PNAS
www.pnas.org

Supplementary Information for

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

Yun Nan[a], Li Liu[a], Eveline Geiser[b,c,d], Hua Shu[a], Chen Chen Gong[b], Qi Dong[a], John D.E. Gabrieli[b,e], and Robert Desimone[b]

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

Yun Nan, Ph.D.
Email: nany@bnu.edu.cn

Robert Desimone, Ph.D.
Email: desimone@mit.edu

This PDF file includes:

Supplementary text
Figs. S1 to S4
Tables S1 to S3
References for SI reference citations

www.pnas.org/cgi/doi/10.1073/pnas.1808412115

Supplementary Information Text

SI Results

1 Time window analysis of the difference waveforms

For lexical tone condition, significant session by group interactions were found in the intervals of 300-350 ms ($F(2, 71) = 8.62$, $P < 0.001$), 350-400 ms ($F(2, 71) = 7.07$, $P = 0.002$), and 400-450 ms ($F(2, 71) = 6.62$, $P = 0.002$). The main effects of session were significant in the intervals of 500-550 ms ($F(1, 2) = 4.58$, $P = 0.036$), 550-600 ms ($F(1, 2) = 10.43$, $P = 0.002$), and 600-650 ms ($F(1, 2) = 6.73$, $P = 0.011$). Simple effect analyses suggested that from 300 ms through 450 ms, piano group showed significant session effects with larger positive-going deflections for post-test than pre-test (all *P*s < 0.01), whereas reading and control groups showed no session effects for 300-350 ms and 350-400 ms (all *P*s > 0.1) and opposite session effects for 400-450 ms interval (both *P*s < 0.05). The main session effects in the time windows of 500-550 ms, 550-600 ms, and 600-650 ms revealed significantly larger negative deflections at post-test than pre-test for all the three groups (all *P*s < 0.01).

For musical pitch condition, marginal to significant session by group interactions were found in the intervals of 250-300 ms ($F(2, 71) = 2.90$, $P = 0.062$), 300-350 ms ($F(2, 71) = 4.25$, $P = 0.018$), and 450-500 ms ($F(2, 71) = 3.30$, $P = 0.043$). No main effect of session was observed for any time window. Simple effect analyses revealed that for the 250-300 ms and 300-350 ms intervals, piano group showed significant session effects (both *P*s < 0.01) with larger positivities at post-test than pre-test, whereas the reading and control groups showed no such effects (all *P*s > 0.1). For the 450-500 ms interval, only the control group showed a significant session effect ($P < 0.05$) with more negative-going deflections observed at post-test than pre-test, whereas the piano and reading groups did not have such effects (both *P*s > 0.1).

The mismatch responses (the lexical tone pMMR from 300-450 ms, the lexical tone LDN from 500-650 ms, and the musical pitch pMMR from 250-350 ms) for further analysis were defined because they originated from at least two consecutive time windows in the corresponding time window analysis（1）.

2 Developmental effects in lexical tone LDN amplitudes

The piano group, similar to the reading and control groups, showed larger lexical tone LDNs at post-test (-1.84 μV) than pre-test (-0.79 μV) (a significant main effect of session, $F(1, 2) = 9.49$, $P = 0.003$, $\eta^2 = 0.10$). The amplitudes of the lexical tone LDNs were significantly largest over frontal electrodes (-1.78 μV), larger over fronto-central electrodes (-1.39 μV), and smallest over central electrodes (-0.78 μV), all *P*s < 0.05 (a significant main effect of area, $F(2, 142) = 42.59$, $P < 0.001$). In terms of hemispheric distribution, the lexical tone LDNs were significantly larger at the left (-1.49 μV) than those at the midline sites (-1.20 μV), $P < 0.01$ (a significant main effect of hemisphere, $F(2, 142) = 4.30$, $P = 0.017$) . No amplitude differences were observed between the left and the right (-1.25 μV) or between the right and the midline, both *P*s > 0.1.

3 Individual difference: good versus poor pitch learners

Among the thirty participants within the piano group, fifteen were categorized into *good pitch learner* group and the other fifteen *poor pitch learner* group based on a median split of the behavioral pitch discrimination enhancements at post-test than pre-test. As a result, the good pitch learners were those who obtained greater increments in pitch discrimination performance after piano training, whereas the poor pitch learners showed little progress, $t(28) = 7.03$, $P < 0.001$. Otherwise, these two groups were matched on all related measures except for the attention scores at pre-test. The good pitch learners started out with higher attention scores than the poor pitch learners, $P < 0.001$.

As shown in Figure S4, compared to poor pitch learners, good pitch learners showed much larger behavioral improvement in word discrimination test, $t(28) = 2.12$, $P = 0.043$. This was mainly due to the more significant gain for the good learners than the poor learners in word discrimination based on lexical tones ($P = 0.022$). By contrast, both good and poor learners showed similar improvements in word discrimination based on consonants and vowels (both Ps > 0.1). Notably, the increments of the lexical tone pMMR across the nine electrodes were not statistically different between these two groups, $P > 0.1$. However, the good pitch learners demonstrated significantly larger increments in the pitch pMMR over the nine electrodes than the poor pitch learners, $F(1, 28) = 5.29$, $P = 0.029$. Moreover, although the good pitch learners showed better attention scores than the poor pitch learners at pre-test, those two groups ended up with the same level of attention scores after piano training ($P > 0.1$). Note that the good pitch learners' advantage over poor pitch learners in behavioral word discrimination and electrical responses to musical pitch deviations still held when the attention scores at pre-test was taken as a covariate, suggesting that the different attention scores at pre-test cannot account for the resulting group differences in both behavioral and electrical measures.

SI Discussion

The development of pitch processing as reflected by the mismatch responses

Sound frequency discrimination abilities drastically develop between the age of 4 to 8 years（2）. As expected, lexical tone LDN clearly increased in amplitude with age across all tested children and independent of the applied training: at post-test both the two training groups and the control group exhibited significantly larger lexical tone LDNs than at pre-test. This LDN increment cannot be attributed to test-retesting effects, as there is no such effect in the musical pitch condition. Thus, the increased LDN in lexical tone condition likely reflects developmental enhancements in lexical tone discrimination that had taken place independently of any training. Increase in LDN to lexical tone processing is a hallmark of healthy language development and developmental deficits reflected in reduced LDNs were observed across childhood and into adolescence（3）. The increased lexical tone LDNs together with the musical training related pMMR enhancements in both musical pitch and lexical tone conditions provide evidence on the complexity of pitch development as reflected by the mismatch responses in these 4-5 year olds. The fact that musical training enhanced lexical tone pMMRs whereas natural language development increased lexical tone LDNs suggests that music learning and natural language maturation affect different time windows of mismatch responses in lexical tone discrimination.

Although both the pMMR (sometimes as P3a) and the LDN were often seen in developmental training studies（4−6）, like the MMN, their functional significance is still unclear. The critical question is how the bottom-up processing and top-down functions interact in the processes as indexed by these components. Conventionally, the MMN is regarded as a pre-attentive component of the mismatch responses（7）, whereas the pMMRs and LDN are taken as reflecting attentional and executive functions involved in the discriminative processes（5, 6）. However, accumulating research starts to recognize a more balanced interplay between the bottom-up and top-down processes as indexed by these components. For the MMN component, although it is usually elicited without overt attention, a recent study showed that it could be modulated by attention（8）, which is in contrast to the pre-attentive claim of the MMN. For the pMMR and the LDN, more research has shown that these two components are critically dependent upon the deviant features（3, 9−12）, suggesting bottom-up contributions in the related processes. Together these evidences are in agreement with the fact that the mammalian auditory system constitutes numerous efferent projections from the cortex to all the hierarchical stages of the auditory system（13−16）, supporting the notion that auditory processing is indeed a result of dynamic interplays between feedforward and feedback processing at different hierarchical levels（17, 18）.

Our current data support the postulation that the pMMRs and LDNs are late manifestations of the MMN. Except for the musical pitch pMMRs, the lexical tone pMMRs and LDNs in both conditions exhibited fronto-central distribution, similar to the MMNs. These same topographic patterns indicate that the late mismatch responses such as the pMMRs and LDNs may well have the same auditory origin as the MMNs（7）. Indeed, a recent study examined the pathophysiological mechanisms underlying MMN and P3a（19）in schizophrenia patients and found that these two components share at least 50% of their generation mechanisms（19）. However, the timing differences in mismatch responses may signal important developmental stages（20）. Future work is necessary to understand the developmental dynamics of these different mismatch responses.

Individual difference on pitch plasticity induced by musical training

As noted in Herholz & Zatorre（21）and Zatorre（22）, pre-existing individual differences and training-induced plasticity are not mutually exclusive. In the current longitudinal study, relative to reading training and no-contact controls, piano training resulted in increased electrical responses to lexical tone and musical pitch changes as well as the accompanied higher behavioral improvements in word discrimination. However, individual differences did fine-tune the detailed profiles of learning in the piano group. According to the behavioral pitch discrimination sensitivities, the good pitch learners demonstrated significantly larger increments in musical pitch pMMRs and word discrimination performance than the poor pitch learners after piano training. Although began with relatively lower attention scores at pre-test, after piano training, the poor pitch learners caught up with the good pitch learners at post-test. Moreover, they demonstrated a similar level of enhancements in behavioral word discrimination based on consonants and vowels and in electrical responses to lexical tone changes as the good learners after piano training.

However, the exact factors that drove these individual differences are less clear. Note that at pre-test, the poor pitch learners were comparable to the good pitch learners on all the related measures except for the attention scores, but the lower attention score at pre-test could not account for poor pitch learners' less enhancements in musical pitch pMMRs and word discrimination performance associated with piano learning than the good pitch learners. In prior work, similar individual differences were explored for pitch memory（23）and pitch discrimination training（24）. These individual variations in the learning outcomes were linked to differential plastic responses of supramarginal and auditory areas to the respective trained tasks. Notably, pre-existing functional sensitivity of auditory cortex to pitch differences may predict the individual outcome of micromelody training（25）. However, please note that epigenetic and other environmental factors except for musical training might also contribute to the currently observed individual learning rates and a more detailed analysis of these factors await further research.

Nonetheless, these results suggest that musical training is beneficial not only for the whole group, but more importantly also for the individuals. Both the good and the poor pitch learners demonstrated gains after piano training, although some were in slightly different aspects. More importantly, the fact that the poor pitch learners achieved a similar level of electrical enhancements in discriminating lexical tones as the good pitch learners suggests that musical training improves neural speech processing unanimously for the whole group, regardless of individual’s music learning rates. This holds great promise for using musical training to promote language abilities in both healthy and at-risk children.

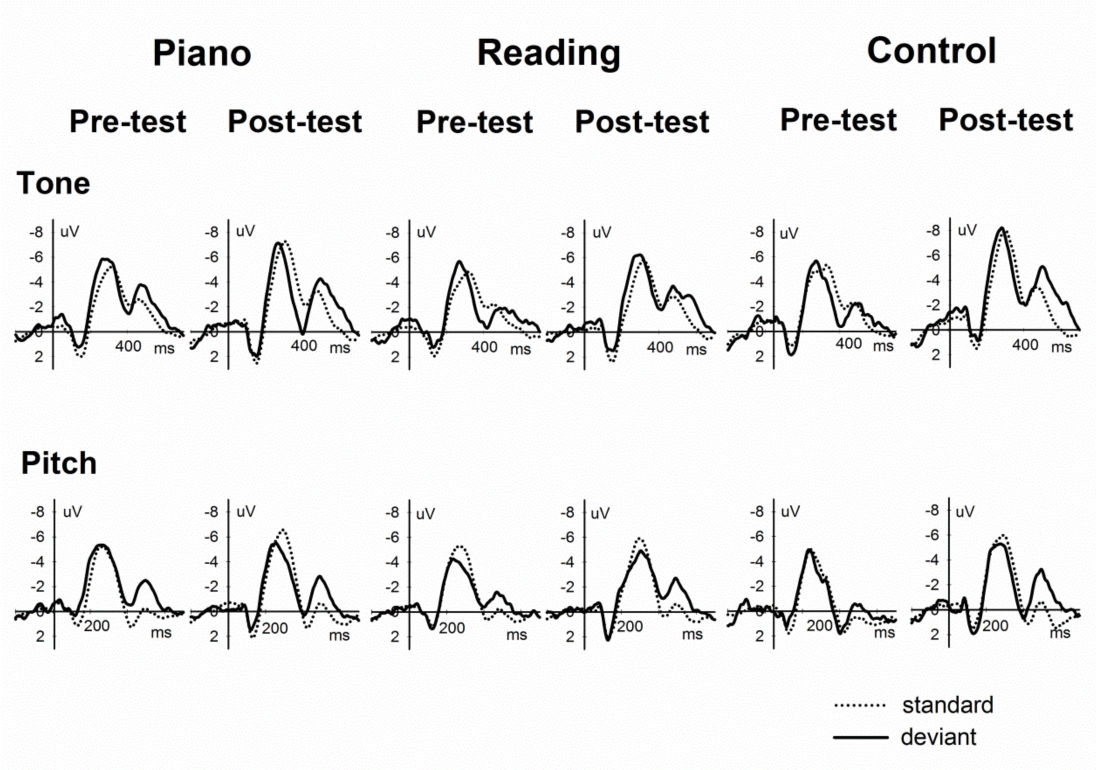

Fig. S1. The grand average waveforms elicited by standards (dotted line) and deviants (solid line) for lexical tone and musical pitch conditions at Fz for the three groups.

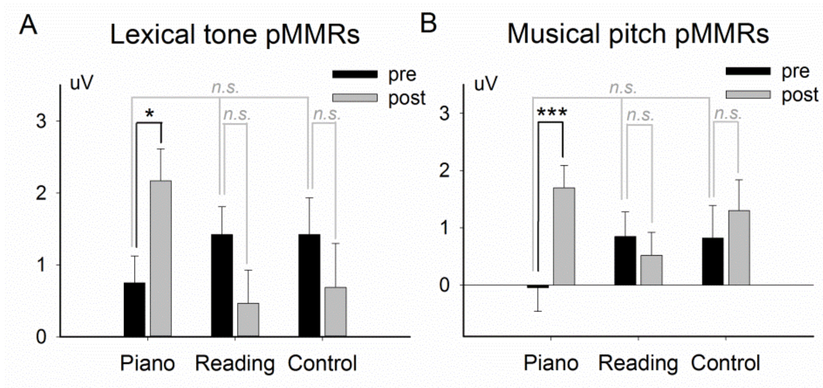

Fig. S2. Mean positive mismatch response (pMMR) amplitudes over the nine electrodes at post-test relative to pre-test for the three groups. The left is lexical tone condition (A) and the right musical pitch condition (B). The piano group showed significant pMMR enhancements at post-test relative to pre-test for both lexical tone and musical pitch conditions. No such effects were observed for the reading and control groups. **P* < 0.05; ****P* < 0.001; *n.s.* indicates no significant difference. Error bar indicates standard error.

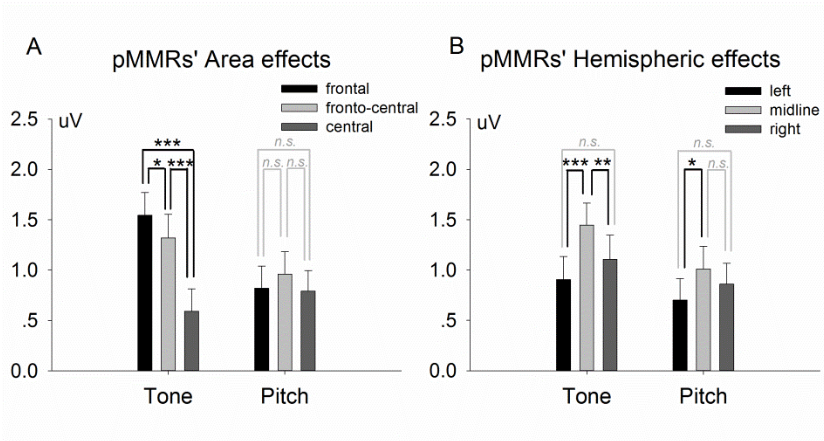

Fig. S3. The area and hemispheric effects for positive mismatch responses (pMMRs). An area effect was observed only in lexical tone pMMRs (A), whereas pMMRs from lexical tone and musical pitch conditions both showed hemispheric effects (B). Data were lumped together across groups and sessions (pre- and post-tests). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; *n.s.* indicates no significant difference. Error bar indicates standard error.

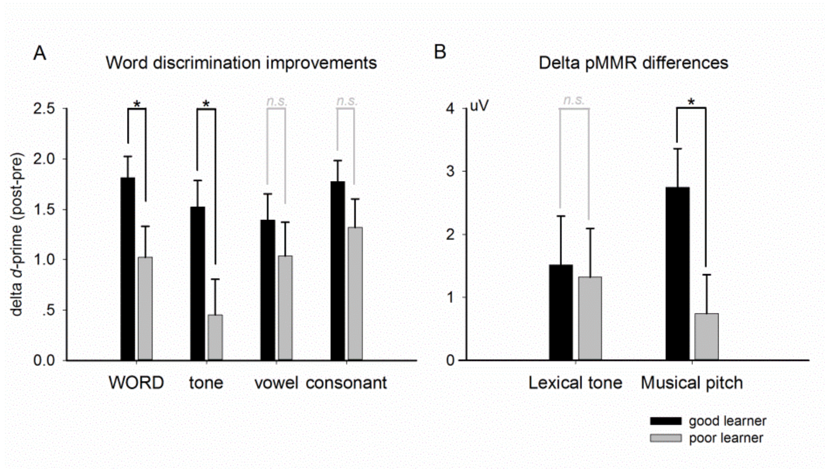

Fig. S4. The comparison between good and poor pitch learners. The left is word discrimination performance (A). The right is delta pMMR differences in lexical tone and musical pitch conditions over the nine electrodes (B). *$P < 0.05$; *n.s.* indicates no significant difference. Error bar indicates standard error.

Table S1. Word pairs used in behavioral discrimination task.

Conditions	Same		Different	
Tone	ta1	ta1	ting1	ting2
	cong2	cong2	sui2	sui1
	shui3	shui3	sheng1	sheng3
	dai4	dai4	ba3	ba1
	hua1	hua1	xin1	xin4
	da2	da2	zhi4	zhi1
	mi3	mi3	you2	you3
	fang4	fang4	qi3	qi2
	zhou1	zhou1	shi2	shi4
	yun2	yun2	guo4	guo2
	lao3	lao3	ci3	ci4
	si4	si4	li4	li3
Consonant	zhua1	zhua1	shen1	zhen1
	dang1	dang 1	shuo1	duo1
	he2	he2	cheng2	neng2
	xue2	xue2	nin2	min2
	ma3	ma3	gei3	mei3
	nv3	nv3	zhao3	zao3
	sui4	sui4	ri4	shi4
	dian4	dian4	lu4	bu4
Vowel	chi1	chi1	fei1	fen1
	qu1	qu1	jin1	jia1
	nan2	nan2	yu2	you2

	wen2	wen2	tai2	tan2
	mai3	mai3	gan3	gai3
	zong3	zong3	bai3	bei3
	song4	song4	hui4	hou4
	kuai4	kuai4	dao4	da4
Practice	hen3	hen3	wu2	wu3
	da1	da1	zhang1	fang1
	ming2	ming2	qing1	qi1

Word pairs based on tones were constructed to cover all possible combinations of tone pairs (1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 2 vs. 4, 3 vs. 4) with two different orders (e.g., 1 vs. 2 and 2 vs. 1). Word pairs based on consonants and vowels included evenly distributed four lexical tones. After practice, all word pairs were delivered in a random order in the word discrimination test.

Table S2. Mean raw scores and age-corrected scaled scores for the five subtests of the Chinese version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI).

subtests	All three groups (N=74)			Piano (*n*=30)		Reading (*n*=28)		Control (*n*=16)	
	pre	post	*P*	pre	post	pre	post	pre	post
Raw score (SD)									
Vocabulary	**18.0 (4.4)**	**19.9 (5.1)**	**0.007**	18.0 (3.7)	20.0 (5.2)	18.0 (5.1)	19.8 (5.7)	18.2 (4.3)	19.9 (4.2)
Similarities	**13.4 (3.9)**	**15.89 (3.2)**	**0.001**	14.0 (4.1)	15.7 (3.5)	12.8 (4.1)	16.2 (3.4)	13.4 (3.1)	15.6 (2.5)
Animal house	**36.6 (11.2)**	**46.3 (6.9)**	**0.001**	38.3 (12.7)	48.0 (6.9)	34.4 (9.9)	44.8 (6.5)	37.2 (10.3)	45.7 (7.3)
Picture completion	**12.0 (2.9)**	**14.3 (2.9)**	**0.001**	12.6 (3.4)	14.9 (3.1)	11.9 (2.5)	13.8 (3.0)	11.3 (2.8)	14.0 (2.8)
Block design	**11.0 (4.4)**	**16.5 (2.0)**	**0.001**	11.4 (4.8)	16.9 (2.3)	10.6 (4.5)	15. 9 (1.9)	10.9 (3.5)	16.9 (1.3)
Scaled score (SD)									
Vocabulary	12.6 (2.4)	11.8 (2.7)	0.022	12.5 (1.7)	11.8 (3.0)	12.8 (2.9)	11.8 (2.6)	12.4 (2.4)	11.6 (2.3)
Similarities	14.5 (2.7)	14.7 (2.6)	0.575	14.8 (2.6)	14.8 (2.7)	14.2 (3.0)	15.0 (2.9)	14.3 (2.2)	14.1 (2.0)

Animal house	11.7 (2.4)	11.2 (2.1)	0.017	12.1 (2.9)	11.8 (2.4)	11.4 (1.9)	10.8 (1.7)	11.6 (2.0)	10.6 (1.8)
Picture completion	12.4 (2.2)	11.8 (2.3)	0.062	12.8 (2.4)	12.4 (2.7)	12.5 (2.1)	11.5 (2.1)	11.6 (1.9)	11.3 (1.7)
Block design	**13.0 (3.0)**	**14.7 (2.0)**	**0.001**	13.2 (3.5)	15.3 (2.1)	12.9 (2.9)	14.2 (2.0)	12.7 (2.1)	14.4 (1.9)

SD indicates standard deviation. Significant session effects (e.g., significant improvements from pre- to post-test) and their *P* values obtained with two-way ANOVAs with session and group as factors are shown in bold (Bonferroni corrected $P < 0.05$).

Table S3. The demographic characteristics and the pre-test word and pitch discrimination performance for the dropouts from the piano, reading, and control groups.

Demographic characteristics	Piano (n=10)	Reading (n=10)	Control (n=24)	Group difference (P)
Mean age, months (SD)	56.2 (8.5)	53.6 (3.1)	56.8 (2.6)	0.199
Male/female	6/4	7/3	11/13	0.336
IQ (SD)	125.0 (12.7)	115.0 (12.4)	122.3 (14.6)	0.249
Digit span (SD)	12.8 (2.9)	11.5 (3.9)	12.2 (3.7)	0.719
Flanker (SD)	84.7 (8.1)	66.5 (18.5)	81.5 (17.6)	0.029*
Word *d*-prime (SD)	1.6 (0.7)	1.2 (0.9)	1.8 (1.0)	0.234
Pitch *d*-prime (SD)	1.0 (1.0)	0.8 (1.0)	1.2 (1.2)	0.168

The group comparisons among the three groups yielded no significant group difference except for the Flanker scores, where the dropouts from the reading group were lower than the dropouts from the other two groups (reading vs. piano, $P = 0.048$; reading vs. control, $P = 0.055$; piano vs. control, $P = 1$). Note that the lower Flanker scores in the reading dropouts might have imposed little bias to the current results because in the analyses reported in the present study, the remainders in each group were again matched on all demographic and pre-test measures. IQ indicates the intelligence quotient. Digit span scores are raw scores of the correct responses. The performance on the Flanker test is the percent correct responses. SD indicates the standard deviation. The P value for group difference in gender (male/female) was obtained using Chi-Square Tests, whereas the P values for other variables were obtained using one-way ANOVA. *$P < 0.05$.

References

1. Lee CY et al. (2012) Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia* 50(14):3228-3239.
2. Jenson KJ， Neff DL (1993) Development of Basic Auditory Discrimination in Preschool Children. *Psychological Science* 4(2):104-107.
3. 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(46):15578-15584.
4. 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(2):282-297.
5. Moreno S， Lee Y， Janus M， Bialystok E (2015) Short-term second language and musical training induces lasting functional brain changes in early childhood. *Child Dev.* 86(2):394-406.
6. Shestakova A， Huotilainen M， Ceponiene R， Cheour M (2003) Event-related potentials associated with second language learning in children. *Clin.Neurophysiol.* 114(8):1507-1512.
7. Naatanen R， Paavilainen P， Rinne T， Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin.Neurophysiol.* 118(12):2544-2590.
8. Auksztulewicz R， Friston K (2015) Attentional Enhancement of Auditory Mismatch Responses: a DCM/MEG Study. *Cereb.Cortex*. 25(11):4273-4283.
9. 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(2):402-416.
10. Putkinen V， Tervaniemi M， Saarikivi K， Huotilainen M (2015) Promises of formal and informal musical activities in advancing neurocognitive development throughout childhood. *Ann.N.Y.Acad.Sci.* 1337:153-162.
11. 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(4):654-661.
12. Kushnerenko E et al. (2007) Processing acoustic change and novelty in newborn infants. *Eur.J.Neurosci.* 26(1):265-274.
13. Diamond IT， Jones EG， Powell TP (1969) The projection of the auditory cortex upon the diencephalon and brain stem in the cat. *Brain Res.* 15(2):305-340.
14. Mulders WH， Robertson D (2000) Evidence for direct cortical innervation of medial olivocochlear neurones in rats. *Hear.Res.* 144(1-2):65-72.
15. Weedman DL， Ryugo DK (1996) Projections from auditory cortex to the cochlear nucleus in rats: synapses on granule cell dendrites. *J.Comp Neurol.* 371(2):311-324.
16. Winer JA， Larue DT， Diehl JJ， Hefti BJ (1998) Auditory cortical projections to the cat inferior colliculus. *J.Comp Neurol.* 400(2):147-174.
17. Buzsaki G， Draguhn A (2004) Neuronal oscillations in cortical networks. *Science* 304(5679):1926-1929.
18. Giraud AL， Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat.Neurosci.* 15(4):511-517.
19. Leitman DI et al. (2010) Sensory deficits and distributed hierarchical dysfunction in schizophrenia. *Am.J.Psychiatry* 167(7):818-827.
20. Yang MT et al. (2015) Attention deficits revealed by passive auditory change detection for pure tones and lexical tones in ADHD children. *Front Hum.Neurosci.* 9:470.

21. Herholz SC， Zatorre RJ (2012) Musical training as a framework for brain plasticity: behavior, function, and structure. *Neuron* 76(3):486-502.
22. Zatorre RJ (2013) Predispositions and plasticity in music and speech learning: neural correlates and implications. *Science* 342(6158):585-589.
23. Gaab N， Gaser C， Schlaug G (2006) Improvement-related functional plasticity following pitch memory training. *Neuroimage.* 31(1):255-263.
24. Jancke L， Gaab N， Wustenberg T， Scheich H， Heinze HJ (2001) Short-term functional plasticity in the human auditory cortex: an fMRI study. *Brain Res.Cogn Brain Res.* 12(3):479-485.
25. Zatorre RJ， Delhommeau K， Zarate JM (2012) Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Front Psychol.* 3:544.