CellPress

Current Biology

Article

Language experience predicts music processing in a half-million speakers of fifty-four languages

Jingxuan Liu,[1,2,5,8,*] Courtney B. Hilton,[3,4,6,8,*] Elika Bergelson,[2] and Samuel A. Mehr[3,4,7,9,*]
[1]Columbia Business School, Columbia University, 665 W 130th Street, New York, NY 10027, USA
[2]Department of Psychology & Neuroscience, Duke University, 417 Chapel Drive, Durham, NC 27708, USA
[3]Yale Child Study Center, Yale University, 300 George Street #900, New Haven, CT 06511, USA
[4]School of Psychology, University of Auckland, 23 Symonds Street, Auckland 1010, New Zealand
[5]Twitter: @LiuJingxuan02
[6]Twitter: @courtneybhilton
[7]Twitter: @samuelmehr
[8]These authors contributed equally
[9]Lead contact
*Correspondence: jl6297@gsb.columbia.edu (J.L.), courtney.hilton@auckland.ac.nz (C.B.H.), sam@yale.edu (S.A.M.)
https://doi.org/10.1016/j.cub.2023.03.067

SUMMARY

Tonal languages differ from other languages in their use of pitch (tones) to distinguish words. Lifelong experience speaking and hearing tonal languages has been argued to shape auditory processing in ways that generalize beyond the perception of linguistic pitch to the perception of pitch in other domains like music. We conducted a meta-analysis of prior studies testing this idea, finding moderate evidence supporting it. But prior studies were limited by mostly small sample sizes representing a small number of languages and countries, making it challenging to disentangle the effects of linguistic experience from variability in music training, cultural differences, and other potential confounds. To address these issues, we used web-based citizen science to assess music perception skill on a global scale in 34,034 native speakers of 19 tonal languages (e.g., Mandarin, Yoruba). We compared their performance to 459,066 native speakers of other languages, including 6 pitch-accented (e.g., Japanese) and 29 non-tonal languages (e.g., Hungarian). Whether or not participants had taken music lessons, native speakers of all 19 tonal languages had an improved ability to discriminate musical melodies on average, relative to speakers of non-tonal languages. But this improvement came with a trade-off: tonal language speakers were also worse at processing the musical beat. The results, which held across native speakers of many diverse languages and were robust to geographic and demographic variation, demonstrate that linguistic experience shapes music perception, with implications for relations between music, language, and culture in the human mind.

INTRODUCTION

From infancy and early childhood, we are surrounded by people speaking and singing.[1–9] This immersion continues throughout the lifespan and is reinforced through our own language and music production.

Human perception readily adapts to these soundscapes: early speech experiences tune our hearing to the speech contrasts of our native language(s),[10–12] and musical experiences during the same time period are thought to have similar "perceptual narrowing" effects, biasing listeners' interpretations of musical rhythm and pitch based on their own musical cultures.[13,14] These effects may cross domains. While music training has minimal causal effects on high-level cognitive skills,[15,16] it may sharpen lower-level aspects of speech processing[17,18] and auditory perception.[19] In the opposite direction, enhanced experience with the types of linguistic pitch used in tonal languages has been argued to shape pitch processing in music.[20–22]

Here, we study the latter possibility, to examine the effects of language experience on music processing, with a focus on pitch. Languages are often classified into three distinct categories based on their use of pitch: tonal, non-tonal, and pitch-accented. While all spoken languages convey information via pitch, tonal languages, which represent over half the world's languages (including many East Asian, Southeast Asian, and African languages),[23] use pitch lexically at the word level. The consonants and vowels of a tonal language syllable are pronounced in conjunction with different pitch levels or shapes to signify different meanings.[24,25] A canonical example is the Mandarin syllable *ma*, which has different meanings depending on its tonal contour (i.e., level, rising, falling-rising, or falling). This property requires pitch sensitivity in both speakers and listeners, lest one scold (*mà*) one's mother (*mā*) instead of one's horse (*mǎ*).

The lexical use of pitch in tonal languages is distinct from how pitch is otherwise used in speech. For example, many languages use pitch to convey affect[26]; to cue non-lexical meaning (e.g., helping to differentiate between questions and statements)[27,28]; to emphasize information[29]; to cue sentence structure with metrical stress patterns,[30] supporting comprehension[31]; and/or as a cue to speech categories, as in infant- or child-directed speech.[32] While these many uses of pitch are typical of speech in both tonal and non-tonal languages (e.g., many Indo-European, South Asian, or Australian languages),[33] in non-tonal languages, pitch is never used lexically to denote word meanings. Last, pitch-accented languages form an intermediate category with limited or mixed use of lexical pitch (e.g., Croatian)[25]; note, however, that whether pitch-accented languages form a coherent standalone category or whether they are better considered on a spectrum between tonal and non-tonal languages, with some mixed cases, is a matter of debate.[34–36] As pitch-accented languages are not our primary focus, here we treat them as a separate group from tonal and non-tonal languages, but also conduct some analyses at the language level rather than the language-type level.

The special role of pitch in tonal languages has motivated the hypothesis that speaking a tonal language sharpens pitch perception in a domain-general fashion. Indeed, compared to speakers of non-tonal languages, native speakers of at least some East and/or Southeast Asian tonal languages not only better discriminate the tones of their native language and those of other tonal languages they do not speak,[37–39] but also may have stronger categorical perception for non-speech pitch patterns generally. Speakers of East/Southeast Asian tonal languages also have distinct neural responses to pitch in brain areas associated with early auditory processing.[37,40–43]

Might domain-general auditory processing advantages transfer to enhanced pitch processing in music? Many studies have tested this question by comparing native speakers of tonal and non-tonal languages on a variety of musical pitch perception tasks. Some studies report that tonal language speakers excel at discriminating melodic patterns[20,44–49] or at discerning fine-grained pitch difference either in isolation or in the context of detuned intervals, contours, and melodies.[21,46,50,51] But other studies fail to replicate these patterns, both for melodic discrimination[52–54] and fine-scale pitch discrimination.[22,41,49,54–56] Some studies even find that tonal language speakers have *more* trouble distinguishing musical pitch contours, suggesting that lexical tone experience could interfere with pitch perception in some contexts.[52,53,55,57]

More generally, because the vast majority of participants in these studies were native speakers of a small number of tonal and non-tonal languages from two non-overlapping geographic areas (East Asia for tonal languages, with most participants being native speakers of Mandarin or Cantonese; North America for non-tonal languages, with most participants being English speakers), it remains unclear whether patterns of results across these studies reflect effects of tonal versus non-tonal language experience in general, effects of growing up in an East Asian versus Western culture, or some interaction between the two.

In this paper, we first assess the current degree of evidence, via meta-analysis, for an effect of tonal language experience on music processing. Then, we report new data from a massive online experiment that recruited a global sample, to directly measure the relation between linguistic experience and music perception across many tonal, non-tonal, and pitch-accented languages.

RESULTS

Meta-analytic effects of tonal language experience on music processing ability

We aggregated summary information from 20 prior studies of music perception and tonal language experience (see STAR Methods for the search criteria) and studied them with random-effects meta-analysis models. Because there is no consensus in the literature about what specific music processing ability to expect a tonal language advantage on, we grouped the prior studies into three rough categories: melody, fine-grained pitch, and rhythm. We built separate meta-analytic models for each group.

The results are in Figure 1. The overall effect size estimate suggests that native speakers of tonal languages have a statistically significant advantage in melodic processing (standardized mean difference: 0.501, 95% CI = [0.192, 0.81], $p = 0.004$). Native tonal language speakers, however, did not differ statistically from non-tonal speakers for either fine-grained pitch processing (standardized mean difference: 0.262, 95% CI = [−0.015, 0.538], $p = 0.062$) or rhythm processing (standardized mean difference: −0.008, 95% CI = [−0.14, 0.125], $p = 0.893$). The meta-analyses also identified three serious concerns, however, which preclude any generalized claim about the effects of tonal language experience on music processing ability.

First, and most importantly, prior studies sample tonal languages narrowly, typically comparing Mandarin or Cantonese speakers from mainland China to English speakers from the United States. Of the tonal language speakers in prior studies, approximately 92% spoke Mandarin or Cantonese, and of the non-tonal speakers, approximately 85% spoke English. As such, no claim about the effects of language experience on music perception on the basis of the prior literature is justifiable because it is not clear whether prior effects generalize beyond a few frequently studied languages.

Second, the majority of prior studies have low statistical power, due their small sample sizes, which may produce unreliable group-level estimates of effects and increase the risk of bias.[58] We estimated the power of each prior study to detect effects of $d = 0.5$, a "medium" size effect comparable to the meta-analytic effect estimated for melodic discrimination tasks (Table 1). Across the three categories of music processing tasks, power was low (for studies of melodic discrimination, median power = 0.49; for fine-grained pitch discrimination, median power = 0.29; for rhythm tasks, median power = 0.48).

Third, participants' musical training experience has rarely been accounted for in the meta-analyzed studies. At best, this contributes additional unsystematic variation within a sample, reducing statistical power. Access to musical training and the form of this musical training, however, may also vary systematically between countries,[59] potentially leading to biased estimates of music perception abilities. Because the vast majority of participants in prior studies of tonal language experience and music perception ability include native speakers of two languages (i.e., Mandarin and Cantonese) from one country, this

Study	Measure	Tonal Sample	Non−tonal Sample	Power	Estimate	Weight
Ngo et al. (2016)	Melody combined	8 Vietnamese	8 English	15%	1.71 [0.56, 2.86]	4%
Chen et al. (2016)	Musical Ear Test	42 Mandarin	42 Dutch	62%	0.96 [0.51, 1.41]	9%
Choi (2021)	Melodic subset	30 Cantonese	30 English	48%	0.94 [0.41, 1.47]	8%
Alexander et al. (2008)	Melodic discrimination	14 Mandarin	14 English	25%	0.90 [0.13, 1.68]	6%
Hove et al. (2010)	Relative pitch recognition	10 'Chinese'; 14 Hmong	14 'Caucasian US students'	30%	0.68 [0.00, 1.35]	7%
Swaminathan et al (2021)	Melody test	104 Chinese languages; 3 Vietnamese	366 English; 42 various non−tonal languages	100%	0.54 [0.32, 0.75]	12%
Wong et al. (2012)	Out−of−key detection	408 Cantonese	154 English/French	100%	0.51 [0.32, 0.70]	12%
Bradley (2016)	Melodic comparison	26 Mandarin; 15 Yoruba	26 English	50%	0.49 [−0.01, 1.00]	9%
Chen et al. (2016)	Melody combined	42 Mandarin	42 Dutch	62%	0.44 [0.01, 0.88]	9%
Stevens et al. (2013)	Interval discrimination	24 Thai	24 English	40%	0.20 [−0.37, 0.76]	8%
Zheng & Samuel (2018)	Melody comparison	24 Mandarin	24 English, 24 Korean	51%	−0.13 [−0.62, 0.36]	9%
Peretz et al. (2011)	Melody combined	18 Mandarin; 4 Vietnamese; 2 Cantonese	25 'non−tonal' (likely French/English)	40%	−0.48 [−1.05, 0.09]	8%
					0.50	
Pfordresher & Brown (2009), Study 1	Interval discrimination	6 Vietnamese; 4 Mandarin; 2 Cantonese	12 English	22%	1.16 [0.30, 2.03]	4%
Bidelman et al. (2013)	F0 difference limens	18 Cantonese	18 English	31%	1.01 [0.31, 1.70]	5%
Bidelman et al. (2011), Study 2	Chordal detuning discrimination	5 Mandarin	5 English	11%	0.92 [−0.39, 2.23]	2%
Pfordresher & Brown (2009), Study 2	Interval discrimination	11 Mandarin	10 English; 1 Portuguese	20%	0.84 [−0.03, 1.71]	4%
Hutka et al. (2015)	F0 difference limens	18 Cantonese	21 English	33%	0.80 [0.15, 1.45]	5%
Giuliano et al. (2011)	Pitch discrimination	16 Mandarin	16 'Non−tonal'	29%	0.65 [−0.05, 1.35]	5%
Bent et al. (2006)	Nonspeech pitch discrimination	13 Mandarin	13 English	22%	0.59 [−0.21, 1.39]	4%
Tong et al. (2018)	Pitch height discrimination task	15 Cantonese	15 English	26%	0.56 [−0.17, 1.29]	5%
Bidelman et al. (2013)	Melody discrimination	18 Cantonese	18 English	31%	0.44 [−0.23, 1.12]	5%
Wong et al. (2012), Study 2	Mistuning detection	22 Cantonese	26 English/French	39%	0.42 [−0.15, 1.00]	5%
Jasmin et al. (2021)	Pitch Discrimination	31 Cantonese	36 English; 26 Spanish	61%	0.41 [−0.03, 0.84]	6%
Bidelman et al. (2011)	F0 discrimination	11 Mandarin	11 English	20%	0.37 [−0.48, 1.21]	4%
Stevens et al. (2013)	Frequency discrimination	24 Thai	24 English	40%	0.35 [−0.22, 0.92]	5%
Giuliano et al. (2011)	Interval discrimination	16 Mandarin	16 'Non−tonal'	28%	0.27 [−0.43, 0.98]	5%
Zheng & Samuel (2018)	Pitch precision	24 Mandarin	24 English, 24 Korean	51%	0.12 [−0.37, 0.62]	6%
Wong et al. (2012), Study 1	Mistuned detection	408 Cantonese	154 English/French	100%	0.00 [−0.19, 0.19]	7%
Pfordresher & Brown (2009), Study 2	Note discrimination	11 Mandarin	10 English; 1 Portuguese	20%	−0.18 [−1.02, 0.66]	4%
Pfordresher & Brown (2009), Study 1	Note discrimination	6 Vietnamese; 4 Mandarin; 2 Cantonese	12 English	22%	−0.21 [−1.02, 0.59]	4%
Tong et al. (2018)	Pitch interval discrimination task	15 Cantonese	15 English	26%	−0.37 [−1.09, 0.35]	5%
Peretz et al. (2011)	Pitch change detection	18 Mandarin; 4 Vietnamese; 2 Cantonese	25 'non−tonal' (likely French/English)	40%	−0.48 [−1.05, 0.09]	5%
Chang et al. (2016)	Melodic tone discrimination	17 Mandarin	19 English	31%	−1.64 [−2.40, −0.88]	4%
					0.26	
Ngo et al. (2016)	Rhythm	8 Vietnamese	8 English	15%	0.80 [−0.22, 1.82]	1%
Peretz et al. (2011)	Rhythm	18 Mandarin; 4 Vietnamese; 2 Cantonese	25 'non−tonal' (likely French/English)	40%	0.24 [−0.32, 0.81]	5%
Choi (2021)	Rhythm subset	30 Cantonese	30 English	48%	0.03 [−0.47, 0.54]	6%
Swaminathan et al (2021)	Rhythm test	104 Chinese languages; 3 Vietnamese	366 English; 42 various non−tonal languages	100%	0.01 [−0.20, 0.22]	33%
Wong et al. (2012), Study 1	Off−beat detection (Study1)	408 Cantonese	154 English/French	100%	−0.02 [−0.21, 0.16]	43%
Wong et al. (2012), Study 2	Off−beat detection (Study2)	22 Cantonese	26 English	39%	−0.20 [−0.77, 0.37]	5%
Jasmin et al. (2021)	Duration Discrimination	31 Mandarin	62 English and Spanish	61%	−0.22 [−0.65, 0.22]	8%
					−0.01	

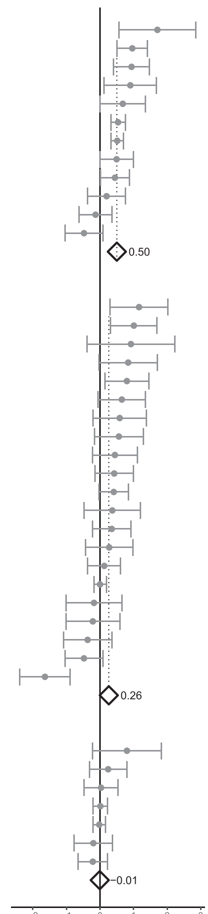

Figure 1. Meta-analytic evidence for the effects of tonal language experience on music processing
Meta-analytic effects estimated for melody processing, fine-grained pitch processing, and rhythm processing, aggregated from 20 prior studies. Results show a statistically significant advantage in melody processing for speakers of tonal languages compared to non-tonal languages ($p = 0.005$) of roughly half a standard deviation in size. Fine-grained pitch processing and rhythm processing did not show statistically significant differences ($p > 0.05$). These findings are, however, limited by the lack of diversity in sampling of languages and countries, and mostly small sample sizes. The table contains detailed information about the study and measures used, and the language composition of the participants in each study is included in the table. The figure embedded within the table shows the estimated effects for each study individually (modeled as random effects), with 95% confidence intervals indicated by the cross bars. The bolded diamond-shaped points represent the overall fixed-effect estimates for each of the three study categories (melody, fine-grained pitch, and rhythm).

issue presents a substantial threat to the validity of prior findings: any systematic biases could produce effects erroneously attributed to differences in tonal language experience rather than cultural experience.

Thus, the meta-analysis demonstrates evidence for a potential effect of tonal language experience on melodic discrimination ability and finer-grained pitch discrimination ability, but a lack of linguistic diversity, low statistical power, and high likelihood of systematic bias in the samples studied warrant caution and call into question the reproducibility and generalizability of these findings.[60,61] These issues can be addressed by studying many native speakers of many languages, from different countries, with and without music training experience, all of whom complete the same assessments of music processing ability.

Citizen-science experiment

We report such a test here, using methods of gamification and citizen science[62–65] to recruit 493,100 people from 203 countries across the globe, including 34,034 native speakers of 19 tonal languages, 16,868 native speakers of 6 pitch-accented languages, and 442,198 native speakers of 29 non-tonal languages (Figure 2; see STAR Methods for detailed information about the sample).

Tonal language experience shapes music processing

Native speakers of tonal languages had a reliable advantage in melodic discrimination compared to speakers of non-tonal and pitch-accented languages (Figure 3; full statistical reporting is in Table 1), with an effect size of substantive practical significance ($\beta = 0.216$, $t = 4.681$, $p < 0.001$), roughly half the size of the effect of having taken music lessons (an experience that one should reasonably expect to directly improve music perception ability). This result replicates the first meta-analytic result, previously shown mainly in Mandarin and Cantonese speakers compared to English speakers (see Meta-analytic effects of tonal language experience on music processing ability section),

Table 1. Fixed effects from random-effects models for each of the three musical tasks

	Melodic discrimination			Mistuning perception			Beat alignment		
Term	β	SE	*t*	β	SE	*t*	β	SE	*t*
Language: Tonal	0.216	0.046	4.681***	−0.087	0.044	−1.961	−0.225	0.032	−7.035***
Language: Pitch-accented	0.009	0.059	0.156	0.03	0.058	0.523	0.091	0.039	2.333*
Music lessons: Yes	0.501	0.003	147.329***	0.417	0.003	149.319***	0.274	0.003	86.837***
Age	0.004	0.000	25.854***	0.003	0.000	21.329***	−0.002	0.000	−15.374***
Gender: Male	0.107	0.003	34.964***	−0.042	0.003	−16.706***	0.142	0.003	50.006***
Gender: Other	0.044	0.012	3.707***	−0.072	0.01	−7.328***	0.046	0.011	4.095***
Tonal × music lessons	−0.032	0.013	−2.448*	−0.016	0.011	−1.539	0.059	0.012	4.833***
Pitch-accented × music lessons	0.084	0.017	5.058***	0.055	0.014	4.009***	−0.009	0.015	−0.587
Intercept	−0.196	0.026	−7.56***	0.113	0.025	4.526***	−0.022	0.018	−1.229

The models estimate the effect of natively speaking a tonal or pitch-accented language on each of the three music perception tests, relative to speaking a non-tonal language, while adjusting for age, gender, whether the person has had music lessons, and the interaction between having had music lessons and the effect of language type. Random intercepts for the participant's native language and their country of residence are also included in the model and visualized in Figure 4. The focal effects of speaking a tonal language for each of the musical tests also robustly replicate when additionally adjusting for income (Table S3), education (Table S4), or world region (Table S5). ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

demonstrating that the tonal language advantage for melodic discrimination generalizes to many *additional* languages.

Also consistent with the meta-analysis, there was no clear advantage for tonal language speakers on the vocal mistuning task ($\beta = -0.087$, $t = -1.961$, $p = 0.055$; Table 1), despite the fact that the task *did* reliably demonstrate clear performance differences between those with and without music training ($\beta = 0.417$, $t = 149.319$, $p < 0.001$). In other words, while some experiences (e.g., musical training) do shape fine-grained pitch perception in the context of vocal mistuning, there seems to be minimal or no effect of tonal language experience on this ability.

Last, contrary to the meta-analytic results, which found no effect of tonal language experience on rhythm perception, native speakers of tonal languages performed *worse* in beat perception relative to the non-tonal and pitch-accented groups ($\beta = -0.225$, $t = -7.035$, $p < 0.001$). Like the melodic discrimination effect, the size of this effect was large, approaching the size of effects of having taken music lessons (Figure 3).

Language-type effects are consistent across languages

To examine the degree to which the main effects held across the different tonal languages studied, we first examined language-wise estimates derived from the mixed-effects model. These showed a high degree of consistency in main effects within each language group (Figure 4). For example, on the melodic discrimination task, 19 of 19 tonal language groups had an estimated advantage over the non-tonal language average ($p < 0.05$), whereas 19 of 19 had an estimated disadvantage on the beat perception task ($p < 0.05$).

We then tested whether speakers of tonal and non-tonal languages could be distinguished on the basis of only their music perception scores, using a permuted linear discriminant function analysis[66] (Figure S2; STAR Methods). The results robustly replicated the main findings. Speakers of tonal and non-tonal languages could be reliably distinguished on average for both melodic discrimination ($p < 0.001$) and beat alignment ($p = 0.004$) tasks, but not for the mistuning perception task ($p = 0.099$).

We then repeated these analyses to compare pitch-accented to non-tonal languages. Here, the approach failed to replicate the small advantage to pitch-accented language speakers on the beat alignment task ($p = 0.059$) identified by our mixed-effect modeling analysis (Table 1) but did find a small advantage for vocal mistuning ($p = 0.045$). We take this mixed result as cause to not interpret either effect for speakers of pitch-accented languages, given both the less clear theoretical basis for such an effect and the noted ambiguity of the pitch-accented language type.[36]

Effects of tonal language experience are not attributable to measured third variables

We tested whether the observed language-type effects were driven by systematic variability across participants of three types.

First, we tested whether the socioeconomic status of participants was associated with their task scores, as socioeconomic status may vary systematically with both country and native language in our data, and may mediate the opportunities a person has for musical development. For participants who reported living in the United States, we collected additional demographic information. To assess how socioeconomic status may mediate the main findings, we analyzed this subset of participants who reported income information ($n = 82{,}727$). Despite being restricted to the United States, this sub-sample was nonetheless linguistically diverse, including 3,335 speakers of 14 tonal languages, 189 speakers of 6 pitch-accented languages, and 79,203 speakers of 29 non-tonal languages. We analyzed these data using a mixed-effect model of the same structure used in the main analysis, except with an additional fixed-effect term for income. The results show that while income does positively predict performance, such that those with higher incomes tended to perform better, the tonal language melodic discrimination advantage and beat alignment disadvantage held robustly after accounting for these income effects (full statistical reporting is in Table S3).

Second, we examined whether systematic variation in education between speakers of different languages may have

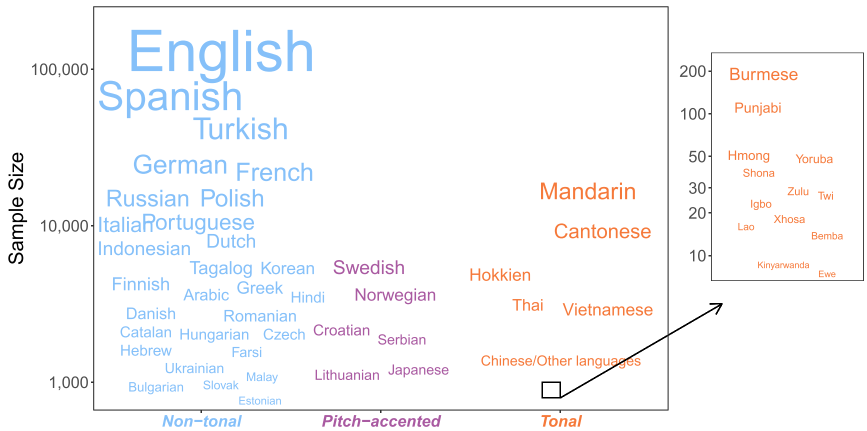

Figure 2. Sample sizes for each language, grouped by language type

The font of each language's name is scaled proportionally to that language's sample size. Horizontal positions are jittered to improve readability. The right panel shows additional tonal languages with smaller sample sizes included in the analyses to increase the diversity of the tonal language sample.

driven the observed differences in their performance. Among participants who reported their educational background (n = 477,906), we again conducted a mixed-effect analysis of the same structure as the main analysis with an additional fixed term for participants' education level. While education positively predicted performance on all three tasks, the main language-type effects replicated (Table S4).

Last, we tested whether proximity to "Western" culture could explain the main findings, as the mistuning perception and beat alignment task stimuli used Western-style popular music; while globalized musical styles make this unlikely, it is possible that the degree of familiarity with this musical style systematically varies between tonal and non-tonal language speakers, which could confound the main findings. We compared the results of the main model in two subsets of the participants (combined n = 211,256): those who resided in a primarily non-English-speaking Eastern country (China, Taiwan, Hong Kong, Thailand, and Vietnam) versus an English-speaking Western country (United States, United Kingdom, New Zealand, Australia, and Canada). The main language-type effects replicated, and the results (Table S5) showed only a small, significant effect of region (West versus East) on the mistuning perception task.

Taken together, these exploratory analyses show that while socioeconomic status, education, and exposure to Western culture can correlate with participant performance, these potential confounders cannot account for the observed language-type effects reported here.

DISCUSSION

We found a clear link between linguistic experience and music processing abilities: native speakers of tonal languages performed better than native speakers of non-tonal languages on a task that required discriminating changes in melodic patterns and worse on a task requiring the perception of a beat. In each case, the effect size associated with being a tonal language speaker was roughly half as large as the effect of receiving music lessons, indicating an effect of substantive practical significance. There was no effect of language experience on the perception of fine-grained pitch, however, despite the fact that musical training was reliably associated with increased ability for this task. This is consistent with the fact that tonal languages do not require distinguishing fine-grained pitch patterns.

Our results are likely to generalize, given that they held across thousands of native speakers of 19 tonal languages and hundreds of thousands of native speakers of 29 non-tonal languages, each sampled from over 100 countries around the world. The tonal languages include not only East Asian languages (e.g., Mandarin, Cantonese) commonly examined by prior studies, but also Southeast Asian and African languages (e.g., Burmese, Igbo, Shona) that have rarely or never been studied in the context of music perception research. The non-tonal languages in our study include speakers of languages as diverse as Arabic, Catalan, Hindi, and Ukrainian, along with non-tonal languages from South/Southeast Asia, such as Indonesian, Tagalog, and Malay. Inclusion of these Asian languages constitutes an especially strong test, since their speakers may be more likely to share cultural similarities to speakers of tonal languages in those regions, thereby helping to rule out cultural confounds.

Our results also help clarify the previously mixed pattern of results concerning the effects of linguistic experience on music processing across different tasks and samples. For example, an advantage for tonal language speakers in melodic pattern processing is consistent with the majority of previous studies,[20,44–46,48,49] though not all of them.[52–54] Meanwhile, similar levels of performance on the fine-grained pitch task between tonal and non-tonal language speakers are supported by some prior studies,[41,53–56] but not other studies that point to a tonal advantage/disadvantage on related tasks.[21,22,50–52,57] Last, while rhythmic abilities in tonal language speakers have been less studied,[49,67] a disadvantage in beat discrimination is consistent with recent work showing that tonal speakers give more weight to pitch cues than duration cues; this weighting cuts across auditory domains.[68] By leveraging a consistent set of tasks and a large sample size, our results make clear that speaking a tonal language has a measurable connection to music perception ability—but not a uniform one, as this linguistic experience produces both positive and negative effects on perceptual abilities of different types.

Why might tonal language experience have these specific effects on music perception? The answer may lie in the shared mechanisms and neural processing resources associated with auditory perception, whether applied to language or music.[17,19,21,27,69–71] Both tonal languages and music rely on specialized pitch-based sound categories (tone contours or

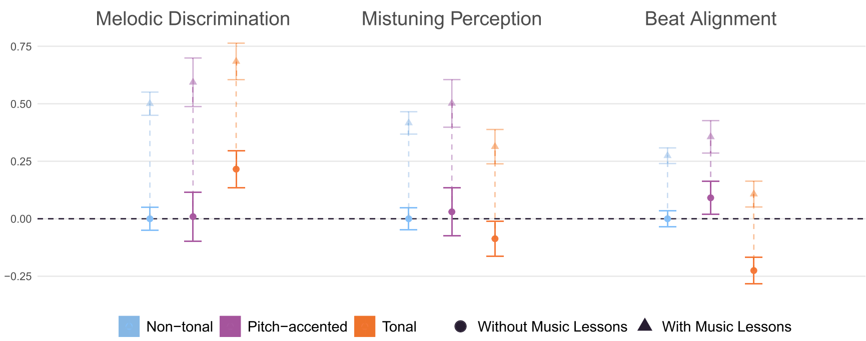

Figure 3. Native speakers of tonal languages have an advantage in melodic discrimination and a disadvantage in beat alignment

The solid dots show the estimated effects of language type, marginalizing over the average proportions for ages and gender, on each of the three music perception abilities tested, but without the estimated effects of having had music lessons. The additive marginal effects of having music lessons are displayed with the faded triangles, providing a comparison point for the language-type effect sizes. After marginalization, for ease of interpretability, a scalar transformation was applied to the coefficients such that the "Non-tonal" and "No music lessons" coefficients were equal to zero. Error bars represent 95% confidence intervals of the mean. The dotted horizontal black line indicates the baseline ($y = 0$). Full statistical reporting of the model's untransformed fixed effects is in Table 1.

levels in speech; pitch motifs and melodies in music). If these categories are learned and processed through shared, domain-general mechanisms, then improving the efficiency of these mechanisms through practice in either domain should result in mutual improvements.[27,57,69,72–74] But not all is shared. Music relies more on fine-grained pitch structure, even compared to tonal languages, as it is required to support the processing of pitch in the context of a tonal hierarchy.[75–78] This may explain why tonal language experience did not have an effect on fine-grained pitch processing in music.

Our results do not explain, however, *how* these shared mechanisms might be improved by experience. One possibility is that language experience could shape domain-general perceptual strategies regarding inferences about high-level perceptual categories on the basis of low-level cues: acquired perceptual biases (i.e., from tonal language experience) may aid the processing of some stimuli while worsening the processing of others. In speech, listeners give more perceptual weight to cues that are more informative in discriminating contrasts that are salient in their native language,[68,79] and tonal language speakers rely more heavily on pitch to categorize and produce speech stress when acquiring a non-tonal L2 language compared to native speakers.[80,81]

Similarly, people with pitch perception deficits learn to compensate for their deficits by giving more weight to durational cues when decoding speech prosody.[68,82] Recent evidence suggests that similar effects emerge in music perception: Mandarin speakers have difficulty *ignoring* pitch cues relative to English and Spanish speakers, who have been found to more frequently make decisions based on duration cues.[68] In turn, this is consistent with theories of the overlapping mechanisms of basic auditory perception.[17,18,70,83,84] Our findings unite these results and show their generality.

While the scope of our data collection allowed for analysis of music processing abilities in thousands of native speakers of six pitch-accented languages, the findings concerning these speakers were murky and failed to replicate across different analysis approaches. Within-language group variability was also high (Figure 4) for pitch-accented languages, suggesting no common group advantage/disadvantage across speakers of pitch-accented languages. This, of course, is complicated by the inherently fuzzier nature of classification of pitch-accented languages, relative to tonal languages.[34–36] Further work that codifies that nature of both linguistic and musical pitch use across this set of generally understudied languages may provide more clarity. We encourage interested readers to re-analyze the open-access data reported here, using alternative classifications of languages in the "pitch-accented" category.

We note several other limitations. First, while we accounted for how much musical training participants had, we did not measure how long they engaged with this training, its intensity, or its type. As a result, our estimates of the effect of musical training have greater uncertainty (although the analyses for participants with no musical training, which largely replicate the main effects, help to mitigate this concern). Second, participants only reported their first language, so we were unable to examine the effects of bilingualism or multilingualism,[85–87] or assess whether fluently speaking both tonal and non-tonal languages (e.g., Mandarin and English) might have contributed additional variability in our results. Third, there are a host of other unmeasured cultural, environmental, and genetic factors that surely affect musical abilities. Moreover, these likely interact with each other, complicating causal inferences from the observational data we collected (see, e.g., recent findings that genetics and musical experience both influence linguistic tone perception in Cantonese).[88] These and other limitations will be best addressed through a variety of methodologies, including more targeted, smaller-scale approaches (including fieldwork) that complement the broad web-based citizen-science approach used here.

In sum, our results show that across a range of geographic and demographic contexts, linguistic experience alters music perception ability in reliable (but not unitary) fashions. This implies that substantively different domains of auditory perception recruit at least some shared processing resources, which themselves are shaped by auditory experience.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

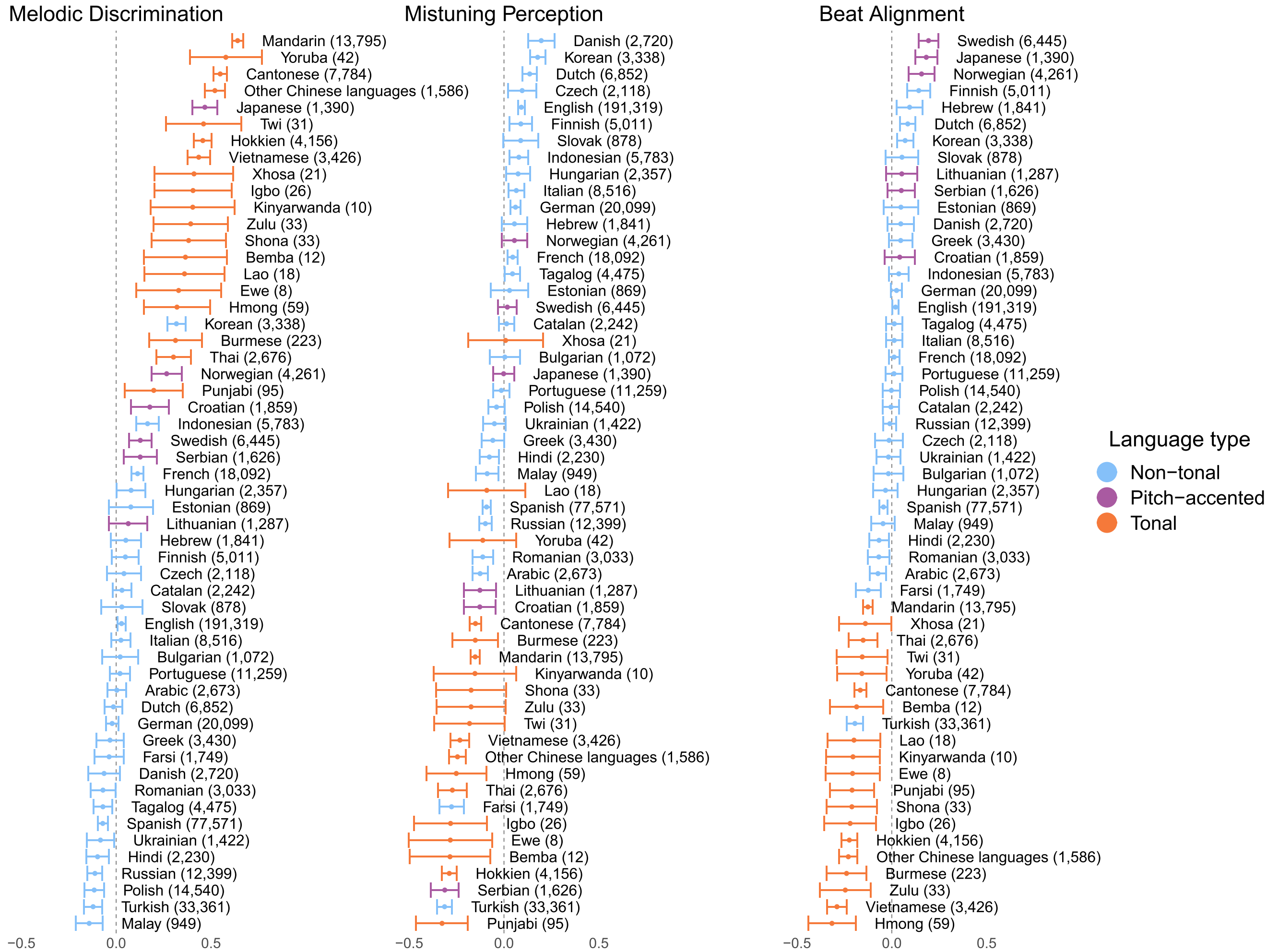

Figure 4. The main effects are consistent across 19 tonal languages

The forest plot displays the estimated average performance for each individual language, after adjusting for the effect of music lessons, age, and gender. The solid points denote random-effect estimates for each language, derived from the mixed-effects models reported in Table 1 and Figure 3; the error bars denote 95% confidence intervals; and the text annotations specify the languages and language-wise sample sizes (in brackets).

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
  - Validation of self-reported headphone use
- METHOD DETAILS
  - Meta-analysis selection criteria
  - Experimental design and materials
  - Melodic discrimination task
  - Mistuning perception task
  - Beat alignment task
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Meta-analysis data processing and analysis
  - Analysis strategy for experimental results
  - Permuted linear discriminant function analysis
  - Results from the preregistered analysis approach

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.03.067.

ACKNOWLEDGMENTS

This research was supported by the Duke Career Center Internship Funding Program for Undergraduate Students (J.L.), the Harvard Data Science Initiative (S.A.M.), the National Institutes of Health Director's Early Independence Award DP5OD024566 (S.A.M. and C.B.H.), the Royal Society of New Zealand Te Apārangi Rutherford Discovery Fellowship RDF-UOA2103 (S.A.M.), and the Royal Society of New Zealand Te Apārangi Marsden Fund Standard Grant MFP-UOA2133 (C.B.H. and S.A.M.). We thank the participants; P. Harrison and D. Müllensiefen for sharing code and assisting with the implementation of their music perception tasks; T. Bent, G. Bidelman, E. Bradley, A. Bradlow, D. Chang, W. Choi, R. Giuliano, M. Hove, S. Hutka, M. Ngo, S. Nguyen, I. Peretz, P. Pfordresher, S.X. Tong, S. Swaminathan, Y. Wang, N. Wicha, and L. Zhang for providing supplementary data and/or associated information for the meta-analysis; J. Simson for technical and research assistance; the members of The Music Lab for discussion and feedback on the citizen-science platform, the experiment, and the manuscript; and three anonymous reviewers, whose constructive feedback shaped this paper substantially.

AUTHOR CONTRIBUTIONS

Conception, J.L. and S.A.M.; experimental design and implementation, S.A.M.; preregistration and planned analyses, J.L., C.B.H., E.B., and S.A.M.; participant recruitment, data management, and data processing, S.A.M., C.B.H., and J.L.; analysis and visualization, J.L. and C.B.H., with contributions from E.B. and S.A.M.; meta-analysis data collection, J.L.; meta-analysis models, J.L. and C.B.H.; writing, J.L., C.B.H., E.B., and S.A.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 18, 2021
Revised: February 8, 2023
Accepted: March 23, 2023
Published: April 26, 2023

REFERENCES

1. Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., and Tor, S. (2019). Day by day, hour by hour: naturalistic language input to infants. Dev. Sci. *22*, e12715.
2. Eibl-Eibesfeldt, I. (1979). Human ethology: concepts and implications for the sciences of man. Behav. Brain Sci. *2*, 1–26.
3. Mehr, S.A., Singh, M., Knox, D., Ketter, D.M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A.A., Hopkins, E.J., et al. (2019). Universality and diversity in human song. Science *366*, 957–970.
4. Mehr, S.A., Krasnow, M.M., Bryant, G.A., and Hagen, E.H. (2020). Origins of music in credible signaling. Behav. Brain Sci. *44*, 1–41.
5. Bonneville-Roussy, A., Rentfrow, P.J., Xu, M.K., and Potter, J. (2013). Music through the ages: trends in musical engagement and preferences from adolescence through middle adulthood. J. Pers. Soc. Psychol. *105*, 703–717.
6. Konner, M. (2010). The Evolution of Childhood: Relationships, Emotion, Mind (Belknap Press of Harvard University Press).
7. Mendoza, J.K., and Fausey, C.M. (2021). Everyday music in infancy. Dev. Sci. *24*, e13122.
8. Mehr, S.A. (2014). Music in the home: new evidence for an intergenerational link. J. Res. Music Educ. *62*, 78–88.
9. Yan, R., Jessani, G., Spelke, E.S., de Villiers, P., de Villiers, J., and Mehr, S. (2021). Across demographics and recent history, most parents sing to their infants and toddlers daily. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20210089.
10. Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. Nat. Rev. Neurosci. *5*, 831–843.
11. Werker, J.F., and Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behav. Dev. *7*, 49–63.
12. Polka, L., and Werker, J.F. (1994). Developmental changes in perception of nonnative vowel contrasts. J. Exp. Psychol. Hum. Percept. Perform. *20*, 421–435.
13. Hannon, E.E., and Trehub, S.E. (2005). Metrical categories in infancy and adulthood. Psychol. Sci. *16*, 48–55.
14. Lynch, M.P., Eilers, R.E., Oller, D.K., and Urbano, R.C. (1990). Innateness, experience, and music perception. Psychol. Sci. *1*, 272–276.
15. Sala, G., and Gobet, F. (2020). Cognitive and academic benefits of music training with children: a multilevel meta-analysis. Mem. Cognit. *48*, 1429–1441.
16. Mehr, S.A., Schachner, A., Katz, R.C., and Spelke, E.S. (2013). Two randomized trials provide no consistent evidence for nonmusical cognitive benefits of brief preschool music enrichment. PLoS One *8*, e82007.
17. Patel, A.D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. Front. Psychol. *2*, 142.
18. Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. Nat. Neurosci. *10*, 420–422.
19. Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. Nat. Rev. Neurosci. *11*, 599–605.
20. Bradley, E.D. (2016). Phonetic dimensions of tone language effects on musical melody perception. Psychomusicology *26*, 337–345.
21. Bidelman, G.M., Hutka, S., and Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. PLoS One *8*, e60676.
22. Pfordresher, P.Q., and Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. Atten. Percept. Psychophys. *71*, 1385–1398.
23. Yip, M. (2002). Tone (Cambridge University Press).
24. Pike, K.L. (1948). Tone Languages (University of Michigan Press).
25. van der Hulst, H. (2011). Pitch accent systems. In The Blackwell Companion to Phonology (Wiley-Blackwell), pp. 1–24.
26. Cowen, A.S., Laukka, P., Elfenbein, H.A., Liu, R., and Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. Nat. Hum. Behav. *3*, 369–382.
27. Patel, A.D. (2008). Music, Language, and the Brain (Oxford University Press).
28. Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., and Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. Neuroimage *28*, 417–428.
29. Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. Lang. Cognit. Process. *25*, 1044–1098.
30. Wagner, M., and McAuliffe, M. (2019). The effect of focus prominence on phrasing. J. Phonetics *77*, 100930.
31. Hilton, C.B., and Goldwater, M.B. (2021). Linguistic syncopation: Meter-syntax alignment affects sentence comprehension and sensorimotor synchronization. Cognition *217*, 104880.
32. Hilton, C.B., Moser, C.J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C.M., Simson, J., Knox, D., Glowacki, L., Alemu, E., et al. (2022). Acoustic regularities in infant-directed speech and song across cultures. Nat. Hum. Behav. *6*, 1545–1556.
33. Maddieson, I. (2013). Tone. In The World Altas of Language Structures Online.
34. Gussenhoven, C. (2004). The Phonology of Tone and Intonation (Cambridge University Press).
35. Hyman, L.M. (2006). Word-prosodic typology. Phonology *23*, 225–257.
36. Hyman, L.M. (2007). How (not) to do phonological typology: the case of pitch-accent. UC Berkeley Phonology Lab Annual Reports *3*, 213–238.
37. Krishnan, A., Gandour, J.T., and Bidelman, G.M. (2010). The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguistics *23*, 81–95.
38. Li, X., and Gao, Z. (2018). The effect of language experience on lexical tone perception. J. Acoust. Soc. Am. *144*, 1865.
39. Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., and Wang, W.S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. J. Phonetics *38*, 616–624.
40. Bidelman, G.M., Gandour, J.T., and Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. J. Cogn. Neurosci. *23*, 425–434.
41. Bidelman, G.M., Gandour, J.T., and Krishnan, A. (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. *77*, 1–10.

42. Bidelman, G.M., and Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. Neuroimage *120*, 191–200.

43. Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cogn. Brain Res. *25*, 161–168.

44. Alexander, J.A., Bradlow, A.R., Ashley, R.D., and Wong, P.C.M. (2008). Music melody perception in tone-language- and nontone-language speakers. J. Acoust. Soc. Am. *124*, 2495.

45. Choi, W. (2021). Musicianship influences language effect on musical pitch perception. Front. Psychol. *12*, 712753.

46. Chen, A., Liu, L., and Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. Lang. Cogn. Neurosci. *31*, 751–760.

47. Ngo, M.K., Vu, K.-P.L., and Strybel, T.Z. (2016). Effects of music and tonal language experience on relative pitch performance. Am. J. Psychol. *129*, 125–134.

48. Swaminathan, S., Kragness, H.E., and Schellenberg, E.G. (2021). The Musical Ear Test: norms and correlates from a large sample of Canadian undergraduates. Behav. Res. Methods *53*, 2007–2024.

49. Wong, P.C.M., Ciocca, V., Chan, A.H.D., Ha, L.Y.Y., Tan, L.-H., and Peretz, I. (2012). Effects of culture on musical pitch perception. PLoS One *7*, e33424.

50. Giuliano, R.J., Pfordresher, P.Q., Stanley, E.M., Narayana, S., and Wicha, N.Y.Y. (2011). Native experience with a tone language enhances pitch discrimination and the timing of neural responses to pitch change. Front. Psychol. *2*, 146.

51. Hutka, S., Bidelman, G.M., and Moreno, S. (2015). Pitch expertise is not created equal: cross-domain effects of musicianship and tone language experience on neural and behavioural discrimination of speech and music. Neuropsychologia *71*, 52–63.

52. Peretz, I., Nguyen, S., and Cummings, S. (2011). Tone language fluency impairs pitch discrimination. Front. Psychol. *2*, 145.

53. Zheng, Y., and Samuel, A.G. (2018). The effects of ethnicity, musicianship, and tone language experience on pitch perception. Q. J. Exp. Psychol. *71*, 2627–2642.

54. Stevens, C.J., Keller, P.E., and Tyler, M.D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. Psychol. Music *41*, 59–74.

55. Bent, T., Bradlow, A.R., and Wright, B.A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. J. Exp. Psychol. Hum. Percept. Perform. *32*, 97–103.

56. Tong, X., Choi, W., and Man, Y.Y. (2018). Tone language experience modulates the effect of long-term musical training on musical pitch perception. J. Acoust. Soc. Am. *144*, 690–697.

57. Chang, D., Hedberg, N., and Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. J. Acoust. Soc. Am. *139*, 2432–2447.

58. Kraemer, H.C., Gardner, C., Brooks, J.O., III, and Yesavage, J.A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. Psychol. Methods *3*, 23–31.

59. Campbell, P.S., and Wiggins, T. (2012). The Oxford Handbook of Children's Musical Cultures (Oxford University Press).

60. Blasi, D.E., Henrich, J., Adamou, E., Kemmerer, D., and Majid, A. (2022). Over-reliance on English hinders cognitive science. Trends Cogn. Sci. *26*, 1153–1170.

61. Yarkoni, T. (2022). The generalizability crisis. Behav. Brain Sci. *45*, e1.

62. Hilton, C.B., and Mehr, S.A. (2022). Citizen science can help to alleviate the generalizability crisis. Behav. Brain Sci. *45*, e21.

63. Huber, B., and Gajos, K.Z. (2020). Conducting online virtual environment experiments with uncompensated, unsupervised samples. PLoS One *15*, e0227629.

64. Hartshorne, J.K., de Leeuw, J.R., Goodman, N.D., Jennings, M., and O'Donnell, T.J. (2019). A thousand studies for the price of one: accelerating psychological science with Pushkin. Behav. Res. Methods *51*, 1782–1803.

65. Long, B., Simson, J., Buxó-Lugo, A., Watson, D.G., and Mehr, S.A. (2023). How games can make behavioural science better. Nature *613*, 433–436.

66. Mundry, R., and Sommer, C. (2007). Discriminant function analysis with nonindependent data: consequences and an alternative. Anim. Behav. *74*, 965–976.

67. Zhang, L., Xie, S., Li, Y., Shu, H., and Zhang, Y. (2020). Perception of musical melody and rhythm as influenced by native language experience. J. Acoust. Soc. Am. *147*, EL385–EL390.

68. Jasmin, K., Sun, H., and Tierney, A.T. (2021). Effects of language experience on domain-general perceptual strategies. Cognition *206*, 104481.

69. Asaridou, S.S., and McQueen, J.M. (2013). Speech and music shape the listening brain: evidence for shared domain-general mechanisms. Front. Psychol. *4*, 321.

70. Peretz, I., Vuvan, D., Lagrois, M.É., and Armony, J.L. (2015). Neural overlap in processing music and speech. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140090.

71. Asano, R., Boeckx, C., and Seifert, U. (2021). Hierarchical control as a shared neurocognitive mechanism for language and music. Cognition *216*, 104847.

72. Delogu, F., Lampis, G., and Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: may melodic ability improve learning of tonal languages by native nontonal speakers? Cogn. Process. *7*, 203–207.

73. Delogu, F., Lampis, G., and Belardinelli, M.O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. Eur. J. Cognit. Psychol. *22*, 46–61.

74. McMullen, E., and Saffran, J.R. (2004). Music and language: a developmental comparison. Music Perception *21*, 289–311.

75. Albouy, P., Benjamin, L., Morillon, B., and Zatorre, R.J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science *367*, 1043–1047.

76. Krumhansl, C.L. (2004). The cognition of tonality as we know it today. J. N. Music Res. *33*, 253–268.

77. Peretz, I., Champod, A.S., and Hyde, K. (2003). Varieties of musical disorders. The montreal battery of evaluation of amusia. Ann. N. Y. Acad. Sci. *999*, 58–75.

78. Zatorre, R.J., Belin, P., and Penhune, V.B. (2002). Structure and function of auditory cortex: music and speech. Trends Cogn. Sci. *6*, 37–46.

79. Schertz, J., and Clare, E.J. (2020). Phonetic cue weighting in perception and production. Wiley Interdiscip. Rev. Cogn. Sci. *11*, e1521.

80. Wang, Q. (2008). L2 stress perception: the reliance on different acoustic cues. In Proceedings of Speech Prosody, pp. 135–138.

81. Yu, V.Y., and Andruski, J.E. (2010). A cross-language study of perception of lexical stress in English. J. Psycholinguist. Res. *39*, 323–344.

82. Jasmin, K., Dick, F., Holt, L., and Tierney, A.T. (2020). Tailored perception: individuals' speech and music perception strategies fit their perceptual abilities. J. Exp. Psychol. Gen. *149*, 914–934.

83. Patel, A.D. (2012). The OPERA hypothesis: assumptions and clarifications. Ann. N. Y. Acad. Sci. *1252*, 124–128.

84. Tierney, A., Krizman, J., Skoe, E., Johnston, K., and Kraus, N. (2013). High school music classes enhance the neural processing of speech. Front. Psychol. *4*, 855.

85. Krizman, J., Marian, V., Shook, A., Skoe, E., and Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. Proc. Natl. Acad. Sci. USA *109*, 7877–7881.

86. Liu, L., and Kager, R. (2017). Enhanced music sensitivity in 9-month-old bilingual infants. Cogn. Process. *18*, 55–65.

87. Liu, L., Chen, A., and Kager, R. (2020). Simultaneous bilinguals who do not speak a tone language show enhancement in pitch sensitivity but not in executive function. Linguist. Approaches Biling. *12*, 310–346.

88. Wong, P.C.M., Kang, X., Wong, K.H.Y., So, H.-C., Choy, K.W., and Geng, X. (2020). ASPM-lexical tone association in speakers of a tone language: direct evidence for the genetic-biasing hypothesis of language evolution. Sci. Adv. *6*, eaba5090.

89. Harrison, P.M.C., Collins, T., and Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: item response theory, computerised adaptive testing, and automatic item generation. Sci. Rep. *7*, 3618.

90. Larrouy-Maestri, P., Harrison, P.M.C., and Müllensiefen, D. (2019). The mistuning perception test: a new measurement instrument. Behav. Res. Methods *51*, 663–675.

91. Harrison, P.M.C., and Müllensiefen, D. (2018). Development and validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). Sci. Rep. *8*, 12395.

92. de Leeuw, J.R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. Behav. Res. Methods *47*, 1–12.

93. Harrison, P. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. J. Open Source Softw. *5*, 2088.

94. Maddieson, I., Flavier, S., and Pellegrino, F. (2014). LAPSyD: Lyon-Albuquerque phonological systems databases. Version 1.0.

95. Moran, S., and McCloy, D. (2019). PHOIBLE 2.0 (Max Planck Institute for the Science of Human History).

96. Ameka, F.K. (2001). Ideophones and the nature of the adjective word class in Ewe. Typol. Stud. Lang. *44*, 25–48.

97. Bailey, L.M. (1988). A non-linear analysis of pitch accent in Swedish. Lingua *75*, 103–124.

98. Eme, C.A., and Odinye, I.S. (2022). Phonology of standard Chinese and Igbo: Implications for Igbo students learning Chinese. Nkuzi Omumu Asusu Journal *1*, 26–37.

99. Evans, J., Yeh, W.-C., and Kulkarni, R. (2018). Acoustics of tone in Indian Punjabi. Trans. Philol. Soc. *116*, 509–528.

100. Goldsmith, J., and Mpiranya, F. (2011). Rhythm, quantity and tone in the Kinyarwanda verb. In Tones and Features, J.A. Goldsmith, E. Hume, and L. Wetzels, eds. (De Gruyter), pp. 25–49.

101. Hamann, S., and Kula, N.C. (2015). Bemba. J. Int. Phonetic Assoc. *45*, 61–69.

102. Inkelas, S., and Zec, D. (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. Language *64*, 227–248.

103. Jefferies, A.A. (1990). Beyond tone: functions of pitch in Shona.

104. Manyah, K.A. (2006). Relation between tone and vowel quality in Twi. In International Symposium on Tonal Aspects of Languages.

105. Niesler, T., Louw, P., and Roux, J. (2005). Phonetic analysis of Afrikaans, English, Xhosa and Zulu using South African speech databases. South. Afr. Ling. Appl. Lang. Stud. *23*, 459–474.

106. van der Hulst, H., Goedemans, R., and van Zanten, E. (2010). A Survey of Word Accentual Patterns in the Languages of the World (De Gruyter Mouton).

107. Westermeyer, R., and Westermeyer, J. (1977). Tonal language acquisition among lao children. Anthropol. Ling. *19*, 260–264.

108. Woods, K.J.P., Siegel, M.H., Traer, J., and McDermott, J.H. (2017). Headphone screening to facilitate web-based auditory experiments. Atten. Percept. Psychophys. *79*, 2064–2072.

109. Borenstein, M., Hedges, L.V., Higgins, J.P.T., and Rothstein, H.R. (2009). Introduction to Meta-analysis (John Wiley & Sons).

110. Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using Lme4. J. Stat. Softw. *67*.

111. Austin, P.C., and Stuart, E.A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat. Med. *34*, 3661–3679.

112. Stuart, E.A. (2010). Matching methods for causal inference: a review and a look forward. Stat. Sci. *25*, 1–21.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data from the experiments and for the meta-analysis	This paper; Zenodo	https://doi.org/10.5281/zenodo.7614188
Software and algorithms		
Code for analysis and data visualization	This paper; Github	https://github.com/themusiclab/language-experience-music
Code for melodic discrimination task	Harrison et al., 2017[89]; Github	https://github.com/pmcharrison/mdt
Code for mistuning perception task	Larrouy-Maestri et al., 2019[90]; Github	https://github.com/pmcharrison/mpt
Code for beat alignment task	Harrison & Müllensiefen[91]; Github	https://github.com/pmcharrison/cabat
R 4.2.2	Comprehensive R Archive Network	https://cran.r-project.org/
jsPsych 6.1.0	de Leeuw 2015[92]	https://www.jspsych.org
psychTestR	Harrison, 2020[93]; Github	https://github.com/pmcharrison/psychTestR
Pushkin 0.0.1 (modified version)	Hartshorne et al., 2019[64]; Github	https://github.com/pushkin-consortium/pushkin
Other		
Preregistration	Open Science Framework	https://osf.io/xurdb
Web-based citizen science experiment	This paper	https://themusiclab.org/quizzes/miq

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the Lead Contact Samuel Mehr, sam@yale.edu.

Materials availability

Readers can try out the experiment at https://themusiclab.org/quizzes/miq; code for each of the three tasks is available at https://github.com/pmcharrison/mpt, https://github.com/pmcharrison/mdt, and https://github.com/pmcharrison/cabat.

Data and code availability

A reproducible version of this manuscript, including all data and code, is available on Github: https://github.com/themusiclab/language-experience-music. Code in this repository automatically downloads the raw data, which can also be found separately on Zenodo: https://doi.org/10.5281/zenodo.7614188. The preregistration can be found on OSF: https://osf.io/xurdb.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Participants were visitors to the citizen-science website https://themusiclab.org who completed a set of three music perception tasks presented as an online game (*Test Your Musical IQ*). We did not recruit participants directly; rather, they visited the site after hearing about it organically (e.g., via Reddit posts, YouTube clips, Twitch streams). Participants gave informed consent under an ethics protocol approved either by the Harvard University Committee on the Use of Human Subjects (protocol IRB2017-1206) or the Yale University Human Research Protection Program (protocol 2000033433). Our analysis was preregistered and deviations are noted in the "Results from the preregistered analysis approach" section. See "resource availability" for more details.

We studied 609,658 participants who completed all three tasks, who had no missing or internally conflicting data, and who reported being a native speaker of one of the 40 most commonly spoken languages among all participants *or* were a native speaker of one of 14 tonal languages that have been less commonly studied in the context of music perception, such as Yoruba, Xhosa, and Lao (n.b., in the preregistration, we initially planned to only study speakers of the 40 most commonly spoken languages in the available data. However, based on advice from reviewers, we subsequently opted to include additional participants to increase the diversity of the tonal languages studied). Data were collected between Nov 8th, 2019 and Nov 12th, 2022.

We excluded participants that (a) had participated in the experiment on another occasion, to avoid any effects of learning (n = 33,020); (b) reported a hearing impairment (n = 85,963); (c) reported their age as below 8 years or above 90 years (n = 1,515); (d) reported an age of music lessons onset that was either below 2 years or above 90 years (n = 1,383); (e) reported a music lesson onset age that was greater than their self-reported age ($n = 370$); (f) reported that they were participating in a noisy environment and were not wearing headphones ($n = 1{,}910$; see "validation of self-reported headphone use" section for analysis of a manipulation check to test whether participants were actually wearing headphones). 7,107 participants were excluded for meeting more than one of the above criteria. The exclusion criteria were preregistered.

The resulting sample ($n = 493{,}100$; see Figure 2 and Table S1) included 34,034 native speakers of 19 tonal languages; 16,868 native speakers of 6 pitch-accented languages; and 442,198 native speakers of 29 non-tonal languages.

Languages were primarily classified based on the World Atlas of Language Structures[33] and the Lyon-Albuquerque Phonological Systems Database.[94] Languages that are not present in either database were classified according to information from the Phonetics Information Base and Lexicon Database[95] or other sources from the linguistics literature.[96–107] A summary of all languages studied here, with further classification details and language-wise sample sizes, is in Table S1.

In addition to participants' music perception scores, we collected demographic information (gender, age, whether or not the participant had taken music lessons, and the age at onset of those lessons). These data are reported in Table S2.

Validation of self-reported headphone use

Participants who self-reported that they were wearing headphones completed a 6-trial headphone detection task[108] designed to be easy for participants wearing headphones and difficult for those listening on free-field speakers. Out of the 346,562 participants who indicated wearing headphones, 323,947 had clean and usable headphone detection data. The distribution of scores for these participants (Figure S1) was strongly left-skewed with the median participant scoring 5.14 of 6 (100%) correct. This implies that the bulk of participants who self-reported wearing headphones were, in fact, wearing headphones.

METHOD DETAILS

Meta-analysis selection criteria

We included in the meta-analysis only studies that examined the pitch-processing ability of native tonal and non-tonal language speakers via behavioral measures. We searched for studies on Google Scholar using the terms *(tone language OR tonal language) AND (musical pitch perception)* and inspected the first 200 results. In addition, we conducted forward and backward cross-referencing of a review article on the link between musical and linguistic pitch.[69] In the identified studies, we excluded those that (1) focused exclusively on absolute pitch, amusia, categorical perception, or cross-modal abilities (e.g., identifying visual representations of musical intervals); (2) studied only musicians; and/or (3) recruited only children under the age of 8. Only studies that were published and written in English were included. In all but one case, the participants studied were native speakers of the tonal language in question. One study[48] had some non-native speakers in the tonal language group. However, that study's inclusion or exclusion from the meta-analyses did not substantively affect the estimates of meta-analytic effect sizes.

Experimental design and materials

Participants completed three music perception tasks measuring ability in *melodic discrimination*,[89] *mistuning perception*,[90] and *beat alignment*.[91] As in the original tasks, each subtask was presented adaptively via psychTestR.[93] To minimize the duration of the experiment, we fixed the length of each subtask at 15 trials, the minimum number of trials with acceptably low mean standard errors, according to the original task designs. Demographic items were presented via jsPsych.[92] The citizen-science platform distributed the experiments using a modified pre-release version of pushkin.[64] Readers can try the three music perception tasks at https://themusiclab.org/quizzes/miq.

Melodic discrimination task

The melodic discrimination task assesses the ability to detect differences in relative pitch between melodic patterns via a three-alternative forced-choice design (3AFC). Participants listened to three transpositions of the same melody and were asked to choose the version in which a relative pitch interval was altered (the 'odd one out'). The materials for the melodic discrimination task were generated algorithmically to achieve precise theory-based gradations in task difficulty as predicted by a cognitive model of melodic processing.[89] These melodies were synthesized using MIDI with an identical piano timbre and a tempo of 120 beats per minute.

Mistuning perception task

The mistuning perception task assesses the ability to identify small-pitch differences in vocal tuning relative to other musical parts using a two-alternative forced-choice design (2AFC). Participants listened to two versions of a short musical excerpt, and in one of the excerpts the vocal track was detuned from the background music. Participants were asked to identify the detuned version. The materials for the mistuning perception task were derived from 37 royalty free excerpts of popular music songs that represent several styles. For each excerpt, the detuned versions were generated where the vocal part was pitch-shifted either up or down by an amount between 10 to 100 cents in 5-cent increments.[90]

Beat alignment task

The beat alignment task assesses the ability to detect synchronization between a click track and some music via a two-alternative forced-choice design (2AFC). Participants listened to two versions of the same musical excerpt, both accompanied by a click track.

One of the click tracks was misaligned by a constant proportion of the beat periodicity (where the proportion is within the range $0 < P \leq 0.5$, where P is the beat periodicity) and participants were asked to identify the track that was correctly aligned. The materials for the beat alignment task were derived from 32 excerpts from the 'Audio Network' production music library and consist of a variety of musical genres and meters. These tracks were then overlaid with the click track, consisting of sine tones with a frequency of 1000Hz and a duration of 20ms with a 10ms fade out.[91]

QUANTIFICATION AND STATISTICAL ANALYSIS

Meta-analysis data processing and analysis

To mitigate the confounding effect of musical training, we excluded data from musicians when separate groups of musicians/non-musicians were recruited. The remaining studies that did not distinguish between musicians and non-musicians had participants with either minimal musical training or did not differ across tonal/non-tonal groups (For Jasmin et al. (2021),[68] which studied several participants with lengthy exposure to musical training, we excluded those with more than three years of musical training and recalculated the relevant statistics). Additionally, we removed measures regarding long-term memory for melody, measures regarding memory for isolated pitches (rather than melodic patterns), or that used speed of processing as the dependent variable. For the remaining pitch-relevant tasks, we classified them into two categories: melodic pattern discrimination and fine-tuned pitch discrimination. Melodic discrimination includes tasks that involve recognizing different note combinations, while fine-tuned pitch discrimination includes tasks that concerns discerning fine-grained pitch differences. Additionally, we aggregated rhythm-related tasks (mostly used as control measures) across the studies into a separate rhythm category. Our classification scheme results in 12 melodic discrimination effect sizes, 20 fine-tuned pitch discrimination effect sizes, and 7 rhythm effect sizes. Mean and standard deviation or Cohen's *d* were collected for all the relevant tasks.

Data from studies/tasks that contained multiple tonal/non-tonal language groups (e.g., separate English and Korean groups) or reported their data at sub-task levels (e.g., 1/4 and 1/2 semitone for the pitch discrimination task) were further processed to produce a composite score for each tonal/non-tonal group or task. Specifically, for studies that contained multiple tonal/non-tonal groups, we used the formula $\overline{X}_{pooled} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$ and $SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2 + \frac{n_1 n_2}{n_1 + n_2}(\overline{X}_1 - \overline{X}_2)^2}{(n_1 + n_2 - 1)}}$ to calculate the combined mean and standard deviation for the group. Meanwhile, for studies that reported data at sub-task levels, we used the formula $\overline{X}_{pooled} = \frac{1}{n}\sum_{i=1}^{n}\overline{X}_i$ and $SD_{pooled} = \sqrt{\left(\frac{1}{n}\right)^2 \left(\sum_{i=1}^{n} Var_i + \sum_{i \neq j}(r_{ij}\sqrt{Var_i}\sqrt{Var_j}\right)}$ (assuming correlation between the sub-task conditions equals 1, since they are targeting the same ability) to calculate a combined mean and standard deviation for the task for tonal and non-tonal language speakers.[109] This step produced one entry for each task, which we treat as our unit of analysis.

From the mean and standard deviation, we calculated Hedge's *G* for each task and used it as our effect size measure. We then ran a random-effects model for each category of the tasks. Please refer to https://github.com/themusiclab/language-experience-music/blob/main/analysis/meta-analysis.Rmd for commented code of the analysis.

Analysis strategy for experimental results

To test whether musical abilities differ reliably as a function of native language type (i.e., tonal vs. pitch-accented vs. non-tonal), we used mixed-effects linear regression adjusting for age, gender, whether the participant had taken music lessons (yes or no), and the interaction between language-type and music-lessons; with random-effects for language and country. Non-tonal language, female gender, and no-music-lessons were used as the reference levels for the fixed-effects. Stated in terms of {lmer} pseudo-code,[110] this model is as follows:

$$\text{Performance} \sim \text{Language type} * \text{Music lessons} + \text{Gender} + \text{Age} + (1|\text{language}) + (1|\text{country})$$

The random-effect structure is particularly helpful for correcting for sampling imbalances, ensuring that no particular languages or countries can dominate the overall effect, while also allowing us to model variation across languages and countries directly (see, e.g., Figure 4).

Note that this analysis approach deviates from our preregistered analysis plan, which involved applying a linear regression model at several sampling levels, without random effects. We made this change on the suggestion of a reviewer and given the utility of mixed-effects models in measuring cultural variation.[62] For transparency, we report analyses and results from the preregistered approach in the 'Results from the preregistered analysis approach' section; these largely replicate the main results.

Permuted linear discriminant function analysis

To test whether speakers of tonal and non-tonal languages could be distinguished on the basis of only their music perception scores, we used a linear discriminant function analysis. To compensate for the multi-level structure of our data (scores nested within languages nested within language types), rather than a standard discriminant function analysis, we used a non-parametric permutational approach[66] and ran it on the subset of the participants who indicated having not received musical training and who were native speakers of languages with a sample size of at least 500. From this dataset, we drew nested random samples of 5 tonal languages and 5 non-tonal languages; for each of the sampled languages, we sampled 100 participant scores with replacement and shuffled the assignment of whether that participant was marked "tonal" or "non-tonal". We trained a linear discriminant function on 30% of that sample, balancing across languages, and then used the trained model to predict whether the remaining 70% of the held-out sample was marked "tonal" or "non-tonal", yielding a percentage-correct score. We repeated the process 10,000 times to construct the null distribution for each of the three music perception tasks (see Figure S2). To estimate the actual classification performance, the same nested sampling process was repeated 100 times for non-shuffled data, from which we obtained the mean proportions of correct classification for each task.

Results from the preregistered analysis approach

In our preregistration (https://osf.io/xurdb), we specified an exploratory-confirmatory approach. For both sets of data, we planned OLS regression models exploratory (n = 183,530) and confirmatory (n = 307,419) samples, controlling for age, gender, and music lesson (yes or no). In addition, to further reduce the confounding effect of covariates, we planned OLS regression models on three alternative samples, using different approaches to control for differences in music lesson experience, gender, and age (coarsened into 10-year bands). The three versions of the data were a 1:1 exactly matched sample, a 1:1 exactly matched sample with only participants who did not receive music lessons, and an inverse-probability weighted sample.[111,112] The same simple linear regression model *Performance* $\sim$ *Language type* was planned for each. Results from these exploratory and confirmatory analyses are presented in Table S6 for transparency. The main findings from the exploratory dataset replicated in the confirmatory dataset.

There are several limitations, however, that complicate the OLS results. First, the large sample size drove almost all effects to statistical significance in a way that may not translate to practical significance. Second, the imbalanced representation of different languages (e.g., dominance of English speakers in the non-tonal language group) may bias our estimates via cultural confounds endemic to the dominant languages in each group. These limitations motivated our adoption of mixed-effect models as the main analysis.