



# Musical training sharpens and bonds ears and tongue to hear speech better

Yi Du<sup>a,b,c,1</sup> and Robert J. Zatorre<sup>b,d</sup>

<sup>a</sup>CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China 100101; <sup>b</sup>Montréal Neurological Institute, McGill University, Montréal, QC, Canada H3A 2B4; <sup>c</sup>Department of Psychology, University of Chinese Academy of Sciences, Beijing, China 100049; <sup>d</sup>International Laboratory for Brain, Music and Sound Research, Montréal, QC, Canada H3A 2B4; and <sup>e</sup>Centre for Research on Brain, Language and Music, Montréal, QC, Canada H3A 2B4

Edited by Dale Purves, Duke University, Durham, NC, and approved November 6, 2017 (received for review July 9, 2017)

**The idea that musical training improves speech perception in challenging listening environments is appealing and of clinical importance, yet the mechanisms of any such musician advantage are not well specified. Here, using functional magnetic resonance imaging (fMRI), we found that musicians outperformed nonmusicians in identifying syllables at varying signal-to-noise ratios (SNRs), which was associated with stronger activation of the left inferior frontal and right auditory regions in musicians compared with nonmusicians. Moreover, musicians showed greater specificity of phoneme representations in bilateral auditory and speech motor regions (e.g., premotor cortex) at higher SNRs and in the left speech motor regions at lower SNRs, as determined by multivoxel pattern analysis. Musical training also enhanced the intrahemispheric and interhemispheric functional connectivity between auditory and speech motor regions. Our findings suggest that improved speech in noise perception in musicians relies on stronger recruitment of, finer phonological representations in, and stronger functional connectivity between auditory and frontal speech motor cortices in both hemispheres, regions involved in bottom-up spectrotemporal analyses and top-down articulatory prediction and sensorimotor integration, respectively.**

musical training | speech in noise perception | auditory–motor integration | multivoxel pattern classification | functional connectivity

**M**usical training is associated with pervasive plasticity in human brains (1). However, does playing an instrument make us better able to understand speech in challenging listening environments? If so, which aspects of neural organization related to musical training contribute to this advantage? This is an important neurobiological question and also has clinical significance because speech in noise perception deficits disproportionately affect the elderly (2), children with language-related learning disorders (3), and those with hearing loss (4).

Many studies have found that musicians outperform nonmusicians in perceiving speech in background noise (5–10). These findings have led to the suggestion that musical training might be used to delay or counteract age-related declines in speech in noise perception (11). Due to the partial overlap between neural circuits dedicated to music and language (12), musical training is thought to strengthen the neurobiological and cognitive underpinnings of both music and speech processing. However, the musical training-related enhancement of speech in noise perception has been disputed (13, 14), on the grounds that auditory working memory, nonverbal IQ, and other factors may confound the results. In fact, speech in noise perception is a complex multifaceted process (10), which is not yet well characterized and may contribute to the heterogeneity of findings.

The ability to track and understand speech amid competing sound sources is supported by both the fidelity of bottom-up sensory encoding of target speech (15–17) and higher-level cognitive processes such as auditory working memory and selective attention (6, 18). If musical training improves auditory processing abilities (e.g., pitch, timing, and timbre) and cognitive skills, as well as their interaction, it might free up resources that could then be dedicated to flexibly adapting strategies to specific

task demands (2, 19). Additionally, auditory and motor functions are tightly coupled in the production and perception of both speech (20) and music (21). Auditory feedback is essential for real-time adjustment of motor commands in speech articulation and musical performance. In parallel, portions of Broca's area and premotor cortex (PMC) are believed to constrain the interpretation of speech and music in a predictive manner, especially under adverse listening contexts (21–24). According to the OPERA (overlap, precision, emotion, repetition, attention) hypothesis (25), because speech and music share overlapped networks and mechanisms of auditory–motor interaction and playing music places higher demands on the precision of processing than does speech, enhanced sensorimotor integration via years of musical practice would benefit speech in noise processing.

Here we first demonstrate a musician advantage on identifying syllables in noise when higher-order cognitive factors like auditory working memory and nonverbal IQ were controlled. Then, we disentangle the possible sources of such an advantage via complementary neuroimaging analyses that looked at regional activity, decoding of speech representations, and functional connectivity. We find that musical training enhances both bottom-up auditory encoding and top-down speech motoric predictions, as well as the auditory–motor integration that differentially contributes to speech in noise advantage depending on the noise intensity in the background.

## Results

**Behaviors.** Fifteen musicians (see Table S1 for details) and 15 nonmusicians identified English phoneme tokens (/ba/, /ma/, /da/, and /ta/) either alone or embedded in broadband noise at multiple

### Significance

**Musical training is a good thing, but does it benefit us in hearing speech in noise as has been suggested? If so, what is the brain mechanism behind this? Here we provide evidence of musician advantage on speech in noise perception at not only the behavioral level but also the level of neural representations of phonemes and functional connectivity. Results implicate better speech encoding in both auditory and speech motor regions, as well as stronger cross-modal auditory–motor integration in musicians than nonmusicians when processing speech, especially in noisy conditions. The idea that musical training improves speech in noise perception by enhancing auditory–motor integration is intriguing, with applications in alleviating speech perception difficulties in aging populations and hearing disorders.**

Author contributions: Y.D. and R.J.Z. designed research; Y.D. performed research; Y.D. analyzed data; and Y.D. and R.J.Z. wrote the paper.

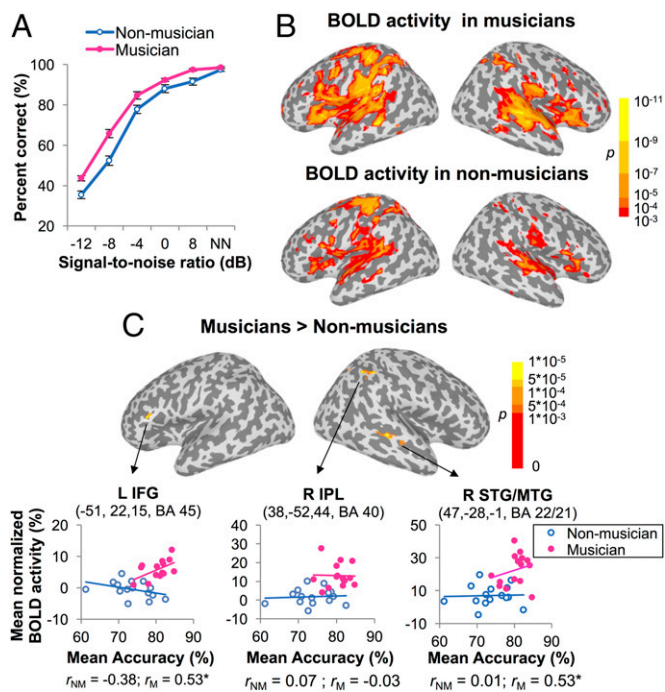
The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

<sup>1</sup>To whom correspondence should be addressed. Email: duy@psych.ac.cn.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1712231114/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1712231114/-DCSupplemental).



**Fig. 1.** Behavioral performance and blood oxygenation level-dependent (BOLD) activity. (A) Identification accuracy across syllables as a function of SNR in musicians and nonmusicians. NN represents the NoNoise condition. Error bars indicate SEM. (B) BOLD activity across conditions in musicians and non-musicians ( $P_{FWE} < 0.001$ ). (C) Regions where musicians showed stronger activity than nonmusicians regardless of conditions ( $P_{FWE} < 0.001$ ) and brain-behavior correlations between the mean activity across conditions in those regions and the mean accuracy in two groups. The coordinates are in Talairach space. \* $P < 0.05$  by Pearson's correlations.

signal-to-noise ratios (SNRs) (-12, -8, -4, 0, and 8 dB) during MRI scanning. The two groups did not differ in age ( $t_{28} = -0.55$ ,  $P = 0.59$ ), years of postsecondary education ( $t_{28} = 0.11$ ,  $P = 0.91$ ), pure-tone average thresholds ( $t_{28} = -0.11$ ,  $P = 0.91$ ), auditory working memory as measured by the forward and backward digit span subtest of the Wechsler Adult Intelligence Scale (26) ( $t_{28} = 0.25$ ,  $P = 0.81$ ), or nonverbal IQ as measured by Cattell's culture fair intelligence test (scale 3 A, ref. 27) ( $t_{28} = 0.51$ ,  $P = 0.62$ ) (Table S2).

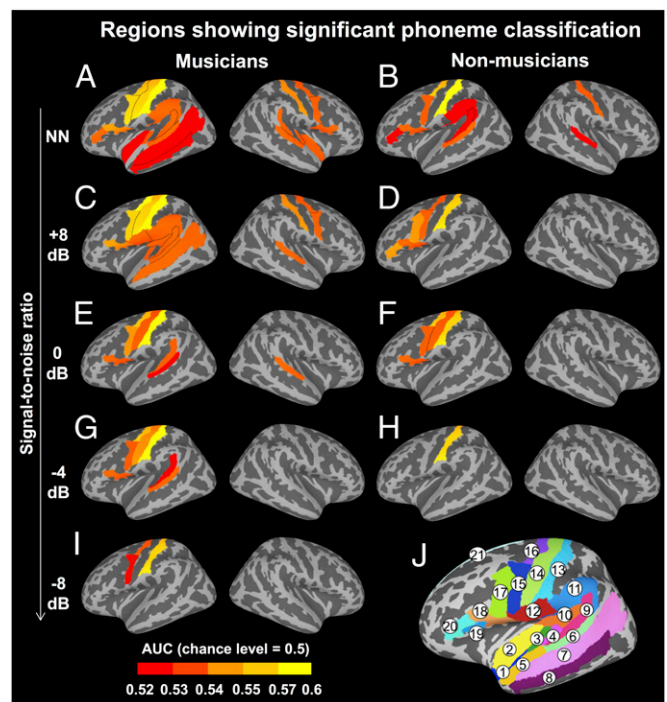
A mixed-effects ANOVA on arcsine-transformed accuracy revealed better performance in musicians than nonmusicians ( $F_{1,28} = 12.42$ ,  $P = 0.001$ ,  $\eta^2 = 0.31$ ). Accuracy elevated with increasing SNR in both groups ( $F_{5,24} = 260.79$ ,  $P < 0.001$ ,  $\eta^2 = 0.98$ ), without a significant group  $\times$  SNR interaction ( $F_{5,24} = 2.38$ ,  $P = 0.07$ ,  $\eta^2 = 0.33$ ; Fig. 1A). Accuracy across conditions did not correlate with digit span score ( $r = 0.24$ ,  $P = 0.21$ ,  $n = 30$ ) or culture fair intelligence score ( $r = 0.28$ ,  $P = 0.13$ ,  $n = 30$ ), nor with years of musical practice ( $r = 0.33$ ,  $P = 0.24$ ,  $n = 15$ ) or the age of training onset ( $r = -0.36$ ,  $P = 0.19$ ,  $n = 15$ ) in musicians. The two groups did not differ in their reaction time ( $F_{1,28} = 0.06$ ,  $P = 0.81$ , mixed-effects ANOVA).

**Effects on Regional Activity.** Collapsing across listening conditions, both groups showed widespread activation of bilateral perisylvian areas and thalamus, as well as left motor and somatosensory regions when identifying syllables relative to the intertrial baseline [Fig. 1B; familywise error-corrected  $P$  ( $P_{FWE}$ )  $< 0.001$ ]. Compared with nonmusicians, musicians showed significantly stronger activity in Broca's area of left inferior frontal gyrus (IFG, BA 45), right inferior parietal lobule (IPL, BA 40), and right superior and middle temporal gyri (STG/MTG, BA22/21) (Fig. 1C and Table S3;  $P_{FWE} < 0.001$ ). Within those three regions of interest (ROIs), the mean activity across conditions in left IFG and right STG/

MTG (both  $r = 0.53$ ,  $P = 0.043$ ) positively correlated with the mean behavioral accuracy in musicians but not so in nonmusicians (IFG:  $r = -0.38$ ,  $P = 0.16$ ; STG/MTG:  $r = 0.01$ ,  $P = 0.99$ ).

**Effects on Speech Representations.** We further assessed how musical expertise affects the decoding of speech representations via multivoxel pattern analysis (MVPA), which can detect fine-scale spatial patterns instead of mean neural activity elicited by different phonemes. MVPA was performed within 42 individually defined anatomical ROIs in both hemispheres, selected independently because they are critical for speech processing according to Neurosynth (Materials and Methods and Fig. 2J). Those 42 ROIs largely overlap with the activation map elicited by the syllable in noise perception task in our participants (Fig. S1), thus validating the selection process.

When nonmusicians identified syllables presented alone, significant phoneme classification [area under the curve (AUC)  $> 0.5$  chance level, one-sample  $t$  tests with FDR-corrected  $P < 0.05$ ] was observed in posterior STG (pSTG) and postcentral gyrus (postCG) bilaterally and planum temporale (PT), supramarginal gyrus (SMG), precentral gyrus (preCG), pars opercularis (POP), and pars triangularis (PTr) of Broca's area in the left hemisphere (Fig. 2B and Table S4). In contrast, for musicians at NoNoise condition, phoneme representations could be reliably classified in more widely distributed auditory and frontal motor regions than those found in nonmusicians (more so in the right hemisphere), including additional bilateral Heschl's gyrus (HG) and inferior



**Fig. 2.** Phoneme classification performance. (A-I) Regions with significant phoneme classification (AUC  $> 0.5$ , one-sample  $t$  test with FDR-corrected  $P < 0.05$ ) at each SNR in musicians (Left) and nonmusicians (Right). (J) Speech-relevant anatomical ROIs used in multivoxel pattern analysis. The ROI mask consisting of 21 left and 21 right ROIs was created by intersecting a Neurosynth automated metaanalysis (search term: speech) and 152 Freesurfer anatomical ROIs (aparc 2009 atlas). 1, planum polare; 2, inferior insula; 3, Heschl's gyrus; 4, transverse temporal sulcus; 5, anterior superior temporal gyrus; 6, posterior superior temporal gyrus; 7, superior temporal sulcus; 8, middle temporal gyrus; 9, planum temporale; 10, posterior lateral fissure; 11, supramarginal gyrus; 12, subcentral gyrus/sulcus; 13, postcentral insula; 14, central sulcus; 15, ventral precentral gyrus; 16, dorsal precentral gyrus; 17, inferior precentral sulcus; 18, inferior frontal gyrus, pars opercularis; 19, anterior lateral fissure; 20, inferior frontal gyrus, pars triangularis; 21, superior frontal gyrus.

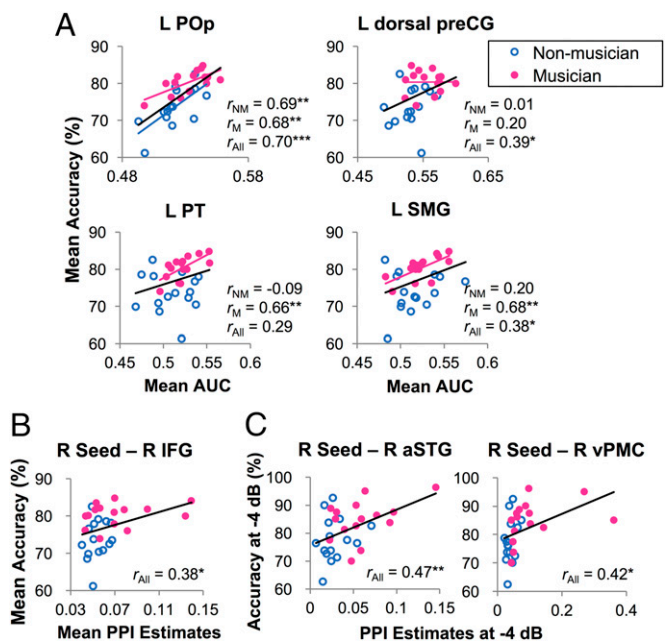
insula, left superior temporal sulcus (STS), MTG and central sulcus, and right PT, preCG, and POP (Fig. 2A and Table S4).

Musicians also had increased resilience to noise interference on phoneme representations. For nonmusicians, weak noise (SNR = 8 dB) disrupted phoneme classification in bilateral auditory regions, with significant decoding revealed only in left frontal motor and somatosensory regions including POP, PT, inferior precentral sulcus, preCG, and postCG (Fig. 2D and Table S4). Phoneme specificity in those regions was mainly retained at 0 dB SNR, disappeared except in left postCG at -4 dB SNR, and was not detected at all when the noise further increased (Fig. 2F and H and Table S4). For musicians, phoneme classification was significant in bilateral pSTG, preCG, and postCG, as well as in left HG, STS, PT, SMG, central sulcus, and POP at 8 dB SNR (Fig. 2C and Table S4), which did not dramatically differ from what was observed when the noise was absent. Significant phoneme decoding was further found in right pSTG until 0 dB SNR; in left pSTG, PT, postCG, central sulcus, preCG, and POP at 0 and -4 dB SNRs; and in left postCG and preCG at -8 dB SNR (Fig. 2E, G, and I and Table S4).

The musician advantage on phoneme decoding was statistically quantified by a mixed-effects ANOVA that revealed significant group difference on AUC scores ( $F_{1,28} = 4.60, P = 0.04, \eta^2 = 0.14$ ) without any significant interaction between group and ROI ( $F_{41,1148} = 1.20, P = 0.18$ ) or between group and SNR ( $F_{5,140} = 0.94, P = 0.46$ ). However, it appeared that training enhancement on phoneme specificity hierarchically moved from bilateral auditory cortices when the noise was weak to the speech motor regions when the noise was intensive. For instance, phoneme decoding was revealed in right pSTG until 0 dB SNR and in left pSTG until -4 dB SNR in musicians, whereas decoding was not reliable in nonmusicians' auditory regions once the noise was present. Although phoneme classification was absent in speech motor regions when SNR < 0 dB in nonmusicians, it was significant in left POP until -4 dB SNR and in dorsal/ventral preCG until -8 dB SNR in musicians. Thus, musical training was associated with improvement in both auditory decoding and motor prediction-based decoding of speech signals in a pattern-specific fashion, which interacted with the noise intensity in the background.

Moreover, improved phoneme representations in speech motor regions and auditory-motor interfaces (e.g., left PT and SMG) predicted better behavioral performance. Within the 21 ROIs with significant phoneme decoding at NoNoise condition in musicians (Fig. 2A), the overall accuracy positively correlated with the overall AUC scores in left POP ( $r = 0.70$ , uncorrected  $P = 0.000$ , FDR-corrected  $P < 0.05$ ), left dorsal preCG ( $r = 0.39$ , uncorrected  $P = 0.03$ ), and left SMG ( $r = 0.38$ , uncorrected  $P = 0.039$ ) across all of the subjects (Fig. 3A). For musicians alone, AUC scores in left POP, left PT, and left SMG predicted accuracy (all  $r > 0.66$ , uncorrected  $P < 0.007$ , FDR-corrected  $P < 0.05$ ). Such a correlation was also found in nonmusicians' left POP ( $r = 0.69$ , uncorrected  $P = 0.005$ , FDR-corrected  $P < 0.05$ ).

**Effects on Functional Connectivity.** Generalized psychophysiological interaction (gPPI; ref. 28) was used to investigate how the functional connectivity between auditory and motor regions was modulated by musical experience and the listening environment (i.e., SNR). gPPI is configured to evaluate how brain regions interact in a context-dependent manner, i.e., the modulation of functional connectivity by a psychological or behavioral context, when there are more than two task conditions. Here an auditory seed combining pSTG and PT Freesurfer ROIs (aparc 2009) in each hemisphere was selected because the two regions together act as auditory-motor interfaces during speech perception and are most likely to send and receive connections to and from motor regions (29). A group  $\times$  SNR mixed-effects ANOVA on PPI estimates revealed significantly ( $P_{FWE} < 0.01$ ) stronger functional connectivity between the left auditory seed and left dorsal PMC (dPMC) and right IFG (including both BA 44 and BA 45) in musicians than nonmusicians. For the right auditory seed, stronger connectivity was found with right HG, right IFG



**Fig. 3.** Correlations between phoneme classification performance, functional connectivity, and behavior performance. (A) Regions showing significant correlation between the mean AUC scores and the mean behavioral accuracy across SNRs. (B) Functional connectivity showing significant correlation between the mean PPI estimates and the mean accuracy across SNRs. (C) Functional connectivity showing significant correlations between the PPI estimates and accuracy at -4 dB SNR. \* $P < 0.05$ , \*\* $P < 0.01$ , and \*\*\* $P < 0.001$  by Pearson's correlations, uncorrected.

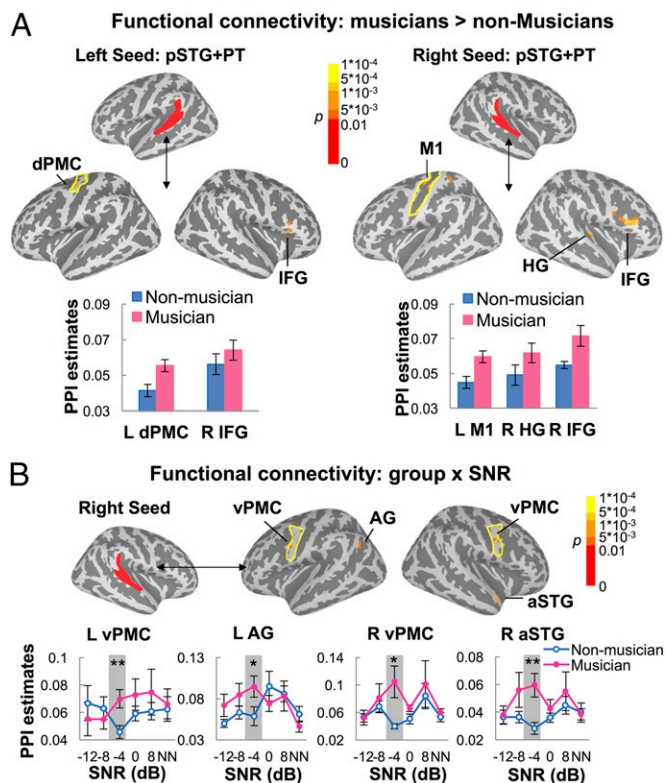
(BA 44, BA 45), and left primary motor cortex (M1) regions in musicians than nonmusicians ( $P_{FWE} < 0.01$ ; Fig. 4A and Table S5). Relative to musicians, nonmusicians showed stronger functional connectivity between bilateral auditory seeds and right cerebellum (Table S5).

A significant group by SNR interaction was revealed between the right auditory seed with right anterior STG (aSTG), bilateral ventral PMC (vPMC), and left angular gyrus (AG) ( $P_{FWE} < 0.01$ ; Fig. 4B and Table S5). Further independent samples  $t$  tests showed significantly stronger connectivity in musicians than nonmusicians at -4 dB SNR between the right auditory seed and all four regions: right aSTG ( $t_{28} = 3.27, P = 0.003$ ), left vPMC ( $t_{28} = 2.96, P = 0.006$ ), right vPMC ( $t_{28} = 2.73, P = 0.01$ ), and left AG ( $t_{28} = 2.09, P = 0.046$ ). Only the connectivity with right aSTG and left vPMC passed correction for multiple comparisons. Thus, musicians showed strengthened functional connectivity between auditory regions and some of the motor regions (e.g., left dPMC) regardless of SNR and enhanced connectivity with other motor areas (e.g., left vPMC) when the noise was relatively intense.

Moreover, stronger functional connectivity between the right auditory seed and right IFG predicted higher behavioral accuracy across conditions and subjects ( $r = 0.38$ , uncorrected  $P = 0.039$ ; Fig. 3B). At -4 dB SNR, functional connectivity between the right auditory seed and right aSTG and right vPMC not only was stronger in musicians than nonmusicians (Fig. 4B), but also positively correlated with accuracy across subjects (right aSTG:  $r = 0.47$ , uncorrected  $P = 0.009$ , FDR-corrected  $P < 0.05$ ; right vPMC:  $r = 0.42$ , uncorrected  $P = 0.02$ ; Fig. 3C). No behavior-PPI correlation was found for each group alone, nor did any correlation reach significance between PPI and classification performance in any region.

## Discussion

This study investigates whether and how long-term musical training contributes to enhanced speech perception in noisy environments. Using a syllable in noise identification task, we did



**Fig. 4.** Training effects on functional connectivity as measured with gPPI. (A) Regions where musicians, relative to nonmusicians, showed significantly stronger ( $P_{FWE} < 0.01$ , mixed-effects ANOVA) PPI with the left (Left) or right (Right) auditory seed (individually defined anatomical ROIs: pSTG + PT). Yellow line shows the contour of anatomically defined dorsal precentral gyrus (Left) and central sulcus (Right). (B) A significant group  $\times$  SNR interaction ( $P_{FWE} < 0.01$ , mixed-effects ANOVA) was revealed for the PPI between the right auditory seed and four regions (bilateral vPMC, left AG, and right aSTG). Yellow line shows the contour of anatomically defined ventral precentral gyrus. \* $P < 0.05$ , \*\* $P < 0.008$  by independent-sample  $t$  tests, uncorrected. Error bars indicate SEM.

observe a musician benefit behaviorally. Moreover, relative to nonmusicians, improved performance in musicians was paralleled by (i) increased activity of Broca's area in left IFG and right auditory cortex, (ii) higher specificity of phoneme representations in both auditory and motor regions in both hemispheres, and (iii) stronger intrahemispherical and interhemispherical functional connectivity between auditory and motor regions. Our findings suggest that musical training may enhance auditory encoding, speech motor prediction, and auditory-motor integration that together contribute to superior speech perception in adverse listening conditions.

**Musician Benefit and Cognitive Factors.** Debates persist regarding whether musical expertise shapes speech perception in challenging listening environments. Two recent studies failed to reveal significant differences between musicians and nonmusicians in understanding sentences masked by speech-shaped noise or speech babble (13, 14). Still, a musician advantage has been found over a variety of masking conditions and timescales from phonemes to sentences (5–10). The discrepancy between studies might be due to sampling error, musician heterogeneity, or tasks that could be solved using multiple cues (e.g., spatial difference) that may close the gap between musicians and nonmusicians (10). In the present study, musicians outperformed nonmusicians in identifying syllables embedded in broadband noise at all SNRs but not in quiet (Fig. 14), supporting the notion that musicianship enhances resistance to noise. The task used involves

no informational masking and relies little on executive functions such as auditory selective attention, working memory, and cognitive control because listeners do not need to segregate or suppress irrelevant sound or hold information in mind. Although enhanced higher-order cognitive processes are found to mediate improved speech in noise perception in musicians at the sentence level (5, 6, 14), they cannot account for the musician benefit here due to the task nature and balanced cognitive scores (auditory working memory and nonverbal IQ) between groups. Although there is likely an interplay between genetic and other predispositions with experience-dependent plasticity of brain circuitry that gives rise to the training effects (30), our findings strongly indicate that musical expertise can boost speech in noise perception which is grounded in enhanced neural processing of speech at the phonemic and syllabic levels, irrespective of higher-order cognitive factors.

No training effect was found for reaction time, which fits with previous findings (8), although most studies did not report it. Additional contrasts on BOLD activity, phoneme classification performance and functional connectivity between participants with shorter and longer reaction time did not reveal any significant effect either. Thus, musical training may improve speech discrimination at the perceptual level but may not accelerate the response selection and decision making process. Different from expectations (5, 13), accuracy here did not correlate with years of practice or the age of start. However, the present study was not designed to examine effects of age of start, because only musicians who started training before 7 were recruited; also, to test effects of years of practice a broader range may be necessary.

**Enhanced Auditory Encoding.** Speech encoding with high fidelity in the presence of noise along the auditory pathway is necessary for matching neural representations of incoming acoustic signals to stored lexical representations (9). Previous studies have underscored the importance of faithful encoding of speech (e.g., frequency following response) in brainstem, thalamus, and cortex for speech in noise perception (17). Musicians show superior frequency following responses, including more robust encoding of speech spectral features and greater neural temporal precision (18, 31). Musicianship also yields coordinated neural plasticity in brainstem and auditory cortex, and such a refined hierarchy of speech representations may provide more precise phonemic templates to linguistic decisions that contribute to better speech perception (31). Here phoneme representations with higher specificity and stronger resistance to noise degradation were revealed in bilateral auditory cortices in musicians than non-musicians (Fig. 2), confirming that enhanced auditory encoding may partially explain the musician benefit on speech in noise perception. Notably, it is bilateral pSTG, the auditory-motor interface in the auditory dorsal stream (29), where the phoneme representations were sharpened in noisy conditions by musical experience. Additionally, musicians compared with nonmusicians showed stronger recruitment of the right auditory cortex in which the activity scaled with performance.

**Improved Speech Motor Prediction.** Motor contributions to speech perception in disambiguating phonological information under adverse listening contexts are increasingly emphasized recently (24). For instance, more robust phoneme representations in speech motor regions are suggested to compensate for noise-impooverished speech representations in auditory cortex (22), and such a mechanism becomes more critical for counteracting the age-related declines in speech perception for older listeners (23). Because playing music is one of the most complex sensory-motor activities, requiring precise timing of several hierarchically organized actions as well as precise control over pitch interval production (21), plasticity in the motor network is commonly found in musicians (32, 33). It is plausible that the motor plasticity may bolster musicians' ability to generate more accurate articulatory predictions in a cross-domain fashion that enhances the top-down motoric modulation of speech perception,

particularly when the auditory system cannot adequately parse speech signals due to noise masking. Indeed, musicians exhibited stronger activity of Broca's area, as well as higher and more robust phoneme specificity in the speech motor system than nonmusicians. Specifically, training effects on phoneme representations were found in right PMC and IFG (the counterpart of Broca's area, BA44) when the noise was absent or weak and in left PMC and Broca's area (BA44) when the noise was relatively strong (Fig. 2). Moreover, how well phonemes were represented in Broca's area (BA44) and auditory-motor interfaces (left PT and SMG) predicted task performance in musicians, and phoneme specificity in left PMC correlated with behavioral accuracy across subjects (Fig. 3A), suggesting a direct link between improved speech motoric representations and better performance.

**Strengthened Auditory-Motor Integration.** Years of musical training should affect not only the auditory and motor modalities separately but also their interactions. In fact, musicians show better organized white matter connections (34), greater resting-state connectivity (35), and higher intercorrelations of cortical thickness (32) between auditory and frontal motor cortices. In line with the literature, here musicians showed enhanced intrahemispherical and interhemispherical functional connectivity between auditory and speech motor regions (e.g., left dPMC and bilateral vPMC) (Fig. 4). Furthermore, strengthened auditory-motor connectivity was associated with improved speech in noise perception because subjects with higher connectivity strength between the right auditory seed and right frontal motor regions (e.g., IFG and vPMC) performed better (Fig. 3B and C). It is proposed that the vPMC and dPMC are implicated in direct and indirect sensorimotor transformations, respectively (21). Mapping auditory features to motor commands when learning to play an instrument engages the vPMC (36) and dPMC (37–39). The dPMC is putatively involved in extracting higher-order sound features to implement temporally organized actions and allow for predictability in perception (21, 39). Listening to speech may not only entail activation of articulatory programs enabled through vPMC links but also engage a neural circuit—in which dPMC is a crucial node—that sets up temporal expectancies in understanding speech. Our findings suggest that musical training differentially affects the functional connectivity of dPMC and vPMC with auditory regions depending on the listening contexts. It is likely that musicians may have enhanced auditory-motor interplay in temporal prediction of speech signals regardless of the listening condition (40), which, however, needs to be tested in future. In parallel, musicians may have strengthened direct mapping of articulatory predictions to auditory inputs, and this function seems to reach its peak at mediate SNRs (e.g., -4 dB) when motoric representations excel auditory representations in specificity as shown before (22, 23).

**Task Difficulty and Hemispheric Asymmetry.** It is reported that the speech motor system is recruited to a greater extent at more challenging conditions (22–24). Here training-related improvement on speech motoric and auditory representations interacted with task difficulty. Specifically, enhanced speech representations dominated musicians' bilateral auditory regions (more prominent in the left hemisphere) at high SNRs (NoNoise ~ -4 dB), whereas it targeted right motor regions at high SNRs but left motor regions at low SNRs (-4 ~ -8 dB). This indicates that musicians may benefit from dynamic speech decoding strategies from relying on refined auditory cues to strengthened motor predictions as difficulty increases. Thus, musician advantage on speech in noise perception has different contributors from auditory ventral and dorsal streams depending on the listening context.

Another finding is that listeners recruit the right hemisphere to a different extent according to their experience. Compared with nonmusicians, musicians had increased activity of right auditory cortex and higher phoneme specificity in right auditory and motor regions, as well as stronger functional connectivity between right auditory and bilateral motor regions. This fits with

previous findings that highlight greater training-related plasticity in right compared with left auditory cortex (1, 32). Such an increased contribution by right auditory cortex and the bilaterally organized auditory-motor integration network in musicians may provide superiority in processing speech in noise compared with the left-lateralized one in nonmusicians (20).

In sum, this study demonstrates musical training-related benefits of speech in noise perception, which is grounded on increased recruitment of auditory and motor cortices, enhanced auditory decoding and motor prediction, plus strengthened auditory-motor integration. Moreover, the auditory and motor contributions to the musician advantage are dynamically weighted according to the task difficulty. Musical training thereby has great potentials to set up the communicating brain for healthy aging and hearing disorders (2, 11).

## Materials and Methods

**Subjects.** Fifteen musicians ( $21.4 \pm 2.7$  y, seven females) and 15 nonmusicians ( $22.1 \pm 4.4$  y, seven females) gave informed written consent approved by the McGill University Health Centre Research Ethics Board to participate in the study. All subjects were healthy right-handed native English speakers with normal pure-tone thresholds at both ears ( $<25$  dB HL for 250–8,000 Hz). Musicians had started training before age 7, had at least 10 y of musical training, and reported practicing consistently ( $\geq 3$  times per week) over the past 3 y (Table S1). Nonmusicians reported less than 1 y of musical experience, which did not occur in the year before the experiment.

**Stimuli and Task.** The stimuli were four naturally produced American English consonant-vowel syllables (/ba/, /ma/, /da/, and /ta/), spoken by a female talker. Each token was 500 ms in duration and matched for average root-mean-square sound pressure level (SPL). A 500-ms white noise segment (4-kHz low-pass, 10-ms rise-decay envelope) starting and ending simultaneously with syllables was used as the masker. Sounds were played by a TDT (Tucker-Davis Technologies) RX-6 real-time processor and presented via MRI-compatible Sensimetrics S14 insert earphones (Sensimetrics Corporation) with Comply foam tips, which maximally attenuate scanner noise by 40 dB. The syllables were fixed at 85-dB SPL; the noise level was adjusted to generate five SNRs (-12, -8, -4, 0, and 8 dB) and the NoNoise condition.

During scanning, 96 stimuli (four trials per syllable per noise condition) were randomly presented in each block with an average interstimulus interval of 4 s (2–6 s, 0.5-s step), and five blocks were given in total. Subjects were asked to identify syllables as fast as possible by pressing one of four keys on a parallel four-button pad using their right hand (index to little fingers in response to /ba/, /da/, /ma/, and /ta/ sequentially).

**Data Acquisition and Preprocessing.** Imaging data were collected using a 3.0-T MRI system (Siemens Magnetom Trio) with a 32-channel head coil. T1-weighted anatomical images were acquired using a magnetization-prepared rapid acquisition gradient echo sequence (sagittal orientation, 192 slices, repetition time (TR) = 2,300 ms, echo time (TE) = 2.98 ms, field of view (FOV) = 256 mm, voxel size =  $1 \times 1 \times 1$  mm). T2\*-weighted functional images were acquired with a continuous multiband-accelerated echo planar imaging sequence (multiband factor = 4, 40 slices, TR = 636 ms, TE = 30 ms, flip angle =  $90^\circ$ , FOV = 192 mm, voxel size =  $3 \times 3 \times 3$  mm). The fMRI data were preprocessed using Analysis of Functional NeuroImages (AFNI) software, including slice timing correction, spatial alignment, image coregistration and normalization. The preprocessed images were then analyzed by General Linear Model (GLM) and MVPA.

**General Linear Model Analysis.** Multiple-regression modeling was performed using the AFNI program 3dDeconvolve. Data were fit with different regressors for four syllables and six noise conditions. The predicted activation time course was modeled as a gamma function convolved with the canonical hemodynamic response function. For each SNR, the four syllables were grouped and contrasted against the baseline (no-stimulus intertrial intervals), as GLM revealed similar activity across syllables. Contrast maps were normalized to Talairach stereotaxic space and spatially smoothed using a Gaussian filter (FWHM = 6.0 mm). Individual maps at each SNR were then subjected to a mixed-effects ANOVA to test the random effects for each group and the main effect of group on BOLD activity. Multiple comparisons were corrected using 3dClustSim with 1,000 Monte Carlo simulations. This yielded a  $P_{FWE} < 0.001$  by using an uncorrected  $P < 0.001$  and removing clusters  $<20$  voxels for the group mean activation maps in both groups (Fig. 1B). For the group difference map, this yielded a  $P_{FWE} < 0.001$ , with an uncorrected  $P < 0.001$ , and cluster size  $\geq 3$  voxels (Fig. 1C).

Results were projected onto a cortical inflated surface template using surface mapping (SUMA) with AFNI.

**Multivoxel Pattern Analysis.** Given the likelihood of high intersubject anatomical variability and fine spatial scale of phoneme representations, pattern classifiers were trained to discriminate neural patterns associated with different phonemes and then tested on independent trials within anatomically defined ROIs. To do so, univariate trialwise  $\beta$  coefficients were first estimated using AFNI program 3dLSS (Least Square Sum regression; ref. 41). Then Freesurfer's automatic anatomical parcellation (aparc2009; ref. 42) algorithm was used to define a set of 152 cortical and subcortical ROIs from individual's anatomical scan. STG was further divided into equational anterior and posterior portions, and preCG was divided into equational dorsal and ventral parts to dissociate their potentially different contributions to speech perception. Next, 21 left and 21 right ROIs sensitive to speech production and perception were selected (Fig. 2J). This was done by intersecting a metaanalytic mask on Neurosynth (43) (search term: speech) with the Freesurfer mask defined in Montreal Neurological Institute space. MVPA were then carried out in the volumetric space within the 42 ROIs at each SNR, using shrinkage discriminant analysis (R package sda; ref. 44) followed by fivefold cross-validation. A multiclass AUC measure computed as the average of all of the pairwise two-class AUC scores was used as an index of classification performance (SI Text).

Significance of classification was evaluated by one-sample  $t$  tests in each ROI at each SNR, where the null hypothesis assumed a theoretical chance

AUC of 0.5. Multiple comparisons were corrected by using a FDR  $q = 0.05$ . AUC scores were also subjected to a mixed-effects ANOVA to evaluate the group difference in classification. Results were then projected on the parcellated cortical inflated map associated with the Freesurfer average template (fsaverage) using SUMA.

**Psychophysiological Interaction.** gPPI analysis (28) was implemented to evaluate the SNR-dependent interaction of the functional connectivity between the auditory seeds and all other regions in the brain. The auditory seeds were defined by merging the PT and pSTG labels from Freesurfer anatomical parcellation and converted to the volumetric space. gPPI was conducted in the volumetric space, and the results were projected on the template surface for visualization. A mixed-effects ANOVA on gPPI estimates tested the main effect of group and the group  $\times$  SNR interaction for each seed, separately. Multiple comparisons were corrected by 3dClustSim using an uncorrected  $P < 0.01$ , which yielded a  $P_{FWE} < 0.01$  by removing clusters  $< 6$  and 10 voxels for the group difference map for the left and right auditory seed, respectively (Fig. 4A) and removing clusters  $< 6$  voxels for the group by SNR interaction map for the right auditory seed (Fig. 4B).

**ACKNOWLEDGMENTS.** This research was supported by grants from the National Natural Science Foundation of China (31671172) and the Thousand Young Talent Plan (to Y.D.) and the Canadian Institutes of Health Research (Foundation Grant) and an infrastructure grant from the Canada Fund for Innovation (to R.J.Z.).

- Herholz SC, Zatorre RJ (2012) Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76:486–502.
- Zendel BR, Alain C (2012) Musicians experience less age-related decline in central auditory processing. *Psychol Aging* 27:410–417.
- Ziegler JC, Pech-Georgel C, George F, Alario FX, Lorenzi C (2005) Deficits in speech perception predict language learning impairment. *Proc Natl Acad Sci USA* 102:14110–14115.
- Wilson RH, McArdle RA, Smith SL (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856.
- Parbery-Clark A, Skoe E, Lam C, Kraus N (2009) Musician enhancement for speech-in-noise. *Ear Hear* 30:653–661.
- Strait DL, Kraus N (2011) Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Front Psychol* 2:113.
- Swaminathan J, et al. (2015) Musical training, individual differences and the cocktail party problem. *Sci Rep* 5:11628.
- Varnet L, Wang T, Peter C, Meunier F, Hoen M (2015) How musical expertise shapes speech perception: Evidence from auditory classification images. *Sci Rep* 5:14489.
- Zendel BR, Tremblay CD, Belleville S, Peretz J (2015) The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J Cogn Neurosci* 27:1044–1059.
- Coffey EBJ, Mogilever NB, Zatorre RJ (2017) Speech-in-noise perception in musicians: A review. *Hear Res* 352:49–69.
- Alain C, Zendel BR, Hutka S, Bidelman GM (2014) Turning down the noise: The benefit of musical training on the aging auditory brain. *Hear Res* 308:162–173.
- Peretz J, Vuvan D, Lagrois ME, Armony JL (2015) Neural overlap in processing music and speech. *Philos Trans R Soc Lond B Biol Sci* 370:20140090.
- Ruggles DR, Freyman RL, Oxenham AJ (2014) Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One* 9:e86980.
- Boebinger D, et al. (2015) Musicians and non-musicians are equally adept at perceiving masked speech. *J Acoust Soc Am* 137:378–387.
- Anderson S, Kraus N (2010) Sensory-cognitive interaction in the neural encoding of speech in noise: A review. *J Am Acad Audiol* 21:575–585.
- Du Y, Kong L, Wang Q, Wu X, Li L (2011) Auditory frequency-following response: A neurophysiological measure for studying the "cocktail-party problem". *Neurosci Biobehav Rev* 35:2046–2057.
- Coffey EBJ, Chapesiuk AMP, Herholz SC, Baillet S, Zatorre RJ (2017) Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Front Neurosci* 11:479.
- Kraus N, Strait DL, Parbery-Clark A (2012) Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Ann N Y Acad Sci* 1252:100–107.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11:599–605.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: Auditory-motor interactions in music perception and production. *Nat Rev Neurosci* 8:547–558.
- Du Y, Buchsbaum BR, Grady CL, Alain C (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci USA* 111:7126–7131.
- Du Y, Buchsbaum BR, Grady CL, Alain C (2016) Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat Commun* 7:12241.
- Skipper JI, Devlin JT, Lametti DR (2017) The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang* 164:77–105.
- Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front Psychol* 2:142.
- Wechsler D (1997) *Wechsler Adult Intelligence Scale* (The Psychological Corporation, San Antonio), 3rd Ed, pp 1–237.
- Cattell RB (1957) *Culture Fair Intelligence Test, a Measure of "g": Scale 3, Forms A and B* (High School Pupils and Adults of Superior Intelligence) (Inst Pers Ability Test, Champaign, IL).
- McLaren DG, Ries ML, Xu G, Johnson SC (2012) A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage* 61:1277–1286.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *J Cogn Neurosci* 15:673–682.
- Zatorre RJ (2013) Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science* 342:585–589.
- 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.
- Bermudez P, Lerch JP, Evans AC, Zatorre RJ (2009) Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex* 19:1583–1596.
- Gaser C, Schlaug G (2003) Brain structures differ between musicians and non-musicians. *J Neurosci* 23:9240–9245.
- Halwani GF, Loui P, Rüber T, Schlaug G (2011) Effects of practice and experience on the arcuate fasciculus: Comparing singers, instrumentalists, and non-musicians. *Front Psychol* 2:156.
- Palomar-García MÁ, Zatorre RJ, Ventura-Campos N, Bueichekú E, Ávila C (2017) Modulation of functional connectivity in auditory-motor networks in musicians compared with nonmusicians. *Cereb Cortex* 27:2768–2778.
- Lahav A, Saltzman E, Schlaug G (2007) Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *J Neurosci* 27:308–314.
- Herholz SC, Coffey EB, Pantev C, Zatorre RJ (2016) Dissociation of neural networks for predisposition and for training-related plasticity in auditory-motor learning. *Cereb Cortex* 26:3125–3134.
- Lega C, Stephan MA, Zatorre RJ, Penhune V (2016) Testing the role of dorsal pre-motor cortex in auditory-motor association learning using transcranial magnetic stimulation (TMS). *PLoS One* 11:e0163380.
- Chen JL, Rae C, Watkins KE (2012) Learning to play a melody: An fMRI study examining the formation of auditory-motor associations. *Neuroimage* 59:1200–1208.
- Schön D, Tillmann B (2015) Short- and long-term rhythmic interventions: Perspectives for language rehabilitation. *Ann N Y Acad Sci* 1337:32–39.
- Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage* 59:2636–2643.
- Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53:1–15.
- Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 8:665–670.
- Ahdesmäki M, Strimmer K (2010) Feature selection in omics prediction problems using cat scores and false non-discovery rate control. *Ann Appl Stat* 4:503–519.