










# Large-scale multi-site study shows no association between musical training and early auditory neural sound encoding

Received: 12 August 2024

Accepted: 11 July 2025

Published online: 08 August 2025

 Check for updates

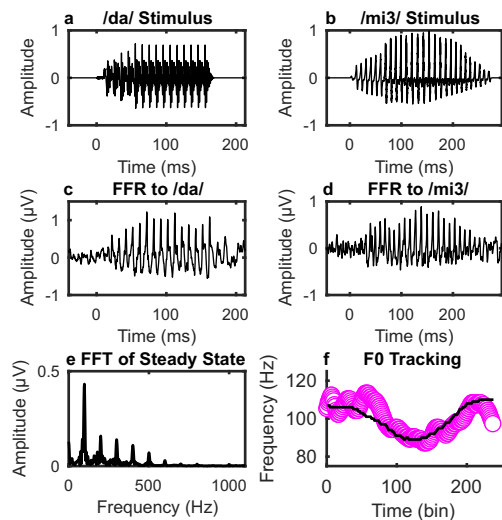
Kelly L. Whiteford <sup>1,12</sup>✉, Lucas S. Baltzell<sup>2</sup>, Matt Chiu <sup>3,13</sup>, John K. Cooper<sup>4</sup>, Stefanie Faucher<sup>4</sup>, Pui Yii Goh<sup>1</sup>, Anna Hagedorn<sup>5</sup>, Vanessa C. Irsik<sup>6,7</sup>, Audra Irvine<sup>8</sup>, Sung-Joo Lim<sup>2,14</sup>, Juraj Mesik <sup>1</sup>, Bruno Mesquita <sup>6</sup>, Breanna Oakes<sup>5</sup>, Neha Rajappa<sup>1</sup>, Elin Roverud<sup>2</sup>, Amy E. Schrlau<sup>4</sup>, Stephen C. Van Hedger <sup>6,7</sup>, Hari M. Bharadwaj <sup>5,9,15</sup>, Ingrid S. Johnsrude<sup>6,7,10</sup>, Gerald Kidd Jr<sup>2</sup>, Anne E. Luebke<sup>4,11</sup>, Ross K. Maddox<sup>4,11,12</sup>, Elizabeth W. Marvin <sup>3</sup>, Tyler K. Perrachione<sup>2</sup>, Barbara G. Shinn-Cunningham <sup>8</sup> & Andrew J. Oxenham <sup>1</sup>

Musical training has been reported to be associated with enhanced neural processing of sounds, as measured via the frequency following response (FFR), implying the potential for human subcortical neural plasticity. We conducted a large-scale multi-site preregistered study ( $n > 260$ ) to replicate and extend the findings underpinning this important relationship. We failed to replicate any of the major findings selected for replication that were published previously in smaller studies. Musical training was not associated with enhanced neural encoding strength of a speech stimulus (/da/) in babble, whether measured via the spectral or temporal representations of the FFR. Similarly, the strength of neural tracking of a speech sound with a dynamic pitch trajectory was not related to either years of musical training or age of onset of musical training. Our findings provide no evidence for associations between early auditory neural responses and either musical training or musical ability.

Music is universal across human societies and can serve multiple functions<sup>1</sup>. In many Western cultures, parents may seek early musical education for their children, sometimes in the hope that the skills learned from musical training may transfer to other aspects of life. A considerable body of research has examined the potential link between such early training and enhanced perceptual and cognitive

skills, with mixed results<sup>2–7</sup>. Although there is strong consensus that musicians tend to perform well on measures that are directly relevant to playing a musical instrument, such as melody<sup>8</sup> and pitch perception<sup>8,9</sup> and fine-grained motor coordination<sup>10</sup>, there is less consensus on the degree and magnitude of advantages related to complex listening outside of the musical environment (e.g., speech in

<sup>1</sup>Department of Psychology, University of Minnesota, Minneapolis, MN, USA. <sup>2</sup>Department of Speech, Language, and Hearing Sciences, Boston University, Boston, MA, USA. <sup>3</sup>Eastman School of Music, University of Rochester, Rochester, NY, USA. <sup>4</sup>Department of Biomedical Engineering, University of Rochester, Rochester, NY, USA. <sup>5</sup>Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, IN, USA. <sup>6</sup>Centre for Brain and Mind, University of Western Ontario, London, ON, Canada. <sup>7</sup>Department of Psychology, University of Western Ontario, London, ON, Canada. <sup>8</sup>Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA. <sup>9</sup>Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, USA. <sup>10</sup>School of Communication Sciences and Disorders, University of Western Ontario, London, ON, Canada. <sup>11</sup>Department of Neuroscience, University of Rochester, Rochester, NY, USA. <sup>12</sup>Present address: Department of Otolaryngology-Head and Neck Surgery, Kresge Hearing Research Institute, University of Michigan, Ann Arbor, MI, USA. <sup>13</sup>Present address: Conservatory of Performing Arts, Baldwin Wallace University, Berea, OH, USA. <sup>14</sup>Present address: Department of Psychology, Binghamton University, Binghamton, NY, USA. <sup>15</sup>Present address: Department of Communication Science and Disorders, University of Pittsburgh, Pittsburgh, PA, USA. ✉e-mail: [klwhit@umich.edu](mailto:klwhit@umich.edu)



**Fig. 1 | Stimuli and sample neural responses.** The acoustic waveforms for /da/ (a) and /mi3/ (b) are shown, along with example FFRs in the time domain (c, d) and spectral domain (e, f) for one individual participant. The fast Fourier transform (FFT) of the steady-state portion of the FFR to /da/ (e) shows the spectral encoding for the F0 (100 Hz) and upper harmonics. Previous research found that musicians had stronger spectral encoding for the upper harmonics to /da/ presented in babble, relative to non-musicians, and stronger stimulus-to-response correlations between the vowel portion of the /da/ stimulus (50–170 ms) and the steady-state portion of the neural response (60–180 ms)<sup>13</sup>. The F0 of the /mi3/ stimulus (f, black line) varies over time (pink: neural F0 tracking). The original study found that musicians had stronger stimulus-to-response correlations between the F0-trajectory of the stimulus and the neural response than non-musicians, and that F0-tracking fidelity was related to the age of onset of musical training.

noise) or cognitive function<sup>2</sup>. There are, however, numerous reports that musical training from an early age is associated with enhanced responses to sound at different levels of the auditory pathways, from the auditory brainstem<sup>11–17</sup> to the midbrain and cortex<sup>8,18–21</sup>, which may serve as a neural substrate of the perceptual advantages of musicianship.

A widely reported brain-behavior association is between musical training and the frequency-following response (FFR), a scalp-recorded potential, typically measured using electroencephalography (EEG). The fidelity of neural encoding can be assessed either in the time domain or in the spectral domain, with the time waveform of the neural response bearing visual similarity to the original stimulus (Fig. 1). The FFR is thought to reflect stimulus-entrained neural responses to periodic sounds, primarily from subcortical levels of the auditory system, but also potentially reflecting some cortical contributions, especially at frequencies of 100 Hz and below<sup>22,23</sup>. Several studies have reported that musicians exhibit a more faithful representation of the stimulus frequency in the FFR of sounds (spectral encoding; Fig. 1e, f) than non-musicians, either for the fundamental frequency (F0) or the upper harmonics, and that musicians' spectral encoding is more robust to interference from background noise<sup>11,12,15–17,24</sup>. Such findings have important implications: if the relationship is causal, it implies the potential for plasticity of subcortical auditory function, at least in childhood (when musical training typically begins). Even if the results from these cross-sectional studies (i.e., studies where musicianship is a self-selected group) do not stem from musical-training-induced changes in subcortical auditory responses, an alternative interpretation, linking the strength of responses in the early stages of the auditory pathways (perhaps genetically influenced) to future musical success, would also be highly intriguing.

Enhanced FFRs in musicians have been demonstrated in numerous studies, using both speech and non-speech sounds<sup>11,12,15–17,24</sup>. The

most commonly used stimuli are speech sounds that include vowels with a constant pitch (e.g., /da/, Fig. 1a) and speech sounds that vary in pitch over time (e.g., the Mandarin utterance, /mi3/, which dips and rises in F0; Fig. 1b, f, black line)<sup>25</sup>. One well-cited study found that musicians had stronger spectral encoding for the upper harmonics of /da/ presented in background noise than non-musicians, and that the correlation between the time waveform of the neural response and the stimulus waveform itself was greater in musicians than in non-musicians<sup>13</sup>. Both findings indicated that musicians' FFRs were more robust to the presence of background noise than those of non-musicians. Another influential study showed that native English speaking musicians had a stronger representation of the time-varying F0 for a Mandarin word presented in quiet<sup>15</sup>, as quantified by the correlation between the F0 contour of the stimulus and neural response, than non-musicians. Importantly, the strength of the neural representations of the stimulus F0 was negatively correlated with the age at which musicians began their musical training. This relationship suggests that the earlier onset of musical training produces stronger neural responses. Together, these two studies have had a major impact, as evidenced by the high number of citations (a combined total of 1904, as of the time of writing), and have provided the foundation for a subsequent body of positive evidence for enhanced early neural encoding of sound in musicians<sup>26</sup>.

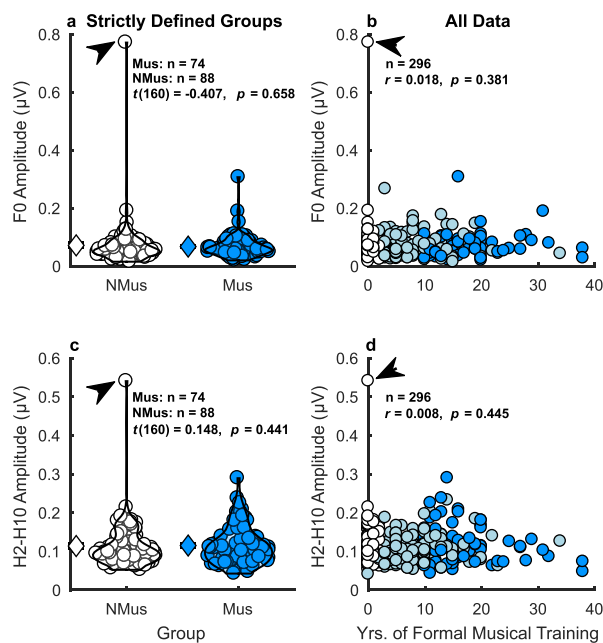
Although the existing evidence has been interpreted as supporting the notion that musicians exhibit advantages in the neural encoding of sound in the early stages of the auditory system<sup>26</sup>, several factors complicate this interpretation and the generalizability of any such advantage. First, most reports have been based on relatively small samples of listeners<sup>13,15,27</sup> with dichotomous samples that often represent extreme ends of the musical spectrum (i.e., people with little or no musical training compared to musicians with many years of experience). Limiting the sample in this manner increases the likelihood that between-group differences will be detected, but such disparate groups may differ in many other ways besides musicianship, such as socioeconomic status or personality<sup>2,28</sup>, which in turn limits the generalizability of the findings<sup>29</sup>. Moreover, most studies have been conducted on young (college-aged) adults, further limiting the generalizability, although a growing number of studies have examined aging effects<sup>30–35</sup>. Another complicating factor is that the measures that revealed differences between musicians and non-musicians sometimes differed between studies<sup>11,13,14,35</sup>—even when the same stimuli were used (see Supplementary Table 1), increasing the potential for false positives. Lastly, the definition of the terms “musician” and “non-musician” has varied between studies, leading to the possibility that any differences in outcomes between studies may reflect, in part, differences in years and nature of training, age training began, and the degree to which musical training or activity is maintained.

In the present study, we attempted to replicate two of the most widely cited findings of a musician advantage for early neural encoding of speech sounds, described above<sup>13,15</sup>, across a large sample of participants at six different sites. In addition to conducting the replications, we extended these studies in several important ways: (1) All sites conducted both studies, allowing for a high-powered aggregate sample. (2) All participants at each site were asked to take part in both studies, allowing for the strength of neural encoding between studies to be compared within the same participants. (3) Both age and musical training varied continuously, increasing the generalizability of the findings, relative to previous studies, while still allowing for a direct replication by maintaining a subset of participants who fit the most stringent criteria for the definitions of musicians and non-musicians. (4) The methods and primary analyses were preregistered before beginning data collection<sup>36</sup>, limiting researcher degrees of freedom when analyzing the data. (5) All data used in analyses are openly available for further exploratory analysis in future studies.

## Results

### Spectral encoding of the syllable /da/ in multi-talker babble

The first study we sought to replicate<sup>13</sup> compared the neural encoding of the syllable /da/ (Fig. 1a, c, e) embedded in multi-talker babble, in 16 musicians and 16 non-musicians, as they passively watched a silent video. In that study, musicians were found to have enhanced neural encoding of the vowel portion of the FFR for the upper harmonics (H2-H10) but not the F0. We first conducted a direct replication in a subset of our participants that represented the extreme ends of the musical training spectrum (74 musicians and 88 non-musicians). As shown in Fig. 2a, c, there was no significant difference in spectral encoding as indexed by FFRs between musicians and non-musicians for either the F0 [ $t(160) = -0.407$ ,  $p = 0.658$ , CI  $[-0.021, \infty]$ ,  $d = -0.064$ ] or upper harmonics [ $t(160) = 0.148$ ,  $p = 0.441$ , CI  $[-0.013, \infty]$ ,  $d = 0.023$ ] of the same /da/ syllable used in the original study. The lack of effect persisted after removing one non-musician outlier who had unusually strong spectral encoding [Supplementary Fig. 1; F0:  $t(159) = 0.684$ ,  $p = 0.248$ , CI  $[-0.005, \infty]$ ,  $d = 0.108$ ; Upper harmonics:  $t(159) = 0.909$ ,  $p = 0.182$ , CI  $[-0.005, \infty]$ ,  $d = 0.144$ ]. Bayesian analyses indicated the data were 7.81 and 5.24 times more likely to occur under the null hypothesis than the alternative hypothesis, for encoding the F0 ( $BF_{+0} = 0.128$ , % error  $< 0.0001$ ) and upper harmonics ( $BF_{+0} = 0.191$ , % error  $= -0.057$ ), respectively. Excluding the outlier non-musician also provided moderate support that the data originated under the null (F0:  $BF_{+0} = 0.315$ , % error  $< 0.0001$ ; upper harmonics:  $BF_{+0} = 0.404$ , % error  $< 0.0001$ ). The results remained robust across a wide range of widths of the Bayesian prior (Supplementary Fig. 2). These results



**Fig. 2 | No relationship between musical training and spectral encoding for /da/ syllable in background noise.** Musical training is not associated with enhanced spectral encoding of the F0 (a, b) or upper harmonics (c, d) of /da/. No effect was found when comparing the strictly defined groups of musicians and non-musicians [a:  $t(160) = -0.407$ ,  $p = 0.658$ , CI  $[-0.021, \infty]$ ,  $d = -0.064$ ; c:  $t(160) = 0.148$ ,  $p = 0.441$ , CI  $[-0.013, \infty]$ ,  $d = 0.023$ ], and no relationship was observed in the broader sample between neural encoding strength and years of musical training (b:  $r = 0.018$ ,  $p = 0.381$ , CI  $[-0.078, 1]$ ; d:  $r = 0.008$ ,  $p = 0.445$ , CI  $[-0.088, 1]$ ). The null effects of musical training remained after removing one outlier non-musician with unusually strong spectral encoding (indicated by arrow; see Supplementary Fig. 1 for plots with outlier removed). For (b, d), the preregistered value of  $\alpha = 0.017$  was applied. Statistical tests were one-tailed. Black outlines: 1D kernel density estimates (KDEs); Diamonds: Average data; Circles: individual data; NMus: Non-musicians (white); Mus: Musicians (dark blue). Participants in neither of the strictly defined groups are shown in light blue. Error bars are  $\pm 1$  standard error of the mean (SEM).

provide no evidence for a relationship between musical training and the FFR for /da/ in background multi-talker babble for either the F0 or the upper harmonics.

Previous research has found that both age and sex can systematically affect FFR spectral amplitudes. Specifically, spectral amplitudes have been found to decrease with age<sup>37</sup> and to be smaller in males than in females, at least for the upper harmonics<sup>38</sup>. These factors may have added unexplained variance to our data, potentially obscuring a musician effect. While we did not collect information regarding sex at birth, we did collect self-reported gender identity and used this measure in exploratory analyses. Across the full sample of participants, we confirmed that age was related to poorer spectral encoding of both the F0 and upper harmonics of /da/ in babble (F0:  $r = -0.246$ ,  $p < 0.0001$ , CI  $[-1, -0.154]$ ; upper harmonics:  $r = -0.173$ ,  $p = 0.001$ , CI  $[-1, -0.079]$ ; Supplementary Fig. 3). There was no difference in spectral encoding between males and females for the F0 or the upper harmonics either before [F0:  $t(158) = -0.326$ ,  $p = 0.745$ , CI  $[-0.025, 0.018]$ ,  $d = -0.053$ ; upper harmonics:  $t(158) = 0.287$ ,  $p = 0.774$ , CI  $[-0.015, 0.02]$ ,  $d = 0.046$ ] or after [F0:  $t(157) = 1.33$ ,  $p = 0.186$ , CI  $[-0.004, 0.019]$ ,  $d = 0.215$ ; upper harmonics:  $t(157) = 1.37$ ,  $p = 0.174$ , CI  $[-0.004, 0.023]$ ,  $d = 0.221$ ] excluding the male (non-musician) outlier with very strong spectral encoding (shown by the arrow in Fig. 2). Gender differences remained non-significant when using age as a covariate [F0:  $F(1157) = 0.173$ ,  $p = 0.678$ , CI  $[-0.016, 0.025]$ ,  $\eta_p^2 = 0.001$ ; upper harmonics:  $F(1157) = 0.05$ ,  $p = 0.823$ , CI  $[-0.019, 0.015]$ ,  $\eta_p^2 = 0.0003$ ; outlier removed: F0:  $F(1156) = 1.53$ ,  $p = 0.219$ , CI  $[-0.018, 0.004]$ ,  $\eta_p^2 = 0.01$ ; upper harmonics:  $F(1156) = 1.69$ ,  $p = 0.195$ , CI  $[-0.022, 0.005]$ ,  $\eta_p^2 = 0.011$ ]. Our subsamples of highly experienced musicians and inexperienced non-musicians did not differ significantly in age [ $t(160) = -1.3$ ,  $p = 0.196$ , CI  $[-7.93, 1.64]$ ,  $d = -0.205$ ; outlier excluded:  $t(159) = -1.36$ ,  $p = 0.176$ , CI  $[-8.1, 1.49]$ ,  $d = -0.215$ ; two-tailed tests] or gender [ $\chi^2(1) = 0.004$ ,  $p = 0.948$ , Cramer's  $V = 0.005$ ; outlier excluded:  $\chi^2(1) = 0.024$ ,  $p = 0.878$ , Cramer's  $V = 0.012$ ]. Exploratory Analysis of Covariates (ANCOVAs) confirmed that musicianship was not related to enhanced spectral encoding for either the F0 [F(1156) = 0.347,  $p = 0.557$ , CI  $[-0.014, 0.027]$ ,  $\eta_p^2 = 0.002$ ] or upper harmonics [F(1156) = 0.022,  $p = 0.882$ , CI  $[-0.016, 0.018]$ ,  $\eta_p^2 = 0.0001$ ], even after adjusting for effects of age and gender. Notably, there was a small but significant interaction between group and gender [F0:  $F(1155) = 4.2$ ,  $p = 0.042$ ,  $\eta_p^2 = 0.026$ ; upper harmonics:  $F(1155) = 4.7$ ,  $p = 0.032$ ,  $\eta_p^2 = 0.029$ ]. ANCOVAs assume parallel slopes, meaning the model may not accurately convey effects of musicianship when attempting to denoise variance in spectral encoding due to gender<sup>39</sup>. Removing the non-musician (male) outlier with the strongest spectral encoding resulted in an interaction for encoding the F0 [F0:  $F(1154) = 4.02$ ,  $p = 0.047$ ,  $\eta_p^2 = 0.025$ ] but did not reach significance for the upper harmonics [F(1154) = 3.36,  $p = 0.069$ ,  $\eta_p^2 = 0.021$ ]. Examining the individual data bifurcated by group and gender revealed that this interaction was primarily driven by a female musician outlier with the second strongest spectral encoding (Supplementary Fig. 4) and was no longer significant after exclusion [F0:  $F(1153) = 3.16$ ,  $p = 0.078$ ,  $\eta_p^2 = 0.02$ ; upper harmonics:  $F(1153) = 2.91$ ,  $p = 0.09$ ,  $\eta_p^2 = 0.019$ ]. The primary findings of no musician advantage in spectral encoding remained after excluding both outliers [F0:  $F(1154) = 0.013$ ,  $p = 0.908$ , CI  $[-0.009, 0.01]$ ,  $\eta_p^2 < 0.0001$ ; upper harmonics:  $F(1154) = 0.203$ ,  $p = 0.653$ , CI  $[-0.016, 0.01]$ ,  $\eta_p^2 = 0.001$ ].

The original study<sup>13</sup> also reported that musicians had significantly greater stimulus-to-response correlations than non-musicians, suggesting that musicians' neural encoding of speech sounds was more robust to noise than that of non-musicians. We attempted to replicate this finding by conducting an independent-samples  $t$ -test on the z-transformed stimulus-to-response correlations between musicians and non-musicians. We found no musician advantage using our pre-planned analyses (Fig. 3a), with the average trend in the opposite direction to that predicted [ $t(160) = -0.716$ ,  $p = 0.763$ , CI  $[-0.013, \infty]$ ,



$d = -0.113$ ]. The Bayes factor was  $BF_{+0} = 0.106$  (error =  $-0.002$ ), meaning the data are 9.43 times more likely to come from the null than the alternative hypothesis, providing moderately strong support for the null (Supplementary Fig. 5). Controlling for age and gender did not change the interpretation of the results [ $F(1,156) = 0.538$ ,  $p = 0.464$ , CI  $[-0.007, 0.015]$ ,  $\eta_p^2 = 0.003$ ]. Our analysis followed the traditional approach of selecting the time-lag between stimulus and response that produced the greatest correlation, whereas the original study limited the stimulus lag times to the range of 6.9–10.9 ms prior to adjusting for site-specific delays between the onset of the trigger and the arrival time of the stimulus at the ear canal. An exploratory analysis using the original study's approach also showed no musician advantage [ $t(160) = 0.074$ ,  $p = 0.471$ , CI  $[-0.017, \infty]$ ,  $d = 0.012$ ] with moderate support for the null hypothesis ( $BF_{+0} = 0.18$ , error  $-0.04\%$ ). The effect of musicianship again remained non-significant after controlling for age and gender [ $F(1156) = 0.025$ ,  $p = 0.875$ , CI  $[-0.023, 0.02]$ ,  $\eta_p^2 = 0.0002$ ].

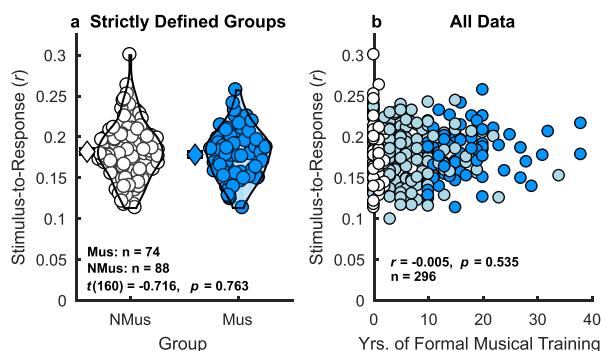
Next, we tested whether years of formal musical training, as measured across the entire cohort of participants, was correlated with enhanced neural encoding of /da/ in babble, quantified as: (1) the

strength of F0 encoding (Fig. 2b), (2) the strength of encoding the upper harmonics (Fig. 2d), and (3) the transformed stimulus-to-response correlation (Fig. 3b). None of these preregistered hypotheses were tested in the original study<sup>13</sup>, but they are extensions of the underlying hypothesis that musical training is associated with enhanced neural representation of speech in noisy backgrounds. A Bonferroni-corrected criterion for significance ( $\alpha = 0.017$ ) was preregistered. Years of formal musical training was not related to enhanced spectral encoding for the F0 ( $r = 0.018$ ,  $p = 0.381$ , CI  $[-0.078, 1]$ ; non-musician outlier excluded:  $r = 0.094$ ,  $p = 0.054$ , CI  $[-0.002, 1]$ , Supplementary Fig. 1b) or upper harmonics ( $r = 0.008$ ,  $p = 0.445$ , CI  $[-0.088, 1]$ ; non-musician outlier excluded:  $r = 0.048$ ,  $p = 0.204$ ,  $[-0.048, 1]$ , Supplementary Fig. 1d). Similarly, years of musical training did not counteract the interference effects of background noise on speech encoding, as quantified via the stimulus-to-response correlation ( $r = -0.005$ ,  $p = 0.535$ , CI  $[-0.101, 1]$ ). An exploratory analysis limiting the cross-correlations to the site-specific adjusted lag windows of 6.9–10.9 ms, as in the original study, also showed no relationship between the adjusted stimulus-to-response correlation and years of formal musical training ( $r = 0.029$ ,  $p = 0.312$ , CI  $[-0.067, 1]$ ). Further exploratory partial correlations controlling for age and gender confirmed no significant effects of years of musical training (Supplementary Fig. 6).

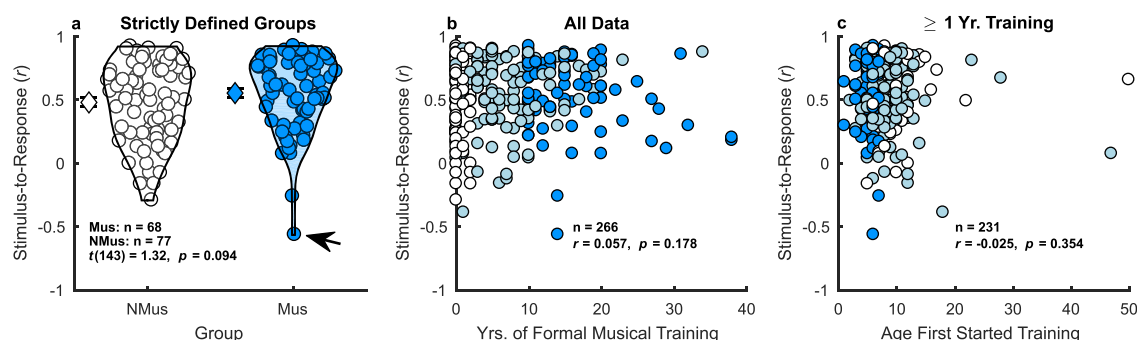
### Neural encoding of linguistic pitch contours

The original study<sup>15</sup> measured EEG responses while participants listened passively to the Mandarin word /mi3/ as they watched a silent video. The stimulus varied in F0 over time between 89 and 110 Hz (Fig. 1f). The original study<sup>15</sup> found that the F0 stimulus-to-response correlation (i.e., the Pearson correlation between the F0 contour of the stimulus and the neural response) was significantly greater in 10 musicians than in 10 non-musicians. We attempted to replicate this finding on the subset of our participants that met the strict criteria of musician and non-musician (68 musicians and 77 non-musicians; Fig. 4a) and found no evidence for a musician advantage [ $t(143) = 1.32$ ,  $p = 0.094$ , CI  $[-0.025, \infty]$ ,  $d = 0.22$ ], with the data about 1.4 times more likely to occur under the null than the alternative hypothesis ( $BF_{+0} = 0.715$ , error  $< 0.0001\%$ , Supplementary Fig. 7). The effect remained non-significant after excluding a musician-group outlier with the poorest neural encoding [ $t(142) = 1.63$ ,  $p = 0.052$ , CI  $[-0.002, \infty]$ ,  $d = 0.273$ ] but resulted in a Bayes factor very close to 1 ( $BF_{+0} = 1.14$ , error  $< 0.0001\%$ ), implying no strong evidence for either the null or the alternative hypothesis.

To determine if age and gender were masking any effect of musicianship on F0-tracking fidelity, an exploratory analysis tested for a musician advantage while adjusting for age and gender. There was



**Fig. 3 | Musical training is not related to more robust encoding of speech in background sounds.** This interpretation remained the same whether comparing encoding in strictly defined musicians vs. non-musicians [ $t(160) = -0.716$ ,  $p = 0.763$ , CI  $[-0.013, \infty]$ ,  $d = -0.113$ ] (a) or assessing encoding across the full cohort of participants, treating years of formal musical training as a continuous variable ( $r = -0.005$ ,  $p = 0.535$ , CI  $[-0.101, 1]$ , preregistered  $\alpha = 0.017$ ) (b). The untransformed stimulus-to-response correlations are plotted for visualization purposes; the z-transformed data were used in analyses. Statistical tests were one-tailed. Black outlines: 1D KDEs; Diamonds: Average data; Circles: Individual data; NMus: Non-musician (white); Mus: Musician (dark blue). Participants in neither of the strictly defined groups are shown in light blue. Error bars:  $\pm 1$  SEM.



**Fig. 4 | Neural encoding for time-varying F0 is unrelated to duration and age of onset of musical training.** No significant difference was observed between the strictly defined groups of musicians and non-musicians [ $t(143) = 1.32$ ,  $p = 0.094$ , CI  $[-0.025, \infty]$ ,  $d = 0.22$ ] (a). There was also no relationship between neural tracking of the F0 and years of formal musical training ( $r = 0.057$ ,  $p = 0.178$ , CI  $[-0.045, 1]$ ) (b) or age of onset of musical training ( $r = -0.025$ ,  $p = 0.354$ , CI  $[-1, 0.084]$ ) (c). In (b, c),

the preregistered criterion for significance included Bonferroni correction for two comparisons ( $\alpha = 0.025$ ). Statistical tests were one-tailed. Black outlines: 1D KDEs; Diamonds: Average data; Circles: Individual data; NMus: Non-musician (white); Mus: Musician (dark blue). Participants in neither of the strictly defined groups are shown in light blue; Arrow indicates the outlier. Error bars:  $\pm 1$  SEM.

still no benefit of musicianship on F0 tracking, either with [ $F(1,139)=1.29$ ,  $p=0.257$ , CI  $[-0.239, 0.064]$ ,  $\eta_p^2=.009$ ] or without [ $F(1,138)=2.09$ ,  $p=0.151$ , CI  $[-0.256, 0.04]$ ,  $\eta_p^2=0.015$ ] the musician outlier with the poorest encoding included.

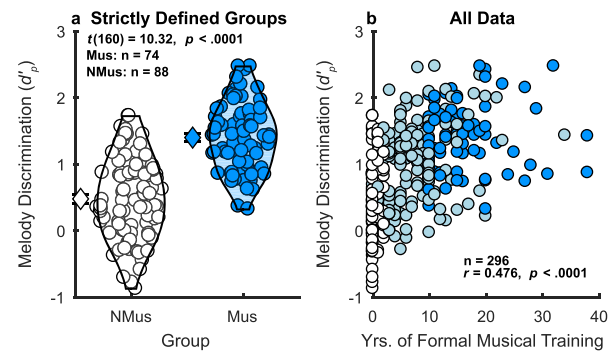
The original study<sup>15</sup> found that the age of onset of musical training as well as years of musical training were correlated with the fidelity of F0 tracking, assessed via the F0 stimulus-to-response correlation. We used the full cohort of participants (excluding those with 0 years of musical training) to test whether age of onset of training was related to F0 tracking (Fig. 4c). We also tested the hypothesis that F0 tracking improves with years of musical training by calculating the correlation between years of formal musical training and the F0 stimulus-to-response correlation (including those with no musical training, as in the original study<sup>15</sup>; Fig. 4b). The preregistered criterion for significance included Bonferroni correction for two comparisons ( $\alpha=0.025$ ). Unlike the original study, which included 16 participants, we found no relationship between the age of onset of musical training and the fidelity of neural encoding of time-varying stimulus F0 (Fig. 4c;  $r=-0.025$ ,  $p=0.354$ , CI  $[-1, 0.084]$ ), with the data 8.77 times more likely to occur under the null than the alternative hypothesis ( $BF_0=0.114$ ; Supplementary Fig. 8). Furthermore, the relationship between years of formal musical training and the fidelity of F0 tracking was not significant (Fig. 4b;  $r=0.057$ ,  $p=0.178$ , CI  $[-0.045, 1]$ ), with moderate evidence that the data originate under the null hypothesis ( $BF_{+0}=0.192$ ). Exploratory partial correlations controlling for age and gender (Bonferroni-corrected  $\alpha=0.025$ ; Supplementary Fig. 9) confirmed no relationship between age of onset of musical training and F0 stimulus-to-response correlations ( $r_p=-0.004$ ,  $p=0.476$ , CI  $[-0.125, 0.135]$ ) or years of formal musical training and F0 stimulus-to-response correlations ( $r_p=0.057$ ,  $p=0.178$ , CI  $[-0.057, 0.185]$ ). However, F0 tracking did worsen with age ( $r=-0.173$ ,  $p=0.002$ , CI  $[-1, -0.073]$ ), consistent with the expected degradation of spectral neural encoding with age<sup>37</sup>. There was no relationship between gender and F0 tracking [ $t(260)=0.901$ ,  $p=0.368$ , CI  $[-0.058, 0.156]$ ].

### Comparing neural responses between measures

Most participants ( $n=263$ ; Mus = 68; NMus = 74) completed both the /da/ test and the /mi3/ test and met all inclusion criteria for analyses (see Methods). We compared the strength of neural encoding between tests to examine whether neural tracking for the F0 of speech in quiet is related to FFRs for encoding for speech in babble (Supplementary Fig. 10). Exploratory analyses demonstrated that the stimulus-to-response correlation for encoding the F0 of /mi3/ in quiet was weakly related to F0 spectral encoding for /da/ in babble ( $r=0.133$ ,  $p=0.016$ , CI  $[0.032, 1]$ , Bonferroni-corrected  $\alpha=0.0167$ ), but this effect was driven by one outlier non-musician with very strong encoding for both measures (outlier removed:  $r=0.072$ ,  $p=0.122$ , CI  $[-0.03, 1]$ ). There was no association between neural encoding for /mi3/ in quiet and the upper harmonics of /da/ in babble ( $r=0.071$ ,  $p=0.125$ , CI  $[-0.031, 1]$ ; outlier removed:  $r=0.009$ ,  $p=0.442$ , CI  $[-0.093, 1]$ ), and there was still no association once one additional outlier musician, with poor F0 tracking but strong encoding for the upper harmonics, was removed ( $r=0.068$ ,  $p=0.138$ , CI  $[-0.035, 1]$ ). While measures between studies were generally unrelated to one another—potentially because dynamic F0 tracking is somewhat different from overall strength in spectral encoding, or because the stimuli were different—we did find a relationship between strength of spectral encoding for the F0 versus upper harmonics within the same stimulus (/da/ in babble:  $r=0.585$ ,  $p<0.0001$ , CI  $[0.514, 1]$ ; outlier non-musician removed:  $r=0.3$ ,  $p<0.0001$ , CI  $[0.204, 1]$ ).

### Does musical ability account for failures to replicate?

Our criteria for the groups of “musician” and “non-musician” were at least as strict as in both original studies, so the failure to replicate any

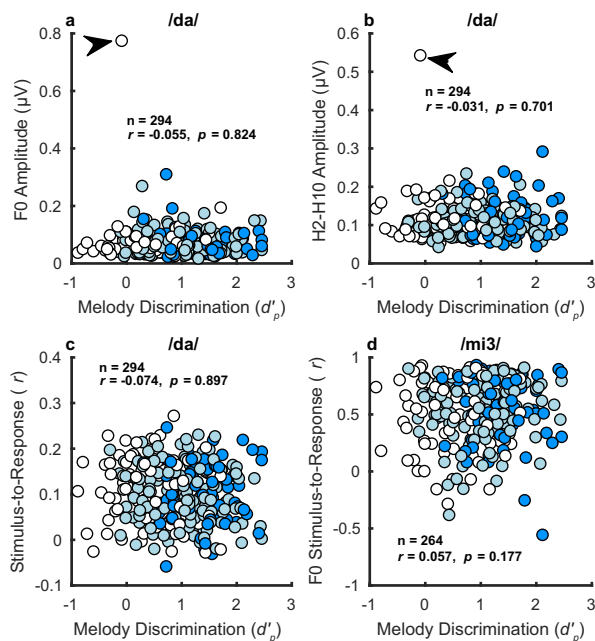


**Fig. 5 | Objective measure of musical ability is related to self-reported musical training.** Musicians had significantly better melody discrimination than non-musicians [ $t(160)=10.32$ ,  $p<0.0001$ , one-tailed test, CI  $[0.774, \infty]$ ,  $d=1.63$ ] (a), and this finding generalized across the full cohort of participants when treating years of formal musical training as a continuous variable ( $r=0.476$ ,  $p<0.0001$ , one-tailed test, CI  $[0.399, 1]$ ) (b). Black outlines: 1D KDEs; Diamonds: Average data; Circles: Individual data; NMus: Non-musician (white); Mus: Musician (dark blue). Participants in neither of the strictly defined groups are shown in light blue. Error bars:  $\pm 1$  SEM.

musician advantages in neural encoding is unlikely to be due to our group definitions. However, musical expertise or aptitude, rather than years of training, may be a more sensitive measure for detecting differences in neural encoding of sound<sup>40</sup>. We tested this hypothesis directly by correlating an objective measure of musical ability (same/different melody discrimination)<sup>41</sup> with our four measures of neural encoding fidelity: (1) spectral encoding for the F0, (2) spectral encoding of the upper harmonics for /da/ in babble, (3) stimulus-to-response correlations for /da/, and (4) F0 tracking for /mi3/ in quiet. Consistent with prior work<sup>41</sup>, we found that musicians had better melody discrimination than non-musicians [ $t(160)=10.3$ ,  $p<0.0001$ , CI  $[0.774, \infty]$ ,  $d=1.63$ ], and this relationship generalized across the full sample of participants when treating years of formal musical training as a continuous variable ( $r=0.476$ ,  $p<0.0001$ , CI  $[0.399, 1]$ ; Fig. 5). However, there was no relationship between the objective measure of musical ability and any of the four measures of sound neural encoding (Fig. 6; see Supplementary Fig. 11 for results with the non-musician outlier excluded).

### Redefining the noise floor: further exploratory analyses

Using the analysis methods outlined in the previous studies, we were unable to replicate any of the preregistered musician advantages in sound neural encoding. One potential limitation of the present study pertains to the definitions of the noise floor (see “Exclusion criteria” section). All prior analyses used the same definitions as the original studies: For the /mi3/ measure, the spectral signal-to-noise ratio (SNR) of the F0 candidate in each windowed bin of the stimulus response had to be  $>1$  (see “Calculating F0 tracking” subsection) and for the /da/ measure there was no SNR exclusion criterion. It is possible these exclusion criteria may not have been stringent enough. While there are many possible ways to differentiate an FFR signal from noise, we chose to use the SNR between the root-mean-square (RMS) amplitude of the neural response to the stimulus and the RMS amplitude of the pre-stimulus period, where no sound was presented, as described in a previous study<sup>25</sup>. Our rationale was that the neural activity when sound is presented should be larger than during silence, and if not, then the neural response is unlikely to primarily reflect stimulus encoding. Using this measure, an SNR = 1 would mean the neural activity during the stimulus is the same magnitude as during silence. All but five participants in the /da/ study and all the participants of the /mi3/ study had RMS amplitude SNRs of 1 or more. However,  $n=54$  /da/ and  $n=21$  /mi3/ participants had SNRs  $<1.5$ , which is a more conservative noise



**Fig. 6 | Musical ability is not related to sound encoding as measured with the scalp-recorded FFR.** Musical ability is unrelated to sound neural encoding for /da/ in babble (a–c) or /mi3/ in quiet (d). Musical ability was assessed using an objective measure of melody discrimination<sup>41</sup> and quantified using a non-parametric estimate of sensitivity,  $d'_p$ <sup>62,63</sup>, where higher  $d'_p$  indicates better performance (a:  $r = -0.055$ ,  $p = 0.824$ , CI  $[-0.15, 1]$ ; b:  $r = -0.031$ ,  $p = 0.701$ , CI  $[-0.127, 1]$ ; c:  $r = -0.074$ ,  $p = 0.897$ , CI  $[-0.196, 1]$ ; d:  $r = 0.057$ ,  $p = 0.177$ , CI  $[-0.044, 1]$ ). All tests were one-tailed and Bonferroni-corrected for four comparisons ( $\alpha = 0.0125$ ). Note that the y-axis in (c) is the adjusted stimulus-to-response correlation. In (c, d), analyses were conducted on the z-transformed stimulus-to-response correlations. White circles: Non-musicians; Dark blue circles: Musicians; Light blue circles: Participants in neither of the strictly defined groups; Arrow points to outliers.

floor criterion proposed previously<sup>25</sup>. We reanalyzed all preregistered analyses, both direct replication and correlational, using this more conservative noise-floor criterion, with participants with SNRs below the noise-floor criterion excluded. For the /mi3/ measure, we used this criterion in addition to the spectral noise floor criterion of the original study. Even with this more conservative noise floor criterion for inclusion, there were still no significant effects of neural encoding related to musical training (Supplementary Figs. 12–16; Supplementary Texts 1 and 2).

## Discussion

Our large-scale replication and extension tested the robustness and generalizability of the widely cited claim that musicians have enhanced neural encoding of sound, as assessed via EEG using the FFR—an electrophysiological index of the fidelity of neural encoding in the early stages of the auditory pathways<sup>26</sup>. The results were consistent across all five direct replication analyses (Table 1): There were no significant musician advantages, despite using the same stimuli, methods, and analyses as in the original studies. Musicians did not exhibit stronger sound spectral encoding for the upper harmonics of speech in background sounds, nor did they demonstrate enhanced tracking for dynamic changes in linguistic F0 for speech sounds in quiet. Neural encoding in musicians was also not more robust to the effects of background noise than that of non-musicians. An extension of the analyses to include years of formal musical training as a continuous variable, which increased the statistical power and the generalizability of the results, also revealed no relationship between musical training and any of our preregistered measures of neural encoding of sound. The conclusions remained the same under additional exploratory

analyses that used a more stringent criterion for the noise floor than the original studies.

A separate set of exploratory analyses, controlling for potential effects of age and gender, confirmed no significant effects of musical training on neural encoding of sound. The lack of an effect of gender in the present study is at odds with prior work showing stronger spectral encoding for the upper harmonics in females than males<sup>38</sup>, which may change with changes in estrogen levels<sup>42</sup>. The lack of an interaction between age and musical training is especially important, as musical training has been proposed to potentially counteract the age-related declines of the fidelity of neural encoding of sound<sup>4,30</sup>, and even speech perception in background noise<sup>34</sup>. Our results confirmed that spectral encoding tends to degrade with age, but that this effect appears to be larger and more consistent for the encoding of the F0 than for the encoding of the upper harmonics. Our findings suggest that poorer FFR spectral encoding with age may not be offset by musical experience, although this question can only be fully answered using large-scale, longitudinal designs with random assignment to music training versus a non-music, active control task. There have been some longitudinal studies examining the effects of various types of training on FFR encoding<sup>4,43–45</sup>, but they have been limited in scope and sample size, with mixed results, suggesting a need for high-powered samples with preregistered analyses and appropriate control groups in any future research on this topic.

A previous review of perceptual and neural associations with musical training noted that conclusions were mixed with respect to behavioral outcomes in speech-in-noise perception but that all studies examining neural advantages in musicians had reported at least one significant effect<sup>26</sup>. Given the seeming robustness of the results, it may be surprising that our large-scale study failed to replicate any of the original findings. This apparent discrepancy has several possible underlying causes. First, the positive results have come from a relatively limited number of laboratories, and different measures have been used to demonstrate the musician advantage in the FFR, even when the same stimuli were used (see Supplementary Table 1). For example, one study<sup>14</sup>, using the same /da/ stimulus as in the present study, found a musician encoding advantage in quiet, but only for the F0 and not the upper harmonics (see Supplementary Table 1 for notes on specific methodological differences, as this study compared F0 encoding in predictable vs. variable contexts), whereas another study using the same stimulus<sup>35</sup> reported a musician advantage for the upper harmonics, but not the F0, both with and without background noise. Two studies<sup>11,12</sup> using a longer-duration /da/ stimulus (350 ms; no background sound) at the same F0 (100 Hz) found musicians had enhanced encoding for the F0 but not the upper harmonics. Thus, while every study cited in the previous review<sup>26</sup> (and listed in Supplementary Table 1) found at least one significant neural musician advantage, the nature of the advantage has varied between studies. In small sample studies, even seemingly minor alterations to analysis choices used can create vibration effects, whereby the magnitude of the observed effect can change substantially with subtle changes to the analyses<sup>46</sup>. There is still scientific value to conducting exploratory EEG analyses, but the interpretation of such findings should be tempered and considered not just in the context of the previous results, but also in the context of whether the analysis pipelines differ in seemingly innocuous but potentially conclusion-altering ways.

A second, related possible explanation for why we did not replicate previous findings is that the sample sizes from past studies in this field of research have all been relatively small ( $n < 30$  per group). Small samples have low power (and increased Type II error rates), which, as a related consequence, means that when significant effects are reported, they are less likely to reflect true effects, and the reported effect sizes are more likely to be inflated<sup>46,47</sup>. Finally, and perhaps most importantly, there is no single agreed-upon analytical



**Table 1 | Key findings selected for direct replication, along with the effect sizes from the present study**

Reported musician advantage	Publication	Sample size		# of citations	Original effect size	Current study effect size
		Mus	NMus			
Enhanced spectral encoding of upper harmonics to /da/ in babble.	Parbery-Clark et al. <sup>13</sup>	<i>n</i> = 74 (16)	<i>n</i> = 88 (15)	527	<i>d</i> = 1.04	<i>d</i> = 0.023
Less degradative effects of babble on encoding speech (/da/).	Parbery-Clark et al. <sup>13</sup>	<i>n</i> = 74 (16)	<i>n</i> = 88 (15)	527	<i>d</i> = 1.05	<i>d</i> = 0.012
Enhanced F0 encoding for linguistic pitch in quiet.	Wong et al. <sup>15</sup>	<i>n</i> = 68 (10)	<i>n</i> = 77 (10)	1377	<i>d</i> = 1.11	<i>d</i> = 0.22
Years of musical training related to higher fidelity of neural encoding for dynamic changes in F0.	Wong et al. <sup>15</sup>	<i>n</i> = 266 (20)		1377	<i>r</i> = 0.456	<i>r</i> = 0.05
Starting a musical instrument at an earlier age related to enhanced F0 tracking.	Wong et al. <sup>15</sup>	<i>n</i> = 231 (16)		1377	<i>r</i> = −0.502	<i>r</i> = −0.025

Cohen's *d* was estimated directly from the original studies using the reported group means and standard deviations, if available; otherwise, *d* was calculated from the reported *t* value and degrees of freedom. The number of citations listed for each paper corresponds to Google Scholar on June 23rd, 2025. All five direct replication analyses failed to replicate. *n*: Current study sample sizes; Mus = musicians; NMus = non-musicians; Parentheses: Original study sample size.

technique for examining the fidelity of sound encoding using EEG. Researchers thus have many degrees of freedom related to the number of ways they can test their hypotheses<sup>48</sup>, and not all of the analyses may be reported in the final publication. Testing the same hypothesis in many different ways increases the chance of a false positive (Type I error), highlighting the importance of preregistration of hypotheses, methods, and planned analyses prior to data collection.

One alternative reason for the failure to replicate could be related to the reliability of the FFR measures used. If test-retest reliability were low, then any relationship between the FFR and any other measures, including musical training, may be obscured by measurement noise. Reliability issues are unlikely to account for the null effects in this study for several reasons: First, we were able to detect the expected decrease in FFR with age, as shown by prior studies<sup>37</sup>, demonstrating that the FFR is correlated with at least one measure. Second, prior to beginning data collection, we measured FFRs in both studies from the same pilot participant at all six sites and UMN twice. Reliability was high (Supplementary Figs. 17 and 18; Supplementary Tables 2,3), ranging from  $0.918 \leq r \leq 0.955$  for the /da/ stimulus and  $0.647 \leq r \leq 0.893$  for /mi3/. Importantly, the within-site reliability was within the range of the between-site reliability, meaning that noise due to between-site differences is unlikely to account for the failure to replicate. These pilot analyses are consistent with prior research suggesting the FFR is highly stable upon test-retest<sup>49</sup> and generalizes this finding across sites. Confirming this expected consistency, we conducted split-half reliability analyses between the first and second half of each individual's recordings in the full dataset, and found adequate reliability (Supplementary Figs. 19 and 20), with values of  $r = 0.634$  for /mi3/ and  $0.858 \leq r \leq 0.951$  for /da/.

Another alternative reason for the failure to replicate is that the present study may have one or more false negatives (Type II errors), despite the large *n* and corresponding sensitivity to small effect sizes. In fact, all five direct replication analyses demonstrate small but non-significant effects in the expected direction. Based on the small effect sizes reported in Table 1, if a musician advantage is present in the full population of musicians and non-musicians, then the effects seem likely to be too small to meaningfully influence perception and behavior on an individual level. While traditional frequentist statistics can tell us whether we fail to reject the null hypothesis, Bayes factor (BF) analyses can assess evidence for the null hypothesis. Our BF analyses generally provided moderate evidence that the data originated under the null hypothesis for all direct replication analyses, with the exception of the test comparing F0 tracking in musicians versus non-musicians. But even here, there was no clear evidence for either the alternative or the null hypothesis.

Lastly, it is possible that we selected the wrong musician advantages to test. For example, a number of studies have claimed that musicians exhibit stronger and shorter-latency responses relative to non-musicians<sup>11–13,30,35</sup>. The majority of studies examining response latencies used an expert peak picker to manually identify the peaks in the early neural responses. It would be difficult to directly replicate such analyses, as different findings between studies could be related to differences in the expertise or strategy of the peak picker(s), making a null finding difficult to interpret. Future analyses of our publicly available data could involve efforts to automate the process of peak-picking<sup>50</sup> or formally assess the reproducibility of outcomes across different peak pickers<sup>51</sup>.

Our large-scale study tested the replicability and generalizability of the widely cited finding that musical training is related to enhanced early neural encoding of sound. Our results do not support this claim. While the neural locus of the FFR is debated, and may contain contributions from cortical sources<sup>18,22,23,52</sup>, especially when measured with MEG<sup>22,23</sup>, the evidence suggests that the dominant sources of EEG measurements are subcortical for frequencies greater than 100 Hz, and originate primarily in the inferior colliculus<sup>53–55</sup>. One important direct implication is that subcortical neural structures that produce phase-locked responses to sound may have limited plasticity, even following many years of intensive musical training starting at an early age. There is some evidence, however, for associations between cortical structure and function and musical training<sup>56</sup>, although this claim is also not without some controversy<sup>57</sup>. In either case, most studies so far have been cross-sectional, making it difficult to determine whether any differences are causally related to the musical training<sup>2</sup>. Future studies, using large-sample longitudinal designs with participants randomly assigned to a musical training intervention or a non-music, active control, will be important for determining whether cortical differences in sound neural encoding are caused by music training, due to pre-existing differences, or a combination of the two.

Using sample sizes that were more than four times those of the original studies, with preregistered methods and analyses, and data collected in six laboratories, we showed that the widely cited finding that musicians have enhanced subcortical responses to sound failed to replicate. In an extension of the original studies, we also found no relationship between the fidelity of neural encoding in the sound-evoked FFR and years of formal musical training. Further exploratory analyses showed that musical training did not offset the age-related deterioration in the spectral encoding of sound. In all, none of our replication, extended, or exploratory analyses provided support for a relationship between musical training and sound neural encoding via the scalp-recorded FFR.

Musical training is not related to an enhancement in early neural encoding of sound as assessed via the scalp-recorded FFR.

Nevertheless, there are, of course, many important reasons why learning and playing music remains a valuable endeavor, including social connection, emotional regulation, or simply the enjoyment of music for its own sake<sup>58</sup>.

Methods

Recruitment and eligibility

Only participants who completed the full online screening, did not report encountering any audio issues, passed the auditory attention check, and indicated they would like to be contacted to participate in future lab studies were eligible for the laboratory portion of the study. Occasionally, a potential participant met the recruitment criteria except that they reported audio issues, in which case, the researcher could invite them to redo the melody portion of the online screening in the lab. Online participants who reported a history of hearing loss (unless they were age 40 or older, in which case some high-frequency hearing loss was allowed; see Table 2), neurological conditions, proficiency in a tonal language (such as Mandarin or Cantonese), or who were not native speakers of North American English (i.e., did not live from birth through age 5 in a household where North American English was the primary spoken language) were ineligible for the laboratory portion of the study.

To ensure an adequate representation across age ranges, participants for the full lab study were recruited at each site to be roughly evenly distributed in each decade of age (20s–60s, with ages 18 and 19 grouped in the 20s decade), based on the age reported in the online screening. Each site aimed to recruit 60 participants, with at least 25% of participants with no more than 2 years of any musical training and no ongoing music performance activities (i.e., they reported that they did not currently play a musical instrument, including voice), and at least 25% of participants who started playing their first musical instrument or voice by the age of 7, had completed at least 10 years of formal musical training, and reported that they currently played a musical instrument. Formal musical training was defined as group or private lessons, excluding standard elementary school activities. The remaining participants had varying amounts of formal musical training. In this way, we could assess the effect of the number of years of formal musical training as a continuous variable, while still being able to perform dichotomous comparisons (musician vs. non-musician) with at least half of our overall sample. Sites aimed to have roughly even and uniform distribution of ages and gender between these two groups, as with all other participants.

Participants

All participants in the present study also took part in a number of behavioral tests in the lab (not reported here). Participants recruited for the in-person portion of the study underwent a pure-tone audiometric screening at octave frequencies between 125 and 8000 Hz, as well as at 6000 Hz. Because age and high-frequency hearing loss covary<sup>59,60</sup>, the maximum allowable hearing loss was titrated per decade, so that participants under the age of 40 were required to have audiometric thresholds  $\leq 20$  dB hearing level (HL) across all tested frequencies, but older adults could have more high-frequency loss (specified in Table 2). All participants were required to meet the

audiometric criteria in both ears to participate. A total of 296 participants (115 male, 177 female, 4 who did not identify as male or female), including 74 musicians and 88 non-musicians (as defined in the “Recruitment and eligibility” section), completed the syllable-in-noise study, and 295 completed the linguistic pitch study. Thirty of the participants for the linguistic pitch study did not meet the pre-specified criterion for analyses (i.e., at least one binned fast Fourier transform (FFT) analysis was in the noise floor, as defined by Wong et al.<sup>15</sup>) or were unable to return to the lab to redo the study, so their data was excluded from this task, leaving a total of 265 participants (106 male; 155 female; 4 non-binary). Participant age ranged from 18 to 69 years for both studies. Most participants completed both studies; reasons for missing data are described in each site’s corresponding EEG log (<https://doi.org/10.7910/DVN/SKOWNK>)<sup>61</sup>. The total number of participants per site for each measure, including their musical status, is provided in Supplementary Table 4.

All participants provided written informed consent and were given monetary compensation or course credit for their in-person participation. All study protocols were approved by the Institutional Review Board at the corresponding university site prior to any data collection: Boston University (4942E), Carnegie Mellon University (STUDY2018\_00000367), Purdue University (I609018209), University of Minnesota (0605S85872 and 1306S37081), University of Rochester (STUDY00004020), and by the Nonmedical Research Ethics Board of the University of Western Ontario (NMREB 112604).

Online screening

Before completing the laboratory portion of the study, all participants remotely completed an initial online screening. The purpose of the screening was to aid in recruitment of qualified participants for the lab-based portion of the study (e.g., based on age, years of formal musical training, etc.), acquire an objective measure of melody perception abilities, and obtain survey information on factors that may co-vary with musicianship (e.g., personality) for use in possible exploratory analyses.

The online screening was administered through Qualtrics, with each site completing recruitment and online data collection under the purview of its own Institutional Review Board. All screening participants provided informed consent online. Participants were not compensated for participation in the screening, but they had the option to enroll in a drawing for a chance to win a gift card as an incentive for participating. All personal identifiers were removed from online data before sharing between sites, so that only the subject ID number linked the online to the laboratory data. The online measures are described below in the order in which they appeared.

**Age.** Participants were asked to select their age from a drop-down menu. A reported age of younger than 18 or older than 89 led to termination of the screening. Only those aged 69 or younger were invited to participate in the EEG studies reported here.

**Level adjustment.** To help ensure that the stimuli for the online listening tasks were audible but not too loud, participants were presented with noise and instructed to adjust their volume so that it was at

Table 2 | Audiometric threshold criteria

Age group	0.125–1 kHz	2 kHz	4 kHz	6 kHz	8 kHz
18–39	$\leq 20$ dB HL	$\leq 20$ dB HL	$\leq 20$ dB HL	$\leq 20$ dB HL	$\leq 20$ dB HL
40–49	$\leq 20$ dB HL	$\leq 20$ dB HL	$\leq 20$ dB HL	$\leq 30$ dB HL	$\leq 30$ dB HL
50–59	$\leq 20$ dB HL	$\leq 25$ dB HL	$\leq 30$ dB HL	$\leq 40$ dB HL	$\leq 40$ dB HL
60–69	$\leq 20$ dB HL	$\leq 30$ dB HL	$\leq 40$ dB HL	–	–

All participants who completed the lab portion of the study met the audiometric threshold criteria specified for their age group in each ear.



an audible but comfortable level. The stimulus was white noise, bandpass filtered between 200 and 1000 Hz, so that the frequency spectrum was comparable to that used in the melody task.

**Attention check.** This task helped exclude participants who were not attending or who did not have properly functioning audio on their devices. To pass the attention check, participants were required to answer at least 3 of 4 trials correctly. Each trial consisted of a short sequence of 1-kHz pure tones. Each tone within a sequence was 400 ms in duration with 50-ms raised-cosine onset and offset ramps, and each tone was separated by 500 ms of silence. Participants were instructed that each trial contains between 0 and 9 tones, and their task was to report the number of tones they heard by selecting the corresponding number from a drop-down menu. Because there were 10 options for each trial, the probability of passing the screening by chance was very low (0.0037). To minimize the duration of the task, each trial only had 1, 2, 3, or 4 tones in a sequence, with each tone-sequence option presented once. Trials had a fixed duration of 5 s, so that the entire task could be completed in less than half a minute. No feedback was provided.

**Melody discrimination.** Stimuli were from the Melody subtest of the Full Profile of Music Perception Skills, with methods as described in Law and Zentner<sup>41</sup>. During each trial, participants first heard a reference melody twice in a row, followed by a comparison melody. The task was to determine whether the comparison melody was the same as or different from the reference melody, with participants selecting their answer from five possible options: “Definitely Same,” “Probably Same,” “I Don’t Know,” “Probably Different,” or “Definitely Different.” One practice trial was provided, followed by 18 data trials. Participants did not receive feedback, but they did receive their total composite score at the end of the task. The composite score provided to participants was calculated using weighted responses as described in Law and Zentner<sup>41</sup>, with confident correct responses (“Definitely Same” or “Definitely Different”) receiving 1 point, less-confident correct responses (“Probably Same” or “Probably Different”) receiving 0.5 point, and incorrect responses or “I Don’t Know” receiving 0 points. For analyses, melody discrimination performance was calculated using a bias-free estimate of sensitivity,  $d'_p$ , as recommended by Strauss et al.<sup>62</sup> and Whiteford et al.<sup>63</sup>.

**Survey questions.** A number of survey questions assessed factors related to demographics, musical engagement, socio-economic status, and hearing status. Self-report of any audio issues during the melody task was also gathered. The full set of questions is available at <https://osf.io/duq34/><sup>36</sup>.

**Big Five Personality Inventory.** The 44-item Big Five Personality Inventory was administered to assess personality<sup>64,65</sup>. This was a self-report questionnaire, where each item was rated on a 5-point response scale.

### Stimuli and procedures

Each study was designed with the purpose of measuring the same effect *in principle* as the original study. In some instances, small methodological deviations were employed for practical reasons, including to decrease the total time of the study or to increase the feasibility of the measure to be collected consistently at multiple sites.

All sites ran the tests in the same order as described below. Tests that were skipped or needed to be rerun due to technical or researcher error were noted in a study log and, whenever possible, the participant returned to complete tests with missing data. Other post-registration changes to the methods are described in Supplementary Text 3.

**EEG: syllable in noise.** To assess the fidelity of neural encoding for speech in noise, we measured EEG responses to the speech syllable

/da/ (Fig. 1, top row) in multi-talker babble. The /da/ had a 100-Hz F0 with 170-ms duration, presented at 80 dB SPL, as used by Parbery-Clark et al.<sup>13</sup>. Both the /da/ and multi-talker babble were generously shared by the principal investigator of the original study. The multi-talker babble had a 37.41-s duration (not 45 s, as mistakenly reported in the original study<sup>13</sup>) and looped continuously throughout the task at 10 dB below the level of the speech syllable. Participants listened passively to /da/ in multi-talker babble over insert earphones (see Table 3) while watching a silent video in a sound-attenuating booth. The /da/ was presented at alternating polarities over two blocks of 3000 trials each (6000 trials total), with an ISI of about 83 ms, so that each block lasted ~13 min. Participants were allowed short breaks between blocks and instructed to remain still during stimulus presentation. Data were acquired with the electrode systems and sampling rates listed in Table 3 with earlobe references. Sites with BioSemi systems ensured that the magnitude of the offset voltages were  $< \pm 30$  mV before beginning data collection.

**EEG: linguistic pitch.** The Mandarin word /mi3/ (the 3 denoting a dipping tone; Fig. 1b, f), which means “rice,” was presented bilaterally at 70 dB SPL. Methods were adapted from Wong et al.<sup>15</sup>, and the stimulus was generously provided by the original authors. The word was originally recorded by a native Mandarin speaker and then adjusted in Praat<sup>66</sup> to have a duration of 278.5 ms and an F0 contour ranging from 89 to 110 Hz. Participants passively listened to 4800 repetitions of /mi3/, divided into two blocks of 2400 trials and presented at alternating polarities (2400 trials per polarity). The ISI was about 83 ms, so that the entire task had a duration of about 30 min (15 min per block). The same electrode system, sampling rate, and active and reference electrodes were used as in the syllable-in-noise task.

### Hardware, software, and materials

With the exception of the online screening, all auditory stimuli were controlled via MATLAB (R2016b). Code for tests and analyses is available on GitHub<sup>57,68</sup> and linked to the project website on the Open Science Framework (OSF; <https://osf.io/duq34/>)<sup>36</sup>. The stimuli for the EEG studies may be available upon reasonable request by contacting the authors of the original studies<sup>13,15</sup>. The principal investigator from Law and Zentner<sup>41</sup> should be contacted for requests to use the melody discrimination stimuli. Table 3 shows the equipment used at each test site.

### Sharing of data

Each site was responsible for quality checking its data before sharing it with the first author to ensure it was formatted in a manner consistent with the other sites. Each site was also responsible for maintaining a detailed log of each EEG session, including explanations for missing data and technical issues.

### Exclusion criteria

Only those who met the criteria listed in the “Recruitment eligibility” and “Participants” sections took part in the study. Missing data from one or more tasks (e.g., from dropping out of the study or researcher error) resulted in exclusion from the corresponding analyses; whenever possible, the participant was rerun on tests with missing data.

EEG data were excluded if there were less than 60% usable trials for any reason, such as a reduced number of stimulus presentations due to technical issues, researcher error, or an excessive number of artifacts. If a participant did not have enough usable stimulus presentations, they were rerun on the corresponding task whenever possible. EEG data from the /mi3/ test were excluded if the spectral SNR was too poor to estimate F0 tracking in the sliding FFT analysis. This occurred if the spectral magnitude of the EEG response for all frequencies was within the noise floor, as defined by the original study (see “Calculating F0 tracking” section, below). There was no

Table 3 | Equipment used at each data-collection site

Site	Boston University (BU)	Carnegie Mellon University (CMU)	Purdue University (PU)	University of Minnesota (UMN)	University of Rochester (UR)	University of Western Ontario (UWO)
EEG System	Biosemi ActiveTwo	Biosemi ActiveTwo	Biosemi ActiveTwo	Biosemi ActiveTwo	BrainVision ActiChamp with EP-Preamps	Biosemi ActiveTwo
EEG Sampling Rate	16.384 kHz	16.384 kHz	16.384 kHz	16.384 kHz	25 kHz	16.384 kHz
EEG Earphones	Etymotic ER3C	Etymotic ER2	Etymotic ER2	Etymotic ER1	Etymotic ER2	Etymotic ER2
# of EEG Channels	32	32	32	32	2	16
EEG Stimulus Delivery System	RME Fireface UCX	RME Fireface UCX	TDT S3 (RZ6)	TDT S3 (RP2)	RME Babyface Pro	RME Fireface 400
EEG Stimulus Presentation Sampling Rate	44.1 kHz	44.1 kHz	24.414 kHz	24.414 kHz	48 kHz	48 kHz
Other Notes	Forehead ground. EEG channels correspond to Cz-Left Auricular and Cz-Right Auricular.					

preregistered SNR criterion for exclusion for the /da/ measure, as in the original study.

**Exploratory noise floor exclusion criteria.** The exploratory noise floor exclusion criterion was defined as RMS amplitude SNRs  $\leq 1.5$ , as described in Skoe and Kraus<sup>25</sup>, where the SNR was calculated by taking the ratio between the RMS amplitude of the entire average stimulus response (after summing across polarities) and the RMS amplitude of the pre-stimulus period (after summing across polarities). All participants with SNRs  $\leq 1.5$  were excluded from the corresponding exploratory analyses (Supplementary Figs. 12–16, Supplementary Texts 1 and 2).

EEG data analyses: syllable in noise

All single-channel preprocessing and analyses were the same as reported in Parbery-Clark et al.<sup>13</sup>, unless otherwise stated. The recordings from one site (UR) were down-sampled so that EEG data from all sites had the same sampling rate; this was not done in the original study but was necessary due to equipment differences. All recordings were bandpass filtered between 70 and 2000 Hz (12 dB/octave with zero-phase shift) and then epoched from –40 to 213 ms, where 0 ms corresponds to the stimulus onset. Trials with activity  $\geq \pm 35 \mu\text{V}$  were treated as artifacts and removed from analyses. Epochs were baseline-corrected based on the mean potential in the pre-stimulus period. The average response was taken across the trials at each polarity, and this average was summed across polarities to minimize the stimulus artifact and the cochlear microphonic<sup>25,69</sup>.

**Between- and within-site reliability.** Prior to beginning data collection, we assessed the between-site reliability by recording the study coordinator’s FFRs at all six sites. For comparison, within-site data were collected on the same pilot subject at UMN twice on two separate days. Reliability was quantitatively assessed using cross-correlation, where each site’s FFR (after summing the average FFR across the two polarities) was correlated with a time-shifted version of another site’s FFR for all possible lag times. The reliability estimate corresponded to the maximum cross-correlation value. All reliability analyses were previously reported in our preregistration. Neural responses in both the time and frequency domains demonstrate high between-site consistency (Supplementary Fig. 16), despite site differences in study staff and equipment. The maximum cross-correlation between sites was high, ranging from  $0.918 \leq r \leq 0.955$  (Supplementary Table 2). The within-site cross-correlation was of a comparable magnitude to the between-site cross-correlations ( $r = 0.945$ ), confirming that the /da/-evoked FFR has high test-retest reliability when measuring the same participant across multiple sessions<sup>49</sup>, even with different equipment and study staff.

**Spectral encoding.** The musician’s advantage for enhanced spectral encoding to the vowel portion of /da/ in babble was assessed using the same FFT analysis methods as the original study<sup>13</sup>. An FFT of the steady-state portion of the EEG response (60–180 ms) was calculated for each subject, with zero padding added. The strength of harmonic encoding for the first 10 harmonics (with the first harmonic corresponding to the F0) was estimated by calculating the average spectral amplitude within 60-Hz-wide frequency bins that were centered around each harmonic. To estimate the overall strength of encoding of the upper harmonics for each subject, the average spectral amplitudes for harmonics 2–10 were summed.

**Stimulus-to-response correlations.** The preregistered stimulus-to-response correlation analysis assessed cross-correlations at all possible lag times (slightly different from the original study<sup>13</sup>, which limited the stimulus lag to 8–12 ms) between the vowel portion of the /da/

stimulus (50–170 ms) without multi-talker babble and the steady-state portion of the neural response (60–180 ms), defined in the same manner as used in the FFT analysis. The maximum correlation across all lag times is referred to as the stimulus-to-response correlation, with stimulus-to-response correlations closer to 0 indicating poorer neural representations of /da/. Because correlations do not adhere to assumptions of normality, they were transformed using Fisher's *r*-to-*z* transformation before conducting analyses; this transformation was not used by the original study<sup>13</sup>.

An exploratory analysis was also conducted to more precisely match the lag times tested in Parbery-Clark et al.<sup>13</sup> and to account for any fixed delay between the onset of the trigger and the arrival time of the stimulus at the ear canal, which varied between sites due to equipment differences. This included accounting for any fixed delay between the onset of the trigger and the onset of the stimulus (e.g., due digital-to-analog conversion time) as well as the time it takes the stimulus to travel the length of the earphone tubes. The stimulus lag used in the cross-correlation analysis by the original study was 8–12 ms, which included a 1.1 ms fixed delay. We therefore limited the lag time to 6.9–10.9 ms for all sites, and then added each individual site's fixed delay time to quantify the site-specific lag window. The cross-correlation was conducted on the neural response to the vowel (50–200 ms) and the zero-padded vowel portion of the stimulus within each site-specific lag window. The maximum correlation within the site-specific lag window for each subject is referred to as the adjusted stimulus-to-response correlation.

### EEG data analysis: linguistic pitch

Recordings were bandpass filtered between 80 and 1000 Hz (12 dB/octave using zero-phase shift) and then epoched from –45 to 295 ms, where 0 ms corresponds to the stimulus onset. All other data preprocessing was as described for the syllable-in-noise task. The F0-tracking analysis used by Wong et al.<sup>15</sup> was chosen for replication because it demonstrated the most consistent evidence of a musician advantage. Preprocessing and analyses were consistent with the original study unless otherwise stated.

**Between- and within-site reliability.** Prior to beginning data collection, the study coordinator was run on the /mi3/ measure at each site to examine between-site consistency, as well as UMN twice, on two separate days, as an estimate of within-site reliability (Supplementary Fig. 17). Reliability was calculated by cross-correlating the entire FFR response with the time-shifted FFR from another session, as described for the /da/ measure. The peak cross-correlation between test-retest at UMN was  $r = 0.838$ , falling within the range of peak cross-correlations between sites ( $0.674 \leq r \leq 0.893$ ; Supplementary Table 3). The magnitude of between- and within-site reliability for the /mi3/ stimulus was slightly smaller than the /da/ stimulus, perhaps reflecting the more dynamic nature of the stimulus F0.

**Calculating F0 tracking.** The strength of F0 tracking was estimated by performing a sliding FFT analysis on the EEG response for each subject over the entire FFR period, after accounting for any fixed delay between the onset of the trigger and the arrival time of the stimulus at the ear canal. This included accounting for any fixed delay between the onset of the trigger and the onset of the stimulus (e.g., due digital-to-analog conversion time) as well as the time it takes the stimulus to travel the length of the earphone tubes. The time-averaged EEG response was segmented into 40-ms bins, with each bin spaced 1-ms apart. 238 bins in total were used, and a Hanning window was applied to each bin. To estimate the spectral content of each bin, an FFT on the windowed bin was conducted, with zero padding out to 1 s. The F0 of each bin was defined as the frequency with the greatest spectral magnitude within  $\pm 0.5$  octave of the mean stimulus F0 (100 Hz). The latter criterion was not described in the

original study, but we decided to include it based on correspondence with the first two authors of the original study for advice on how to ensure F0 tracking corresponds to the F0 rather than the upper harmonics. Any frequencies with spectral magnitudes that were not above the noise floor (spectral SNR > 1) were excluded as possible F0 candidates, and the next greatest spectral magnitude (with spectral SNR > 1) was chosen as the F0. Spectral SNR was calculated by first performing a Hanning-windowed FFT on the average pre-stimulus period, when no stimulus was present, and then calculating the ratio between the F0 candidate in the 40-ms windowed stimulus bin and the peak amplitude of that same frequency in the pre-stimulus period. This method for estimating the noise floor was not fully described in the original study<sup>15</sup> but is described in Skoe and Kraus<sup>25</sup> and is believed to be the method used by the original study.

The fidelity of F0 tracking was measured by comparing the F0 tracking of the EEG response to the estimated F0 of the stimulus. The /mi3/ stimulus was down-sampled to 16.384 kHz, and the same sliding FFT analysis was performed on the stimulus to assess the degree to which the FFR response matched the stimulus F0, with the first bin in the analysis beginning at time 0. The F0 within each bin was defined as the frequency with the peak magnitude within  $\pm 0.5$  octaves around the true approximate stimulus F0 (100 Hz); defining such a frequency window was necessary (but not described in the original study), as the second harmonic has a greater spectral magnitude than the F0 for a majority of the stimulus. The degree to which F0 neural encoding matches the stimulus F0, or the F0 stimulus-to-response correlation, was calculated using Pearson's *r* for each subject. Because correlation coefficients do not follow a normal distribution, each *r* was transformed to *z* using Fisher's *r*-to-*z* transformation before conducting the statistical tests described below, consistent with the original study.

### Statistical analyses

To ensure high-powered analyses for all tests, the data were aggregated across sites. All group comparisons used the same, relatively strict definition of musician and non-musician, as described in the "Recruitment and eligibility" section. This ensured that the definition of musician was generally as strict, if not more so, as the definition used in the original studies. Outliers were identified by visual inspection, and analyses were conducted both with and without outliers. Direct replication analyses used the same statistical tests as the original study. CIs correspond to 95% confidence intervals. The CIs for *t*-tests are CIs on the mean difference between groups, whereas ANCOVA CIs are on the simple mean difference contrast between groups after adjusting for covariates. CIs for partial correlations were calculated in JASP using 1000 bootstraps<sup>70</sup>. Analyses that treated years of formal musical training as a continuous variable were conducted across the full cohort of participants unless otherwise stated. As the overarching finding across the original studies was that musicianship provides an advantage to sound processing, all corresponding significance tests relating to musicianship were one-tailed with  $\alpha = 0.05$  unless otherwise noted, consistent with our preregistration protocol. In contrast, any tests of differences relating to gender were two-tailed, as there was no strong hypothesis why males or females should have stronger responses. Exploratory analyses that used age and gender as covariates were conducted using ANCOVA with Type III sums of squares for between-group comparisons and partial correlations for continuous comparisons, with gender dummy coded as 1 = female and 2 = male. Participants who did not identify as either male or female were excluded from all gender analyses due to the small sample size of this group ( $n = 4$ ). The ANCOVA statistical assumptions of linearity and homogeneity of regression slopes were tested by visually inspecting scatterplots, including the regression lines between the covariate and dependent variables for each group. Homogeneity of regression slopes was also tested by ensuring the interaction between the group



and covariate had a  $p$  value  $> 0.05$ . Homogeneity of variances was tested using Levene's test of equality of error variances. Data were analyzed in MATLAB 2016b and JASP<sup>70</sup>.

Exploratory BF hypothesis tests supplement the direct replication analyses in order to assess the support for the alternative hypothesis (i.e., musicians are better than non-musicians) versus the null hypothesis using the reporting standards outlined in van Doorn et al.<sup>71</sup>. Between-groups comparisons were assessed using Bayesian independent-samples  $t$ -tests with a truncated Cauchy prior distribution ( $r = \frac{1}{\sqrt{2}}$ ) so that only positive effect sizes were examined. Bayesian Pearson correlations used a uniform prior. Robustness was assessed across a wide range of prior widths, with results reported in the supplementary materials. BF Supplementary figures were created using JASP version 0.18.3 and compiled using Adobe Illustrator CS6.

Any changes to the preregistered analysis plan that occurred post-registration are described in Supplementary Text 4.

**Spectral encoding for /da/: group comparisons.** The overall strength of spectral encoding in musicians and non-musicians was assessed using two independent-samples  $t$ -tests, one for the F0 and one for the upper harmonics. This analysis is identical to the original study but deviates slightly from our preregistered plan (see Supplementary Text 4).

**F0 tracking for /mi3/: group comparisons.** The hypothesis that musicians would have better F0 tracking than non-musicians was tested using an independent-samples  $t$ -test on the z-transformed F0 stimulus-to-response correlations between the two groups.

**Musical ability and sound neural encoding fidelity.** Melody discrimination performance was scored by calculating  $d'_p$ , a non-parametric estimate of sensitivity. This changed from our preregistered plan to calculate the weighted composite scores, due to recommendations from Strauss et al.<sup>62</sup> and Whiteford et al.<sup>63</sup> to avoid conflating sensitivity with response bias. All correlations with musical ability were exploratory analyses that predicted a positive relationship between sound neural encoding and melody discrimination and therefore also used one-tailed tests, as with the preregistered analyses. The criterion for significance was Bonferroni-corrected for four comparisons ( $\alpha = 0.0125$ ). The /da/ stimulus-to-response correlations used in these analyses used the fixed lag window to match Parbery-Clark et al.<sup>13</sup>.

**Split-half reliability of neural encoding.** After filtering, artifact rejection, and baseline correction, the remaining individual data were split into first and second halves separately for each stimulus polarity. The average response was taken across the trials at each polarity, and this average was summed across polarities for the first and the second halves separately. All spectral encoding analyses and stimulus-to-response correlations were conducted in the same manner as the preregistered analyses, with the exception that only the fixed lag window was used for the /da/ stimulus-to-response correlations. Note that the spectral SNR criterion for the /mi3/ F0-tracking analyses was used on the full dataset, rather than each half of the data separately; this ensured all participants in the full analysis were used in the split-half reliability analysis.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data analyzed in this study are available on the Harvard Dataverse (<https://doi.org/10.7910/DVN/SKOWNK>)<sup>61</sup>.

## Code availability

Experiment code is available on Github (<https://github.com/kellywhiteford/Musicianship>)<sup>67</sup> and OSF<sup>36</sup>. All Matlab code used to preprocess, plot, and analyze the data is available on Github (<https://github.com/kellywhiteford/MusicianshipEEG>)<sup>68</sup>.

## References

- Mehr, S. A. et al. Universality and diversity in human song. *Science* **366**, eaax0868 (2019).
- Schellenberg, E. G. & Lima, C. F. Music training and nonmusical abilities. *Annu. Rev. Psychol.* **75**, 87–128 (2024).
- Costa-Giomi, E. The effects of three years of piano instruction on children's cognitive development. *J. Res. Music Educ.* **47**, 198–212 (1999).
- Dubinsky, E., Wood, E. A., Nespoli, G. & Russo, F. A. Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Front. Neurosci.* **13**, 1153 (2019).
- Schellenberg, E. G. Music lessons enhance IQ. *Psychol. Sci.* **15**, 511–514 (2004).
- Degé, F. Music lessons and cognitive abilities in children: How far transfer could be possible. *Front. Psychol.* **11**, 557807 (2021).
- Bigand, E. & Tillmann, B. Near and far transfer: Is music special?. *Mem. Cogn.* **50**, 339–347 (2022).
- Schneider, P. et al. Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nat. Neurosci.* **5**, 688–694 (2002).
- Madsen, S. M. K., Whiteford, K. L. & Oxenham, A. J. Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Sci. Rep.* **7**, 12624 (2017).
- Hyde, K. L. et al. Musical training shapes structural brain development. *J. Neurosci.* **29**, 3019–3025 (2009).
- Musacchia, G., Sams, M., Skoe, E. & Kraus, N. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proc. Natl. Acad. Sci. USA* **104**, 15894–15898 (2007).
- Musacchia, G., Strait, D. & Kraus, N. Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* **241**, 34–42 (2008).
- Parbery-Clark, A., Skoe, E. & Kraus, N. Musical experience limits the degradative effects of background noise on the neural processing of sound. *J. Neurosci.* **29**, 14100–14107 (2009).
- Parbery-Clark, A., Strait, D. L. & Kraus, N. Context-dependent encoding in the auditory brainstem subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* **49**, 3338–3345 (2011).
- Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T. & Kraus, N. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**, 420–422 (2007).
- Bidelman, G. M., Krishnan, A. & Gandour, J. T. Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *Eur. J. Neurosci.* **33**, 530–538 (2011).
- Bidelman, G. M. & Krishnan, A. Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Res.* **1355**, 112–125 (2010).
- Coffey, E. B. J., Musacchia, G. & Zatorre, R. J. Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *J. Neurosci.* **37**, 830–838 (2017).
- Bianchi, F. et al. Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. *NeuroImage* **163**, 398–412 (2017).
- Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., Baillet, S. & Zatorre, R. J. Neural correlates of early sound encoding and their

- relationship to speech-in-noise perception. *Front. Neurosci.* **11**, 1–14 (2017).
21. Du, Y. & Zatorre, R. J. Musical training sharpens and bonds ears and tongue to hear speech better. *Proc. Natl. Acad. Sci. USA* **114**, 13579–13584 (2017).
  22. Coffey, E. B. J., Herholz, S. C., Chapesiuk, A. M. P., Baillet, S. & Zatorre, R. J. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* **7**, 11070 (2016).
  23. Coffey, E. B. J. et al. Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* **10**, 5036 (2019).
  24. Rodrigues, M. et al. Frequency following response and musical experience: A review. *J. Hear. Sci.* **9**, 9–16 (2019).
  25. Skoe, E. & Kraus, N. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* **31**, 302–324 (2010).
  26. Coffey, E. B. J., Mogilever, N. B. & Zatorre, R. J. Speech-in-noise perception in musicians: a review. *Hear. Res.* **352**, 49–69 (2017).
  27. Anderson, S., Parbery-Clark, A., White-Schwoch, T. & Kraus, N. Aging affects neural precision of speech encoding. *J. Neurosci.* **32**, 14156–14164 (2012).
  28. Corrigan, K. A., Schellenberg, E. G. & Misura, N. M. Music training, cognition, and personality. *Front. Psychol.* **4**, 222 (2013).
  29. Preacher, K. J., Rucker, D. D., MacCallum, R. C. & Nicewander, W. A. Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol. Methods* **10**, 178–192 (2005).
  30. Parbery-Clark, A., Anderson, S., Hittner, E. & Kraus, N. Musical experience offsets age-related delays in neural timing. *Neurobiol. Aging* **33**, 1483.e1–1483.e4 (2012).
  31. Zendel, B. R. & Alain, C. The influence of lifelong musicianship on neurophysiological measures of concurrent sound segregation. *J. Cogn. Neurosci.* **25**, 503–516 (2013).
  32. Zendel, B. R., West, G. L., Belleville, S. & Peretz, I. Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiol. Aging* **81**, 102–115 (2019).
  33. Fleming, D., Belleville, S., Peretz, I., West, G. & Zendel, B. R. The effects of short-term musical training on the neural processing of speech-in-noise in older adults. *Brain Cogn.* **136**, 103592 (2019).
  34. Zendel, B. R. & Alain, C. Musicians experience less age-related decline in central auditory processing. *Psychol. Aging* **27**, 410–417 (2012).
  35. Parbery-Clark, A., Anderson, S., Hittner, E. & Kraus, N. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. *Front. Aging Neurosci.* **4**, 1–12 (2012).
  36. Whiteford, K. et al. Association of musical training with auditory and speech neural coding and perception. <https://doi.org/10.17605/OSF.IO/DUQ34> (2019).
  37. Clinard, C. G., Tremblay, K. L. & Krishnan, A. R. Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hear. Res.* **264**, 48–55 (2010).
  38. Krizman, J., Skoe, E. & Kraus, N. Sex differences in auditory subcortical function. *Clin. Neurophysiol.* **123**, 590–597 (2012).
  39. Miller, G. A. & Chapman, J. P. Misunderstanding analysis of covariance. *J. Abnorm. Psychol.* **110**, 40–48 (2001).
  40. Mankel, K. & Bidelman, G. M. Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proc. Natl. Acad. Sci. USA* **115**, 13129–13134 (2018).
  41. Law, L. N. C. & Zentner, M. Assessing musical abilities objectively: construction and validation of the Profile of Music Perception Skills. *PLoS ONE* **7**, e52508 (2012).
  42. Krizman, J., Rotondo, E. K., Nicol, T., Kraus, N. & Bieszczad, K. M. Sex differences in auditory processing vary across estrous cycle. *Sci. Rep.* **11**, 22898 (2021).
  43. Lau, B. K., Ruggles, D. R., Katyal, S., Engel, S. A. & Oxenham, A. J. Sustained cortical and subcortical measures of auditory and visual plasticity following short-term perceptual learning. *PLOS ONE* **12**, e0168858 (2017).
  44. Reetzke, R., Xie, Z., Llanos, F. & Chandrasekaran, B. Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Curr. Biol.* **28**, 1419–1427.e4 (2018).
  45. Carcagno, S. & Plack, C. J. Subcortical plasticity following perceptual learning in a pitch discrimination task. *J. Assoc. Res. Otolaryngol.* **12**, 89–100 (2011).
  46. Button, K. S. et al. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
  47. Ioannidis, J. P. A. Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
  48. Luck, S. J. in *An introduction to the event-related potential technique* 309–340 (MIT Press, 2014).
  49. Song, J. H., Nicol, T. & Kraus, N. Test–retest reliability of the speech-evoked auditory brainstem response. *Clin. Neurophysiol.* **122**, 346–355 (2011).
  50. Manta, O. et al. Development and evaluation of automated tools for auditory-brainstem and middle-auditory evoked potentials waves detection and annotation. *Brain Sci.* **12**, 1675 (2022).
  51. Trübutschek, D. et al. EEGManyPipelines: a large-scale, grassroots multi-analyst study of electroencephalography analysis practices in the wild. *J. Cogn. Neurosci.* **36**, 217–224 (2024).
  52. Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E. & Johnsrude, I. S. Attentional modulation of envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. *J. Assoc. Res. Otolaryngol.* **19**, 83–97 (2018).
  53. Shinn-Cunningham, B., Varghese, L., Wang, L. & Bharadwaj, H. Individual differences in temporal perception and their implications for everyday listening. in *The Frequency-Following Response* Vol. 61 (eds Kraus, N., Anderson, S., White-Schwoch, T., Fay, R. R. & Popper, A. N.) 159–192 (Springer International Publishing, 2017).
  54. Bidelman, G. M. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage* **175**, 56–69 (2018).
  55. Bharadwaj, H. M. et al. Non-invasive assays of cochlear synaptopathy—candidates and considerations. *Neuroscience* **407**, 53–66 (2019).
  56. Schneider, P. et al. Neuroanatomical disposition, natural development, and training-induced plasticity of the human auditory system from childhood to adulthood: a 12-year study in musicians and nonmusicians. *J. Neurosci.* **43**, 6430–6446 (2023).
  57. Riegel, J., Schüller, A. & Reichenbach, T. No evidence of musical training influencing the cortical contribution to the speech-FFR and its modulation through selective attention. *eNeuro* <https://doi.org/10.1101/2024.07.25.605057> (2024).
  58. Rentfrow, P. J. The role of music in everyday life: current directions in the social psychology of music. *Soc. Personal. Psychol. Compass* **6**, 402–416 (2012).
  59. Whiteford, K. L., Kreft, H. A. & Oxenham, A. J. Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age. *J. Assoc. Res. Otolaryngol.* **18**, 619–633 (2017).
  60. Moscicki, E. K., Elkins, E. F., Baur, H. M. & McNamara, P. M. Hearing loss in the elderly: an epidemiologic study of the Framingham Heart Study Cohort. *Ear Hear.* **6**, 184–190 (1985).
  61. Whiteford, K. L. et al. Musicianship EEG dataset. Harvard Dataverse. <https://doi.org/10.7910/DVN/SKOWNK> (2025).
  62. Strauss, H., Reiche, S., Dick, M. & Zentner, M. Online assessment of musical ability in 10 minutes: development and validation of the Micro-PROMS. *Behav. Res. Methods* <https://doi.org/10.3758/s13428-023-02130-4> (2023).

63. Whiteford, K. L., Goh, P. Y., Stevens, K. L. & Oxenham, A. J. Dissociating sensitivity from bias in the Mini Profile of Music Perception Skills. *JASA Express Lett.* **3**, 094401 (2023).
  64. John, O. P., Donahue, E. M. & Kentle, R. L. *The Big Five Inventory—Versions 4a and 54* (University of California, Berkeley, Institute of Personality and Social Research, 1991).
  65. John, O. P., Naumann, L. P. & Soto, C. J. Paradigm shift to the integrative Big-Five Trait Taxonomy: history, measurement, and conceptual issues. in *Handbook of personality: Theory and research* (eds John, O. P., Robins, R. W. & Pervin, L. A.) 114–158 (Guilford Press, 2008).
  66. Boersma, P. & Weenink, D. PRAAT [Computer software]. (2004).
  67. Whiteford, K. L., Mesik, J., Maddox, R. K. & Oxenham, A. J. Large-scale multi-site study shows no association between musical training and early auditory neural sound encoding. Musicianship Experiment Code: <https://github.com/kellywhiteford/Musicianship>. <https://doi.org/10.5281/zenodo.15707010> (2025).
  68. Whiteford, K. L. Large-scale multi-site study shows no association between musical training and early auditory neural sound encoding. Musicianship EEG Analysis Code: <https://github.com/kellywhiteford/MusicianshipEEG>. <https://doi.org/10.5281/zenodo.15707281> (2025).
  69. Aiken, S. J. & Picton, T. W. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* **245**, 35–47 (2008).
  70. JASP Team. JASP (Version 0.18.3): <https://jasp-stats.org/> (2024).
  71. van Doorn, J. et al. The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon. Bull. Rev.* **28**, 813–826 (2021).
- T.K.P., and B.G.S.C. edited the protocol. K.L.W., L.S.B., V.C.I., A.I., S.J.L., J.M., H.M.B., and R.K.M. set up and piloted the experiment at their respective sites. K.L.W. trained study staff in the protocol and assisted with level calibration at each site. K.L.W., L.S.B., M.C., J.K.C., S.F., P.Y.G., A.H., V.C.I., A.I., S.J.L., B.M., B.O., N.R., E.R., A.E.S., S.C.V.H., and A.E.L. contributed to recruitment and/or data collection. K.L.W., J.K.C., P.Y.G., V.C.I., A.I., S.J.L., and B.M. contributed substantially to data quality management. K.L.W. preprocessed and analyzed the data. K.L.W. and A.J.O. wrote the paper. K.L.W. curated data and code for open source use. All authors edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-62155-5>.

**Correspondence** and requests for materials should be addressed to Kelly L. Whiteford.

**Peer review information** *Nature Communications* thanks Glenn Schellenberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

## Acknowledgements

This research was supported by funding from the National Science Foundation [NSF-BCS grant 1840818, awarded to A.J.O., H.M.B., G.K. Jr., A.E.L., R.K.M., E.W.M., T.K.P., and B.G.S.] and the National Institutes of Health [R01 DC005216, awarded to A.J.O.]. I.S.J. and trainees V.I., B.M., and S.C.V.H. were funded by the Canada First Research Excellence Fund Award “BrainsCAN” (2017–2023) to Western University. We thank Yaminah Carter and the numerous undergraduate researchers who assisted in data collection and/or data quality management, including Penelope Corbett, Angela Sim, and Kara Stevens. We would also like to thank Nina Kraus for providing the stimuli for the EEG experiments, Marcel Zentner for sharing the melody discrimination stimuli, Erika Skoe and Patrick Wong for advice related to the FO tracking analysis, David Purcell for assisting with level calibration at UWO, and Dorea Ruggles for providing an earlier version of the EEG code for UMN.

## Author contributions

A.J.O., H.M.B., I.S.J., G.K.Jr., A.E.L., R.K.M., E.W.M., T.K.P., and B.G.S.C. provided funding and supervision at their respective sites. A.J.O. and K.L.W. designed the study. K.L.W. and A.J.O. wrote the preregistered protocol. A.J.O., K.L.W., H.M.B., I.S.J., G.K.Jr., A.E.L., R.K.M., E.W.M.,