



# Piano training enhances the neural processing of pitch and improves speech perception in Mandarin-speaking children

Yun Nan<sup>a,1</sup>, Li Liu<sup>a</sup>, Eveline Geiser<sup>b,c,d</sup>, Hua Shu<sup>a</sup>, Chen Chen Gong<sup>b</sup>, Qi Dong<sup>a</sup>, John D. E. Gabrieli<sup>b,e</sup>, and Robert Desimone<sup>b,1</sup>

<sup>a</sup>State Key Laboratory of Cognitive Neuroscience and Learning, International Data Group (IDG)/McGovern Institute for Brain Research, Beijing Normal University, 100875 Beijing, China; <sup>b</sup>McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139; <sup>c</sup>Neuropsychology and Neurorehabilitation Service, The Laboratory for Investigative Neurophysiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland; <sup>d</sup>Radiodiagnostic Service, The Laboratory for Investigative Neurophysiology, University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland; and <sup>e</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139

Contributed by Robert Desimone, May 23, 2018 (sent for review September 4, 2016; reviewed by Aniruddh D. Patel, Larry R. Squire, Adam Tierney, and Robert J. Zatorre)

**Musical training confers advantages in speech-sound processing, which could play an important role in early childhood education. To understand the mechanisms of this effect, we used event-related potential and behavioral measures in a longitudinal design. Seventy-four Mandarin-speaking children aged 4–5 y old were pseudorandomly assigned to piano training, reading training, or a no-contact control group. Six months of piano training improved behavioral auditory word discrimination in general as well as word discrimination based on vowels compared with the controls. The reading group yielded similar trends. However, the piano group demonstrated unique advantages over the reading and control groups in consonant-based word discrimination and in enhanced positive mismatch responses (pMMRs) to lexical tone and musical pitch changes. The improved word discrimination based on consonants correlated with the enhancements in musical pitch pMMRs among the children in the piano group. In contrast, all three groups improved equally on general cognitive measures, including tests of IQ, working memory, and attention. The results suggest strengthened common sound processing across domains as an important mechanism underlying the benefits of musical training on language processing. In addition, although we failed to find far-transfer effects of musical training to general cognition, the near-transfer effects to speech perception establish the potential for musical training to help children improve their language skills. Piano training was not inferior to reading training on direct tests of language function, and it even seemed superior to reading training in enhancing consonant discrimination.**

music | piano | reading | education

**M**usic has been a part of humankind since prehistoric times (1). It appeals to our emotional needs and is enjoyable to people around the world. Recent research has also gradually discovered its great potential in shaping our cognition. Specifically, musical training has been linked to improved linguistic abilities (for a review, see ref. 2). Musicians have demonstrated better language skills compared with untrained individuals (3–6). Longitudinal studies suggest that this musician advantage in speech processing could be induced by musical training. Children and adolescents in controlled longitudinal studies showed enhanced language abilities after musical training as reflected by both behavioral (7) and electrophysiological (8–11) measures. Behavioral results even suggest that a music program produced benefits in phonological awareness for preschoolers similar to those of a matched phonological skill program (7). Twenty weeks of daily 10-min training with both programs significantly increased phonological awareness in preschoolers compared with sports training (7). However, the mechanisms underlying the

observed language benefits of musical training remain unknown. Here we addressed this question in a longitudinal training study.

Music and language share many aspects of sensory, motor, and cognitive processing of sound (12). Consequently, the shared acoustic features of music and speech sound are the likely basis of the cross-domain transfer effects of musical training (13–15). For instance, it is suggested that musical training strengthens the processing of sound attributes present in both music and speech (such as pitch and timbre), thus facilitating a better encoding of speech sound (13). Moreover, any positive effects of musical training on general cognitive abilities such as intelligence quotient (IQ) (16, 17) could lead to better language skills. Recently, it has been further proposed that sound coding and cognition might mutually reinforce each other during musical training (18). To our knowledge, direct evidence favoring either suggestion is

## Significance

**Musical training is beneficial to speech processing, but this transfer's underlying brain mechanisms are unclear. Using pseudorandomized group assignments with 74 4- to 5-year-old Mandarin-speaking children, we showed that, relative to an active control group which underwent reading training and a no-contact control group, piano training uniquely enhanced cortical responses to pitch changes in music and speech (as lexical tones). These neural enhancements further generalized to early literacy skills: Compared with the controls, the piano-training group also improved behaviorally in auditory word discrimination, which was correlated with their enhanced neural sensitivities to musical pitch changes. Piano training thus improves children's common sound processing, facilitating certain aspects of language development as much as, if not more than, reading instruction.**

Author contributions: Y.N., L.L., E.G., H.S., C.C.G., Q.D., J.D.E.G., and R.D. designed research; Y.N., L.L., and H.S. performed research; Y.N., L.L., and C.C.G. contributed new reagents/analytic tools; Y.N. and E.G. analyzed data; and Y.N., E.G., and R.D. wrote the paper.

Reviewers: A.D.P., Tufts University; L.R.S., Veterans Affairs San Diego Healthcare and University of California, San Diego; A.T., Birbeck College London; and R.J.Z., McGill University.

The authors declare no conflict of interest.

Published under the PNAS license.

Data deposition: The teaching material is available for download at the Zenodo repository ([https://zenodo.org/record/1287836#Wx\\_Uf0gvwuU](https://zenodo.org/record/1287836#Wx_Uf0gvwuU), doi:10.5281/zenodo.1287836).

<sup>1</sup>To whom correspondence may be addressed. Email: nany@bnu.edu.cn or desimone@mit.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808412115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1808412115/-DCSupplemental).

Published online June 25, 2018.

lacking (for a review, see ref. 15). Here, we aimed to examine the possible mechanisms underlying the linguistic benefits of musical learning from both perspectives, i.e., common sound processing and general cognition.

In the current study, we pseudorandomly assigned 74 kindergarten children aged 4–5 y to piano or reading training for 6 mo or to a no-contact control group. The three groups were matched regarding general cognitive measures (including IQ, working memory, and attention) and socioeconomic status at the beginning of the study. We measured the effects of musical training on the processing of pitch as indexed by the positive mismatch responses (pMMRs) as well as on general cognitive measures and language abilities. Pitch is a common sound element of both speech and music. In Mandarin, pitch is also used as the lexical tone, which is a salient speech component and is acquired early in life (19). There are four main lexical tones with different pitch inflections: level, mid-rising, dipping, and high-falling. For instance, the syllable *ma* can mean either “mother” or “horse,” depending upon whether the attached lexical tone is level or dipping. Cognitive abilities were measured with tests of IQ, attention, and memory. General language abilities were measured by a behavioral word-discrimination test that included variations of consonants, vowels, and tones.

Both piano and reading training were administered in the kindergarten in 45-min sessions three times a week. The control group simply followed the daily routines in the kindergarten. It served to control for the effects of development/maturation and retesting. The reading training used in the current study was based on the shared reading model (20), which is widely used in China. It offers an interactive reading experience in which the teacher models reading for a group of children by reading words aloud from enlarged texts, and the children read along. Compared with a phonological training program used in a previous study (7), reading represents a common, everyday activity that might improve literacy skills. Furthermore, schools must often make practical decisions about expending resources on the arts, such as music, versus supplemental instruction in the academic subjects that are often of most concern. Reading training also allowed us to test whether reading and piano training shared a neural basis for any observed effects on language. Although reading training was expected to improve language abilities (7), we expected that its impacts on sound processing, if any, would be more limited to the speech domain. By contrast, musical training might affect sound processing, especially pitch processing, across both speech and music domains (6).

## Results

### The Language-Processing Benefits of Piano Training.

**Behavioral pitch and word discrimination.** We first asked if there was a significant difference among the groups in the discrimination of pitch from pretest to posttest. In the pitch-discrimination task, children were asked to make a same/different judgment after they heard a pair of notes which were either the same or different (with interval sizes ranging from zero to three semitones). An ANOVA with session (pre and post) and group (piano, reading, and control) as factors on the pitch discrimination  $d'$  obtained a significant main effect of session,  $F(1, 71) = 35.04$ ,  $P < 0.001$ ,  $\eta^2 = 0.30$ , but no group  $\times$  session interaction or main effect of group (both  $P$ s  $> 0.1$ ) (Table 1). For the results and a discussion of detailed analyses on behavioral pitch-discrimination performance for the piano group, please see [SI Appendix](#). Thus, all three groups appeared to gain similarly in pitch discrimination over the time course of the study.

However, we did find group differences in the ability to discriminate words after training. The piano-training group was expected to show such advantages based on their experience in discriminating sounds in their piano training, and the reading group also might show improvements based on the greater experience in reading stories.

In a word-discrimination task, children were asked to make a same/different judgment after they heard a pair of monosyllabic words which were either the same or different (see [SI Appendix](#), [Table S1](#) for the full list of word pairs). The words could differ by tone, by consonant, or by vowel. The two-way ANOVA on word-discrimination performance,  $d'$ , yielded a significant main effect of session,  $F(1, 71) = 70.50$ ,  $P < 0.001$ ,  $\eta^2 = 0.48$ , and a significant group  $\times$  session interaction,  $F(2, 71) = 3.26$ ,  $P = 0.044$ ,  $\eta^2 = 0.04$ . No main effect of group was observed ( $P > 0.1$ ). All three groups yielded significantly better performance at posttest than at pretest. Step-down one-way ANOVA with group as the factor revealed a main group effect for delta  $d'$  ( $d'$  at posttest minus  $d'$  at pretest) in word discrimination,  $F(2, 71) = 3.26$ ,  $P = 0.044$ ,  $\eta^2 = 0.08$ , agreeing with the observed group  $\times$  session interaction. The piano group showed significantly higher performance enhancement at posttest than the control group ( $P = 0.019$ ). The reading group also trended toward higher performance enhancement than the control group ( $P = 0.102$ ), whereas no significant difference was observed between the reading and piano groups ( $P > 0.1$ ) (Fig. 1 and Table 1).

We further refined the analysis of the performance difference in word discrimination by analyzing behavioral performance separately for differences in tone, consonant, and vowel. These analyses revealed significant differences among the three groups in increased word discrimination at posttest based on distinctions

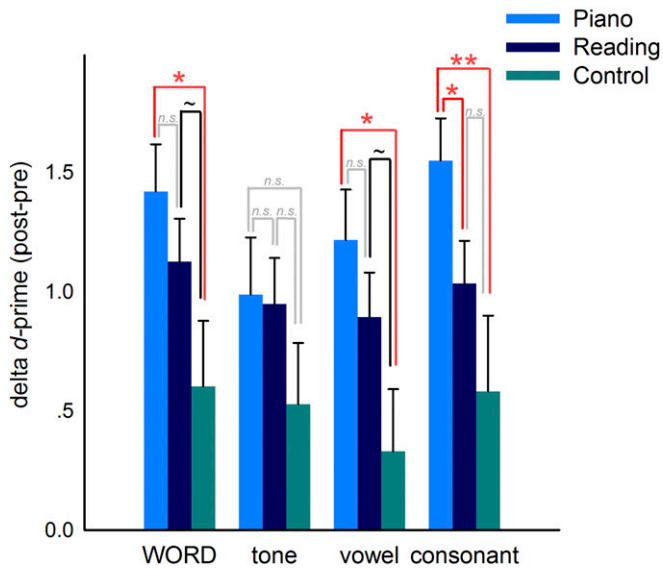
**Table 1. Behavioral discrimination performance**

Discrimination performance	Piano, $n = 30$	Reading, $n = 28$	Control, $n = 16$	Group difference, $P$
<b>Word <math>d'</math></b>				
Pre (SD)	1.4 (1.0)	1.3 (0.9)	1.8 (0.9)	0.374
Post (SD)	2.9 (0.8)	2.5 (1.2)	2.4 (1.1)	0.211
Delta (SD)	1.4 (1.1)	1.1 (0.9)	0.6 (1.1)	0.044*
<b>Pitch <math>d'</math></b>				
Pre (SD)	0.9 (0.9)	0.7 (1.1)	1.0 (0.7)	0.521
Post (SD)	1.9 (1.2)	1.7 (1.1)	1.5 (1.3)	0.523
Delta (SD)	1.1 (1.4)	1.0 (1.2)	0.5 (1.1)	0.372

Behavioral discrimination performance includes  $d'$  results at pre/posttest and delta  $d'$  (post minus pre) for the piano, reading, and control groups in behavioral word and pitch discrimination tests. A group difference in behavioral discrimination was observed for delta word  $d'$ . No such effect was observed for delta pitch  $d'$ .

\* $P < 0.05$ . All  $P$  values were obtained using one-way ANOVA.

## Group differences in word discrimination improvements



**Fig. 1.** Group differences in the improved behavioral performance in word discrimination at posttest compared with pretest. As suggested by the delta  $d'$  ( $d'$  at posttest minus  $d'$  at pretest), the piano group showed significantly more improvement in word discrimination and word discrimination based on vowels than the controls (both  $P$ s < 0.05). The reading group had similar trends ( $P = 0.102$  for word discrimination and  $P = 0.082$  for vowel distinctions) and was statistically indistinguishable from the piano group (both  $P$ s > 0.1). For word discrimination based on consonants, the piano group significantly outperformed the reading group ( $P < 0.05$ ) and the control group ( $P < 0.01$ ); the reading group was not statistically different from the control group ( $P > 0.1$ ). No group difference was observed for word discrimination improvements based on tones. \* $P < 0.05$ ; \*\* $P < 0.01$ ; ~ indicates  $P = 0.1$ ; n.s., no significant difference (approximately  $P > 0.1$ ). Error bars indicate SE.

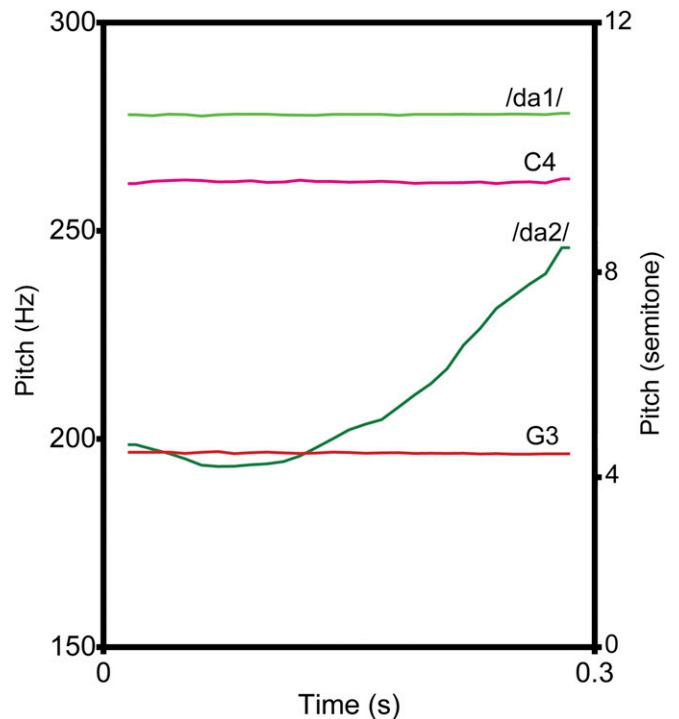
in consonants,  $F(2, 71) = 4.85$ ,  $P = 0.011$ ,  $\eta^2 = 0.12$ , and vowels,  $F(2, 71) = 3.57$ ,  $P = 0.033$ ,  $\eta^2 = 0.09$ , but not in tones,  $F(2, 71) = 1.0$ ,  $P = 0.4$ . For word discrimination based on consonants, the piano group outperformed both the control group ( $P = 0.006$ ) and the reading group ( $P = 0.046$ ) with significantly higher performance enhancement at posttest, whereas the reading group was not significantly different from the passive controls ( $P > 0.1$ ) (Fig. 1). On the other hand, for word discrimination based on vowels, the piano group demonstrated significantly greater enhancement than the controls ( $P = 0.014$ ); the reading group also performed marginally higher than the controls ( $P = 0.082$ ). However, the reading group and the piano group yielded statistically indistinguishable performance ( $P > 0.1$ ) (Fig. 1). In sum, compared with controls, the piano group showed improved auditory word discrimination based on consonants and vowels after training. The reading training may have facilitated word discrimination based on vowels to a similar extent as piano training, but only piano training showed a unique effect in boosting word discrimination based on consonants.

Following the piano group's greater improvements in word discrimination based on consonants compared with the reading group at posttest, we investigated some possible neural bases for the training effects using event-related potentials (ERPs). Stimuli used in the ERP experiment included a pair of monosyllabic words that differed only in lexical tone and a pair of notes that differed in pitch. Fig. 2 depicts a pair of words, /da1/ (which means "to build") and /da2/ (which means "to answer") that differ by tone. As shown in Fig. 2, the word pair and the note pair were matched in term of pitch interval (approximately five semitones). The grand average

waveforms elicited by standards and deviants for lexical tone and musical pitch conditions at electrode Fz for the three groups are presented in *SI Appendix, Fig. S1*. Fig. 3 illustrates difference waveforms at Fz (Fig. 3A), where both the lexical tone pMMRs and the musical pitch pMMRs were maximal, and the topographic maps for the delta pMMRs (Fig. 3B) among the three groups. Furthermore, we explored the possible relationships between the auditory improvements and the neural enhancements.

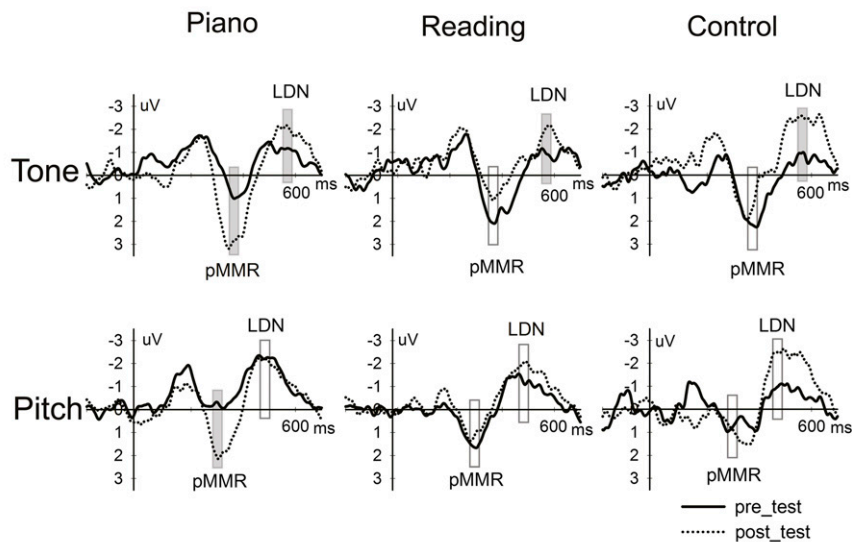
**Lexical tone pMMRs.** A four-way ANOVA of the lexical tone pMMR amplitudes with session (pre and post), area (frontal, fronto-central, and central), hemisphere (left, midline, and right), and group (piano, reading, and control) as factors showed a significant group  $\times$  session interaction,  $F(2, 71) = 6.56$ ,  $P = 0.002$ ,  $\eta^2 = 0.14$ . No session or group effect was observed (both  $P$ s > 0.1). There were main effects for area,  $F(2, 142) = 45.23$ ,  $P < 0.001$ ,  $\eta^2 = 0.19$ , and hemisphere,  $F(2, 142) = 10.02$ ,  $P < 0.001$ ,  $\eta^2 = 0.05$ .

The piano group showed significantly larger pMMR amplitudes at posttest compared with those at pretest,  $F(1, 29) = 6.93$ ,  $P = 0.013$ ,  $\eta^2 = 0.18$  (*SI Appendix, Fig. S24*). No such effects were observed for the reading or the control groups ( $P$ s > 0.1). Thus, piano training was associated with an enhancement of pMMRs. The amplitudes of the lexical tone pMMRs were largest over frontal electrodes (1.55  $\mu$ V) than at the other electrodes and were larger over fronto-central electrodes (1.32  $\mu$ V) than at central electrodes (0.60  $\mu$ V), all  $P$ s < 0.05 (*SI Appendix, Fig. S34*). In terms of hemispheric distribution, the lexical tone pMMRs were significantly larger at midline sites (1.45  $\mu$ V) than at bilateral sites (left = 0.91  $\mu$ V and right = 1.11  $\mu$ V), both  $P$ s < 0.01, whereas no significant difference was observed between bilateral sites ( $P > 0.1$ ) (*SI Appendix, Fig. S3B*).

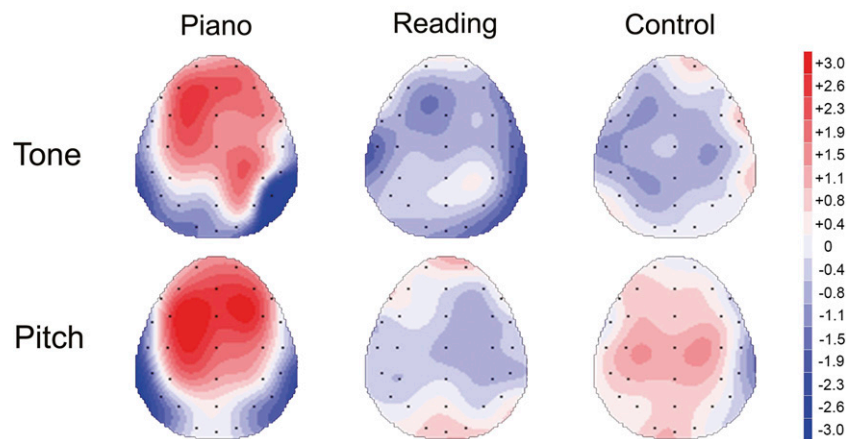


**Fig. 2.** Pitch contours of the stimuli in the EEG tests. The paired monosyllabic words differ only in lexical tone (/da1/ means "to build"; /da2/ means "to answer"), whereas the paired notes differ in pitch. Pitch intervals between the two stimuli in each domain are in similar ranges (approximately five semitones): /da1/ (lime,  $F_0 = 277.93$  Hz), /da2/ (green, mean  $F_0 = 208.93$  Hz), C4 (magenta,  $F_0 = 261.78$  Hz), and G3 (red,  $F_0 = 196.63$  Hz).

## A Difference waves at Fz



## B Topographic maps for the delta pMMRs



**Fig. 3.** Difference waveforms at Fz and the topographic maps for the delta pMMRs among the three groups. (A) Difference waveforms for lexical tone (Upper) and musical pitch (Lower) conditions between the standard and deviant responses (deviant minus standard) for the pretest (solid line) and posttest (dotted line) sessions at Fz for the three groups. The piano group showed significantly enhanced pMMRs at posttest relative to pretest for both lexical tone and musical pitch conditions, whereas the reading and control groups did not have such effects. All three groups demonstrated significantly enlarged LDNs for lexical tone changes (for detailed results of the LDN components, as well as the related discussion, please see *SI Appendix*). No such effects were observed in the musical pitch condition. The 40-ms analysis windows for the peaks of pMMR and LDN are marked by open and gray rectangles. The gray rectangles indicate significant session effects. (B) The topographic maps demonstrate the scalp distributions of the delta pMMR amplitudes (post minus pre) across the lexical tone (Upper) and musical pitch (Lower) conditions in the three groups. The time window is 40 ms. As indicated by the key in the figure, hotter (darker red) areas indicate greater positivity, and cooler (darker blue) areas indicate greater negativity. For both the lexical tone and musical pitch conditions, the topographical maps of the piano group are much hotter than those of the reading and control groups, converging with the unique training effects in the piano group as indexed by the significantly enhanced pMMRs at posttest relative to pretest.

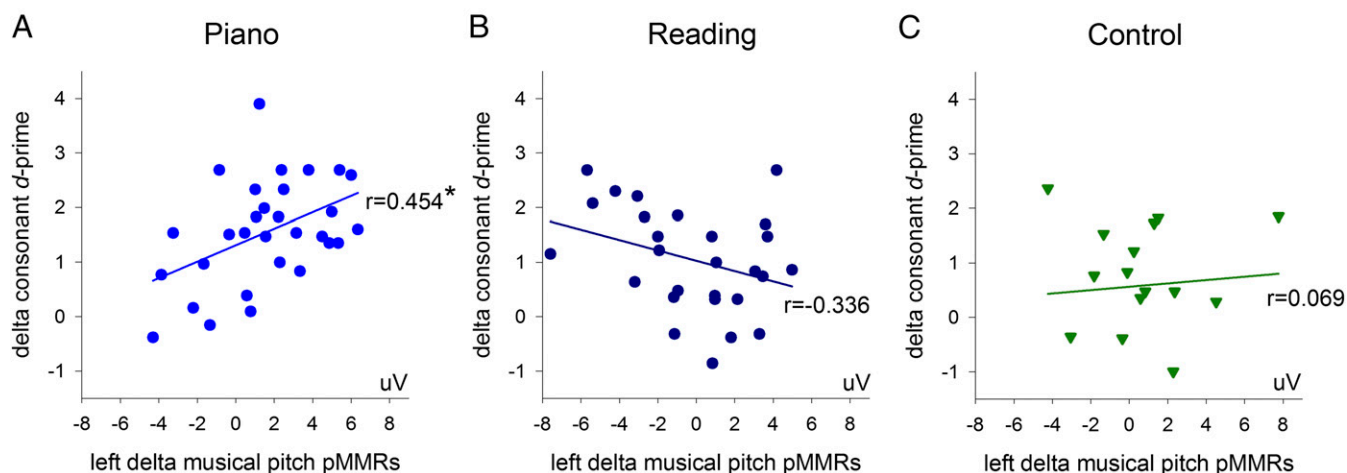
**Musical pitch pMMRs.** As with the lexical pMMR, the four-way ANOVA computed on musical pitch pMMR yielded a significant group  $\times$  session interaction,  $F(2, 71) = 4.06, P = 0.021, \eta^2 = 0.10$ . No other session, group, or area effects were observed (all  $P$ s  $> 0.1$ ). There was a main effect of hemisphere,  $F(2, 142) = 3.97, P = 0.021, \eta^2 = 0.03$ . The piano group showed significantly larger pMMRs at posttest compared with those values at pretest,  $F(1, 29) = 13.97, P < 0.001, \eta^2 = 0.32$  (*SI Appendix, Fig. S2B*). No such effects were observed for the reading group or the controls ( $P$ s  $> 0.1$ ). Across all three groups, the musical pitch pMMRs were significantly larger at midline electrodes (1.01  $\mu$ V) than at left-side electrodes (0.70  $\mu$ V),  $P = 0.02$  (*SI Appendix, Fig. S3B*). However, they were not significantly different between midline

and right-side electrodes (0.86  $\mu$ V) or between the bilateral electrodes (both  $P$ s  $> 0.1$ ).

Thus, the piano-training group showed neural evidence for an improvement in auditory processing with training, even though piano training showed an advantage only in behavioral word discrimination based on consonants but with no clear difference between the piano and reading groups in the other related behavioral improvements (i.e., word discrimination and specifically word discrimination based on vowels).

**Neural and behavioral correlations.** Given that the piano group demonstrated specific training effects in behavioral word discrimination based on consonants and in musical pitch and lexical tone pMMRs, we then examined the neural-behavioral relationships among these measures. Within the piano group, the behavioral

## Correlation between neural and behavioral effects specific to piano training



**Fig. 4.** There was a significant positive correlation between the increased musical pitch pMMR amplitudes over the left-side electrodes and the improved word discrimination performance based on consonants at posttest compared with pretest within the piano-training group (A) but not within the reading (B) or control (C) groups.  $*P < 0.05$ .

improvements in word discrimination based on consonants were significantly correlated with the increments in the musical pitch pMMRs over the left-side electrodes at posttest relative to pretest,  $r(30) = 0.45$ ,  $P = 0.012$  (Fig. 4A). This correlation did not hold over either the midline or the right-side electrodes (both  $P$ s  $> 0.1$ ) within either the reading group (Fig. 4B) or the control group (Fig. 4C) (both  $P$ s  $> 0.1$ ). No correlation was found between word discrimination based on consonants and the increments of the delta lexical tone pMMRs.

**General Cognitive Measures.** All three groups improved equally in IQ at posttest compared with pretest. For the five subtests of the IQ test, the raw scores were all significantly improved at posttest compared with pretest [main effects of session: vocabulary,  $F(1, 71) = 7.68$ ,  $P = 0.007$ ,  $\eta^2 = 0.10$ ; similarities,  $F(1, 71) = 22.75$ ,  $P < 0.001$ ,  $\eta^2 = 0.24$ ; animal house,  $F(1, 71) = 68.12$ ,  $P < 0.001$ ,  $\eta^2 = 0.49$ ; picture completion,  $F(1, 71) = 31.02$ ,  $P < 0.001$ ,  $\eta^2 = 0.30$ ; block design,  $F(1, 71) = 104.50$ ,  $P < 0.001$ ,  $\eta^2 = 0.59$ ]. By contrast, for the scaled scores, only the subtest of block design demonstrated significant improvements at posttest compared with pretest [main effect of session:  $F(1, 71) = 16.36$ ,  $P < 0.001$ ,  $\eta^2 = 0.19$ ]. For the other four subtests, the scaled scores did not show any session effects (all  $P$ s  $> 0.05$ ; please see *SI Appendix, Table S2* for details on IQ results). These data are in line with previous research suggesting that, in comparison with the scaled scores, which are derived from conversion processes involving age-normalization and subtest-standardization, the raw scores are more sensitive to meaningful changes in intellectual functioning with age in longitudinal studies (e.g., ref. 21). Performances for both working memory and attention were also significantly better at posttest than at pretest [main effects of session: working memory,  $F(1, 71) = 22.75$ ,  $P < 0.001$ ,  $\eta^2 = 0.24$ ; attention,  $F(1, 71) = 30.46$ ,  $P < 0.001$ ,  $\eta^2 = 0.30$ ]. However, there were no group differences in these improvements (main effects of group, all  $P$ s  $> 0.1$ ; group  $\times$  session interactions, all  $P$ s  $> 0.1$ ). Thus, the effects of piano training appeared to be specifically in word discrimination and neural measures of auditory processing rather than broad improvements in cognitive function.

### Discussion

One month after the intervention, compared with the reading and the no-contact control groups, the piano group showed sig-

nificantly enlarged mismatch responses to both musical pitch and lexical tone deviations as reflected by the pMMRs. In addition, relative to the control group, the piano group demonstrated significantly enhanced word-discrimination performance in general as well as word discrimination based on vowel differences. The reading group yielded similar trends and was statistically indistinguishable from the piano group. Furthermore, the piano group demonstrated a unique advantage in consonant-based word discrimination relative to the reading and the control groups. This improved consonant-based word discrimination following musical training was correlated with increased musical pitch pMMRs among children within the piano group.

The results provide clear experimental evidence that musical training enhanced neural processing of pitch, a shared sound element in music and speech domains, and caused concurrent benefits in auditory Mandarin word discrimination, measured behaviorally. By contrast, all three groups improved equally on general cognitive measures including IQ, attention, and working memory. The mechanisms underlying musical training's language benefits are thus attributable to musical training's effects on common sound processing across domains rather than on enhanced general cognitive abilities.

### Common Sound Processing: The Mechanisms Underlying the Effects of Musical Training on Speech Perception.

Prior pioneering work showed musical training's benefits in speech and speculated that the shared sound attributes of music and speech (13–15) might be one of the causes for the transfer effect. Here, using data from pMMRs, we provide direct evidence supporting the shared sound-processing hypothesis. Compared with reading and control groups, 6 mo of piano training uniquely enhanced children's neural sensitivity to both musical pitch and lexical tones, based on increased pMMRs to both musical pitch and lexical tone changes. Although sharing a similar time window, amplitude range, and morphology, the pitch change-induced pMMRs between music and speech domains were different in terms of topographic distribution. They were largest over the midline electrodes across domains, but only the speech pMMRs showed an additional frontal dominance (*SI Appendix, Fig. S3*). These results suggest a combination of overlapping and distinct neural systems for pitch processing across domains, consistent with our earlier functional brain imaging results with Mandarin-speaking musicians (22). In that study, although

pitch processing activated the right superior temporal and the left inferior frontal areas for both music and speech among Mandarin musicians, the temporal activation was more sensitive to pitch relationships in music than to those in the speech domain (22). In the same line, Norman-Haignere et al. (23) recently found adjacent regions selective for the processing of pitch, music, and English word sounds in auditory cortex, with some overlap in selectivity across regions. Taking these results together with the present study, we speculate that early musical training affects the development not only of regions specialized for music but also of regions specialized for pitch in general, which are common to both music and tonal languages.

The enlarged lexical tone pMMRs reflect heightened neural sensitivity to lexical tone changes due to piano training, converging with our recent cross-sectional results in adult musicians (6). This result also confirms earlier studies showing improvements in speech processing after musical training (8–11). It is notable that this clear neural effect in speech-tone processing with such a short period of musical training (6 mo) is not common. Recent work shows that, in a nontonal language background such as French, 6 or even 12 mo of musical training did not result in any observable neural or behavioral effect in speech-tone processing among 8-y-olds (10). The authors attributed the lack of training effects on speech tones to the pitch intervals they employed. They used pitch intervals with two or seven semitones, which may have been either too small or too large compared with the five-semitone pitch interval used in the current study. In addition to the acoustic differences, the speech tone changes in the current study were phonologically significant to the participants, whereas the vowel frequency variations employed in Chobert et al. (10) were not. This difference in the higher-order functional significance or the ecological validity of the speech stimuli might have also contributed to the positive result in the present study compared with the negative result in the prior study.

In behavioral word discrimination, the piano-training group showed significant improvements in comparison with the controls. Although the reading group demonstrated only a trend toward improved performance relative to the controls, the effect was in the same direction as the piano group. This strengthens the notion suggested by previous research (7) that musical training can be as effective as phonological programs or reading training in enhancing speech-sound perception. In addition, relative to the reading group, the piano group demonstrated a specific advantage in word discrimination based on consonants. However, no musical training effects on word discrimination based on lexical tones was observed. Behavioral discrimination of musical pitch did not respond to musical training either, indicating a general lack of musical training effects on behavioral pitch sensitivity across both speech and music domains.

The difference between the neural pMMR and behavioral pitch effects of music training might be due to differences in the stimulus characteristics involved in the different measures. To increase the ecological validity of the test, the behavioral discrimination tasks involved stimuli which were more fine-grained (musical pitches with intervals of zero to three semitones) and variable (lexical tones with all the possible contrasts) than those used in the ERP study (five-semitone intervals for both lexical tone and musical pitch conditions; a contrast between tone 1 and tone 2 for the lexical tone condition). Moreover, the discrepancy might also suggest that the brain measures are more sensitive than the behavioral measures to the influence of musical training. The pMMRs represented neural discriminative processes recorded with a passive task, whereas the behavioral task was active. For children, pMMRs seem to be more sensitive than an active discrimination task (e.g., ref. 24).

For behavioral word discrimination, the specific musical training effect was observed for consonants rather than for vowels or lexical tones. This result may be related to the different devel-

opmental trajectories of these speech phonemes. Between 2 and 6 y of age, children undergo a critical period for the development of specific speech components such as lexical tones, vowels, and consonants (19, 25). This is consistent with the cross-group improvement in speech processing found in the present study, where both the behavioral word discrimination and the lexical tone late discriminative negativities (LDNs) (Fig. 3 and *SI Appendix*) were significantly enhanced at posttest compared with pretest across all three groups. Furthermore, earlier evidence (26) suggested that Mandarin-speaking children aged 1–4 y acquired different speech elements in the following order: first tones, then the final consonant or vowel in a syllable, and finally the initial consonant in a syllable. The related consonant variations in the current study were mainly initial consonants (*SI Appendix, Table S1*). Moreover, at pretest children from all three groups were already 4 y old (mean age  $\pm$  SD:  $4.6 \pm 0.3$ ). Their relative developmental order of the speech elements at pretest showed the pattern as suggested by Hua and Dodd (26): lexical tones ( $d'$  mean  $\pm$  SD:  $1.6 \pm 1.1$ ) were significantly ahead of vowels ( $d'$  mean  $\pm$  SD:  $1.3 \pm 1.1$ ),  $P < 0.01$ , and both lexical tones and vowels were significantly ahead of consonants ( $d'$  mean  $\pm$  SD:  $1.0 \pm 0.8$ ), both  $P$ s  $< 0.01$ . In light of this order of phonetic acquisition in Mandarin, the specific piano-training effect on word discrimination based on consonants in the current study might indicate musical training's advantage over reading training in facilitating the behavioral discrimination of the more developmentally demanding speech component, i.e., the initial consonants, in this age range (4–5 y).

Finally, the effect of piano training on consonant discrimination implies that musical training might strengthen the processing of basic temporal and spectral aspects of sound (27) and provide benefits in speech perception, not only in neural lexical tone processing but also in behavioral word discrimination based on nontonal cues. This is consistent with the brain-behavioral correlation we observed, where the behavioral effects of word discrimination based on consonants were predicted by the enhancements of music pMMRs but not by the enhancements of lexical tone pMMRs. The correlation results indicate that, instead of common neural pitch processing, common basic temporal and spectral sound processing more likely underlies the observed brain-behavioral link. However, these propositions raise further questions about the associations between musical training and its effects on specific spectral and temporal sound processing at both the behavioral and neural levels, especially for speech sound (e.g., ref. 28). Future research, preferably with a longitudinal approach, is warranted to better understand these relationships.

Taken together, the observed effects of musical training on speech perception are best thought of as the outcome of complex interactions between stimuli characteristics, task-related factors, musical training, and the respective developmental stage. The present study shows that 6 mo of piano training led to neural and behavioral benefits in speech perception with concurrently enhanced neural processing of musical sound for children age 4–5 y. The advanced musical pitch pMMRs corroborate previous findings in the neural effects of musical sound processing due to musical training (29). This neural link between musical training and the related language benefits was expected but has been missing in previous research. For instance, Moreno et al. (8) examined pitch processing in both speech and music domains for a group of 8-y-olds after 6 mo of musical training and observed enhanced reading and pitch discrimination in speech, but they did not find similar training effects on the neural processing of musical pitch. According to the dual-stream model of language processing in the brain (30), the common sound processing, i.e., the basic spectrotemporal analysis of sound, is carried out in bilateral auditory cortices. When there is an active discrimination task, higher areas such Broca's area and the right prefrontal cortex are also involved in making the judgment, and the resulting neural networks are domain specific (e.g., ref. 31). The lack of effects on

common neural pitch processing in Moreno et al. (8) could thus be attributed to their active congruency judgment task, which was further complicated by two difficulty levels, e.g., strong and weak pitch changes. In contrast, using the passive mismatch paradigm which minimized the influence of task-related factors, the present study demonstrates the neural link, suggesting advanced common sound processing across domains as an important mechanism underlying the speech benefits of musical training.

**The Uniqueness of Musical Training.** The current study extends previous notions that musical training can enhance speech perception to the same extent as phonological programs (7) and shows unique musical training effects on both behavioral and electrophysiological speech-sound processing. Relative to reading training, piano training specifically enhanced behavioral sensitivities to consonants and neural sensitivities to lexical tones. Three factors might underlie these unique effects of musical training. First, the sound processing required for piano training is more fine-grained than that required for reading training. This idea has been proposed by Patel (15) in the extended OPERA hypothesis as music demands greater precision than speech for certain aspects of sound processing. Taking pitch processing, for example, Patel (15) proposed that in music very small pitch differences are perceptually crucial (e.g., an in-key and an out-of-key note can be only one semitone apart), whereas in speech the pitch contrasts are normally more coarse-grained. Second, the focus of piano training is on the precise representation of the sound through multiple inputs, whereas reading training emphasizes mapping the speech sound to its semantic meaning (32). Third, the reward value of musical training (33, 34), as suggested in the OPERA hypothesis (15), might also contribute to the unique neural plastic effects observed in piano training relative to reading training. Piano training may be more pleasurable to young children than reading and thus might engage attention and learning systems to a greater degree.

Piano training in the current study is multimodal in nature. It involves fine-grained multisensory inputs of sound (35, 36), requiring the interaction and cooperation of multiple sensory and motor systems. Similarly, reading training as employed in the current study also involves multiple sensory and motor inputs of sounds, especially because reading the words out loud was explicitly required during training. Nonetheless, piano training outperforms reading training in terms of neural speech processing as well as behavioral word discrimination based on consonants because of the more fine-grained and accurate sound representations.

**No Evidence for Improved General Cognition Following Piano Training in the 4- and 5-y-Olds.** In contrast to the clear evidence for the common sound-processing mechanisms, the current results suggest no training effect on general cognitive measures. After the intervention, all three groups improved similarly in IQ, attention, and working memory, mostly likely due to repetition and maturation effects. According to Barnett and Ceci (37), the observed language effects in the current study could be attributable to near transfer that usually occurs between highly similar domains such as music and language, in contrast to far transfer, which occurs in dissimilar contexts or domains. In the current study, far transfer would apply to the effects of musical training on general cognitive measures. As in the current study, far transfer is not common in training studies (e.g., ref. 37).

Specifically, our failure to find a far transfer in general cognitive functions is consistent with other longitudinal studies of children trained with music (8, 10, 38). However, there is contradictory evidence suggesting that musical training in the forms of instrumental, voice, or even simply listening-based computerized programs could significantly enhance full-scale IQ or verbal IQ (as indicated by the vocabulary subtest) and executive functions in children (16, 17). We did observe general developmental/repetition effects in these cognitive measures across

all three groups. Hence, a ceiling effect is not an issue here. Furthermore, it is unlikely that training duration or intensity explains these contradictory results, because the current study actually employed almost twice the amount of musical training (72 45-min lessons) than the previous studies (36 45-min lessons or 40 45-min lessons) (16, 17). Equally unlikely is the sample size. Although our sample size ( $n = 74$ ) is smaller than Schellenberg's (16) ( $n = 132$ ), it is similar to that of Moreno et al. (17) ( $n = 72$ ). Indeed, with the current sample size, the power was sufficient to detect any medium effect (a power of 0.96 for at least a medium effect size).

**Practical Implications.** Even if music (or reading) training does not generalize to improvements in general cognition, the positive effects of musical training on word discrimination and brain measures of tone processing suggest that music could play an important role in early education. Music and speech development are nonlinear over time and show similar sensitive periods (39, 40). Thus, musical education at an early age is likely to be most important. Indeed, it is widely acknowledged that musical training's effects on speech are maximized with early age of onset (e.g., ref. 3). Finally, as tone is one of the predominant units in speech development and impaired tone processing is a marked characteristic for developmental learning disabilities in both speech (dyslexia) and music (congenital amusia) (41, 42) among tone-language speakers, musical training might ameliorate these conditions by directly strengthening tone processing.

## Conclusion

Using a longitudinal design and pseudorandomized group assignments, our current study reveals a causal link between musical training and the enhanced speech-sound processing at both the neural and behavioral levels. More importantly, by targeting neural pitch processing in both music and speech domains, we are able to reveal one of the important mechanisms underlying musical training's language benefits—strengthened common sound processing.

## Materials and Methods

**Participants.** One hundred twenty nonmusician Mandarin-speaking children at the age of 4–5 y from a public school kindergarten in Beijing were recruited for the current study to form three groups of 40 each. Forty-six children were excluded from the final analysis due to dropouts (44: 10 from the piano group, 10 from the reading group, and 24 from the control group) or excessive artifacts in their electrophysiological data (two from the reading group). Note that more children in the control group dropped out than in the other two groups ( $\chi^2$  tests, both  $P$ s < 0.01), although the overall retention of children in the three groups (62%) was close to 65% as reported in an earlier study (10). Moreover, the dropouts from the control group were similar to the dropouts from the piano and reading groups in terms of the pretest discrimination performance and demographic characteristics (SI Appendix, Table S3). According to information drawn from a questionnaire completed by their parents before the study, none of the remaining 74 children had known neurological or auditory deficits. Among them, four were left-handed. All children were from similar socioeconomic backgrounds as indicated by their parents' education level (the average degree was at the college level for both parents among all three groups) and monthly home income. None of these children had any formal musical training before the study. This research protocol was approved by the Institutional Review Board of Beijing Normal University.

Written informed consents were obtained from the director of the kindergarten, the teachers, and the parents before the study. Both piano and reading training were administered in the kindergarten. At pre- and post-tests, all children received gifts to keep them motivated and engaged in the study.

**Design of the Longitudinal Study.** To control for group differences before the training, the children were pseudorandomly assigned to a piano-training, reading-training, or no-contact (wait-list training) group based on age, gender, and socioeconomic background. The three groups of children were also matched on general cognitive measures, including IQ [the Chinese version of the Wechsler Preschool and Primary Scale of Intelligence (43)], working memory (digit span), and attention (a fish version of the Eriksen

**Table 2. Demographic characteristics of the piano, reading, and control groups**

Demographic characteristics	Piano, <i>n</i> = 30	Reading, <i>n</i> = 28	Control, <i>n</i> = 16	Group difference, <i>P</i>
Mean age, mo (SD)	54.7 (3.3)	54.6 (3.3)	56.6 (4.6)	0.174
Male/female	17/13	19/9	9/7	0.625
IQ (SD)	122.8 (10.9)	120.5 (12.4)	118.8 (9.9)	0.478
Digit span (SD)	12.1 (3.5)	11.3 (3.8)	12.0 (2.7)	0.664
Flanker (SD)	77.9 (18.7)	78.4 (19.9)	82.2 (19.8)	0.756

Digit span scores are raw scores of the correct responses. The performance on the Flanker test is the percent correct responses. The *P* value for group difference in gender (male/female) was obtained using  $\chi^2$  tests; the *P* values for other variables were obtained using one-way ANOVA.

Flanker task) (44). The IQ test was based on five subtests, including vocabulary, similarities, animal house, picture completion, and block design. The piano group comprised 30 children (mean age = 4.6 y, SD = 0.28; one left-handed), the reading group comprised 28 children (mean age = 4.6 y, SD = 0.28; two left-handed), and the control group comprised 16 children (mean age = 4.7 y, SD = 0.38; one left-handed). The demographic characteristics of the groups are shown in Table 2.

Training lasted for 6 mo, from January through July, with a 1-mo break for the Spring Festival Holiday. Both piano and reading training were implemented with established programs with engaging textbooks and experienced teachers. For the no-contact control group, the children and their parents could select either the piano or reading training after the posttest (the end of this study). The pre- and posttests were administered immediately before and 1 mo (summer holiday) after the training. Two experienced teachers were hired for each training method. Sessions were 45 min long, three times per week, and delivered to groups of four to six children. Piano training was based on an established program (methods of C.C.G.) and included basic knowledge about notes, rhythm, and notation. The children were instructed to listen, discriminate, and recognize notes. They were required to play the piano with and without listening to CD recordings. No practice outside of class was required. Reading training was based on the shared-book reading program and emphasized reading and rereading oversized books with enlarged print and illustrations (45). There were no extracurricular music classes for the three groups of children. For reading activities at home, we controlled two related variables, time reading together with the children reported by the parents or other caregivers and the number of books each family had, among the three groups.

### Behavioral Tests.

**Word discrimination test.** A list of monosyllabic words was produced by a female voice actor whose native language was Mandarin Chinese. Recordings occurred in a soundproof booth using a Marantz PMD-620 digital recorder (D&M Professional) at a sampling rate of 44.1 kHz. Eighty-four monosyllabic words were used to construct 56 trials: half with one pair of the same words and the other half with one pair of different words, which differed by one of the elements (tone, consonant, or vowel) (see *SI Appendix, Table S1* for the full list of word pairs). The duration of the employed monosyllabic words ranged from 310 to 610 ms (mean = 470, SD = 72). Within a trial, there was a 500-ms gap between the two words.

**Pitch discrimination test.** All the notes in the pitch discrimination test were computer-synthesized with a piano-like timbre, sampled from the two adjacent octaves C3–B3 and C4–B4. There were 48 trials: one half of the trials contained one pair of identical notes, and the other half comprised two different notes with interval sizes ranging from one to three semitones. All the notes were 300 ms long. Within a trial, there was a 500-ms gap between the two notes.

**Procedure.** All the words and notes were equated for sound intensity with 10-ms linear on/off ramps using Praat (46). Both discrimination tests were completed in a soundproof booth in a single session. The stimuli were presented to the participants binaurally through Sennheiser HD 201 headphones adjusted individually to a comfortable level. Participants were asked to make a same/different judgment after the presentation of each pair of words or notes. The order of the word- and pitch-discrimination tests was counterbalanced across participants. The behavioral tests were delivered using E-prime software (Psychological Software Tools). Before the tests, detailed instructions and warm-up training were delivered to ensure that all participants understood the task.

### EEG Tests.

**Stimuli.** Mandarin words and musical notes were presented during the EEG experiment. Two monosyllabic words, /da1/ and /da2/ (1 = level tone and 2 = rising tone) (Fig. 2), were chosen from the stimuli of the behavioral word-discrimination test. Two notes, C4 and G3, were computer-synthesized with a piano-like timbre. The monosyllables and notes were digitally standardized using Praat (46) with a duration of 300 ms and a sound pressure level of 70 dB.

As shown in Fig. 2, the interval between /da1/ (mean F0 = 277.93 Hz) and /da2/ (mean F0 = 208.93 Hz) was approximately five semitones, which was close to the five-semitone difference between C4 (F0 = 261.78 Hz) and G3 (F0 = 196.63 Hz). This closeness was because the mismatch responses are sensitive to the deviant sizes (10, 47, 48). The two stimuli within each domain thus formed two perceptually comparable standard/deviant pairs for the measurements of the electrophysiological mismatch responses. Note that this five-semitone deviant size was chosen based on Chobert et al. (10), where the authors used either two semitones (for the small deviant size) or seven semitones (for the large deviant size) but found no training-related effects. These authors suggested that their failure to find training effects was due to the deviant sizes being either too small (at two semitones) or too large (at seven semitones) (10).

**Procedures and EEG recording.** Two oddball blocks of lexical tone and two oddball blocks of musical pitch were presented in a counterbalanced way to each participant during the EEG session, following the procedure used in our previous work (6). The two stimuli in speech or music varied in the respective two blocks as either standard or deviant. The stimulus sequence in each block contained a total of 480 stimuli with alternating standards and deviants (80 deviants; deviant probability *P* = 0.17). The deviants occurred pseudo-randomly among the standards, and two adjacent deviants were separated by at least two standards. The interstimulus interval varied randomly between 600 and 800 ms at a step of 1 ms. The stimuli in the EEG session were presented with the E-prime software (Psychological Software Tools).

The participants were seated comfortably in a quiet room and were instructed to ignore the presented sounds while watching a mute cartoon of their choice. The stimuli were presented to the participants binaurally through Sennheiser HD 201 headphones with the volume individually adjusted to a comfortable level. The participants were instructed to minimize head motion and eye blinking during the EEG recording. The EEG session lasted approximately 60 min including preparation and data acquisition.

The EEGs were recorded using a SynAmps EEG amplifier and the Scan 4.5 package (Neuroscan, Inc.) with a 1,000-Hz sampling rate through a bandpass filter of 0.05–200 Hz. Each participant wore a Quick-Cap with 32 tin scalp channels placed according to the international 10–20 system. The ground electrode was placed at the midpoint between FP1 and FP2. Vertical and horizontal electrooculograms were recorded above and below the left eye and next to the outer canthi of both eyes. The reference electrode was placed on the tip of the nose. The impedance of each channel was kept below 5 k $\Omega$ .

### Data Analysis.

**EEG processing.** For the off-line signal processing, the trials with saccades were discarded upon visual inspection. Eye blinks were corrected using the linear regression function provided by the Neurocan software (49). The data were digitally filtered (low-pass filter of 30 Hz, zero-phase, and 24 dB per octave) and were segmented into 900-ms epochs that included a 200-ms prestimulus baseline and 700 ms after the onset of the stimuli. After baseline correction, epochs containing voltage changes that exceeded  $\pm 80$   $\mu$ V at any electrode were excluded from the analysis. The remaining epochs were separately averaged for each stimulus. The average trial numbers for the three groups



across conditions and sessions were 60 (SE = 1.2) for the deviant trials. Additional analyses showed no main effects of group, condition, or session.

**Identification of the specific mismatch responses.** Previous developmental studies have shown that, for both lexical tone and musical pitch discrimination, depending upon the deviant sizes as well as specific paradigms, the mismatch responses include not only the early response, such as the mismatch negativity (MMN) (50), but also the pMMR (19, 51) and the LDN (48, 52) components in a later time window (300 ms post change onset). The MMN that normally occurs between 100 and 300 ms after the deviation stimulus onset has received the most research attention (50). This response is predominantly distributed over the fronto-central scalp and indexes automatic detection of changes based on sensory memory traces of the preceding stimulation (50). Unlike this early negative component, the pMMR and the LDN are more frequently observed in children. The pMMR is referred to as the “P3a” in some studies and is interpreted as an analog of the P3a response in adults, reflecting involuntary attention switching to a task-irrelevant stimulus (53, 54) or a less matured auditory discrimination process (19, 55). The LDN is another negative event-related potential with a peak between 400–600 ms after change onset, presumably indexing attention reorientation after a distracting sound (53) or an immature stage of auditory change processing (56). Together, these three components form a MMN–pMMR–LDN complex related to auditory change discrimination (e.g., ref. 53). Thus, we examined the main effects of session or session  $\times$  group interactions for all these three components in the current study.

A difference waveform for each individual participant was obtained by subtracting the standard waveform from the respective deviant waveform for musical pitch and lexical tone conditions separately. Nine fronto-central electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4) were selected for statistical analysis following the conventional choice for analyzing the mismatch responses based on the typical fronto-central distribution of these responses (50, 57). To examine the possible training effects on each of the three mismatch responses (including MMNs, pMMRs, and LDNs), we first ran a continuous time window analysis with the mean amplitudes in 12 successive 50-ms time windows from 100–700 ms after change onset for musical pitch and lexical tone conditions separately (19, 55, 58). Only the time windows with a significant main effect of session or a session  $\times$  group interaction were reported (for details of the time window analysis, please see *SI Appendix*). The effect observed for the separate mismatch component analysis

was considered meaningful only if it originated from at least two consecutive time windows (i.e., 100 ms) in the corresponding time window analysis (19). Based on the results of the time window analysis, two mismatch components, pMMR and LDN, were identified, and their peaks were defined as the mean values over the following 40-ms windows: 345–385 ms for lexical tone pMMR; 550–590 ms for lexical tone LDN; 285–325 ms for musical pitch pMMR; and 455–495 ms for musical pitch LDN. Here, we report only the pMMR results, which demonstrated training effects for both speech and music in the piano group relative to the other two groups. The LDN results showed developmental effects in the speech condition for all three groups, which are reported in *SI Appendix*. No session-related effects were observed with the MMN responses, which therefore are not reported here.

**Overall data analysis.** Behavioral performance in the word/pitch discrimination task was measured by  $d'$ , calculated from both the hit rates and the false alarms: hits were correct responses identifying pairs of different words/notes as different, and false alarms were incorrect responses identifying pairs of identical words/notes as different. For the refined analysis based on word distinctions in tone, vowel, or consonant, false alarms were incorrect responses identifying pairs of identical words preassigned to each category as different (*SI Appendix, Table S1*). The general cognitive measures and performances of the behavioral tests were analyzed using two-way repeated-measures ANOVAs with session (pre vs. post) as the within-subjects factor and group (piano, reading, and control) as the between-subjects factor. Four-way repeated-measures ANOVAs with area (frontal, fronto-central, and central areas), hemisphere (the left: F3, FC3, C3; the midline: Fz, FCz, Cz; and the right: F4, FC4, C4), and session (pre vs. post) as the within-subjects factors and group (piano, reading, and control) as the between-subjects factor were performed for lexical tone and musical pitch conditions separately to examine the effects of successive time windows and the two mismatch components. Greenhouse–Geisser corrections were performed when necessary. The Bonferroni correction was used for post hoc multiple comparisons.

**ACKNOWLEDGMENTS.** This work was supported by 973 Program Grant 2014CB846103; National Natural Science Foundation of China Grants 31471066, 31521063, 31571155, and 31221003; 111 Project Grant B07008; Beijing Municipal Science and Technology Commission Grant Z151100003915122; the InterDiscipline Research Funds of Beijing Normal University; and Fundamental Research Funds for the Central Universities Grant 2017XTCX04 and 2015KJJC28.

- Conard NJ, Malina M, Münzel SC (2009) New flutes document the earliest musical tradition in southwestern Germany. *Nature* 460:737–740.
- Tierney A, Kraus N (2013) Music training for the development of reading skills. *Prog Brain Res* 207:209–241.
- 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.
- Zuk J, et al. (2013) Enhanced syllable discrimination thresholds in musicians. *PLoS One* 8:e80546.
- Weiss MW, Bidelman GM (2015) Listening to the brainstem: Musicianship enhances intelligibility of subcortical representations for speech. *J Neurosci* 35:1687–1691.
- Tang W, Xiong W, Zhang YX, Dong Q, Nan Y (2016) Musical experience facilitates lexical tone processing among Mandarin speakers: Behavioral and neural evidence. *Neuropsychologia* 91:247–253.
- Degé F, Schwarzer G (2011) The effect of a music program on phonological awareness in preschoolers. *Front Psychol* 2:124.
- Moreno S, et al. (2009) Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cereb Cortex* 19:712–723.
- François C, Chobert J, Besson M, Schön D (2013) Music training for the development of speech segmentation. *Cereb Cortex* 23:2038–2043.
- Chobert J, François C, Velay JL, Besson M (2014) Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cereb Cortex* 24:956–967.
- Tierney AT, Krizman J, Kraus N (2015) Music training alters the course of adolescent auditory development. *Proc Natl Acad Sci USA* 112:10062–10067.
- Patel AD (2008) *Music, Language, and the Brain* (Oxford Univ Press, New York).
- Besson M, Chobert J, Marie C (2011) Transfer of training between music and speech: Common processing, attention, and memory. *Front Psychol* 2:94.
- 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.
- Patel AD (2014) Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hear Res* 308:98–108.
- Schellenberg EG (2004) Music lessons enhance IQ. *Psychol Sci* 15:511–514.
- Moreno S, et al. (2011) Short-term music training enhances verbal intelligence and executive function. *Psychol Sci* 22:1425–1433.
- Kraus N, White-Schwoch T (2016) Neurobiology of everyday communication: What have we learned from music? *Neuroscientist* 23:287–298.
- Lee CY, et al. (2012) Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia* 50:3228–3239.
- Holdaway D (1979) *The Foundations of Literacy* (Ashton Scholastic, Sydney).
- Searcy YM, et al. (2004) The relationship between age and IQ in adults with Williams syndrome. *Am J Ment Retard* 109:231–236.
- Nan Y, Friederici AD (2013) Differential roles of right temporal cortex and Broca's area in pitch processing: Evidence from music and Mandarin. *Hum Brain Mapp* 34:2045–2054.
- Norman-Haignere S, Kanwisher NG, McDermott JH (2015) Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88:1281–1296.
- Davids N, et al. (2011) The nature of auditory discrimination problems in children with specific language impairment: An MMN study. *Neuropsychologia* 49:19–28.
- Singh L, Goh HH, Wewalaarachchi TD (2015) Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition* 142:1–11.
- Hua Z, Dodd B (2000) The phonological acquisition of putonghua (modern standard Chinese). *J Child Lang* 27:3–42.
- Zatorre RJ, Gandour JT (2008) Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* 363:1087–1104.
- Zuk J, et al. (2017) Revisiting the “enigma” of musicians with dyslexia: Auditory sequencing and speech abilities. *J Exp Psychol Gen* 146:495–511.
- Fujioka T, Ross B, Kakigi R, Pantev C, Trainor LJ (2006) One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain* 129:2593–2608.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256:846–849.
- Storkel HL (2009) Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *J Child Lang* 36:291–321.
- Blood AJ, Zatorre RJ (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci USA* 98:11818–11823.
- Salimpoor VN, Benovoy M, Larcher K, Dagher A, Zatorre RJ (2011) Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci* 14:257–262.
- Wan CY, Schlaug G (2010) Music making as a tool for promoting brain plasticity across the life span. *Neuroscientist* 16:566–577.
- Herholz SC, Zatorre RJ (2012) Musical training as a framework for brain plasticity: Behavior, function, and structure. *Neuron* 76:486–502.
- Barnett SM, Ceci SJ (2002) When and where do we apply what we learn? A taxonomy for far transfer. *Psychol Bull* 128:612–637.

38. Hyde KL, et al. (2009) Musical training shapes structural brain development. *J Neurosci* 29:3019–3025.
39. Kuhl PK (2010) Brain mechanisms in early language acquisition. *Neuron* 67:713–727.
40. Penhune VB (2011) Sensitive periods in human development: Evidence from musical training. *Cortex* 47:1126–1137.
41. Zhang Y, et al. (2012) Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *J Child Psychol Psychiatry* 53: 874–882.
42. Nan Y, Sun Y, Peretz I (2010) Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain* 133:2635–2642.
43. Gong YX, Dai XY (1986) China-Wechsler younger children scale of intelligence (C-WYCSI). *Psychol Sci* 2:23–30.
44. Eriksen BA, Eriksen CW (1974) Effects of noise letters upon identification of a target letter in a non-search task. *Percept Psychophys* 16:143–149.
45. Anderson RC, et al. (2002) Shared-book reading in China. *Chinese Children's Reading Acquisition* (Springer, Boston), pp 131–155.
46. Boersma P (2001) Praat, a system for doing phonetics by computer. *Glott Int* 5:341–345.
47. Bishop DV, Hardiman MJ, Barry JG (2010) Lower-frequency event-related desynchronization: A signature of late mismatch responses to sounds, which is reduced or absent in children with specific language impairment. *J Neurosci* 30:15578–15584.
48. Bishop DV, Hardiman MJ, Barry JG (2011) Is auditory discrimination mature by middle childhood? A study using time-frequency analysis of mismatch responses from 7 years to adulthood. *Dev Sci* 14:402–416.
49. Semlitsch HV, Anderer P, Schuster P, Presslich O (1986) A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23:695–703.
50. Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol* 118: 2544–2590.
51. Mahajan Y, McArthur G (2015) Maturation of mismatch negativity and P3a response across adolescence. *Neurosci Lett* 587:102–106.
52. Halliday LF, Barry JG, Hardiman MJ, Bishop DV (2014) Late, not early mismatch responses to changes in frequency are reduced or deviant in children with dyslexia: An event-related potential study. *J Neurodev Disord* 6:21.
53. Shestakova A, Huotilainen M, Ceponiene R, Cheour M (2003) Event-related potentials associated with second language learning in children. *Clin Neurophysiol* 114: 1507–1512.
54. He C, Hotson L, Trainor LJ (2009) Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *Eur J Neurosci* 29:861–867.
55. Shafer VL, Yu YH, Datta H (2010) Maturation of speech discrimination in 4- to 7-yr-old children as indexed by event-related potential mismatch responses. *Ear Hear* 31: 735–745.
56. Putkinen V, Tervaniemi M, Huotilainen M (2013) Informal musical activities are linked to auditory discrimination and attention in 2-3-year-old children: An event-related potential study. *Eur J Neurosci* 37:654–661.
57. Putkinen V, Tervaniemi M, Saarikivi K, Ojala P, Huotilainen M (2014) Enhanced development of auditory change detection in musically trained school-aged children: A longitudinal event-related potential study. *Dev Sci* 17:282–297.
58. Hövel H, et al. (2015) Auditory event-related potentials are related to cognition at preschool age after very preterm birth. *Pediatr Res* 77:570–578.