speech and temper widespread assumptions that music-related neuroplasticity is solely experience-driven. We find that formal music experience is neither necessary nor sufficient to enhance the brain’s neural encoding and perception of speech and other complex sounds.

## Results

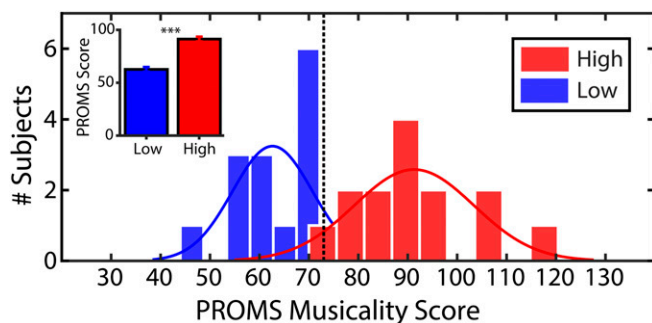
We measured speech-evoked auditory brain responses in young, normal-hearing listeners ( $n = 28$ ) who had minimal ( $<3$  y; average:  $0.7 \pm 0.8$  y) formal musical training and would thus be classified as nonmusicians in prior studies on music-induced neuroplasticity (14, 18, 19). Listeners were divided into low- and high-musicality groups based on an objective battery of musical listening abilities (Profile of Music Perception Skills; PROMS) (30) that included assessment of melody, tuning, accent, and tempo perception (Fig. 1). Groups were otherwise matched in age, socioeconomic status, education, handedness, and years of musical experience (all  $P$ s  $> 0.05$ ) (*Materials and Methods*). Since all participants reported limited to no general music ability, ability to read music, or ability to transcribe a simple melody by ear, no group differences emerged on these self-report measures. As expected based on our group split, highly skilled listeners exhibited better scores on the total and individual PROMS subtests than the low-musicality group (all  $P$ s  $< 0.001$ ). No group differences

were observed on the QuickSIN test (33), a behavioral measure of SIN perception [ $t(26) = 1.27$ ,  $P = 0.22$ ].

We then tested whether musicality was associated with enhanced neural processing for clean and noise-degraded speech as reported in the FFRs of highly trained musicians (14, 18–20). FFRs were recorded while participants passively listened to speech sounds, consistent with previous studies on musicians and speech plasticity (18, 19). Amplitudes and onset latencies of the FFR were used to assess the overall magnitude and temporal precision of listeners’ neural response to speech in the early auditory pathway (Fig. 2). We found that FFR F0 amplitudes, reflecting voice “pitch” coding (10, 11), showed a group  $\times$  noise interaction [ $F(1,26) = 6.42$ ,  $P = 0.018$ ,  $d = 0.99$ ] (Fig. 2E). Tukey adjusted multiple comparisons showed that noise had a degradative effect (clean  $>$  noise) in low-musicality listeners. In stark contrast, speech FFRs in highly musical ears were invariant to noise (clean = noise), indicating superior speech encoding even in challenging acoustic conditions. Harmonic (i.e., H2–H5) amplitudes, reflecting the neural encoding of speech “timbre,” similarly showed a group  $\times$  noise interaction [ $F(1,26) = 7.90$ ,  $P = 0.009$ ,  $d = 1.10$ ]. This effect was attributable to stronger encoding of speech harmonics in noise for the high-scoring PROMS group, whereas no noise-related changes were observed in the low-scoring PROMS group. FFR latency showed a main effect of group [ $F(1,25) = 6.47$ ,  $P = 0.018$ ,  $d = 0.98$ ] where speech responses were earlier (i.e., faster precision) in high PROMS scorers across the board (Fig. 2F). No group differences (or interactions) were observed for rms neural noise [ $F(1,25) = 0.05$ ,  $P = 0.827$ ,  $d = 0.09$ ], an index reflecting the efficiency or quality in auditory neural processing (34, 35). These results suggest the neural encoding of salient speech cues is enhanced in listeners with naturally more skilled listening capacities, paralleling the speech enhancements observed in highly trained musicians (10).

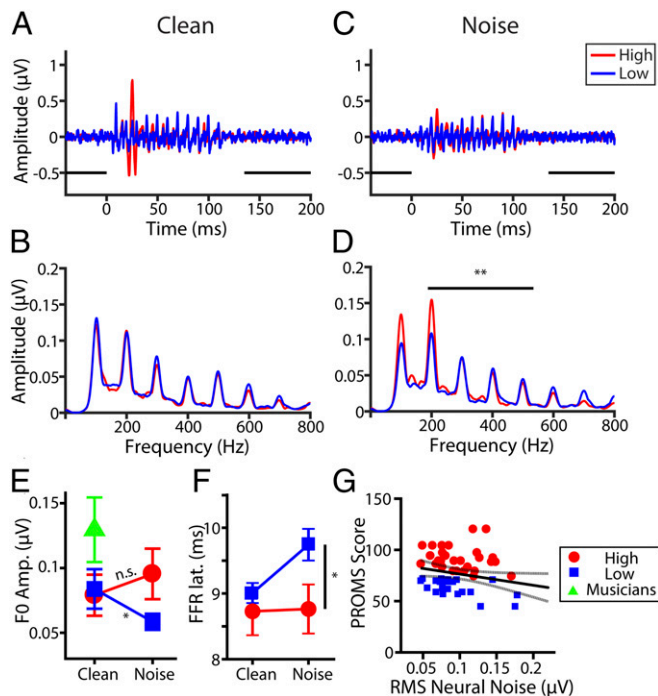
Having established that individuals with musician-like auditory skills show superior neural encoding of speech, we next assessed the relation between brain activity and behavioral auditory abilities. We used generalized linear mixed effects (GLME) model regression to evaluate links between behavioral PROMS scores and neural FFR responses. We found that neural noise (34, 35) was a strong predictor of listeners’ musicality [ $t(54) = -2.61$ ,  $P = 0.012$ ] (Fig. 2G); that is, greater “brain noise” was associated with lower PROMS performance indicative of poorer auditory perceptual skills. Links between total PROMS scores and FFR F0 amplitudes [ $t(54) = 1.61$ ,  $P = 0.11$ ] and latency [ $t(54) = 1.58$ ,  $P = 0.12$ ] were insignificant.

These initial analyses focused exclusively on listeners’ total PROMS score, which taps multiple perceptual dimensions and thus may have masked relations between auditory brain responses and certain aspects of musical listening skills. To tease apart the domains of auditory processing most related to speech-FFR enhancements, we ran separate GLMEs between each of the PROMS subtest scores (i.e., melody, tuning, accent, and tempo) and neural measures (i.e., F0 amplitude, neural noise, latency). Higher tuning scores predicted larger F0 amplitudes [ $t(54) = 2.18$ ,  $P = 0.033$ ] and lower neural noise [ $t(54) = -3.13$ ,  $P = 0.003$ ]. Better tempo scores were predicted by neural noise [ $t(54) = -2.12$ ,  $P = 0.039$ ]; accent scores by FFR latency [ $t(54) = 2.51$ ,  $P = 0.015$ ]. No other FFR and PROMS subtest relationships were significant (all  $P$ s  $> 0.05$ ). We then ranked the regression models by their Akaike information criterion (AIC) to evaluate the relative predictive value of neural FFRs for each auditory subdomain (see *SI Appendix, SI Text* for all AIC values). Both FFR neural noise and F0 amplitudes were best predicted by tuning subtest scores [AICs =  $-188.03$  and  $-156.80$ , respectively]. In contrast, latency showed best correspondence with accent scores [AIC =  $179.22$ ]. This dissociation in brain-behavior relationships suggests that spectral measures of the FFR (neural noise, F0 amplitude) are more associated with perceptual skills related to fine pitch discrimination (11, 36), whereas neural latencies are more strongly associated with timing or rhythmic perception.



**Fig. 1.** PROMS scores reveal that some listeners have highly adept (musician-like) auditory skills despite having no formal music training. Listeners (all nonmusicians) were divided into high- and low-musicality groups based on a median split of scores on the full PROMS battery (dashed vertical line). (Inset) Mean PROMS scores across groups. Error bars =  $\pm 1$  SEM; \*\*\* $P < 0.001$ .





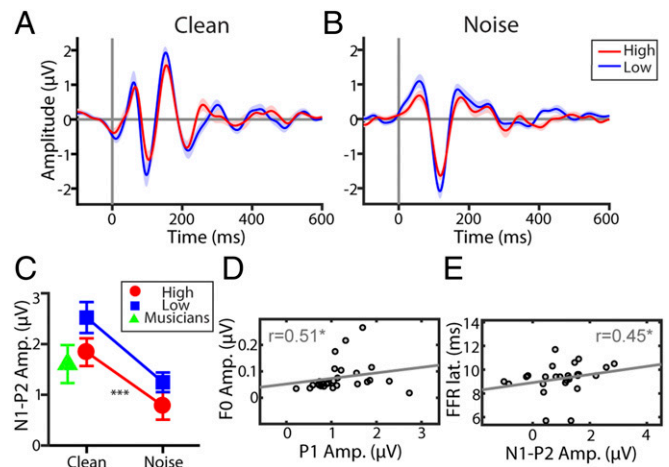
**Fig. 2.** Speech-evoked FFRs reveal neural enhancements in musical ears. FFR waveforms and spectra in the clean (A and B) and noisy (C and D) speech conditions reflecting phase-locked neural activity to the spectrotemporal characteristics of speech. (E) FFR F0 amplitudes. Data from actual trained musicians (40) are shown for comparison. Highly musical listeners exhibited stronger encoding of speech at F0 (voice pitch) and its integer multiple harmonics (timbre) for degraded speech than less musical individuals. Formally trained musicians (40) still exhibit larger FFRs than nonmusicians, regardless of the nonmusicians' inherent musicality (musician data were not available for noise). (F) FFR latency is earlier in high vs. low PROMS scorers. (G) GLME regression relating brain and behavioral measures (aggregating all clean/noise responses;  $n = 56$ ). Individuals with higher levels of intrinsic neural noise are less musical (i.e., have lower PROMS scores). The solid line shows the regression fit; dotted lines indicate the 95% CI interval. Error bars =  $\pm 1$  SEM; \* $P < 0.05$ , \*\* $P < 0.01$ , n.s. = nonsignificant.

Unlike FFRs, cortical ERPs to speech exhibited strong stimulus-related effects of noise but did not show group differences in latency or amplitude (Fig. 3). An ANOVA showed a main effect of stimulus noise on peak negativity (N1)–peak positivity (P2) amplitudes [ $F(1,26) = 35.30$ ,  $P < 0.0001$ ,  $d = 2.33$ ], with weaker responses to noise-degraded speech than to clean speech. There was no group difference in ERP magnitude [ $F(1,25) = 3.42$ ,  $P = 0.0761$ ,  $d = 0.73$ ], nor was there a group  $\times$  noise interaction [ $F(1,26) = 0.30$ ,  $P = 0.587$ ,  $d = 0.22$ ]. Still, N1 and P2 waves showed the expected latency prolongation in noise ( $P_s < 0.0001$ ) (SI Appendix, Fig. S1) (37, 38). Finally, in contrast to FFRs, a GLME showed cortical N1–P2 amplitudes did not predict PROMS scores [ $t(54) = 1.39$ ,  $P = 0.17$ ]. These results indicate that while cortical responses showed an expected noise-related decrement in speech coding (38, 39), they were not modulated by individuals' inherent listening abilities.

Brain–brain correlations further revealed that FFR F0 amplitudes were correlated with ERP P1 amplitudes [ $r = 0.51$ ,  $P = 0.006$ ] for noise-degraded speech (Fig. 3D), replicating prior studies showing that brainstem responses during SIN processing predict cortical responses further upstream (34, 37, 40). Additionally, FFR latency predicted P2 ( $r = 0.40$ ,  $P = 0.0346$ ) and N1–P2 amplitudes ( $r = 0.45$ ,  $P = 0.0151$ ) (Fig. 3E); that is, more sluggish neural encoding in the brainstem was linked with larger cortical activity in response to speech, as observed in lower PROMS scorers (compare Fig. 3E and C). These correspondences were not observed for clean speech (all  $P_s > 0.05$ ), suggesting that brainstem–cortical relationships are most apparent under more taxing listening

conditions (37). We did not observe correlations between neural measures (FFR and ERP) and behavioral QuickSIN scores (all  $P_s > 0.05$ ), but this result might be expected given that FFRs and ERPs were recorded under passive listening conditions.

Previous neuroimaging studies suggest that formal musical training enhances the behavioral and neural encoding of speech (10, 14, 18–20, 40–42), and the majority of studies reporting musician advantages in auditory neural processing have employed the identical FFR methodology used here. An interesting question that emerges from our data, then, is how the neurophysiological enhancements we observe in highly adept nonmusicians (i.e., high PROMS scorers) compare with those reported in actual trained musicians. To address this question, we compared speech FFRs and ERPs from our high PROMS scorers with previously published data from formally trained musicians obtained using identical speech stimuli (40) (parallel data in noise were not available). Despite enhanced auditory function in high vs. low PROMS scorers, musicians with  $\sim 10$  y of formal training (40) exhibited larger speech FFRs than all the nonmusicians in the current sample [one-tailed  $t(24) = 1.89$ ,  $P = 0.03$ ] as well as the high [ $t(24) = 1.76$ ,  $P = 0.04$ ] and low [ $t(24) = 1.61$ ,  $P = 0.05$ ] PROMS groups separately (Fig. 2E). In contrast, cortical ERP amplitudes did not differ between nonmusicians with high PROMS scores and actual musicians [ $t(24) = -0.52$ ,  $P = 0.31$ ]. However, musicians' responses were smaller than those of low PROMS scorers [ $t(24) = -1.96$ ,  $P = 0.03$ ] (Fig. 3C), paralleling the effects observed in some cross-sectional studies comparing musicians and nonmusicians (14). Additional contrasts between QuickSIN scores in our PROMS nonmusician cohorts and those of actual musicians revealed that trained individuals outperformed all nonmusicians behaviorally regardless of the nonmusicians' musicality [ $t(26) = 2.75$ ,  $P = 0.011$ ] (SI Appendix, SI Text). Replicating prior cross-sectional studies (39, 42), musicians' SIN reception thresholds were  $\sim 1.5$ – $2$  dB lower (i.e., better) than the two PROMS groups, who did not differ [ $t(26) = 1.27$ ,  $P = 0.22$ ] (SI Appendix, Fig. S2). Collectively, our findings imply that (i) individuals with highly adept, intrinsic auditory skills but no formal training have enhanced (musician-like) neural processing of speech, but (ii) formal musicianship might provide an additional,



**Fig. 3.** Cortical speech-evoked responses are modulated by noise but not listeners' inherent auditory skills (musicality). (A and B) ERP waveforms for clean (A) and noise-degraded (B) speech. (C) N1–P2 amplitudes and latencies (SI Appendix, Fig. S1) indicate noise-related changes in neural activity but no differences between musicality groups. Musician data shown for comparison are from ref. 40. Trained musicians' N1–P2 amplitudes differ from low PROMS scorers but are similar to those of musical sleepers (high-scoring PROMS group). (D and E) Relationships between brainstem (FFR) and cortical (ERP) measures for noise-degraded speech ( $n = 28$  responses). (D) Larger P1 responses at the cortical level are associated with larger FFR F0 amplitudes. (E) Faster brainstem FFRs are associated with smaller N1–P2 responses, as seen in high PROMS scorers and trained musicians (compare with C). Error bars =  $\pm 1$  SEM; \* $P < 0.05$ , \*\*\* $P < 0.001$ .

experience-dependent “boost” on top of preexisting differences in auditory brain function.

## Discussion

By recording neuroelectric brain responses in highly skilled listeners who lack formal musical training, we provide strong evidence that inherent auditory system function, in the absence of experience, is associated with enhanced neural encoding of speech and auditory perceptual advantages. Our study explicitly shows that certain individuals have musician-like auditory neural function, as has been conventionally indexed via speech FFRs. More broadly, these findings challenge assumptions that the neuroplasticity associated with musical training and speech processing is solely experience-driven (cf. refs. 2, 3, and 18). Importantly, we do not claim that our study negates the possibility that actual musical training can confer experience-dependent auditory plasticity (e.g., refs. 16 and 43–46). Rather, our data argue that preexisting factors may play a larger role in putative links between musical experience and enhanced speech processing than conventionally thought.

Neurological differences among intrinsically skilled listeners were particularly evident in speech FFRs, which showed individuals with higher musicality scores had faster and more robust neural responses to the voice pitch (F0) and timbre (harmonics) cues of speech, even amid interfering noise. In fact, the advantages we find here in nonmusician musical sleepers are remarkably similar to those reported in trained musicians, who similarly show enhanced neural encoding and behavioral recognition of clean and noise-degraded speech (10, 12, 19, 41). Additionally, we found greater neural noise was associated with poorer auditory skills (i.e., lower musicality scores). Neural noise has been interpreted as reflecting the variability in how sensory information is translated across the brain (47). Inasmuch as lesser noise reflects a greater efficiency in auditory processing and perception (34, 35), the higher-quality neural representations we find among high PROMS scorers may allow more veridical readout of signal identity and thus account for their superior behavioral abilities. Our data are also consistent with recent fMRI findings demonstrating that the strength of (passively measured) resting-state connectivity between auditory and motor brain regions before training is related to better musical proficiency in short-term instrumental learning (43). These findings, along with current EEG data, suggest that intrinsic differences in neural function may predict outcomes in a variety of auditory contexts, from understanding someone at the cocktail party, where pitch and timbre cues are vital for understanding a noise-degraded talker (10, 39, 48), to success in music-training programs (43).

One interesting finding was that variations in the neural encoding of speech were better explained by certain perceptual domains (i.e., PROMS subtest scores). Among perceptual subtests, FFR spectral measures were best predicted by tuning scores, whereas accent perception was best predicted by FFR latency measures. The tuning subtest requires detection of a subtle pitch manipulation ( $<1/2$  semitone) within a musical chord (30). Given that FFRs reflect the neural integrity of stimulus properties, higher-fidelity FFR responses (i.e., increased F0 amplitudes, decreased noise, faster latencies) may allow finer discrimination of acoustic details at the behavioral level and account for the superior auditory perceptual skills we find in our individuals with high PROMS scores. In this regard, our data here in musical sleepers aligns closely with recent findings by Nan et al. (16), who showed that short-term musical training enhances the neural processing of pitch and improves speech perception in children following short-term music lessons.

Neurophysiological measures revealed that group differences in speech processing were more apparent in FFRs than in ERP responses. This opposite pattern of effects across neural measures (i.e., larger FFRs and reduced ERPs; compare Fig. 2E with Fig. 3C) is reminiscent of both animal (49) and human (14, 40, 50) electrophysiological studies which show that, even for the same task, neuroplastic changes in brainstem (e.g., FFR) receptive fields tend in the opposite direction of changes in auditory cortex (e.g., ERPs). For a discussion on the neural generators of the FFR, see *SI Appendix, SI Discussion*. Regardless of where

our scalp responses are generated, we can still conclude that among people without formal musical training, certain individuals' brains produce phase-locked neural responses that better capture the acoustic information in speech.

Despite parallels in ERPs between high-scoring PROMS listeners and trained musicians, it is possible that cortical and/or behavioral differences in speech processing emerge only (i) after strong experiential plasticity rather than subtle innate function or (ii) online during tasks requiring top-down processing and/or attention. Indeed, we found behavioral QuickSIN enhancements in musicians but not in musical sleepers (*SI Appendix, Fig. S2*). This suggests that stronger, more protracted experiences might be needed to observe plasticity at later stages of the auditory system and to transfer the effects to speech perception. Structural and functional neuroimaging studies have revealed striking differences in musicians at the cortical level which are predictive of musical and language abilities (e.g., refs. 2, 43, and 51). Thus, musicianship might tune language-related networks of the brain more broadly, beyond those core auditory sensory responses indexed by our FFRs and ERPs. Full-brain imaging of musical sleepers would be a logical next step to further explore the relationship between innate listening abilities and cortical function. Presumably, musical sleepers might show more pervasive cortical differences outside the auditory system as evaluated here. In any case, our data reinforce the fact that neuroplasticity is not an all-or-nothing phenomenon and that it manifests in different gradations at lower (brainstem) vs. higher (cortical/behavioral) levels of the auditory system (14, 40) in accordance with a listener's experience.

It is well established that cortical responses are heavily modulated by attention, and the neuroplastic benefits of musical training are consistently stronger under active than under passive listening tasks (52, 53). Because our study utilized a passive listening paradigm, it is perhaps unsurprising that we failed to find group differences in neural N1–P2 amplitudes among high- and low-musicality individuals if cortical ERP enhancements emerge mainly under states of goal-directed attention (13, 14). Paralleling the ERP data, we similarly found that PROMS groups did not differ in behavioral QuickSIN scores, in contrast to the SIN benefits observed in trained musicians (*SI Appendix, Fig. S2*) (10, 39, 42). One interpretation of these data is that the QuickSIN and ERPs are not sensitive enough to detect the finer individual differences in speech processing among nonmusicians as revealed by FFRs, a putative marker of subcortical processing (7). Indeed, direct comparisons between each class of speech-evoked responses show not only that they are functionally distinct (14, 40) but also that FFRs are more stable than cortical ERPs both within and between listeners (54). The lower variability of brainstem FFRs may offer a better reflection of innate, hardwired processes of audition (55) than the cortex (e.g., ERPs, behavior), which are more malleable and heavily influenced by subject state, attention, and task demands.

While our data provide evidence that natural propensities in auditory skills might account for certain *de novo* enhancements in the brain's speech processing, provocatively, they also show that formally trained musicians (40) have an additional boost in neuro-behavioral function, even beyond those auditory systems deemed inherently superior. These results are consistent with randomized control studies of music-enrichment programs that report treatment effects following 1–2 y of music training (16, 20, 46). Thus, longitudinal studies still provide compelling evidence for brain plasticity associated with musical training (16, 43–46). However, an aspect that remains relatively uncontrolled in previous studies is possible placebo effects. Reminiscent of recent revelations in the cognitive brain-training literature (56), individuals may enter into (music) training expecting to receive benefits, in which case perceptual-cognitive gains may be partially epiphenomenal. Additionally, individuals with undetected superiorities in listening abilities may remain more engaged or motivated in music-training programs, an aspect that may confound outcomes of longitudinal studies.

Still, it is possible that our groups differed on some other environmental factor rather than inherent auditory perceptual skills *per se*. For example, high-scoring PROMS listeners may differ in their



early exposure to music (21), daily recreational music listening, perceptual investment with music, or other types of perceptual-cognitive skills not assessed by the PROMS, which tests only receptive capabilities. However, we note that participants reported minimal to no musical abilities and did not possess absolute (perfect) pitch. Importantly, groups were matched on these self-report measures. While our participants did not personally regard themselves as musical before this study, their perceptual scores and neural responses indicate otherwise. Still, it remains to be seen if the musician-like auditory function observed here is present in certain individuals from birth or emerges over a more protracted time course during normal auditory development. Our data also leave open the possibility that listeners scoring better on the PROMS have more experience or investment in perceptual engagement with music, making it difficult to tease apart whether the effects are truly predispositions or perhaps are partially experience-based. However, animal studies show that passive listening is insufficient to induce auditory neuroplasticity (57), making it unlikely that recreational or informal exposure drives the enhancements in high-scoring PROMS listeners. While the origin of neural differences among musical sleepers is an interesting avenue for future work, we argue that the mere identification of such individuals highlights two more important points: (i) inherent perceptual abilities differentiate people previously considered to be homogeneous nonmusicians, producing responses that mirror those attributed to formal music training; and (ii) the need to consider preexisting factors before claiming that music or other learning activities engender neuroplastic benefit (cf. ref. 56).

In conclusion, our findings reveal that individuals with highly adept listening skills but no formal music training (i.e., musical sleepers) have better auditory system function in the form of more robust and temporally precise neural encoding of speech. These effects closely mirror the experience-dependent plasticity reported in seminal studies on trained musicians. From a nature-vs.-nurture perspective, our study suggests that nature (i.e., preexisting differences in auditory brain function) constrains neurobiological and behavioral responses so that individuals with higher-fidelity auditory neural representations also tend to be better perceptual listeners. Nevertheless, the experience-dependent effects of training seem to “nurture” neurobiological function and provide an additional layer of gain to the sensory processing of communicative signals. Most importantly, our results emphasize the critical need to document a priori listening skills before assessing music’s effects on speech/hearing abilities and claiming experience- or training-related effects.

## Materials and Methods

**PROMS Musicality Test.** We assessed receptive musicality objectively using the brief version of the PROMS (30). See *SI Appendix, SI Text* and refs. 30 and 58 for additional information regarding internal consistency, reliability, and validation of the PROMS. The test battery consists of four subtests tapping listening skills in the domains of melody, tuning, accent, and tempo perception. Each subtest contains 18 trials. Listeners heard two identical sound clips and then were asked if a third probe clip was the same as or different from the previous two. Scores for each subtest were calculated based on accuracy and confidence ratings (i.e., two points were given for correctly reporting “definitely the same” or “definitely different,” one point was given for “probably the same” or “probably different,” and no points were given for “I don’t know” or an incorrect answer). The total PROMS score reflects the combined sum of all subtest scores. A median split of the total score divided participants into two groups (i.e., high-scoring and low-scoring PROMS groups; see Fig. 1).

**Participants.** Twenty-eight young adults (age:  $22.2 \pm 3.1$  y, 23 females) participated in the main experiment. This sample size was determined a priori to match those of comparable studies on musical training and auditory plasticity that have shown effects between musicians and nonmusicians (10, 14, 15, 40). All spoke American English as their first language with no prior tone-language experience and were identified as right-handed according to the Edinburgh Handedness Survey (59). All participants were screened for a history of neuropsychiatric disorders. Formal musical training is thought to enhance the neural encoding of speech (14, 18–20, 40, 41), particularly in noise (10, 42). Hence, all participants were required to have minimal formal musical training (i.e.,  $<3$  y) and no musical training within the past 5 y. Critically, groups did not differ in their years of formal musical training [high-scoring group =  $0.57 \pm 0.63$  y,

low-scoring group =  $0.79 \pm 0.97$  y;  $t(26) = -0.69$ ,  $P = 0.50$ ]. On a seven-point Likert scale (ranging from 0 = no ability to 7 = professional/perfect ability), both groups self-reported minimal music ability [high-scoring group =  $1.6 \pm 1.5$ , low-scoring group =  $1.1 \pm 1.2$ ;  $t(26) = 1.11$ ,  $P = 0.28$ ], ability to read music [high-scoring group =  $0.9 \pm 1.3$ , low-scoring group =  $0.4 \pm 0.5$ ;  $t(26) = 1.37$ ,  $P = 0.18$ ], and ability to transcribe a simple melody given a starting pitch [high-scoring group =  $0.6 \pm 1.3$ , low-scoring group =  $0.3 \pm 0.8$ ;  $t(26) = 0.85$ ,  $P = 0.40$ ]. Additionally, none reported having absolute (perfect) pitch, i.e., the ability to name a note by ear without an external reference tone.

Audiometry confirmed normal hearing in all listeners (i.e., thresholds  $<25$  dB hearing loss, 250–4,000 Hz). Groups were also matched in age [high-scoring group =  $23.2 \pm 3.6$  y, low-scoring group =  $21.2 \pm 2.3$  y;  $t(26) = 1.73$ ,  $P = 0.10$ ], socioeconomic status [scored based on highest level of parental education from 1 (high school without diploma or GED) to 6 (doctoral degree): high-scoring group =  $4.5 \pm 1.4$ , low-scoring group =  $4.1 \pm 1.0$ ;  $t(26) = 0.93$ ,  $P = 0.49$ ], formal education [high-scoring group =  $16.4 \pm 2.7$ , low-scoring group =  $15.6 \pm 2.4$ ;  $t(26) = 0.73$ ,  $P = 0.48$ ], and handedness laterality (59) [high-scoring group =  $85.7 \pm 21.1$ , low-scoring group =  $85.6 \pm 2.8$ ;  $t(26) = 0.02$ ,  $P = 0.986$ ]. Gender was marginally unbalanced between groups (Fisher’s exact test,  $P = 0.041$ ) and was used as a covariate in subsequent analyses. The University of Memphis Institutional Review Board (IRB) approved all experiments involving human subjects in this study. Participants gave written informed consent in compliance with IRB protocol no. 2370.

**Stimuli.** We used a synthetic, 100-ms speech sound (/a/) with an F0 of 100 Hz previously shown to elicit robust group differences in FFRs and ERPs among experienced musicians (40). Following previous studies (10, 19, 20), there was no task during EEG recordings, and participants watched a self-selected movie as they passively listened to speech sounds. In addition to “clean” tokens [signal-to-noise ratio (SNR) =  $\infty$  dB], speech was presented in background noise since musician’s speech enhancements are usually observed under acoustically taxing conditions (10, 39). Noise-degraded speech was created by overlaying continuous, non-time-locked multitalker babble to the clean token at a +10 dB SNR (10). Speech sounds were presented in alternating polarity accordingly to a clustered sequence, with interlaced interstimulus intervals (ISIs) that were optimized for recording both brainstem (ISI = 150 ms; 2,000 trials) and cortical potentials (ISI = 1,500 ms; 200 trials) while minimizing response adaptation (60). Stimulus presentation was controlled by MATLAB 2013b (MathWorks) routed to a TDT RP2 interface (Tucker-Davis Technologies). Tokens were delivered binaurally at an 83-dB sound-pressure level (SPL) through shielded ER-2 earphones (Etymotic Research). Clean and noise blocks were randomized across participants.

**Behavioral SIN Task.** We measured listeners’ speech-reception thresholds in noise using the QuickSIN test (33). Participants were presented six sentences with five key words embedded in four-talker babble noise. Sentences were presented at a 70-dB SPL at decreasing SNRs (steps of  $-5$  dB) from 25 dB (very easy) to 0 dB (very difficult). Listeners scored one point for each correctly repeated keyword. SNR loss (in decibels) was determined as the SNR for 50% criterion performance.

**EEG Recordings.** Neuroelectric activity was recorded differentially between Ag/AgCl disk electrodes placed on the scalp at the high forehead ( $\sim$ Fpz) referenced to linked mastoids (A1/A2; mid-forehead = ground). This montage is optimal for simultaneously recording brainstem and cortical auditory responses (14, 40, 60). Electrode impedance was kept  $\leq 3$  k $\Omega$ . EEGs were digitized at 10 kHz (SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of 4,000 Hz DC. EEGs were then epoched (FFR:  $-40$  to 200 ms; ERP:  $-100$  to 600 ms), baselined, and averaged in the time domain to derive FFRs/ERPs for each condition. Sweeps exceeding  $\pm 50$   $\mu$ V were rejected as artifacts. Responses were then filtered into high- and low-frequency bands to isolate FFRs (85–2,500 Hz) and ERPs (3–25 Hz) (14, 40, 41).

**FFR analysis.** We computed the FFT to measure the F0 or voice pitch coding of each FFR waveform (10, 11). Similarly, timbre encoding was quantified by measuring the mean amplitude of the second through fifth harmonics (H2–H5) (10, 11). Neural noise was calculated as the mean rms amplitude of the prestimulus ( $-40$  to 0 ms) and poststimulus (135–200 ms) intervals, respectively (34, 35, 47, 50) (see Fig. 2A). FFR latency was measured as the maximum cross-correlation between the stimulus and FFR waveforms between 6–12 ms, the expected onset latency of brainstem responses (7, 8, 14). **ERP analysis.** We measured the overall magnitude of the N1–P2 complex, reflecting the early registration of sound in cerebral cortex, as the voltage difference between the two individual waves (14). Individual latencies were measured as the N1 between 80 and 120 ms and P2 between 130 and 180 ms (40).

**Statistics.** Individual *t* tests assessed group differences across various demographic data, PROMS total and subtest scores, and QuickSIN scores. Identical conclusions were obtained with parametric and nonparametric tests for variables that were nonnormally distributed. Unless otherwise noted, we used two-way, mixed-model ANOVAs (group  $\times$  noise level with subjects as a random factor; SAS9.4, GLIMMIX) with gender as a covariate to analyze all neural data. However, gender was not a significant covariate in any of our analyses. Effect sizes for omnibus ANOVAs are reported as Cohen's *d*. Spearman's rho assessed brain-brain correlations between FFR and ERP measures. Conditional studentized residuals confirmed the absence of influential outliers.

GLMEs evaluated relations between behavioral (PROMS musicality scores) and neural responses (FFRs). Subjects were modeled as a random factor nested within group in the regression to model both random intercepts (per subject) and slopes (per group) [e.g., PROMS  $\sim$  FFR<sub>F0</sub> + (sub)group].

Both FFR neural noise and F0 amplitudes were evaluated as neural predictors of listeners' auditory perceptual abilities (PROMS scores). In addition to total scores, we assessed relations between subtest scores (i.e., melody, tuning, accent, tempo) and speech FFRs to determine which aspects of auditory perception (i.e., timing, spectral discrimination, and others) were best predicted by physiological responses. Best predictors were determined as the models minimizing AIC. GLME regressions and visualization were achieved using the *fitlme* and *fitlm* functions in MATLAB, respectively.

**ACKNOWLEDGMENTS.** We thank Caitlin Price for comments on previous versions of this manuscript and Jessica Yoo for assistance with data collection. This work was supported by a grant from the University of Memphis Research Investment Fund and by National Institute on Deafness and Other Communication Disorders of the NIH Grant R01DC016267 (to G.M.B.).

- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11:599–605.
- Herholz SC, Zatorre RJ (2012) Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76:486–502.
- Münste TF, Altenmüller E, Jäncke L (2002) The musician's brain as a model of neuroplasticity. *Nat Rev Neurosci* 3:473–478.
- Moreno S, Bidelman GM (2014) Examining neural plasticity and cognitive benefit through the unique lens of musical training. *Hear Res* 308:84–97.
- Coffey EBJ, Mogilever NB, Zatorre RJ (2017) Speech-in-noise perception in musicians: A review. *Hear Res* 352:49–69.
- Sohmer H, Pratt H, Kinarit R (1977) Sources of frequency following responses (FFR) in man. *Electroencephalogr Clin Neurophysiol* 42:656–664.
- Bidelman GM (2018) Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175:56–69.
- Coffey EB, Herholz SC, Chapesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat Commun* 7: 11070.
- Weiss MW, Bidelman GM (2015) Listening to the brainstem: Musicianship enhances intelligibility of subcortical representations for speech. *J Neurosci* 35:1687–1691.
- Parbery-Clark A, Skoe E, Kraus N (2009) Musical experience limits the degradative effects of background noise on the neural processing of sound. *J Neurosci* 29: 14100–14107.
- Bidelman GM, Krishnan A (2010) Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res* 1355:112–125.
- Kraus N, Skoe E, Parbery-Clark A, Ashley R (2009) Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Ann N Y Acad Sci* 1169:543–557.
- Marie C, Delogu F, Lampis G, Belardinelli MO, Besson M (2011) Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *J Cogn Neurosci* 23:2701–2715.
- Bidelman GM, Alain C (2015) Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci* 35:1240–1249.
- Shahin A, Bosnyak DJ, Trainor LJ, Roberts LE (2003) Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J Neurosci* 23:5545–5552.
- Nan Y, et al. (2018) Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children. *Proc Natl Acad Sci USA* 115: E6630–E6639.
- Jääskeläinen IP, Ahveninen J, Belliveau JW, Raij T, Sams M (2007) Short-term plasticity in auditory cognition. *Trends Neurosci* 30:653–661.
- Wong PC, Skoe E, Russo NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat Neurosci* 10:420–422.
- Musacchia G, Sams M, Skoe E, Kraus N (2007) Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc Natl Acad Sci USA* 104: 15894–15898.
- Tierney AT, Krizman J, Kraus N (2015) Music training alters the course of adolescent auditory development. *Proc Natl Acad Sci USA* 112:10062–10067.
- Trehub SE (2003) The developmental origins of musicality. *Nat Neurosci* 6:669–673.
- Tan YT, McPherson GE, Peretz I, Berkovic SF, Wilson SJ (2014) The genetic basis of music ability. *Front Psychol* 5:658.
- Pullii K, et al. (2008) Genome-wide linkage scan for loci of musical aptitude in Finnish families: Evidence for a major locus at 4q22. *J Med Genet* 45:451–456.
- Ukkola LT, Onkamo P, Raijas P, Karma K, Järvelä I (2009) Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One* 4:e5534.
- Park H, et al. (2012) Comprehensive genomic analyses associate UGT8 variants with musical ability in a Mongolian population. *J Med Genet* 49:747–752.
- Norton A, et al. (2005) Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain Cogn* 59:124–134.
- Honing H (2018) *The Origins of Musicality* (The MIT Press, Cambridge, MA).
- Barros CG, et al. (2017) Assessing music perception in young children: Evidence for and psychometric features of the M-factor. *Front Neurosci* 11:18.
- Gordon EE (1989) *Advanced Measures of Music Audiation* (GIA, Chicago).
- Law LNC, Zentner M (2012) Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One* 7:e52508.
- Seashore CE, Lewis L, Saetveit JG (1960) *Seashore Measures of Musical Talents* (The Psychological Corporation, New York).
- Wallentin M, et al. (2010) The Musical Ear Test, a new reliable test for measuring musical competence. *Learn Individ Differ* 20:188–196.
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116:2395–2405.
- Skoe E, Krizman J, Kraus N (2013) The impoverished brain: Disparities in maternal education affect the neural response to sound. *J Neurosci* 33:17221–17231.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. *J Neurosci* 32:14156–14164.
- Bidelman GM, Krishnan A, Gandour JT (2011) Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur J Neurosci* 33:530–538.
- Parbery-Clark A, Marmel F, Bair J, Kraus N (2011) What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci* 33:549–557.
- Bidelman GM, Davis MK, Pridgen MH (2018) Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear Res* 367: 149–160.
- Parbery-Clark A, Skoe E, Lam C, Kraus N (2009) Musician enhancement for speech-in-noise. *Ear Hear* 30:653–661.
- Bidelman GM, Weiss MW, Moreno S, Alain C (2014) Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci* 40:2662–2673.
- Musacchia G, Strait D, Kraus N (2008) Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear Res* 241:34–42.
- Zendel BR, Alain C (2012) Musicians experience less age-related decline in central auditory processing. *Psychol Aging* 27:410–417.
- Wollman I, Penhune V, Segado M, Carpentier T, Zatorre RJ (2018) Neural network retuning and neural predictors of learning success associated with cello training. *Proc Natl Acad Sci USA* 115:E6056–E6064.
- Hyde KL, et al. (2009) The effects of musical training on structural brain development: A longitudinal study. *Ann N Y Acad Sci* 1169:182–186.
- Jaschke AC, Honing H, Scherder EJA (2018) Longitudinal analysis of music education on executive functions in primary school children. *Front Neurosci* 12:103.
- Kraus N, et al. (2014) Music enrichment programs improve the neural encoding of speech in at-risk children. *J Neurosci* 34:11913–11918.
- Skoe E, Krizman J, Anderson S, Kraus N (2015) Stability and plasticity of auditory brainstem function across the lifespan. *Cereb Cortex* 25:1415–1426.
- Zendel BR, Alain C (2009) Concurrent sound segregation is enhanced in musicians. *J Cogn Neurosci* 21:1488–1498.
- Slee SJ, David SV (2015) Rapid task-related plasticity of spectrotemporal receptive fields in the auditory midbrain. *J Neurosci* 35:13090–13102.
- Bidelman GM, Villafuerte JW, Moreno S, Alain C (2014) Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol Aging* 35:2526–2540.
- Du Y, Zatorre RJ (2017) Musical training sharpens and bonds ears and tongue to hear speech better. *Proc Natl Acad Sci USA* 114:13579–13584.
- Chobert J, Marie C, François C, Schön D, Besson M (2011) Enhanced passive and active processing of syllables in musician children. *J Cogn Neurosci* 23:3874–3887.
- Zendel BR, Alain C (2013) The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *J Cogn Neurosci* 25:503–516.
- Bidelman GM, Pousson M, Dugas C, Fehrenbach A (2018) Test-retest reliability of dual-recorded brainstem versus cortical auditory-evoked potentials to speech. *J Am Acad Audiol* 29:164–174.
- Ehret G (1997) The auditory midbrain, a “shunting yard” of acoustical information processing. *The Central Auditory System*, eds Ehret G, Romand R (Oxford Univ Press, New York), pp 259–316.
- Foroughi CK, Monfort SS, Paczynski M, McKnight PE, Greenwood PM (2016) Placebo effects in cognitive training. *Proc Natl Acad Sci USA* 113:7470–7474.
- Fritz J, Elhilali M, Shamma S (2005) Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res* 206:159–176.
- Kunert R, Willems RM, Hagoort P (2016) An independent psychometric evaluation of the PROMS measure of music perception skills. *PLoS One* 11:e0159103.
- Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Bidelman GM (2015) Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J Neurosci Methods* 241:94–100.