

1 Musical training does not enhance neural sound encoding at early stages of the auditory system: A
2 large-scale multisite investigation

3
4 **Corresponding author:** Kelly L. Whiteford (whit1945@umn.edu), Department of Psychology, University
5 of Minnesota, Minneapolis, MN, USA

6
7 Kelly L. Whiteford¹, Lucas S. Baltzell², Matt Chiu³, John K. Cooper⁴, Stefanie Faucher⁴, Pui Yii Goh¹, Anna
8 Hagedorn⁵, Vanessa C. Irsik^{6,7}, Audra Irvine⁸, Sung-Joo Lim², Juraj Mesik¹, Bruno Mesquita⁶, Breanna
9 Oakes⁵, Neha Rajappa¹, Elin Roverud², Amy E. Schrlau⁴, Stephen C. Van Hedger^{6,7}, Hari M. Bharadwaj^{5,9},
10 Ingrid S. Johnsrude^{6,7,10}, Gerald Kidd Jr.², Anne E. Luebke^{4,11}, Ross K. Maddox^{4,11}, Elizabeth W. Marvin³,
11 Tyler K. Perrachione², Barbara G. Shinn-Cunningham^{8,12}, Andrew J. Oxenham¹

12

13 ¹Department of Psychology, University of Minnesota

14 ²Department of Speech, Language, and Hearing Sciences, Boston University

15 ³Eastman School of Music, University of Rochester

16 ⁴Department of Biomedical Engineering, University of Rochester

17 ⁵Department of Speech, Language, and Hearing Sciences, Purdue University

18 ⁶Centre for Brain and Mind, University of Western Ontario

19 ⁷Department of Psychology, University of Western Ontario

20 ⁸Department of Biomedical Engineering, Carnegie Mellon University

21 ⁹Weldon School of Biomedical Engineering, Purdue University

22 ¹⁰School of Communication Sciences and Disorders, University of Western Ontario

23 ¹¹Department of Neuroscience, University of Rochester

24 ¹²Neuroscience Institute, Carnegie Mellon University

25
26
27
28
29
30
31
32
33
34
35
36

Abstract

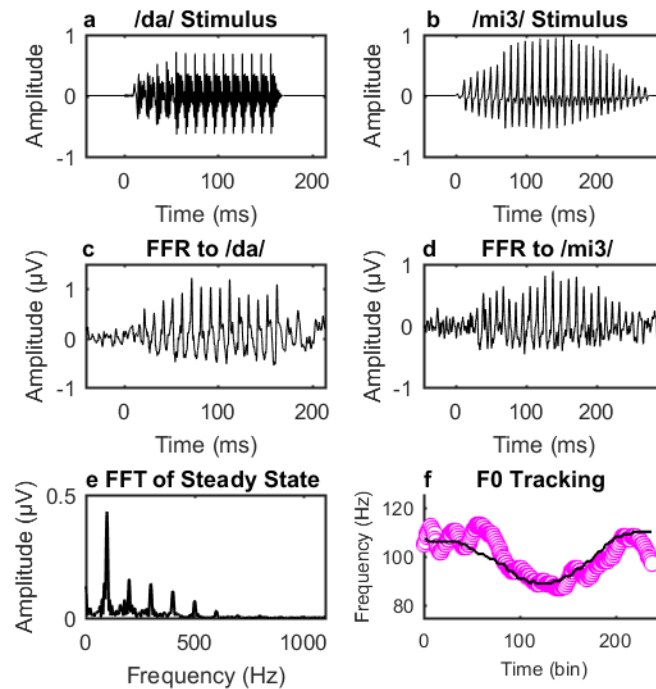
Musical training has been associated with enhanced neural processing of sounds, as measured via the frequency following response (FFR), implying the potential for human subcortical neural plasticity. We conducted a large-scale multi-site preregistered study ($n > 260$) to replicate and extend the findings underpinning this important relationship. We failed to replicate any of the major findings published previously in smaller studies. Musical training was related neither to enhanced spectral encoding strength of a speech stimulus (/da/) in babble nor to a stronger neural-stimulus correlation. Similarly, the strength of neural tracking of a speech sound with a time-varying pitch was not related to either years of musical training or age of onset of musical training. Our findings provide no evidence for plasticity of early auditory responses based on musical training and exposure.

37 Music is universal across human societies and can serve multiple functions¹. In many Western cultures,
38 parents may seek early musical education for their children, sometimes in the hope that the skills
39 learned from musical training may transfer to other aspects of life. A considerable body of research has
40 examined the potential link between such early training and enhanced perceptual and cognitive skills,
41 with mixed results²⁻⁷. However, there is more consensus that musical training from an early age is
42 associated with enhanced responses to sound at different levels of the auditory pathways, from the
43 auditory brainstem⁸⁻¹⁴ to the midbrain and cortex¹⁵⁻¹⁹.

44 A widely reported neural signature of musical training involves the frequency following response
45 (FFR), a scalp-recorded potential measured using electroencephalography (EEG). The FFR is thought to
46 reflect stimulus-entrained neural responses to periodic sounds, primarily from subcortical levels of the
47 auditory system, but also potentially reflecting some cortical contributions, especially at frequencies of
48 100 Hz and below^{20,21}. Several studies have reported that musicians exhibit stronger FFR spectral
49 encoding of sounds than non-musicians either for the fundamental frequency (F0) or the upper
50 harmonics of the stimulus, and that musicians' spectral encoding is more robust to interference from
51 background noise^{8,9,12-14,22}. Such findings have important implications: if the relationship is causal, it
52 implies the potential for plasticity of subcortical auditory function, at least in childhood (when musical
53 training typically begins). Even if the results from these cross-sectional studies do not stem from musical
54 training-induced changes in subcortical auditory responses, an alternative interpretation linking
55 the strength of responses in the early stages of the auditory pathways (perhaps genetically influenced)
56 to future musical success would also be highly intriguing.

57 Enhanced FFRs in musicians have been demonstrated in numerous studies for a variety of
58 stimuli, including speech and non-speech sounds^{8,9,12-14,22}. Two widely used types of stimuli are speech
59 sounds with vowels that are sustained in pitch over time (e.g., /da/, Figure 1a) and speech sounds that
60 vary in pitch over time (e.g., the Mandarin utterance, /mi3/, which dips and rises in F0 over time, Figure

61 1b,f, black line)²³. The fidelity of neural encoding can be assessed either in the time domain (Figure 1c,d)
62 or in the spectral domain (Figure 1e,f). One seminal study found that musicians had stronger spectral
63 encoding for the upper harmonics of /da/ presented in background noise relative to non-musicians¹⁰.
64 Furthermore, the correlation between the averaged time waveform of the neural response and the
65 stimulus waveform was greater in musicians, indicating that musicians' FFRs were more robust to the
66 presence of background noise. Another influential study showed that native English speaking musicians
67 had a stronger representation of the time-varying F0 for a Mandarin word presented in quiet¹², as
68 quantified by the correlation between the F0 contour of the stimulus and neural response (Figure 2f).
69 Importantly, the strength of the neural representations of the stimulus F0 was negatively related to the
70 age at which musicians began their musical training. This relationship suggests that earlier onset of
71 musical training produces stronger neural responses. Together, these two studies have had a major
72 impact, as evidenced by the high number of citations (a combined total of 1770, as of the time of
73 writing), and form the early foundation of the subsequent body of positive evidence for enhanced early
74 neural encoding of sound in musicians²⁴.



75

76 **Figure 1. Stimuli and sample neural responses.** The acoustic waveforms for /da/ (a) and /mi3/ (b) are shown,
77 along with example FFRs in the time (c,d) and spectral domain (e,f) for one individual participant from the present
78 study. The FFT of the steady-state portion of the FFR to /da/ (e) shows the spectral encoding for the F0 (100 Hz)
79 and upper harmonics. Previous research found that musicians had stronger spectral encoding for the upper
80 harmonics to /da/ presented in babble, relative to non-musicians, and stronger stimulus-to-response correlations
81 between the vowel portion of the /da/ stimulus (50–170 ms) and the steady-state portion of the neural response
82 (60–180 ms)¹⁰. The F0 of the /mi3/ stimulus (f, black line) varies in F0 over time (pink: neural F0 tracking). The
83 original study found that musicians (relative to non-musicians) had stronger stimulus-to-response correlations
84 between the F0-trajectory of the stimulus and the neural response and that F0-tracking fidelity was related to age
85 of onset of musical training.

86

87 Although current evidence supports the notion that musicians exhibit advantages in the neural
88 encoding of sound²⁴, several factors complicate the interpretation of this advantage. First, most reports
89 have been based on relatively small samples of listeners^{10,12,25} with dichotomous samples that often

90 represent extreme ends of the musical spectrum (i.e., untrained people compared to musicians with
91 many years of experience). Limiting the sample in this manner is beneficial from an experimenter
92 standpoint, in that it increases the likelihood that between-group differences will be detected, but such
93 disparate groups may differ in many other ways besides musicianship, such as socioeconomic status or
94 personality^{2,26}, limiting the generalizability of the findings²⁷. Moreover, most studies have been
95 conducted on young (college-aged) adults, although there are a growing number of studies examining
96 aging effects^{28–33}. Another complicating factor is that the definition of the terms “musician” and “non-
97 musician” has varied between studies, leading to the possibility that any differences in outcomes
98 between studies may reflect, in part, differences in years and nature of training, age training began, and
99 degree to which musical training or activity is maintained.

100 In the present study, we attempted to replicate the two seminal findings of a musician
101 advantage for spectral neural encoding of speech sounds, described above^{10,12}, across a large sample of
102 participants at six different sites. In addition to replicating, we also extended these studies in several
103 important ways: (1) All sites conducted both studies, allowing for a high-powered aggregate sample. (2)
104 All participants at each site took part in both studies, allowing for the strength of neural encoding
105 between studies to be compared within the same participants. (3) Both age and musical training varied
106 continuously, increasing the generalizability of the findings relative to previous studies, while still
107 allowing for a direct replication by maintaining a subset of participants who fit the most stringent
108 criteria for the definitions of musicians and non-musicians. (4) The methods and primary analyses were
109 preregistered before beginning data collection³⁴, limiting researcher degrees of freedom when analyzing
110 the data. (5) Additional data were collected on factors that have been found to co-vary with musical
111 training, including personality (i.e., the openness to experience factor of the ‘Big 5’ personality scale)
112 and socio-economic status²⁶, which could be used in exploratory analyses or by other researchers in

113 future analyses. The results provide important insights into the replicability and robustness of the
114 musician advantage in early neural processing of sound across the adult lifespan.

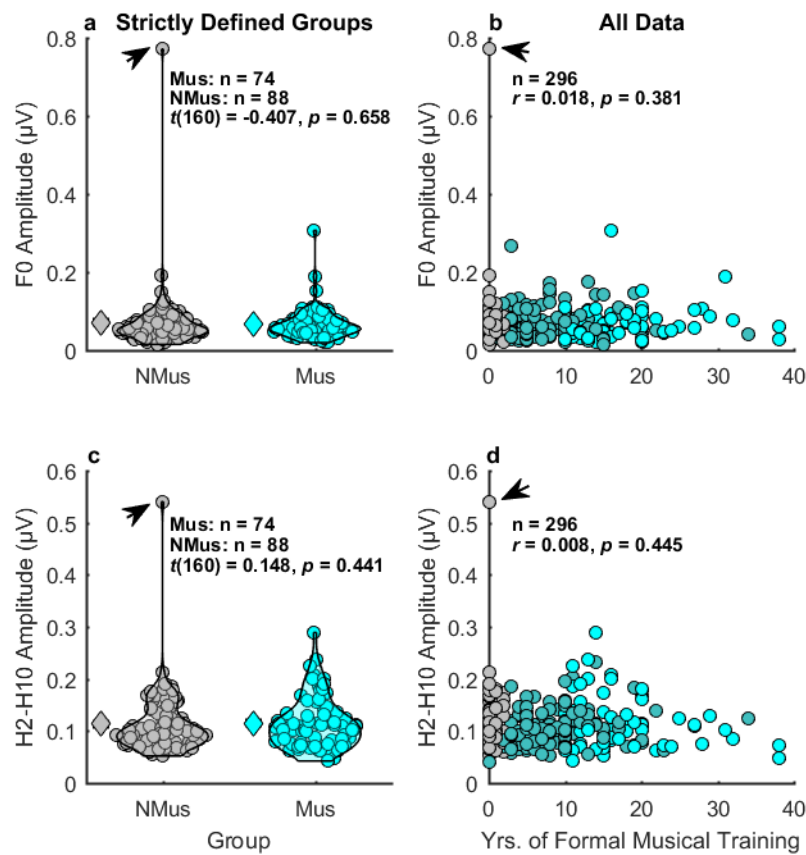
115
116
117

Results

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

119 **Group comparisons.** The first study we sought to replicate¹⁰ compared the neural spectral
120 encoding of the vowel portion of the syllable /da/ (Figure 1a,c,e), embedded in multi-talker babble, in 16
121 musicians and 15 non-musicians, as they passively watched a silent video. In that study, musicians were
122 found to have enhanced encoding of the upper harmonics (H2-H10) but not the F0. We first conducted a
123 direct replication in a subset of our participants that represented the extreme ends of the musical
124 training spectrum (74 musicians and 88 non-musicians). As shown in Figure 2 (a,c), there was no
125 significant difference in spectral encoding as indexed by FFRs between musicians and non-musicians for
126 either the F0 [$t(160) = -.407, p = .658, d = -.064$] or upper harmonics [$t(160) = .148, p = .441, d = .023$] of
127 the same /da/ syllable used by the original study. The lack of effect persisted after removing one non-
128 musician outlier who had unusually strong spectral encoding [Supplementary Figure 1; F0: $t(159) = .684,$
129 $p = .248, d = .108$; Upper harmonics: $t(159) = .909, p = .182, d = .144$]. Bayesian analyses indicated the
130 data were 7.81 and 5.24 times more likely to occur under the null hypothesis than the alternative
131 hypothesis, for encoding the F0 ($BF_{+0} = .128, \% \text{ error} < .0001$) and upper harmonics ($BF_{+0} = .191, \% \text{ error} =$
132 $\sim .057$), respectively. Excluding the outlier non-musician also provided moderate support that the data
133 originated under the null (F0: $BF_{+0} = .315, \% \text{ error} < .0001$; upper harmonics: $BF_{+0} = .404, \% \text{ error} <$
134 $.0001$). The results remained robust across a wide range of widths of the Bayesian prior (Supplementary
135 Figure 2). These results show no relationship between musical training and the FFR for /da/ in
136 background multi-talker babble for neither the F0 nor the upper harmonics.

137 The spectral amplitudes of the FFR decrease with age³⁵, which may have added unexplained
138 variance to our data, potentially obscuring a musician effect. Across the full sample of participants, we
139 confirmed that age was related to poorer spectral encoding of both the F0 and upper harmonics of /da/
140 in babble (F0: $r = -.246$, $p < .0001$, one-tailed; upper harmonics: $r = -.173$, $p = .001$, one-tailed;
141 Supplementary Figure 3). Our subsamples of highly experienced musicians and inexperienced non-
142 musicians did not differ significantly in age [$t(160) = -1.3$, $p = .196$; outlier excluded: $t(159) = -1.36$, $p =$
143 $.176$; two-tailed tests]. Exploratory ANCOVAs also confirmed that musicianship was not related to
144 enhanced spectral encoding for either the F0 [$F(1,159) = .444$, $p = .506$, $\eta_p^2 = .003$] or upper harmonics
145 [$F(1,159) = .004$, $p = .947$, $\eta_p^2 < .0001$] even after adjusting for effects of age. As before, these findings
146 remained the same when excluding the outlier non-musician with the strongest spectral encoding [F0:
147 $F(1,158) = .123$, $p = .726$, $\eta_p^2 = .0008$; upper harmonics: $F(1,158) = .494$, $p = .483$, $\eta_p^2 = .003$].
148



149

150 **Figure 2. Musical training is not related to enhanced spectral encoding of the F0 (a,b) or upper harmonics (c,d) of**

151 **/da/ syllable in background noise.** No effect was found when comparing the strictly defined groups of musicians

152 and non-musicians (a,c), and no relationship was observed in the broader sample between neural encoding

153 strength and years of musical training (b,d). The null effects of musical training remained after removing one

154 outlier non-musician with unusually strong spectral encoding (indicated by arrow; see Supplementary Figure 1 for

155 plots with outlier removed). Black outlines: 1D kernel density estimates (KDEs); Diamonds: Average data; Circles:

156 individual data; NMus: non-musicians (grey); Mus: musicians (cyan). Participants in neither of the strictly defined

157 groups are shown in dark cyan.

158

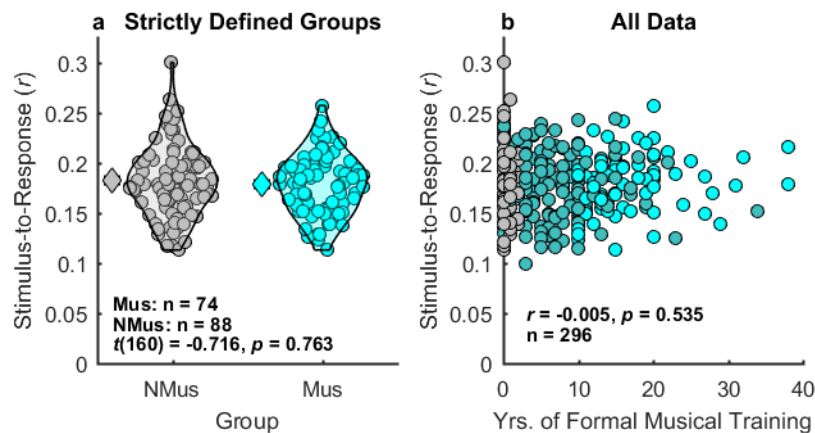
159 The original study¹⁰ also reported that musicians had significantly greater stimulus-to-response

160 correlations than non-musicians, suggesting that musicians' neural encoding of speech sounds was more

161 robust to noise than that of non-musicians. We attempted to replicate this finding by conducting an

162 independent-samples *t*-test on the z-transformed stimulus-to-response correlations between musicians
163 and non-musicians. We found no musician advantage using our pre-planned analyses (Fig. 3a), with the
164 average trend in the opposite direction to that predicted [$t(160) = -.716, p = .763, d = -.113$]. The Bayes
165 factor was $BF_{+0} = .106$ (error = $\sim .002$), meaning the data are 9.43 times more likely to come from the null
166 than the alternative hypothesis, providing moderately strong support for the null (Supplementary Figure
167 4). Our analysis followed the traditional approach of selecting the time-lag between stimulus and
168 response that produced the greatest correlation, whereas the original study limited the stimulus lag
169 times to the range of 6.9–10.9 ms prior to adjusting for site-specific delays between the onset of the
170 trigger and the arrival time of the stimulus at the ear canal. An exploratory analysis using the original
171 study's approach also showed no musician advantage [$t(160) = .074, p = .471, d = .012$] with moderate
172 support for the null hypothesis ($BF_{+0} = .18$, error $\sim .04\%$).

173



174

175 **Figure 3. Musical training is not related to more robust encoding of speech in background sounds.** The
176 untransformed stimulus-to-response correlations are plotted for visualization purposes; the z-transformed data
177 were used in analyses. Black outlines: 1D KDEs; Diamonds: Average data; Circles: individual data; NMus: non-
178 musician (grey); Mus: musician (cyan). Participants in neither of the strictly defined groups are shown in dark cyan.

179

180 **Continuous comparisons.** Next, we tested whether years of formal musical training, as
181 measured across the entire cohort of participants, was correlated with enhanced neural encoding of
182 /da/ in babble, quantified as: (1) the strength of F0 encoding (Figure 2b), (2) the strength of encoding
183 the upper harmonics (Figure 2d), and (3) the transformed stimulus-to-response correlations (Figure 3b).
184 None of these preregistered hypotheses were tested in the original study¹⁰, but they are extensions of
185 the underlying hypothesis that musical training is associated with enhanced neural representation of
186 speech in noisy backgrounds. A Bonferroni-corrected criterion for significance ($\alpha = .017$) was
187 preregistered. Years of formal musical training was not related to enhanced spectral encoding for the F0
188 ($r = .018, p = .381$; non-musician outlier excluded: $r = .094, p = .054$, Supplementary Figure 2) or upper
189 harmonics ($r = .008, p = .445$; non-musician outlier excluded: $r = .048, p = .204$); similarly, years of
190 musical training did not offset the masking effects of background noise on speech encoding, as
191 quantified via the stimulus-to-response correlation ($r = -.005, p = .535$). An exploratory analysis limiting
192 the cross-correlations to the site-specific adjusted lag windows of 6.9-10.9 ms, as in the original study,
193 also showed no relationship between the adjusted stimulus-to-response correlation and years of formal
194 musical training ($r = .029, p = .312$). Further exploratory partial correlations controlling for age confirmed
195 no significant effects of years of musical training (see Supplementary Information).

196

197 **Neural encoding of linguistic pitch contours**

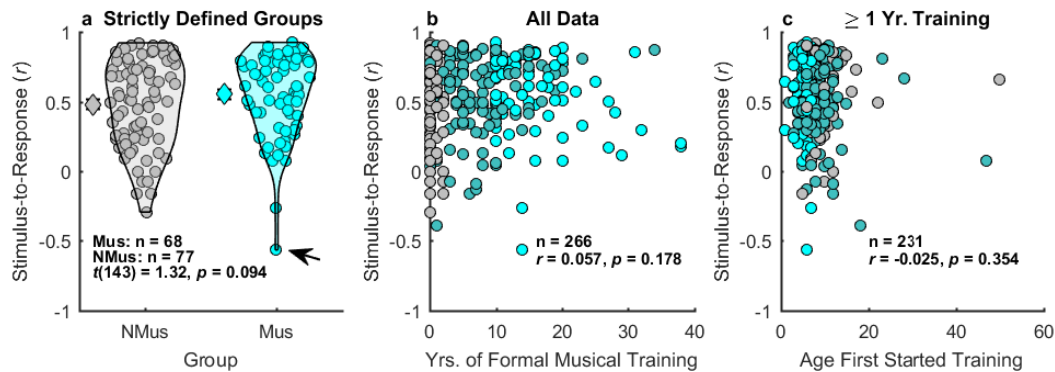
198 **Group comparisons.** The original study¹² measured EEG responses while participants listened
199 passively to the Mandarin word /mi3/ as they watched a silent video. The stimulus varied in F0 over
200 time between 89 to 110 Hz (Figure 1f). The original study found that the F0 stimulus-to-response
201 correlation (i.e., the Pearson correlation between the F0 contour of the stimulus and the neural
202 response) was significantly greater in 10 musicians than in 10 non-musicians. We attempted to replicate
203 this finding on the subset of our participants that met the strict criteria of musician and non-musician

204 (68 musicians and 77 non-musicians; Figure 4a) and found no musician advantage [$t(143) = 1.32, p =$
205 $.094, d = .22$], with the data about 1.4 times more likely to occur under the null than the alternative
206 hypothesis ($BF_{+0} = .715$, error $< .0001\%$, Supplementary Figure 5). The effect remained non-significant
207 after excluding a musician-group outlier with the poorest neural encoding [$t(142) = 1.63, p = .052, d =$
208 $.273$] but resulted in a Bayes factor very close to 1 ($BF_{+0} = 1.14$, error $< .0001\%$), implying no strong
209 evidence for either the alternative or the null hypothesis.

210 To test if age was masking any effect of musicianship on F0-tracking fidelity, an exploratory
211 analysis tested for a musician advantage while adjusting for age. There was still no benefit of
212 musicianship on F0-tracking, either with [$F(1,142) = 1.6, p = .208, \eta_p^2 = .011$] or without [$F(1,141) = 2.47,$
213 $p = .118, \eta_p^2 = .017$] the musician outlier with the poorest encoding.

214

215



216

217 **Figure 4. Neural encoding for time-varying F0 is unrelated to duration and age of onset of musical training. No**

218 significant difference was observed between the strictly defined groups of musicians and non-musicians (a). There
219 was also no relationship between neural tracking of the F0 and years of formal musical training (b) or age of onset
220 of musical training (c). Diamonds: Average data; Circles: individual data; NMus: non-musician (grey); Mus: musician
221 (cyan). Participants in neither of the strictly defined groups are shown in dark cyan; Arrow indicates outlier data
222 point.

223

224 **Continuous comparisons.** The original study¹² found that the age of onset of musical training as
225 well as years of musical training were correlated with the fidelity of F0 tracking, assessed via the F0
226 stimulus-to-response correlation. We used the full cohort of participants (excluding those with 0 years
227 of musical training) to test whether age of onset of training was related to F0 tracking (Figure 4c). We
228 also tested the hypothesis that F0 tracking improves with years of musical training by calculating the
229 correlation between years of formal musical training and the F0 stimulus-to-response correlation
230 (including those with no musical training, as in the original study¹²; Figure 4b). The preregistered
231 criterion for significance included Bonferroni correction for two comparisons ($\alpha = .025$). Unlike the
232 original study, which included 16 participants, we found no relationship between the age of onset of
233 musical training and the fidelity of neural encoding of time-varying stimulus F0 (Figure 4c; $r = -.025$, $p =$
234 $.354$), with the data 8.77 times more likely to occur under the null than the alternative hypothesis ($BF_{-0} =$
235 $.114$; Supplementary Figure 6). Furthermore, the relationship between years of formal musical training
236 and the fidelity of F0 tracking was not significant (Figure 4b; $r = .057$, $p = .178$), with moderate evidence
237 that the data originate under the null hypothesis ($BF_{+0} = .192$). Exploratory partial correlations
238 controlling for age (Bonferroni-corrected $\alpha = .025$) confirmed no relationship between age of onset of
239 musical training and F0 stimulus-to-response correlations ($r_p = -.005$, $p = .469$) or years of formal musical
240 training and F0 stimulus-to-response correlations ($r_p = .057$, $p = .177$). However, F0-tracking did worsen
241 with age ($r = -.173$, $p = .002$), consistent with the expected degradation of spectral neural encoding with
242 age³⁵.

243

244 **Comparing neural responses between measures**

245 Most participants ($n=263$; Mus = 68; NMus = 74) completed both the /da/ test and the /mi3/ test and
246 met all inclusion criteria for analyses (see Methods). We compared the strength of neural encoding
247 between tests to examine whether neural tracking for the F0 of speech in quiet is related to FFRs for

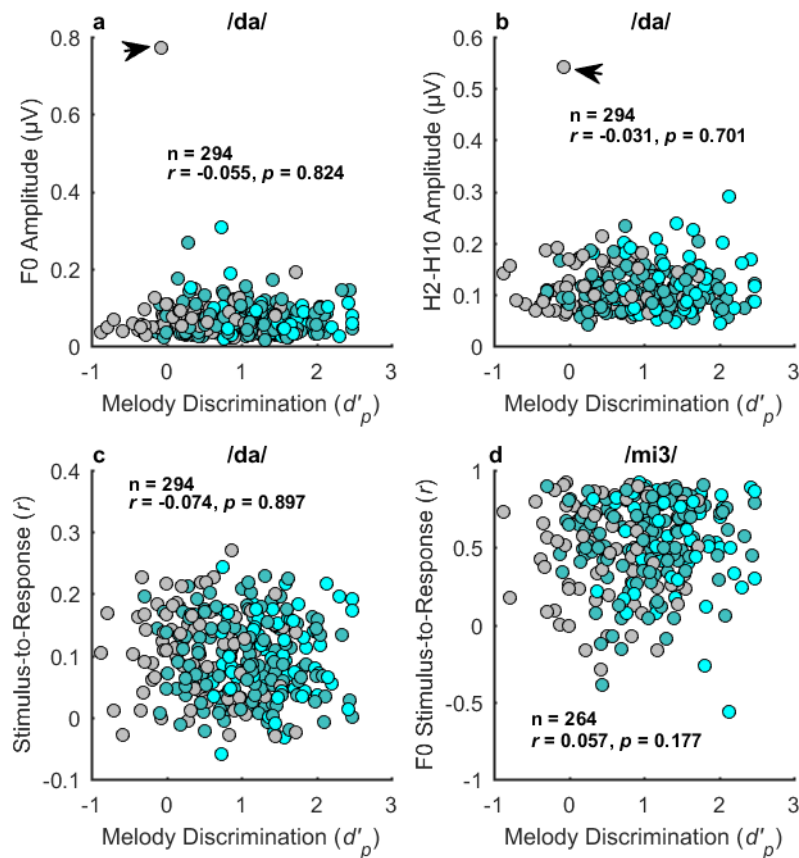
248 encoding for speech in babble (Supplementary Figures 7,8). Exploratory analyses demonstrated that the
249 stimulus-to-response correlation for encoding the F0 of /mi3/ in quiet was weakly related to F0 spectral
250 encoding for /da/ in babble ($r = .133$, $p = .016$, Bonferroni-corrected $\alpha = .0125$), but this effect did not
251 reach significance and was driven by one outlier non-musician with very strong encoding for both
252 measures (outlier removed: $r = .072$, $p = .122$). There was no association between neural encoding for
253 /mi3/ in quiet and the upper harmonics of /da/ in babble ($r = .071$, $p = .125$; outlier removed: $r = .009$, p
254 $= .442$), and there was still no association once one additional outlier musician, with poor F0 tracking but
255 strong encoding for the upper harmonics, was removed ($r = .068$, $p = .138$). While measures between
256 studies were unrelated to one another—potentially because dynamic F0 tracking is somewhat different
257 from overall strength in spectral encoding, or because the stimuli were different—we did find a
258 relationship between strength of spectral encoding for the F0 versus upper harmonics within the same
259 stimulus (/da/ in babble: $r = .585$, $p < .0001$; outlier non-musician removed: $r = .3$, $p < .0001$).

260

261 **Does musical ability account for failures to replicate?**

262 Our criteria for the groups of “musician” and “non-musician” were at least as strict as in both original
263 studies, so the failure to replicate any musician advantages in neural encoding is unlikely to be due to
264 our group definitions. However, musical expertise or aptitude, rather than years of training, may be a
265 more sensitive measure for detecting differences in neural encoding of sound³⁶. We tested this directly
266 by correlating an objective measure of musical ability (same/different melody discrimination)³⁷ with our
267 four measures of neural encoding fidelity: (1) spectral encoding for the F0 and (2) upper harmonics for
268 /da/ in babble, (3) stimulus-to-response correlations for /da/, and (4) F0-tracking for /mi3/ in quiet.
269 There was no relationship between the objective measure of musical ability and any of the four
270 measures of sound neural encoding (Figure 5; see Supplementary Figure 9 for results with the non-
271 musician outlier excluded).

272



273

274 **Figure 5. Musical ability is unrelated to sound neural encoding for /da/ in babble (a-c) or /mi3/ in quiet (d).**

275 Musical ability was assessed using an objective measure of melody discrimination³⁷ and quantified using a non-
 276 parametric estimate of sensitivity, d'_p ^{38,39}, where higher d'_p indicates better performance. Grey circles: non-
 277 musicians; Cyan circles: musicians; Dark cyan circles: Participants in neither of the strictly defined groups; Arrow
 278 points to outlier data.

279

280 **Table 1.** Key findings selected for direct replication, along with the effect sizes from present study. The
 281 number of citations listed for each paper corresponds to Google Scholar on July 16th, 2024. All five
 282 direct-replication analyses failed to replicate. Parentheses: original study sample size.

Reported Musician Advantage	Publication	Sample Size		# of Citations	Original Effect Size	Current Study Effect Size
		Mus.	Non-Mus.			

Enhanced spectral encoding of upper harmonics to /da/ in babble.	Parbery-Clark et al. (2009) ¹⁰	n=74 (16)	n=88 (15)	492	$d = 1.04$	$d = .023$
Less degradative effects of babble on encoding speech (/da/).	Parbery-Clark et al. (2009) ¹⁰	n=74 (16)	n=88 (15)	492	$d = 1.05$	$d = .012$
Enhanced F0 encoding for linguistic pitch in quiet.	Wong et al. (2007) ¹²	n=68 (10)	n=77 (10)	1278	$d = 1.11$	$d = .22$
Years of musical training related to higher fidelity of neural encoding for dynamic changes in F0.	Wong et al. (2007) ¹²	n=266 (20)		1278	$r = .456$	$r = .05$
Starting a musical instrument at an earlier age related to enhanced F0 tracking.	Wong et al. (2007) ¹²	n=231 (16)		1278	$r = -.502$	$r = -.025$

283 *Note.* Cohen's d was estimated directly from the original studies using the reported groups means and
284 standard deviations, if available; otherwise, d was calculated from the reported t -value and degrees of
285 freedom.

286
287

Discussion

288 Our large-scale replication and extension tested the robustness and generalizability of the widely
289 accepted finding that musicians have enhanced neural encoding of sound, as assessed with the FFR – an
290 electrophysiological index of the fidelity of neural encoding in the early stages of the auditory
291 pathways²⁴. The results were consistent across all five direct-replication analyses (Table 1): There were
292 no significant musician advantages, despite using the same stimuli, methods, and analyses as in the
293 original studies. Musicians did not exhibit stronger sound spectral encoding for the upper harmonics of
294 speech in background sounds, nor did they demonstrate enhanced tracking for dynamic changes in
295 linguistic F0 for speech sounds in quiet. Neural encoding in musicians was also not more robust to the
296 effects of background noise than that of non-musicians. An extension of the analyses to include years of
297 formal musical training as a continuous variable, which increased the statistical power and the

298 generalizability of the results, also revealed no relationship between musical training and any of our
299 preregistered measures of neural encoding of sound.

300 Exploratory analyses controlling for potential effects of age also confirmed no significant effects
301 of musical training on neural encoding of sound. This finding is especially important, as musical training
302 has been proposed to potentially counteract the age-related declines of the fidelity of neural encoding
303 of sound^{4,28}, and even speech perception in background noise³². Our results confirmed that spectral
304 encoding tends to degrade with age, but this effect appeared to be larger and more consistent for the
305 encoding of the F0 compared to the upper harmonics. Our findings that poorer spectral encoding with
306 age was unrelated to musical training provide no support for efforts to enhance the FFR in older
307 listeners through musical training interventions^{4,40}.

308 **Why did the effect of musicianship fail to replicate?**

309 A recent review of perceptual and neural associations with musical training noted that conclusions were
310 mixed with respect to behavioral outcomes but that all studies examining neural advantages in
311 musicians had reported at least one significant effect²⁴. Given the seeming robustness of the results, it
312 may be surprising that our large-scale study failed to replicate any of the original findings. This apparent
313 discrepancy may have several possible underlying causes. First, the positive results have come from a
314 relatively limited number of laboratories, and even then the musician advantage has not been
315 consistent among studies. For example, one study¹¹, using the same /da/ stimulus as in the present
316 study, found a musician encoding advantage in quiet, but only for the F0 and not the upper harmonics,
317 whereas another study using the same stimulus³³ reported a musician advantage for the upper
318 harmonics, but not the F0, both with and without background noise. Second, the sample sizes from past
319 studies in this field of research have all been relatively small ($n < 30$), raising the likelihood of false
320 positives^{41,42}. Finally, and perhaps most importantly, there is no single agreed-upon analytical technique
321 for examining the fidelity of sound encoding using EEG. Researchers thus have many degrees of freedom

322 related to the number of ways they can test their hypotheses⁴³, and not all of the analyses may be
323 reported in the final publication. Testing the same hypothesis many different ways increases the chance
324 of a false positive (Type I error), highlighting the importance of preregistration of hypotheses, methods,
325 and planned analyses prior to data collection.

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

337 Lastly, it is possible that we selected the wrong musician advantages to test. For example, a
338 number of studies have claimed that musicians exhibit stronger and shorter-latency responses relative
339 to non-musicians^{8-10,28,33}. The majority of studies examining response latencies used an expert peak-
340 picker to manually identify the peaks in the early neural responses. It would be impossible to directly
341 replicate such analyses, as different findings between studies could be related to differences in the
342 expertise or strategy of the peak picker(s), making a null finding difficult to interpret. Future analyses of
343 our publicly available data could involve efforts to automate the process of peak-picking⁴⁴ or formally
344 assess the reproducibility of outcomes across different peak pickers⁴⁵.

345

346 **Broader Implications**

347 Our study is the first large-scale replication and extension of the widely accepted finding that musical
348 training is related to enhanced neural encoding of sound. Our results do not support this claim. While
349 the neural locus of the FFR is debated, and may contain contributions from cortical sources^{15,20,21,46}, the
350 evidence suggests that the dominant sources are subcortical for frequencies greater than 100 Hz, and
351 originate primarily in the inferior colliculus⁴⁷⁻⁴⁹. One important direct implication is that phase-locked
352 subcortical neural responses to sound are likely not nearly as plastic as previously thought, even
353 following many years of intensive musical training starting at an early age. It may be that cortical
354 structure and function is more susceptible to music-related interventions, although most studies so far
355 have been cross-sectional, making it difficult to determine whether any differences are due to musical
356 training².

357

358 **Conclusions**

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

367 Musical training is not related to an enhancement in early neural encoding of sound.

368 Nevertheless, there are of course many important reasons why learning and playing music remains a

369 valuable endeavor, including social connection, emotional regulation, or simply the enjoyment of music
370 for its own sake⁵⁰.

371

372 **Methods**

373 **Recruitment and eligibility**

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

384 To ensure an adequate representation across age ranges, participants for the full lab study were
385 recruited at each site to be roughly evenly distributed in each decade of age (20s – 60s, with ages 18 and
386 19 grouped in the 20s decade), based on the age reported in the online screening. Each site aimed to
387 recruit 60 participants, with at least 25% of participants with no more than 2 years of any musical
388 training and no ongoing music performance activities (i.e., they reported that they did not currently play
389 a musical instrument, including voice), and at least 25% of participants who started playing their first
390 musical instrument or voice by the age of 7, had completed at least 10 years of formal musical training,
391 and reported that they currently played a musical instrument. Formal musical training was defined as
392 group or private lessons, excluding standard elementary school activities. The remaining participants

393 could have varying amounts of formal musical training. In this way, we could assess the effect of number
394 of years of formal musical training as a continuous variable, while still being able to perform
395 dichotomous comparisons (musician vs. non-musician) with at least half of our overall sample. Sites
396 aimed to have roughly even and uniform distribution of ages and gender between these two groups, as
397 with all other participants.

398

399 **Participants**

400 All participants in the present study also took part in a number of behavioral tests in the lab (not
401 reported here). Participants recruited for the in-person portion of the study underwent a pure-tone
402 audiometric screening at octave frequencies between 125 and 8000 Hz, as well as at 6000 Hz. Because
403 age and high-frequency hearing loss co-vary^{51,52}, the maximum allowable hearing loss was titrated per
404 decade, so that participants under the age of 40 were required to have thresholds ≤ 20 dB hearing level
405 (HL) across all tested frequencies, but older adults could have more high-frequency loss (specified in
406 Table 2). All participants were required to meet the audiometric criteria in both ears to participate. A
407 total of 296 participants (115 male; 177 female; 4 non-binary), including 74 musicians and 88 non-
408 musicians (as defined in “Recruitment and Eligibility” section), completed the syllable-in-noise study,
409 and 295 completed the linguistic pitch study. Thirty of the participants for the linguistic pitch study did
410 not meet the pre-specified criterion for analyses (i.e., at least one binned FFT analysis was in the noise
411 floor, as defined by Wong et al., 2007¹²) or were unable to return to the lab to redo the study, so their
412 data was excluded from this task, leaving a total of 265 participants (106 male; 155 female; 4 non-
413 binary). Participant age ranged from 18 to 69 years for both studies. Most participants completed both
414 studies; reasons for missing data are described in each site’s corresponding EEG log
415 (<https://osf.io/dugq34/>). The total number of participants per site for each measure, including their
416 musical status, are provided in Supplementary Table 1.

417

418 **Table 2.** Audiometric threshold criteria. All participants that completed the lab-portion of the study met
419 the audiometric threshold criteria specified for their age group in each ear.

Age Group	.125 - 1 kHz	2 kHz	4 kHz	6 kHz	8 kHz
18-39	≤ 20 dB HL	≤ 20 dB HL	≤ 20 dB HL	≤ 20 dB HL	≤ 20 dB HL
40-49	≤ 20 dB HL	≤ 20 dB HL	≤ 20 dB HL	≤ 30 dB HL	≤ 30 dB HL
50-59	≤ 20 dB HL	≤ 25 dB HL	≤ 30 dB HL	≤ 40 dB HL	≤ 40 dB HL
60-69	≤ 20 dB HL	≤ 30 dB HL	≤ 40 dB HL	-	-

420

421 All participants provided written informed consent and were given monetary compensation or
422 course credit for their in-person participation. All study protocols were approved by the Institutional
423 Review Board at the corresponding university site prior to any data collection: Boston University
424 (4942E), Carnegie Mellon University (STUDY2018_00000367), Purdue University (1609018209),
425 University of Minnesota (0605S85872 and 1306S37081), University of Rochester (STUDY00004020), and
426 by the Nonmedical Research Ethics Board of the University of Western Ontario (NMREB 112604).

427

428 **Online screening**

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

434 The online screening was administered through Qualtrics, with each site completing recruitment
435 and online data collection under the purview of their own IRB. All screening participants provided
436 informed consent online. Participants were not compensated for participation in the screening, but they
437 had the option to enroll in a drawing for a chance to win a gift card as an incentive for participating. All
438 personal identifiers were removed from online data before sharing between sites, so that only the

439 subject ID number linked the online to the laboratory data. The online measures are described below in
440 the order that they appeared.

441 **Age.** Participants were asked to select their age from a dropdown menu. A reported age of
442 younger than 18 or older than 89 led to termination of the screening.

443 **Level adjustment.** To help ensure that the stimuli for the online listening tasks were audible but
444 not too loud, participants were presented with noise and instructed to adjust their volume so that it was
445 at an audible but comfortable level. The stimulus was white noise, bandpass filtered between 200 and
446 1000 Hz, so that the frequency spectrum was comparable to that used in the melody task.

447 **Attention check.** This task helped exclude participants who were not attending or who did not
448 have properly functioning audio on their device. To pass the attention check, participants were required
449 to answer at least 3 of 4 trials correctly. Each trial consisted of a short sequence of 1-kHz pure tones.
450 Each tone within a sequence was 400 ms in duration with 50-ms raised-cosine onset and offset ramps,
451 and each tone was separated by 500 ms of silence. Participants were instructed that each trial contains
452 between 0-9 tones, and their task was to report the number of tones they heard by selecting the
453 corresponding number from a dropdown menu. Because there were 10 options for each trial, the
454 probability of passing the screening by chance was very low (.0037). To minimize the duration of the
455 task, each trial only had 1, 2, 3, or 4 tones in a sequence, with each tone-sequence option presented
456 once. Trials had a fixed duration of 5 s, so that the entire task could be completed in less than half a
457 minute. No feedback was provided.

458 **Melody discrimination.** Stimuli were from the Melody subtest of the Full Profile of Music
459 Perception Skills (PROMs), with methods as described in Law and Zentner (2012)³⁷. During each trial,
460 participants first heard a reference melody twice in a row, followed by a comparison melody. The task
461 was to determine whether the comparison melody was the same as or different from the reference
462 melody, with participants selecting their answer from five possible options: “Definitely Same”, “Probably

463 Same”, “I Don’t Know”, “Probably Different”, or “Definitely Different”. One practice trial was provided,
464 followed by 18 data trials. Participants did not receive feedback, but they did receive their total
465 composite score at the end of the task. The composite score provided to participants was calculated
466 using weighted responses as described in Law and Zentner (2012)³⁷, with confident correct responses
467 (“Definitely Same” or “Definitely Different”) receiving 1 point, less-confident correct responses
468 (“Probably Same” or “Probably Different”) receiving .5 point, and incorrect responses or “I Don’t Know”
469 receiving 0 points. For analyses, melody discrimination performance was calculated using a bias-free
470 estimate of sensitivity, d'_p , as recommended by Strauss et al. (2023)³⁸ and Whiteford et al. (2023)³⁹.

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

474 **Big Five Personality Inventory.** The 44-item Big Five Personality Inventory was administered to
475 assess personality^{53,54}. This was a self-report questionnaire, where each item is rated on a 5-point
476 response scale.

477

478 **Stimuli and procedures**

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

483 All sites ran the tests in the same order as described below. Tests that were skipped or needed
484 to be rerun due to technical or researcher error were noted in a study log and, whenever possible, the
485 participant returned to complete tests with missing data.

486 **EEG: Syllable in noise.** To assess the fidelity of neural encoding for speech in noise, we
487 measured EEG responses to the speech syllable /da/ (Figure 1, top row) in multi-talker babble. The /da/
488 had a 100-Hz F0 with 170-ms duration, presented at 80 dB SPL, as used by Parbery-Clark et al.¹⁰. Both
489 the /da/ and multi-talker babble were generously shared by the principal investigator of the original
490 study. The multi-talker babble had a 37.41-s duration (not 45 s, as mistakenly reported in the original
491 study¹⁰) and looped continuously throughout the task at 10 dB below the level of the speech syllable.
492 Participants listened passively to /da/ in multi-talker babble over insert earphones (see Table 3) while
493 watching a silent video in a sound-attenuating booth. The /da/ was presented at alternating polarities
494 over two blocks of 3000 trials each (6000 trials total), with an ISI of about 83 ms, so that each block
495 lasted approximately 13 min. Participants were allowed short breaks between blocks and instructed to
496 remain still during stimulus presentation. Data were acquired with the electrode systems and sampling
497 rates listed in Table 3 with earlobe references. Sites with BioSemi systems ensured that the magnitude
498 of the offset voltages were $< \pm 30$ mV before beginning data collection.

499 **EEG: Linguistic pitch.** The Mandarin word /mi3/ (the 3 denoting a dipping tone; Figure 1b,f),
500 which means “rice,” was presented bilaterally at 70 dB SPL. Methods were adapted from Wong et al.¹²,
501 and the stimulus was generously provided by the original authors. The word was originally recorded by a
502 native Mandarin speaker and then adjusted in Praat⁵⁵ to have a duration of 278.5 ms and an F0 contour
503 ranging from 89-110 Hz. Participants passively listened to 4800 repetitions of /mi3/, divided into two
504 blocks of 2400 trials and presented at alternating polarities (2400 trials per polarity). The ISI was about
505 83 ms, so that the entire task had a duration of about 30 min (15 min per block). The same electrode
506 system, sampling rate, and active and reference electrodes was used as in the syllable-in-noise task.

507

508 **Hardware, software, and materials**

509 With the exception of the online screening, all auditory stimuli were controlled via Matlab (R2016b).
 510 Code for tests and analyses is available on GitHub and linked to the project website on the Open Science
 511 Framework (OSF; <https://osf.io/duq34/>). The stimuli for the EEG studies may be available upon
 512 reasonable request by contacting the authors of the original studies^{10,12}. The principal investigator from
 513 Law and Zentner (2012)³⁷ should be contacted for requests to use the melody discrimination stimuli.
 514 Table 3 shows the equipment used at each test site.
 515
 516 **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.

517

518 **Sharing of Data**

519 Each site was responsible for quality checking their data before sharing with the first author to ensure it
520 was formatted in a manner consistent with the other sites. Each site was also responsible for
521 maintaining a detailed log of each EEG session and providing the code they use to clean and format the
522 raw data.

523

524 **EEG Data Analyses: Syllable in Noise**

525 All single-channel preprocessing and analyses were the same as reported in Parbery-Clark et al.¹⁰, unless
526 otherwise stated. The recordings from one site (UR) were down-sampled so that EEG data from all sites
527 had the same sampling rate; this was not done in the original study but was necessary due to equipment
528 differences. All recordings were bandpass filtered between 70-2000 Hz (12 dB/octave with zero-phase
529 shift) and then epoched from -40 – 213 ms, where 0 ms corresponds to the stimulus onset. Trials with
530 activity $\geq \pm 35 \mu\text{V}$ were treated as artifacts and removed from analyses. Epochs were baseline-corrected
531 based on the mean potential in the pre-stimulus period. The average response was taken across the
532 trials at each polarity, and this average was summed across polarities to minimize the stimulus artifact
533 and the cochlear microphonic^{23,56}. Only the spectral analyses were selected for replication with the aim
534 of including fewer hypothesis tests to simplify the analysis plan.

535 **Spectral encoding.** The musician advantage for enhanced spectral encoding to the vowel
536 portion of /da/ in babble was assessed using the same, fast Fourier transform (FFT) analysis methods as
537 the original study¹⁰. An FFT of the steady-state portion of the EEG response (60-180 ms) was calculated
538 for each subject, with zero padding added. The strength of harmonic encoding for the first 10 harmonics
539 (with the first harmonic corresponding to the F0) was estimated by calculating the average spectral

540 amplitude within 60-Hz-wide frequency bins that were centered around each harmonic. To estimate the
541 overall strength of encoding of the upper harmonics for each subject, the average spectral amplitudes
542 for harmonics 2-10 were summed.

543 **Stimulus-to-response correlations.** The preregistered stimulus-to-response correlation analysis
544 assessed cross-correlations at all possible lag times (slightly different the original study¹⁰, which limited
545 the stimulus lag to 8-12 ms) between the vowel portion of the /da/ stimulus (50 – 170 ms) and the
546 steady-state portion of the neural response (60-180 ms), defined in the same manner as used in the FFT
547 analysis. The maximum correlation across all lag times is referred to as the stimulus-to-response
548 correlation, with stimulus-to-response correlations closer to 0 indicating poorer neural representations
549 of /da/. Because correlations do not adhere to assumptions of normality, they were transformed using
550 Fisher's *r*-to-*z* transformation before conducting analyses; this transformation was not used by the
551 original study¹⁰.

552 An exploratory analysis was also conducted to more precisely match the lag times tested in
553 Parbery-Clark et al. (2009)¹⁰ and to account for any fixed delay between the onset of the trigger and the
554 arrival time of the stimulus at the ear canal, which varied between sites due to equipment differences.
555 This included accounting for any fixed delay between the onset of the trigger and the onset of the
556 stimulus (e.g., due digital to analog conversion time) as well as the time it takes the stimulus to travel
557 the length of the earphone tubes. The stimulus lag used in the cross-correlation analysis by the original
558 study was 8-12 ms, which included a 1.1 ms fixed delay. We therefore limited the lag time to 6.9-10.9
559 ms for all sites, and then added each individual site's fixed delay time to quantify the site-specific lag
560 window. The cross-correlation was conducted on the neural response to the vowel (50-200 ms) and the
561 zero-padded vowel portion of the stimulus within each site-specific lag window. The maximum
562 correlation within the site-specific lag window for each subject is referred to as the adjusted stimulus-to-
563 response correlation.

564

565 **EEG Data Analysis: Linguistic Pitch**

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

571 **Calculating F0 tracking.** The strength of F0 tracking was estimated by performing a sliding FFT
572 analysis on the EEG response for each subject over the entire FFR period, after accounting for any fixed
573 delay between the onset of the trigger and the arrival time of the stimulus at the ear canal. This included
574 accounting for any fixed delay between the onset of the trigger and the onset of the stimulus (e.g., due
575 digital to analog conversion time) as well as the time it takes the stimulus to travel the length of the
576 earphone tubes. The time-averaged EEG response was segmented into 40-ms bins, with each bin spaced
577 1-ms apart. 238 bins in total were used, and a Hanning window was applied to each bin. To estimate the
578 spectral content of each bin, an FFT on the windowed bin was conducted, with zero-padding out to 1 s.
579 The F0 of each bin was defined as the frequency with the greatest spectral magnitude within $\pm .5$ octave
580 of the mean stimulus F0 (100 Hz). The latter criterion was not described in the original study, but we
581 decided to include it based on correspondence with the first two authors of the original study for advice
582 on how to ensure F0 tracking corresponds to the F0 rather than the upper harmonics. Any frequencies
583 with spectral magnitudes that were not above the noise floor were excluded as possible F0s. The noise
584 floor was calculated by performing a Hanning-windowed FFT on the average pre-stimulus period, when
585 no stimulus was present. This method for estimating the noise floor was not described in Wong et al.¹²
586 but is described in Skoe and Kraus (2010)²³ and is believed to be the method used by the original study.

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

598

599 **Statistical Analyses**

600 To ensure high-powered analyses for all tests, the data were aggregated across sites. All group
601 comparisons used the same, relatively strict definition of musician and non-musician, as described in the
602 "Recruitment and eligibility" section. This ensured that the definition of musician was generally as strict,
603 if not more so, as the definition used in the original studies. Outliers were identified by visual inspection,
604 and analyses were conducted both with and without outliers. Direct replication analyses used the same
605 statistical tests as the original study. Analyses that treated years of formal musical training as a
606 continuous variable were conducted across the full cohort of participants unless otherwise stated. As
607 the overarching finding across the original studies was that musicianship provides an advantage to
608 sound processing, all corresponding significance tests were one-tailed with $\alpha = .05$ unless otherwise
609 noted. Exploratory analyses that used age as a covariate were conducted using Analysis of Covariance
610 (ANCOVA) for between-group comparisons and partial correlations for continuous comparisons.

611 ANCOVA statistical assumptions of linearity and homogeneity of regression slopes were tested by
612 visually inspecting scatterplots including the regression lines between the covariate and dependent
613 variables for each group. Homogeneity of regression slopes was also tested by ensuring the interaction
614 between the group and co-variate had a p -value $> .05$. Homogeneity of variances was tested using
615 Levene's test of equality of error variances. Data were analyzed in Matlab 2016b and JASP⁵⁷.

616 Exploratory Bayes factor (BF) hypothesis tests supplement the direct replication analyses in
617 order to assess the support for the alternative hypothesis (i.e., musicians are better than non-musicians)
618 versus the null hypothesis using the reporting standards outlined in van Doorn et al. (2021)⁵⁸. Between-
619 groups comparisons were assessed using Bayesian independent-samples t tests with a truncated Cauchy
620 prior distribution ($r = \frac{1}{\sqrt{2}}$) so that only positive effect sizes were examined. Bayesian Pearson
621 correlations used a uniform prior. Robustness was assessed across a wide range of prior widths, with
622 results reported in the supplementary materials. BF supplementary figures were created using JASP
623 version 0.18.3 and compiled using Adobe Illustrator CS6.

624

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

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

632 **Musical ability and sound neural encoding fidelity.** Melody discrimination performance was
633 scored by calculating d_p , a non-parametric estimate of sensitivity. This changed from our preregistered
634 plan to calculate the weighted composite scores, due to recommendations from Strauss et al. (2023)³⁸

635 and Whiteford et al. (2023)³⁹ to avoid conflating sensitivity with response bias. All correlations with
636 musical ability were exploratory analyses that predicted a positive relationship between sound neural
637 encoding and melody discrimination and therefore used one-tailed tests. The criterion for significance
638 was Bonferroni-corrected for four comparisons ($\alpha = .0125$). The /da/ stimulus-to-response correlations
639 used in these analyses used the fixed lag window to match Parbery-Clark et al. (2009)¹⁰.

640

641 **Exclusion criteria**

642 Only those who met the criteria listed in the “Recruitment eligibility” and “Participants” sections took
643 part in the study. Missing data from one or more tasks (e.g., from dropping out of the study or
644 researcher error) resulted in exclusion on the corresponding analyses; whenever possible, the
645 participant was rerun on tests with missing data. Each site has a log of explanations for missing data and
646 technical issues.

647 EEG data were excluded if there were less than 60% usable trials for any reason, such as a
648 reduced number of trials due to technical issues, researcher error, or an excessive number of artifacts. If
649 a participant did not have enough usable trials, they were re-run on the corresponding task whenever
650 possible. EEG data from the /mi3/ test was excluded if the SNR was too poor to estimate F0 tracking in
651 the sliding FFT analysis. This occurred if the spectral magnitude of the EEG response for all frequencies
652 was within the noise floor.

653

654 **Acknowledgments**

655 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. disclose support for the research of
656 this work from the National Science Foundation [NSF-BCS grant 1840818]. A.J.O. discloses support for
657 the research of this work from the National Institutes of Health [R01 DC005216]. I.S.J and trainees V.I.,
658 B.M. and S.vH. were funded by the Canada First Research Excellence Fund Award “BrainsCAN” (2017-
659 2023) to Western University. We thank numerous undergraduate researchers who assisted in data
660 collection and/or data quality management, including Penelope Corbett, Angela Sim, and Kara Stevens.

661

- 662 1. Mehr, S. A. *et al.* Universality and diversity in human song. *Science* **366**, eaax0868 (2019).
- 663 2. Schellenberg, E. G. & Lima, C. F. Music training and nonmusical abilities. *Annual Review of*
664 *Psychology* **75**, 87–128 (2024).
- 665 3. Costa-Giomi, E. The effects of three years of piano instruction on children’s cognitive development.
666 *J. Res. Music Educ.* **47**, 198–212 (1999).
- 667 4. Dubinsky, E., Wood, E. A., Nespoli, G. & Russo, F. A. Short-term choir singing supports speech-in-
668 noise perception and neural pitch strength in older adults with age-related hearing loss. *Front.*
669 *Neurosci.* **13**, (2019).
- 670 5. Schellenberg, E. G. Music lessons enhance IQ. *Psychol. Sci.* **15**, 511–514 (2004).
- 671 6. Degé, F. Music lessons and cognitive abilities in children: How far transfer could be possible. *Front.*
672 *Psychol.* **11**, 557807 (2021).
- 673 7. Bigand, E. & Tillmann, B. Near and far transfer: Is music special? *Mem. Cognit.* **50**, 339–347 (2022).
- 674 8. Musacchia, G., Sams, M., Skoe, E. & Kraus, N. Musicians have enhanced subcortical auditory and
675 audiovisual processing of speech and music. *Proc. Natl. Acad. Sci.* **104**, 15894–15898 (2007).
- 676 9. Musacchia, G., Strait, D. & Kraus, N. Relationships between behavior, brainstem and cortical
677 encoding of seen and heard speech in musicians and non-musicians. *Hear. Res.* **241**, 34–42 (2008).
- 678 10. Parbery-Clark, A., Skoe, E. & Kraus, N. Musical experience limits the degradative effects of
679 background noise on the neural processing of sound. *J. Neurosci.* **29**, 14100–14107 (2009).
- 680 11. Parbery-Clark, A., Strait, D. L. & Kraus, N. Context-dependent encoding in the auditory brainstem
681 subserves enhanced speech-in-noise perception in musicians. *Neuropsychologia* **49**, 3338–3345
682 (2011).
- 683 12. Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T. & Kraus, N. Musical experience shapes human
684 brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* **10**, 420–422 (2007).

- 685 13. Bidelman, G. M., Krishnan, A. & Gandour, J. T. Enhanced brainstem encoding predicts musicians'
686 perceptual advantages with pitch. *Eur. J. Neurosci.* **33**, 530–538 (2011).
- 687 14. Bidelman, G. M. & Krishnan, A. Effects of reverberation on brainstem representation of speech in
688 musicians and non-musicians. *Brain Res.* **1355**, 112–125 (2010).
- 689 15. Coffey, E. B. J., Musacchia, G. & Zatorre, R. J. Cortical correlates of the auditory frequency-following
690 and onset responses: EEG and fMRI evidence. *J. Neurosci.* **37**, 830–838 (2017).
- 691 16. Bianchi, F. *et al.* Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of
692 neuroplasticity in musicians. *NeuroImage* **163**, 398–412 (2017).
- 693 17. Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., Baillet, S. & Zatorre, R. J. Neural correlates of early
694 sound encoding and their relationship to speech-in-noise perception. *Front. Neurosci.* **11**, 1–14
695 (2017).
- 696 18. Du, Y. & Zatorre, R. J. Musical training sharpens and bonds ears and tongue to hear speech better.
697 *Proc. Natl. Acad. Sci.* **114**, 13579–13584 (2017).
- 698 19. Schneider, P. *et al.* Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex
699 of musicians. *Nat. Neurosci.* **5**, 688–694 (2002).
- 700 20. Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S. & Zatorre, R. J. Cortical contributions to
701 the auditory frequency-following response revealed by MEG. *Nat. Commun.* **7**, 11070 (2016).
- 702 21. Coffey, E. B. J. *et al.* Evolving perspectives on the sources of the frequency-following response. *Nat.*
703 *Commun.* **10**, 5036 (2019).
- 704 22. Rodrigues, M. *et al.* Frequency following response and musical experience: A review. *J. Hear. Sci.* **9**,
705 9–16 (2019).
- 706 23. Skoe, E. & Kraus, N. Auditory brain stem response to complex sounds: a tutorial. *Ear Hear.* **31**, 302–
707 324 (2010).

- 708 24. Coffey, E. B. J., Mogilever, N. B. & Zatorre, R. J. Speech-in-noise perception in musicians: A review.
709 *Hear. Res.* **352**, 49–69 (2017).
- 710 25. Anderson, S., Parbery-Clark, A., White-Schwoch, T. & Kraus, N. Aging affects neural precision of
711 speech encoding. *J. Neurosci.* **32**, 14156–14164 (2012).
- 712 26. Corrigall, K. A., Schellenberg, E. G. & Misura, N. M. Music training, cognition, and personality. *Front.*
713 *Psychol.* **4**, 222 (2013).
- 714 27. Preacher, K. J., Rucker, D. D., MacCallum, R. C. & Nicewander, W. A. Use of the extreme groups
715 approach: A critical reexamination and new recommendations. *Psychol. Methods* **10**, 178–192
716 (2005).
- 717 28. Parbery-Clark, A., Anderson, S., Hittner, E. & Kraus, N. Musical experience offsets age-related delays
718 in neural timing. *Neurobiol. Aging* **33**, 1483.e1-1483.e4 (2012).
- 719 29. Zendel, B. R. & Alain, C. The influence of lifelong musicianship on neurophysiological measures of
720 concurrent sound segregation. *J. Cogn. Neurosci.* **25**, 503–516 (2013).
- 721 30. Zendel, B. R., West, G. L., Belleville, S. & Peretz, I. Musical training improves the ability to
722 understand speech-in-noise in older adults. *Neurobiol. Aging* **81**, 102–115 (2019).
- 723 31. Fleming, D., Belleville, S., Peretz, I., West, G. & Zendel, B. R. The effects of short-term musical
724 training on the neural processing of speech-in-noise in older adults. *Brain Cogn.* **136**, 103592 (2019).
- 725 32. Zendel, B. R. & Alain, C. Musicians experience less age-related decline in central auditory processing.
726 *Psychol. Aging* **27**, 410–417 (2012).
- 727 33. Parbery-Clark, A., Anderson, S., Hittner, E. & Kraus, N. Musical experience strengthens the neural
728 representation of sounds important for communication in middle-aged adults. *Front. Aging*
729 *Neurosci.* **4**, 1–12 (2012).
- 730 34. Whiteford, K. *et al.* Association of musical training with auditory and speech neural coding and
731 perception. (2019) doi:10.17605/OSF.IO/DUQ34.

- 732 35. Clinard, C. G., Tremblay, K. L. & Krishnan, A. R. Aging alters the perception and physiological
733 representation of frequency: Evidence from human frequency-following response recordings. *Hear.*
734 *Res.* **264**, 48–55 (2010).
- 735 36. Mankel, K. & Bidelman, G. M. Inherent auditory skills rather than formal music training shape the
736 neural encoding of speech. *Proc. Natl. Acad. Sci.* **115**, 13129–13134 (2018).
- 737 37. Law, L. N. C. & Zentner, M. Assessing musical abilities objectively: Construction and validation of the
738 Profile of Music Perception Skills. *PLoS ONE* **7**, e52508 (2012).
- 739 38. Strauss, H., Reiche, S., Dick, M. & Zentner, M. Online assessment of musical ability in 10 minutes:
740 Development and validation of the Micro-PROMS. *Behav. Res. Methods* (2023) doi:10.3758/s13428-
741 023-02130-4.
- 742 39. Whiteford, K. L., Goh, P. Y., Stevens, K. L. & Oxenham, A. J. Dissociating sensitivity from bias in the
743 Mini Profile of Music Perception Skills. *JASA Express Lett.* **3**, 094401 (2023).
- 744 40. Patel, A. D. Why would musical training benefit the neural encoding of speech? The OPERA
745 hypothesis. *Front. Psychol.* **2**, (2011).
- 746 41. Ioannidis, J. P. A. . Why most published research findings are false. *PLoS Med.* **2**, e124 (2005).
- 747 42. Button, K. S. *et al.* Power failure: Why small sample size undermines the reliability of neuroscience.
748 *Nat. Rev. Neurosci.* **14**, 365–376 (2013).
- 749 43. Luck, S. J. in *An introduction to the event-related potential technique* 309–340 (MIT Press, 2014).
- 750 44. Manta, O. *et al.* Development and evaluation of automated tools for auditory-brainstem and
751 middle-auditory evoked potentials waves detection and annotation. *Brain Sci.* **12**, 1675 (2022).
- 752 45. Trübutschek, D. *et al.* EEGManyPipelines: A large-scale, grassroots multi-analyst study of
753 electroencephalography analysis practices in the wild. *J. Cogn. Neurosci.* **36**, 217–224 (2024).

- 754 46. Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E. & Johnsrude, I. S. Attentional modulation of
755 envelope-following responses at lower (93–109 Hz) but not higher (217–233 Hz) modulation rates. *J.*
756 *Assoc. Res. Otolaryngol.* **19**, 83–97 (2018).
- 757 47. Shinn-Cunningham, B., Varghese, L., Wang, L. & Bharadwaj, H. Individual differences in temporal
758 perception and their implications for everyday listening. in *The Frequency-Following Response* (eds.
759 Kraus, N., Anderson, S., White-Schwoch, T., Fay, R. R. & Popper, A. N.) vol. 61 159–192 (Springer
760 International Publishing, Cham, 2017).
- 761 48. Bidelman, G. M. Subcortical sources dominate the neuroelectric auditory frequency-following
762 response to speech. *NeuroImage* **175**, 56–69 (2018).
- 763 49. Bharadwaj, H. M. *et al.* Non-invasive assays of cochlear synaptopathy – candidates and
764 considerations. *Neuroscience* **407**, 53–66 (2019).
- 765 50. Rentfrow, P. J. The role of music in everyday life: Current directions in the social psychology of
766 music. *Soc. Personal. Psychol. Compass* **6**, 402–416 (2012).
- 767 51. Whiteford, K. L., Kreft, H. A. & Oxenham, A. J. Assessing the role of place and timing cues in coding
768 frequency and amplitude modulation as a function of age. *J. Assoc. Res. Otolaryngol.* **18**, 619–633
769 (2017).
- 770 52. Moscicki, E. K., Elkins, E. F., Baur, H. M. & McNamara, P. M. Hearing loss in the elderly: An
771 epidemiologic study of the Framingham Heart Study Cohort. *Ear Hear.* **6**, 184–190 (1985).
- 772 53. John, O. P., Donahue, E. M. & Kentle, R. L. *The Big Five Inventory -- Versions 4a and 54*. (University of
773 California, Berkeley, Institute of Personality and Social Research, Berkeley, CA, 1991).
- 774 54. John, O. P., Naumann, L. P. & Soto, C. J. Paradigm Shift to the Integrative Big-Five Trait Taxonomy:
775 History, Measurement, and Conceptual Issues. in *Handbook of personality: Theory and research* (eds.
776 John, O. P., Robins, R. W. & Pervin, L. A.) 114–158 (Guilford Press, New York, NY, 2008).
- 777 55. Boersma, P. & Weenink, D. PRAAT. (2004).

- 778 56. Aiken, S. J. & Picton, T. W. Envelope and spectral frequency-following responses to vowel sounds.
779 *Hear. Res.* **245**, 35–47 (2008).
- 780 57. JASP Team. JASP (Version 0.18.3). (2024).
- 781 58. van Doorn, J. *et al.* The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon.*
782 *Bull. Rev.* **28**, 813–826 (2021).
- 783