bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Language experience shapes music processing across 40 tonal, pitch-accented, and non-tonal languages

Jingxuan Liu[1,2], Courtney B. Hilton[2], Elika Bergelson[1], & Samuel A. Mehr[2,3,4,*]

[1]Department of Psychology & Neuroscience, Duke University, Durham, NC 27708, USA
[2]Department of Psychology, Harvard University, Cambridge, MA 02138, USA
[3]Data Science Initiative, Harvard University, Cambridge, MA 02138, USA
[4]School of Psychology, Victoria University of Wellington, Wellington 6012, New Zealand

*Corresponding author. E-mail: sam@wjh.harvard.edu

Tonal languages differ from other languages in their use of pitch (tones) to distinguish words. Some research suggests that the *linguistic* pitch expertise of tonal language speakers may generalize to improved discrimination of some aspects of *musical* pitch: tonal language speakers may therefore have music perception advantages over speakers of other languages. The evidence is mixed, however, as prior studies have studied small numbers of participants in only a few tonal languages and countries, making it challenging to disentangle the effects of linguistic experience from variability in music training experience, cultural differences, and so on. Here, we report an assessment of music perception skill in native speakers of 40 languages, with a preregistered exploratory-confirmatory design, including tonal (e.g., Mandarin, Vietnamese), pitch-accented (e.g., Japanese, Croatian), and non-tonal (e.g., Spanish, Hungarian) languages. Whether or not participants had taken music lessons, native speakers of tonal languages (confirmatory n = 20,102) had an improved ability to discriminate musical melodies. But this improvement came with a trade-off: relative to speakers of pitch-accented (confirmatory n = 9,694) or non-tonal languages (confirmatory n = 242,096), tonal speakers were worse at discriminating fine-scale pitch-tuning and worse at processing the musical beat. These results, which held across 5 tonal languages and were robust to geographic and demographic variation, demonstrate that linguistic experience shapes music perception ability, with implications for relations between music, language, and culture in the human mind.

1 Introduction

From infancy and early childhood, we are surrounded by people speaking (Bergelson et al., 2019; Eibl-Eibesfeldt, 1979; Konner, 2010) and singing (Bonneville-Roussy et al., 2013; Mehr et al., 2020, 2019; Mehr, 2014; Mendoza & Fausey, 2021; Yan et al., 2021). 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) (Kuhl, 2004; Polka & Werker, 1994; Werker & Tees, 1984), 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 (Hannon & Trehub, 2005; Lynch et al., 1990). These effects may cross domains. While music training has minimal causal effects on high-level cognitive skills (Mehr, 2013; Sala & Gobet, 2020), it may sharpen lower-level aspects of speech processing (Patel, 2011; Wong et al., 2007) and auditory perception (Kraus & Chandrasekaran, 2010). In the opposite direction, enhanced experience with the kind of linguistic pitch used in tonal languages has been argued to shape pitch processing in music (Bidelman et al., 2013; Bradley, 2016; Pfordresher & Brown, 2009).

Here, we study the latter possibility, to examine the effects of language experience on music processing, with a focus on pitch. Languages can be 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 Southeast Asian and African languages; Yip

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(2002)] use pitch in a special fashion. In tonal languages, pitch is often used lexically: speaking the same syllable with a different pitch level or shape alters meaning at the word level (Pike, 1948; van der Hulst, 2011). 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 (Cowen et al., 2019); to cue non-lexical meaning [e.g., helping to differentiate between questions and statements; Patel (2008); Tong et al. (2005)]; to emphasize information (Breen et al., 2010); to cue sentence structure with metrical stress patterns (Wagner & McAuliffe, 2019) supporting comprehension (Hilton & Goldwater, 2021); and is a prominent feature of infant-directed speech (Hilton et al., 2021). These many uses of pitch are typical of speech in non-tonal languages [e.g., many Indo-European, South Asian, or Australian languages; Maddieson (2013)], but in these languages, pitch is never used lexically to denote word meanings. A third group, *pitch-accented* languages, is an intermediate category with limited or mixed use of lexical pitch [such as Croatian; van der Hulst (2011)].[1]

The special role of pitch in tonal languages has motivated the hypothesis that speaking a tonal language sharpens pitch perception in a domain-general way. Indeed, compared to speakers of non-tonal languages, native speakers of tonal languages not only better discriminate the tones of their native language and those of other tonal languages they do not speak (Li & Gao, 2018; Peng et al., 2010), but also show stronger categorical perception for non-speech pitch patterns generally (Bent et al., 2006; Bidelman & Lee, 2015). Tonal language speakers also have distinct neural responses to pitch in brain areas associated with early auditory processing (Bidelman et al., 2011a; Bidelman et al., 2011b).

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. The results have been mixed (Table 1). Some studies report that tonal language speakers excel at discriminating pitch patterns in the form of melodies, intervals, and contours (Alexander et al., 2008; Bidelman et al., 2013; Bradley, 2016; Chen et al., 2016; Pfordresher & Brown, 2009; Wong et al., 2012); or at discerning fine-grained pitch difference either in isolation or in the context of detuned melodies (Bidelman et al., 2013; Chen et al., 2016; Giuliano et al., 2011; Hutka et al., 2015). But other studies fail to replicate these patterns, both for melodic discrimination (Giuliano et al., 2011; Stevens et al., 2013; Tong et al., 2018) and fine-scale pitch discrimination (Bent et al., 2006; Bidelman et al., 2011a; Pfordresher & Brown, 2009; Stevens et al., 2013; Tong et al., 2018). 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 (Bent et al., 2006; Chang et al., 2016; Peretz et al., 2011; Zheng & Samuel, 2018).

What explains these conflicting results? First, and most importantly, prior studies sample languages narrowly (i.e., typically comparing Mandarin or Cantonese speakers from mainland China to English speakers from the United States), with only a few exceptions (i.e., Yoruba speakers in Bradley, 2016; Thai speakers in Stevens et al., 2013; Dutch speakers in Chen et al., 2016). This, combined with generally small samples (median prior sample size per language group: $n = 25$; interquartile range: 16-42; minimum: $n = 11$; see Table 1) makes it difficult to rule out confounding effects from factors such as culture and genetics (Deutsch et al., 2006; Hove et al., 2010) or linguistic variability (Evans & Levinson, 2009). Generalization from particular speakers of specific languages to entire groups of languages (i.e., from studies of Mandarin speakers to claims concerning general effects of tonal language experience) require larger samples of both languages and participants (Hilton & Mehr, In Press; Yarkoni, 2020).

[1] Whether pitch-accented language 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 (e.g., Gussenhoven & Gussenhoven, 2004; Hyman, 2009, 2006). Indeed, a category of languages that groups evidently disparate languages together, such as Japanese and Swedish, may not be defensible. 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-group level. We encourage anyone interested in alternative groupings of languages to use our public data and code to re-analyze accordingly.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Language comparison	Effect	Measure	Study	Tonal n	Non-tonal n
Mandarin vs. English	+	Melodic discrimination	Alexander et al. (2008)	14	14
	+	Melodic comparison	Bradley (2016)	26	26
	+	Contour discrimination	Creel et al. (2018)	133	53
	+	Pitch discrimination	Giuliano et al. (2011)	17	16
	−	Melodic identification	Alexander et al. (2008)	14	14
	−	Nonspeech pitch contour identification	Bent et al. (2006)	13	12
	−	Melodic tone discrimination	Chang et al. (2016)	17	38
	∅	Nonspeech pitch discrimination	Bent et al. (2006)	13	12
	∅	F0 discrimination	Bidelman et al. (2011)	11	22
	∅	Chordal detuning discrimination	Bidelman et al. (2011)	11	22
	∅	Melodic tone identification	Chang et al. (2016)	17	38
	∅	Timbre discrimination	Creel et al. (2018)	133	53
	∅	Interval discrimination	Giuliano et al. (2011)	17	16
	∅	Melody comparison	Zheng & Samuel (2018)	60	60
	∅	Pitch precision	Zheng & Samuel (2018)	60	60
Mandarin or Cantonese (mostly) vs English (mostly)	+	Melody test	Swaminathan et al (2021)	107	409
	∅	Rhythm test	Swaminathan et al (2021)	107	409
Mandarin, Cantonese, or Vietnamese vs. English	+	Interval discrimination	Pfordresher & Brown (2009)	12	12
	∅	Note discrimination	Pfordresher & Brown (2009)	12	12
Mandarin vs. English or Portuguese	+	Interval discrimination	Pfordresher & Brown (2009)	11	11
	∅	Note discrimination	Pfordresher & Brown (2009)	11	11
Mandarin vs. English or Korean	−	Melody comparison	Zheng & Samuel (2018)	48	96
	∅	Pitch precision	Zheng & Samuel (2018)	48	96
Mandarin vs. Dutch	+	Montreal Battery of Evaluation of Amusia	Chen et al. (2016)	42	42
	+	Musical Ear Test	Chen et al. (2016)	42	42
Mandarin vs. Japanese	+	Melody discrimination	Zhang et al. (2020)	31	28
	−	Rhythm discrimination	Zhang et al. (2020)	31	28
Mandarin, Cantonese, or Vietnamese vs. Unspecified (mostly French)	−	Pitch change detection	Peretz et al. (2011)	14	25
	∅	Montreal Battery of Evaluation of Amusia	Peretz et al. (2011)	14	25
Cantonese vs. English	+	F0 difference limens	Bidelman et al. (2013)	18	36
	+	Pitch processing speed	Bidelman et al. (2013)	18	36
	+	Pitch memory	Bidelman et al. (2013)	18	36
	+	Melody discrimination	Bidelman et al. (2013)	18	36
	+	F0 DL task	Hutka et al. (2015)	18	42
	∅	Pitch height discrimination task	Tong et al. (2018)	30	30
	∅	Pitch interval discrimination task	Tong et al. (2018)	30	30
	∅	Mistuning detection	Wong et al. (2012)	22	26
	∅	Off-beat detection	Wong et al. (2012)	22	26
Cantonese vs. English or French	+	Out-of-key detection	Wong et al. (2012)	408	154
	∅	Off-beat detection	Wong et al. (2012)	408	154
	∅	Mistuned detection	Wong et al. (2012)	408	154
Vietnamese vs. English	+	Montreal Battery of Evaluation of Amusia	Ngo et al. (2016)	16	16
	∅	Cochran-Weiss-Shanteau index of expertise	Ngo et al. (2016)	16	16
Thai vs. English	∅	Interval discrimination	Stevens et al. (2013)	24	24
	∅	Frequency discrimination	Stevens et al. (2013)	24	24
Yoruba vs. English	+	Melodic comparison	Bradley (2016)	15	26

Table 1. Prior studies on the effects of language experience on music processing have mixed results. Key: Tonal-language advantages indicated by +; disadvantages by −; null results by 0.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Second, participants' musical training experience has rarely been accounted for in prior work. At best, this contributes additional unsystematic variation within a sample, reducing statistical power; but access to musical training may also vary systematically between countries (e.g., Campbell & Wiggins, 2012), potentially leading to biased estimates of music perception abilities. Such biases could erroneously be attributed to differences in tonal language experience, leading to contradictory findings across studies.

Third, the operationalization of "musical pitch perception" and its measurement in prior studies has varied widely, ranging from holistic aspects of melody apprehension to the discrimination of fine-scale pitch differences. This has made it difficult to find consensus on a general claim concerning the relation between music perception ability and language experience.

These issues can be addressed by studying many native speakers of many languages, with and without music training experience, all of whom complete the same assessments of music processing ability. Here, we report a study of the ability to discriminate melodies, detect mistuned singing, and detect misaligned beats, in 468,581 people. Participants self-reported their native language, location, demographics, and degree of musical training, enabling language-wise and language-type-wise analyses of each of these musical abilities and their relations to linguistic and musical experience.

2 Methods

2.1 Participants

Participants were visitors to the citizen-science website https://themusiclab.org who completed a set of three music perception tests 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). All participants gave informed consent under an ethics protocol approved by the Committee on the Use of Human Subjects, Harvard University's Institutional Review Board.

Of the 1,993,012 participants who started the experiment between Nov 8th, 2019 and Aug 21st, 2021, 716,855 completed the experiment. Those who participated in the first 5 months of data collection (n = 297,600, post-exclusion n = 196,689) were used in an exploratory dataset, analyses of which are reported in Text S1.1; these analyses formed the basis for a preregistration of confirmatory analyses, which is available at https://osf.io/xurdb.[2]

The remaining 404,306 participants formed the confirmatory dataset. Based on preregistered exclusion criteria, we removed 42,571 participants with missing and/or internally conflicting demographics data, resulting in 361,735 participants. Of these remaining participants, we then removed those that (a) had participated in the experiment on another occasion, to avoid any effects of learning (n = 22,297); (b) reported a hearing impairment (n = 47,640); (c) reported their age as below 8 years or above 90 years (n = 491); (d) reported an age of music lessons onset that was either below 2 years or above 90 years (n = 496); (e) reported a music lesson onset age that was greater than their self-reported age (n = 106); (f) reported that they were participating in a noisy environment and were not wearing headphones (n = 824; see Text S1.2 for analysis of a manipulation check to test whether participants were actually wearing headphones); or (g) more than one of the above (n = 5,808). Finally, given our planned analyses at the language level, we also excluded participants whose native language was not among the 40 most-represented languages in the exploratory dataset to ensure statistically meaningful comparisons were licensed (n = 12,181 participants; or 206 languages with a median of 9 participants per excluded language).

The native language of each participant was classified as *tonal*, *non-tonal*, or *pitch-accented* based on the Lyon-Albuquerque Phonological Systems Database (Maddieson et al., 2014) and the World Atlas of Language Structures (Maddieson, 2013). Languages that are not present in either database were classified according to information from the Phonetics Information Base and Lexicon Database (Moran & McCloy, 2019) or other

[2]Our preregistration noted that we would study participants whose data were collected between 28 April 2020 and 20 February 2021; we opted to include additional data collected up until the date of submission of this paper, to maximize the confirmatory sample size.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sources from the linguistics literature. No discrepancies were found across these sources; Table S1 contains information about the language classifications.

The post-exclusion confirmatory dataset, which we analyze for the remainder of this paper, thus contains data from 271,892 participants, representing 40 native languages (5 tonal, 6 pitch-accented, and 29 non-tonal; see Figure 1 and Table S1 for specific languages and associated sample sizes). Demographic information is in Table S2.

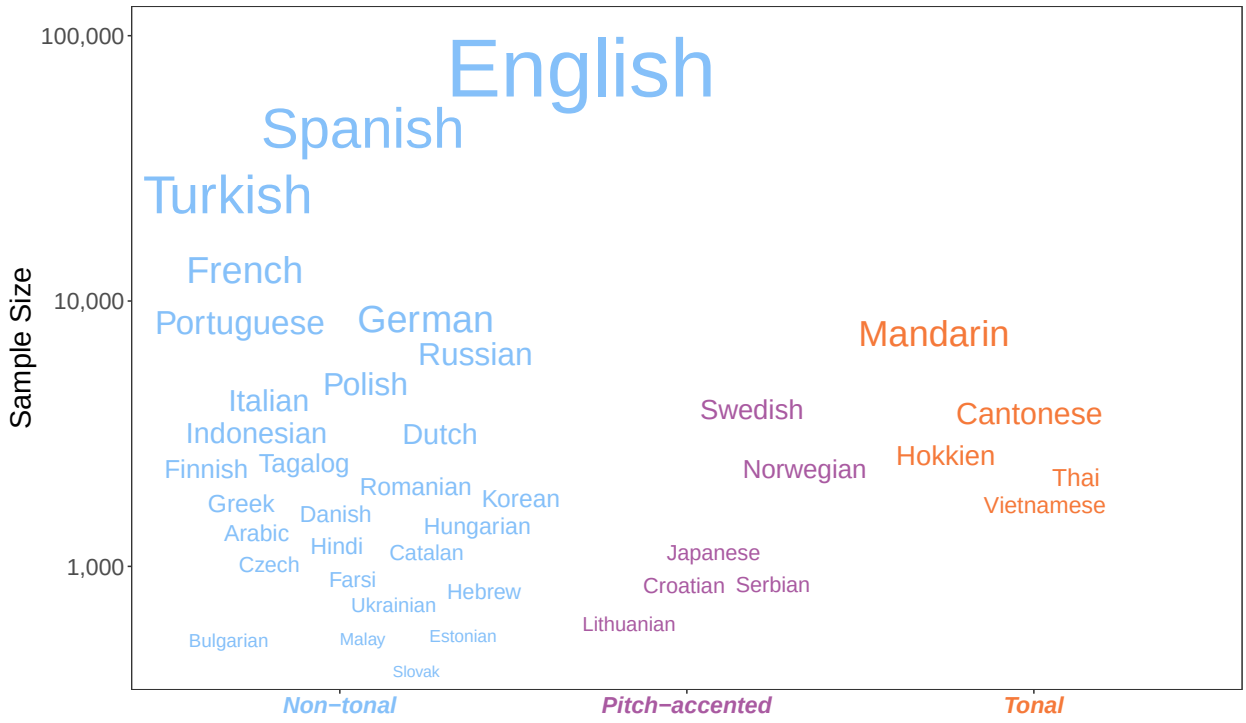

Figure 1. Sample sizes, grouped by language and language type. The font of each language's name is scaled proportionally to that language's sample size in the confirmatory dataset. Horizontal positions are jittered to improve readability.

2.2 Stimuli

Participants completed three music perception tests measuring ability in *melodic discrimination* (Harrison et al., 2017), *mistuning perception* (Larrouy-Maestri et al., 2019), and *beat alignment* (Harrison & Müllensiefen, 2018). The melodic discrimination assesses the ability to detect differences between melodic patterns: participants listened to three transpositions of the same melodic pattern and were asked to choose the version in which one pitch interval was altered (i.e., an oddball task). The mistuning perception test assesses the ability to identify vocal mistuning: participants listened to two versions of a short musical excerpt, one of which had a vocal track that was detuned from the background music, and were asked to identify the out-of-tune version. The beat alignment test assesses the ability to detect correct synchronization between a click track and some music: 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 and participants were asked which example was correctly aligned.

As in the original tests cited above, each of the subtests was presented adaptively using `psychTestR` (Harrison, 2020). To minimize the duration of the experiment, we fixed the length of each subtest at 15 trials, the minimum number of trials with acceptably low mean standard errors, according to the original test designs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3 Analysis strategy

The goal of our analyses was to determine whether musical abilities differ reliably as a function of native language type (i.e., tonal vs. pitch-accented vs. non-tonal). A simple comparison of scores across language types would have been confounded with the degree of musical experience sampled within each language, along with other cultural or personal factors. Instead, following our preregistration, we applied an ordinary-least-squares regression with measured potential-confounders as covariates to our full sample, modeled as $Performance \sim Language\,type + Gender + Age + Music\,lessons$, with non-tonal language, female, and no music lessons as the reference levels.

We then supplemented this model-based approach to dealing with confounding with a design-based approach: creating three additional versions of the data from the full sample, each controlling for confounding in different ways. The *exact match subsample* included participants matched 1-to-1 within each language type (tonal, pitch-accented, and non-tonal) on music lessons, gender, and age (coarsened into 10 year bins). The *no music lessons exact match subsample* further filtered the exact match subsample to only participants who had not received any music lessons. Finally, the *inverse probability weighted sample* included the total sample population, balanced on music lessons, gender, and age using inverse-probability weighting (Austin & Stuart, 2015; Stuart, 2010). These three subsamples were modeled with the same ordinary-least-squares regression: $Performance \sim Language\,type$ (with non-tonal language as the reference level). For brevity, and because these four different models provide robustly convergent evidence, we only report the full sample analysis in the main text; analysis details for each subsample model are reported in Text S1.3 and Tables S3 and S4a-c.

Based on the exploratory analyses (Text S1.1), we expected that native speakers of tonal languages would outperform native speakers of both pitch-accented and non-tonal languages on the melodic discrimination task; but would *underperform* both other groups on both the mistuning perception and beat alignment tasks.

As a final analytic note, because group-level comparisons in a very large sample risks yielding statistically significant effects of tiny size (i.e., practically non-significant), we also preregistered an inference criterion: in the main analyses, we compare the effects of language type *to the effects of having taken music lessons*. This provides a plainly interpretable benchmark: for instance, being a native speaker of a particular language type is associated with effects that will be interpreted as larger, of comparable size, or smaller than the effects of having taken music lessons.

3 Results

The main analyses (Table 2 and Figure 2) supported the exploratory predictions. First, and most strikingly, native speakers of tonal languages had reliable advantages in melodic discrimination, of large practical significance: the effect of tonal language experience on melodic discrimination ability is nearly as large as the effect of having taken music lessons. This effect replicated across all four of our methods to control for confounding (Text S1.3, Tables S3 and S4a-c).

The tonal-language advantage in melodic discrimination traded off with other music perception abilities, however. Native speakers of tonal languages had a practically and statistically significant disadvantage in beat alignment scores, relative to the non-tonal-language group (~1/4th the effect of having taken music lessons; Figure 2), and showed small but inconsistent disadvantages in mistuning perception scores (see also Text S1.3).

Native speakers of pitch-accented languages showed a melodic discrimination advantage over non-tonal speakers, but this effect was far smaller than that of the tonal-language speakers. They performed *better* than non-tonal-language speakers on both the mistuning perception and beat alignment tests, however, consistent with a tonal-language disadvantage on these two tasks. We caution, however, that the category of pitch-accented languages is inherently more ambiguous than that of tonal languages, so we do not interpret these differences further (see also Footnote 1).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Melodic Discrimination			Mistuning Perception			Beat Alignment		
Term	β	SE	t	β	SE	t	β	SE	t
Language: Tonal	0.407	0.007	54.773***	-0.088	0.006	-13.821***	-0.091	0.007	-12.729***
Language: Pitch-accented	0.097	0.010	9.284***	0.173	0.009	19.358***	0.196	0.010	19.565***
Music lessons: Yes	0.555	0.004	133.893***	0.526	0.004	148.547***	0.362	0.004	91.118***
Age	0.007	0.000	35.263***	0.006	0.000	32.883***	0.002	0.000	8.69***
Gender: Male	0.084	0.004	20.777***	-0.048	0.003	-13.903***	0.131	0.004	33.662***
Intercept	-0.320	0.007	-49.177***	-0.039	0.006	-6.985***	-0.186	0.006	-29.775***

Table 2. Regression results for each musical test in the confirmatory sample. ***$p < 0.001$. Note that an indicator variable for the gender "Other" was included as a predictor in the model, but is omitted from this table as it was rarely selected by participants.

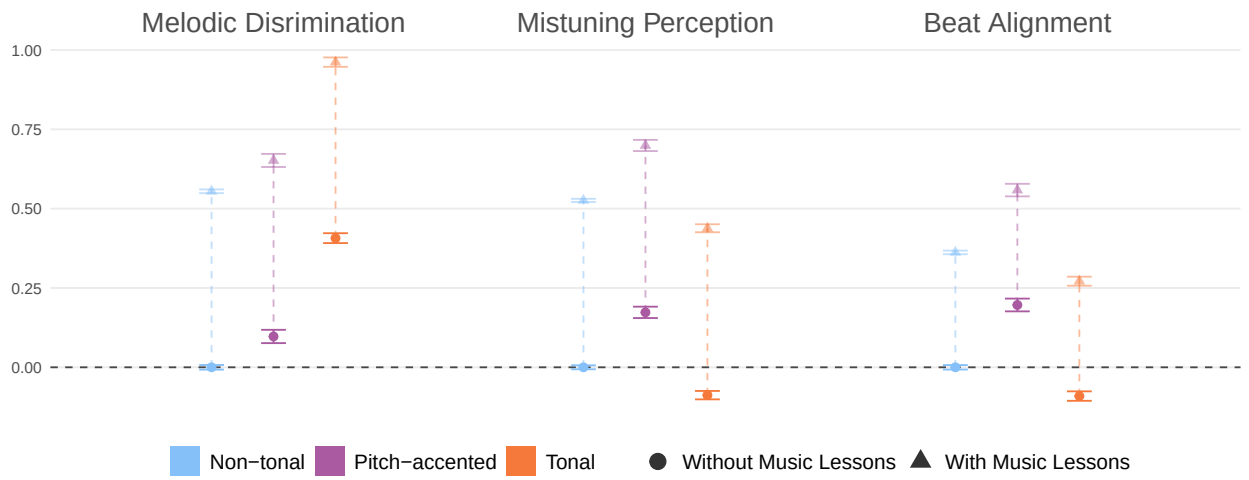

Figure 2. Estimated effects of language type (tonal language: orange, pitch-accented language: purple, non for-tonal language: blue), marginalizing over the average proportions for ages and gender, and shown relative to the marginal effect of additionally having music lessons. After marginalization, for ease of interpretability, a scalar transformation was applied to the coefficients such that "Non-tonal" and "No music lesson" coefficient was equal to zero. Solid circular points indicate marginal estimates of speaking a given language, semi-transparent triangular points above indicate this effect combined with that of having music lessons. Error-bars represent 95% confidence intervals of the mean. The dotted horizontal black line indicates the baseline (y = 0).

In follow-up analyses, we explored the consistency of language-type effects. First, we probed the stability of test performance within speakers of each of the 40 languages studied here, by comparing the mean performance of all speakers of each language across the exploratory and confirmatory datasets. This provides a measure of the stability in the relationship between performance and individual languages. Stability was high (Figure 3): mean language-wise test performance in the exploratory dataset correlated highly with that of the confirmatory dataset (melodic discrimination: $r = 0.92$; beat alignment: $r = 0.91$; mistuning detection: $r = 0.87$; ps < 0.001).

We then tested the consistency of the main effects across the languages that make up each language group (Figure 4). For melodic discrimination, of the 10 highest-performing languages, half were tonal (i.e., all 5 tonal languages studied here), with tonal languages ranked significantly higher than non-tonal languages (Mann-Whitney $U = 126$, $p = 0.007$). The opposite pattern was evident for beat alignment, where the 5 tonal languages were clustered toward the bottom of the distribution ($U = 20$, $p = 0.008$); no significant difference was evident for the mistuning perception test, consistent with the somewhat weaker estimated differences on this test ($U = 42$, $p = 0.149$). The language-wise performance in pitch-accented languages

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

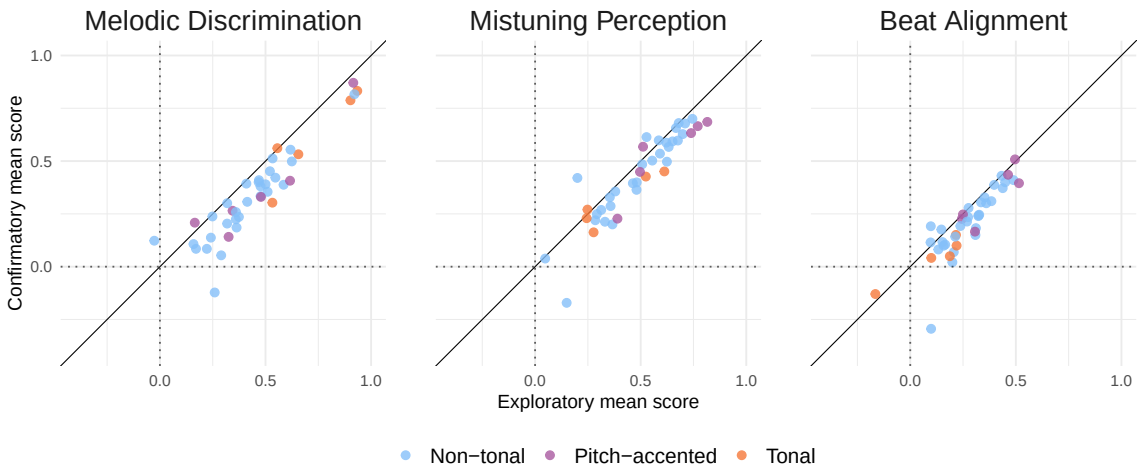

Figure 3. Mean exploratory vs confirmatory scores within each individual language are highly similar, clustered around the y = x line (representing a perfect match) across the three musical tests. Each dot represents an individual language (tonal language: orange, pitch-accented language: purple, non-tonal language: blue).

was far more variable than the tonal languages, perhaps reflecting the fuzzier nature of this category. Larger versions of each column in Figure 4 are presented in Figures S1a-c.

4 Discussion

We found a clear link between linguistic experience and music processing abilities: native speakers of tonal languages performed substantially better than native speakers of non-tonal languages on a test that required discriminating changes in melodic patterns; the effect size of being a tonal language speaker was nearly as large as the effect of receiving music lessons. Notably, our melodic discrimination results did not reflect a general advantage in tonal language speakers' music processing abilities; in fact, tonal language speakers did not show an advantage over speakers of non-tonal or pitch-accented languages on tests probing sensitivity to finer-grained pitch (mistuning) and rhythm (beat alignment). Indeed, on these aspects of music processing, tonal speakers showed a modest *disadvantage*.

Our results are likely to generalize across tonal languages, given that they held in our dataset across multiple tonal languages, each represented by hundreds or thousands of native speakers. They 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 (Alexander et al., 2008; Bidelman et al., 2013; Bradley, 2016; Creel et al., 2018; Pfordresher & Brown, 2009; Swaminathan et al., 2021; Wong et al., 2012), though not all (Bidelman et al., 2011a; Giuliano et al., 2011; Stevens et al., 2013; Tong et al., 2018; Zheng & Samuel, 2018). The modest disadvantage we see on fine-grained pitch tasks too is supported by some prior studies (Bent et al., 2006; Bidelman et al., 2011a; Chang et al., 2016; Peretz et al., 2011; Pfordresher & Brown, 2009; Stevens et al., 2013; Tong et al., 2018; Wong et al., 2012; Zheng & Samuel, 2018) but not others (Bidelman et al., 2013; Giuliano et al., 2011; Hutka et al., 2015). And while rhythmic abilities in tonal language speakers have rarely been studied (see Wong et al., 2012 ; Zhang et al., 2020), 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 (Jasmin et al., 2021). By leveraging a consistent set of tests and a large sample size, our results make clear that speaking a tonal language has a measurable but delimited connection to music skills.

Why might tonal language experience have these specific effects on music perception? Like others (e.g. Bidelman et al., 2013; Patel & Iversen, 2014), we suspect that the answer lies in the shared mechanisms and

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

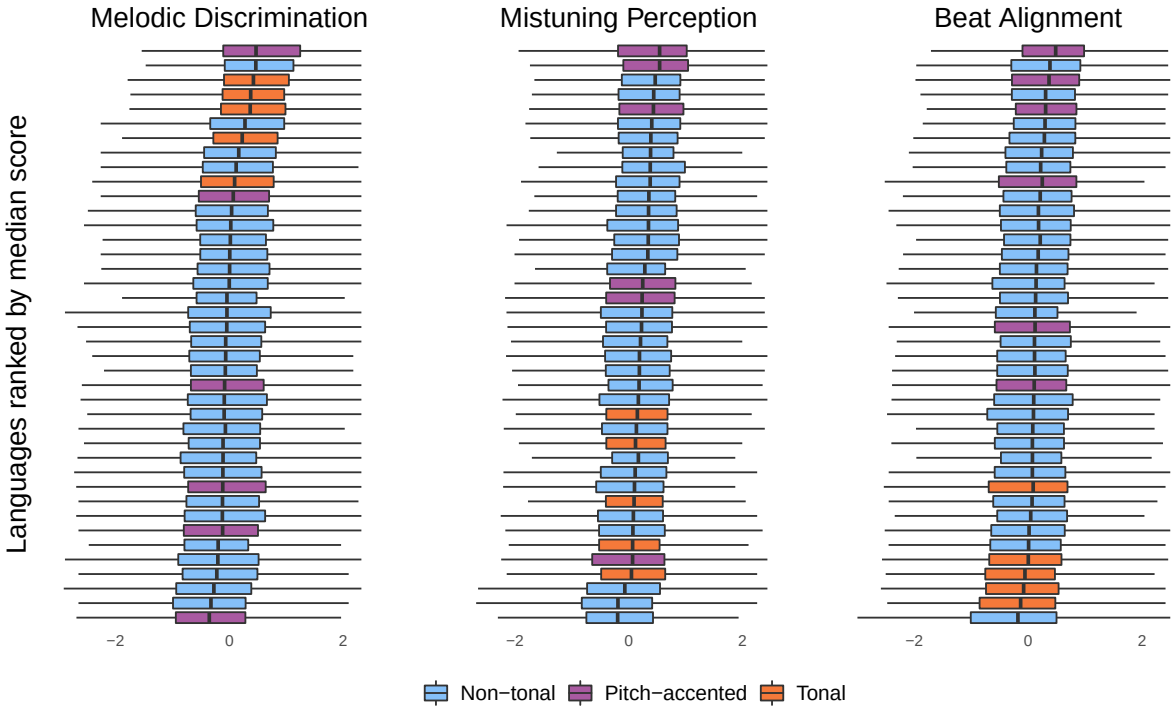

Figure 4. Languages ranked (from high to low) by their median scores on each test. Each boxplot represents a language, with the median score of all native speakers of that language indicated by the vertical black line; the interquartile range by the width of the box, and the range by the horizontal black line. Language types are indicated by the shading (see Legend). To avoid potential confounds of musical experience, only participants who reported not having had musical training are included in this figure. Larger versions of each column are in Figures S1a-c, including language labels and language-wise sample sizes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neural processing resources associated with auditory perception — whether they are applied to language or music. Both tonal languages and music rely on specialized sound categories (tone contours in speech; pitch motifs in music). If these categories are learned through shared, domain-general learning mechanisms, then improving the efficiency of these mechanisms through practice in either domain should result in mutual improvements (Asaridou & McQueen, 2013; Chang et al., 2016; Delogu et al., 2010, 2006; McMullen & Saffran, 2004; Patel, 2008). This idea highlights differences between putatively domain-general auditory processing and putatively domain-specific mechanisms that underlie music processing [e.g., the processing of pitch in the context of a tonal hierarchy; Peretz & Coltheart (2003); Zatorre et al. (2002); Krumhansl (2004)].

It does not explain, however, *how* these learning 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 (Schertz & Clare, 2020), 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 (Wang, 2008; Yu & Andruski, 2010). Similarly, people with pitch perception deficits learn to compensate for their deficits by giving more weight to durational cues when decoding speech prosody (Jasmin, Sun, et al., 2020; Jasmin, Dick, et al., 2020). 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 (Jasmin et al., 2021). In turn, this is consistent with theories of the overlapping mechanisms of basic auditory perception (Patel, 2011; Patel & Iversen, 2014; Tierney et al., 2013; Wong et al., 2007). 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, we hesitate to make any strong claims concerning the link between this type of linguistic experience and music skills. From a statistical standpoint, the evidence for differences in music perceptual abilities in this group were far weaker than those of tonal languages (with much smaller effect sizes; see Figure 2 and Table 2) and far more variable (with large between-language variability within the pitch-accented language group; see Figure 4 and Figures S1a-c). While the results do converge with one prior study (Burnham et al., 2015), which showed that for a variety of pitch tasks, Swedish (pitch-accented) speakers were marginally better than English (non-tonal) speakers, but worse than Thai or Mandarin (tonal) speakers, clearer evidence is necessary for a general claim concerning the effects of pitch-accented language experience on music processing. This, of course, is complicated by the inherently fuzzier nature of classification of pitch-accented languages, relative to tonal languages [see Footnote 1; Gussenhoven & Gussenhoven (2004); Hyman (2006); Hyman (2009)]. We welcome further work that codifies that nature of both linguistic and musical pitch use across this set of generally understudied languages. To this end, we make all our data available through OSF for further analysis by interested parties.

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, or its intensity. 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 (Krizman et al., 2012; Liu et al., 2020; Liu & Kager, 2017), nor assess whether speaking languages that use tone differently (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 [e.g. recent findings that genetics and musical experience both influence linguistic tone perception in Cantonese; Wong et al. (2020)]. These limitations may be addressed in future experiments conducted at smaller scales, but with more precision than a citizen-science approach allows for. For example, targeted analyses of musical abilities in tonal-language speakers who are not typically of East Asian descent, such as Yoruba or Zulu speakers, could help to solidify the generality of the results (thus far, we have only collected data from 40 speakers of these languages via the citizen-science methods reported

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 shared processing resources, which themselves are shaped by auditory experience.

End notes

Supplemental information

The supplemental information includes 3 text sections, 8 tables and 4 figures.

Data, code, and materials availability

A reproducible version of this manuscript, including all data and code, is available at https://github.com/themusiclab/language-experience-music. The preregistration is at https://osf.io/xurdb. Readers can try out the experiment at https://themusiclab.org/quizzes/miq; code for each of the three tests is available at https://github.com/pmcharrison/mpt, https://github.com/pmcharrison/mdt, and https://github.com/pmcharrison/cabat.

Acknowledgments

This research was supported by the Duke University Internship Funding Program (J.L.); the Harvard Data Science Initiative (S.A.M.); and the National Institutes of Health Director's Early Independence Award DP5OD024566 (S.A.M. and C.B.H.). We thank the participants; P. Harrison and D. Müllensiefen for sharing code and assisting with the implementation of their music perception tasks; J. Simson for technical and research assistance; and the members of The Music Lab for discussion and feedback on the citizen-science platform, the experiment, and the manuscript.

Author contributions

Conception: J.L.; S.A.M. Experimental design and implementation: S.A.M. Pre-registration and planned analyses: J.L.; C.B.H; E.B.; S.A.M. Participant recruitment: S.A.M. Analysis and visualization: J.L.; C.B.H.; E.B.; S.A.M. Writing: J.L.; C.B.H.; E.B.; S.A.M.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

S1.1 Details of the exploratory analysis

The exploratory sample (N = 196,689; demographics are in Table S5) consisted of participants who completed the three music perception tests between Nov. 8th, 2019 and Apr 27th, 2020 and passed our exclusion criteria (See Table S5 for breakdown of demographics). We ran the same analyses as in the main text (including the main regression analysis and the three alternate approaches); the results are summarized in Table S6.

S1.2 Validation of self-reported headphone use

Participants who self-reported that they were wearing headphones completed a 6-trial headphone detection task (Woods et al., 2017) designed to be easy for participants wearing headphones and difficult for those listening on free-field speakers. Out of the 194,206 participants who indicated wearing headphones, 165,828 had clean and usable headphone detection data. The distribution of scores for these participants (Figure S2) was strongly left-skewed with the median participant scoring 5.09 of 6 (100%) correct. This implies that the bulk of participants who self-reported wearing headphones were, in fact, wearing headphones.

S1.3 Replications of main analyses at alternate sample levels

We replicated the main results (Table 2 and Figure 2) using three alternate techniques of controlling for confounding (see Methods: these were approaches of exact matching; exact matching while excluding participants who had music lessons; and inverse probability weighting). The results repeated robustly, as summarized in Table S3, and reported fully in Tables S4a-c.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Language	Language Type	Exploratory n	Confirmatory n	Top country/region	Source
Mandarin	Tonal	3917	8991	China	LAPSyD, WALS
Cantonese	Tonal	2701	4398	Hong Kong	WALS
Hokkien	Tonal	1058	3034	Taiwan	WALS
Thai	Tonal	615	2178	Thailand	LAPSyD, WALS
Vietnamese	Tonal	1528	1501	Vietnam	LAPSyD, WALS
Swedish	Pitch-accented	2504	3394	Sweden	LAPSyD, Bailey (1988)
Norwegian	Pitch-accented	1407	2680	Norway	LAPSyD, WALS
Japanese	Pitch-accented	364	1001	Japan	LAPSyD, WALS
Serbian	Pitch-accented	539	973	Serbia	Inkelas & Zec (1988); Van der Hulst et al. (2010)
Croatian	Pitch-accented	762	961	Croatia	Inkelas & Zec (1988); Van der Hulst et al. (2010)
Lithuanian	Pitch-accented	489	685	Lithuania	LAPSyD
English	Non-tonal	70387	98584	United States	LAPSyD, WALS
Spanish	Non-tonal	51054	35796	Spain	LAPSyD, WALS
Turkish	Non-tonal	1776	31618	Turkey	LAPSyD, WALS
French	Non-tonal	5820	11101	France	LAPSyD, WALS
German	Non-tonal	8581	10230	Germany	LAPSyD, WALS
Portuguese	Non-tonal	3887	7129	Brazil	PHOIBLE, Van der Hulst et al. (2010)
Russian	Non-tonal	4820	6285	Russia	LAPSyD, WALS
Polish	Non-tonal	9910	5710	Poland	LAPSyD, WALS
Italian	Non-tonal	3499	4898	Italy	LAPSyD
Dutch	Non-tonal	2726	3651	The Netherlands	LAPSyD
Indonesian	Non-tonal	2098	3613	Indonesia	LAPSyD, WALS
Tagalog	Non-tonal	1390	2830	Philippines	LAPSyD, WALS
Finnish	Non-tonal	2068	2673	Finland	LAPSyD, WALS
Greek	Non-tonal	1324	1980	Greece	LAPSyD, WALS
Korean	Non-tonal	1270	1755	South Korea	LAPSyD, WALS
Romanian	Non-tonal	1145	1697	Romania	WALS
Arabic	Non-tonal	884	1520	Egypt	LAPSyD, WALS
Danish	Non-tonal	1070	1390	Denmark	PHOIBLE, Van der Hulst et al. (2010)
Hungarian	Non-tonal	878	1370	Hungary	LAPSyD, WALS
Hindi	Non-tonal	664	1356	India	LAPSyD, WALS
Czech	Non-tonal	767	1154	Czech Republic	LAPSyD
Farsi	Non-tonal	550	1015	Iran	LAPSyD, WALS
Catalan	Non-tonal	1527	1000	Spain	LAPSyD, WALS
Hebrew	Non-tonal	861	914	Israel	LAPSyD, WALS
Ukrainian	Non-tonal	412	823	Ukraine	PHOIBLE, Van der Hulst et al. (2010)
Bulgarian	Non-tonal	390	590	Bulgaria	LAPSyD, WALS
Estonian	Non-tonal	333	489	Estonia	Van der Hulst et al. (2010)
Malay	Non-tonal	349	476	Malaysia	PHOIBLE
Slovak	Non-tonal	365	449	Slovakia	PHOIBLE, Van der Hulst et al. (2010)

Table S1. The languages studied here, with sample sizes, largest-sample-size country, and the source of the language classification. Abbreviations: WALS (The world atlas of language structures), LAPSyD (Lyon-Albuquerque phonological systems database)

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Demographics	Tonal (n = 20,102)	Non-tonal (n = 242,096)	Pitch-accented (n = 9,694)
Gender			
Female	49.5%	36.4%	34.0%
Male	49.6%	62.1%	65.2%
Other	0.9%	1.4%	0.9%
Mean age (SD)	23.67 (7.91)	24.3 (9.7)	26.56 (10.43)
Music lesson			
Yes	72.0%	67.0%	63.0%
No	28.0%	33.0%	37.0%
Mean age at onset of music lessons (SD)	8.78 (5.31)	10.12 (5.24)	9.72 (4.67)

Table S2. Demographics of participants in the confirmatory dataset, by language type.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Term	Melodic Discrimination	Mistuning Perception	Beat Alignment
Main Analysis			
Language: Tonal	0.407***	-0.088***	-0.091***
Language: Pitch-accented	0.097***	0.173***	0.196***
Matched			
Language: Tonal	0.342***	-0.142***	-0.145***
Language: Pitch-accented	0.076***	0.147***	0.185***
Matched (No Lessons Only)			
Language: Tonal	0.403***	-0.031	-0.111***
Language: Pitch-accented	0.034	0.198***	0.232***
Inverse Probability Weighted			
Language: Tonal	0.336***	-0.127***	-0.152***
Language: Pitch-accented	0.081***	0.152***	0.184***

Table S3. Summary of main results, repeated across the four analysis approaches (*$p < 0.05$, ***$p < 0.001$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	Melodic Discrimination			Mistuning Perception			Beat Alignment		
Term	β	SE	t	β	SE	t	β	SE	t
Language: Tonal	0.342	0.015	22.483***	-0.142	0.013	-11.002***	-0.145	0.014	-10.217***
Language: Pitch-accented	0.076	0.015	5.019***	0.147	0.013	11.353***	0.185	0.014	13.068***
Intercept	0.283	0.011	26.315***	0.429	0.009	46.943***	0.177	0.010	17.715***

Table S4a. Least squares regression outputs for each musical test from the exact match sample (***p $<$ 0.001).

	Melodic Discrimination			Mistuning Perception			Beat Alignment		
Term	β	SE	t	β	SE	t	β	SE	t
Language: Tonal	0.403	0.024	16.865***	-0.031	0.022	-1.377	-0.111	0.025	-4.416***
Language: Pitch-accented	0.034	0.024	1.413	0.198	0.022	8.867***	0.232	0.025	9.27***
Intercept	-0.056	0.017	-3.325***	0.079	0.016	4.999***	-0.053	0.018	-3.006

Table S4b. Least squares regression outputs for each musical test from the exact match sample without music lessons (***p $<$ 0.001).

	Melodic Discrimination			Mistuning Perception			Beat Alignment		
Term	β	SE	t	β	SE	t	β	SE	t
Language: Tonal	0.336	0.008	41.15***	-0.127	0.007	-17.631***	-0.152	0.008	-18.355***
Language: Pitch-accented	0.081	0.011	7.013***	0.152	0.010	16.02***	0.184	0.010	18.74***
Intercept	0.284	0.002	133.194***	0.427	0.002	232.744***	0.184	0.002	91.808***

Table S4c. Least squares regression outputs for each musical test from the inverse probability weighted sample (**p $<$ 0.001).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Demographics	Tonal (n = 9,819)	Non-tonal (n = 180,805)	Pitch-accented (n = 6,065)
Gender			
Female	48.5%	29.6%	29.6%
Male	50.7%	69.3%	69.6%
Other	0.7%	1.2%	0.8%
Mean age (SD)	23.13 (7.59)	25.34 (9.57)	25.39 (9.32)
Music lesson			
Yes	76.0%	69.0%	69.0%
No	24.0%	31.0%	31.0%
Mean age of onset of music lessons (SD)	8.34 (5.08)	10.32 (5.45)	9.76 (4.35)

Table S5. Demographics of the exploratory dataset, by language type.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Term	Melodic Discrimination	Mistuning Perception	Beat Alignment
Main Analysis			
Language: Tonal	0.409***	-0.099***	-0.148***
Language: Pitch-accented	0.107***	0.17***	0.134***
Matched			
Language: Tonal	0.359***	-0.146***	-0.186***
Language: Pitch-accented	0.082***	0.145***	0.121***
Matched (No Lessons Only)			
Language: Tonal	0.49***	-0.079*	-0.183***
Language: Pitch-accented	0.098*	0.114***	0.197***
Inverse Probability Weighted			
Language: Tonal	0.35***	-0.137***	-0.203***
Language: Pitch-accented	0.099***	0.162***	0.127***

Table S6. Beta coefficients for different analysis approaches in the exploratory sample (*$p < 0.05$, ***$p < 0.001$).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

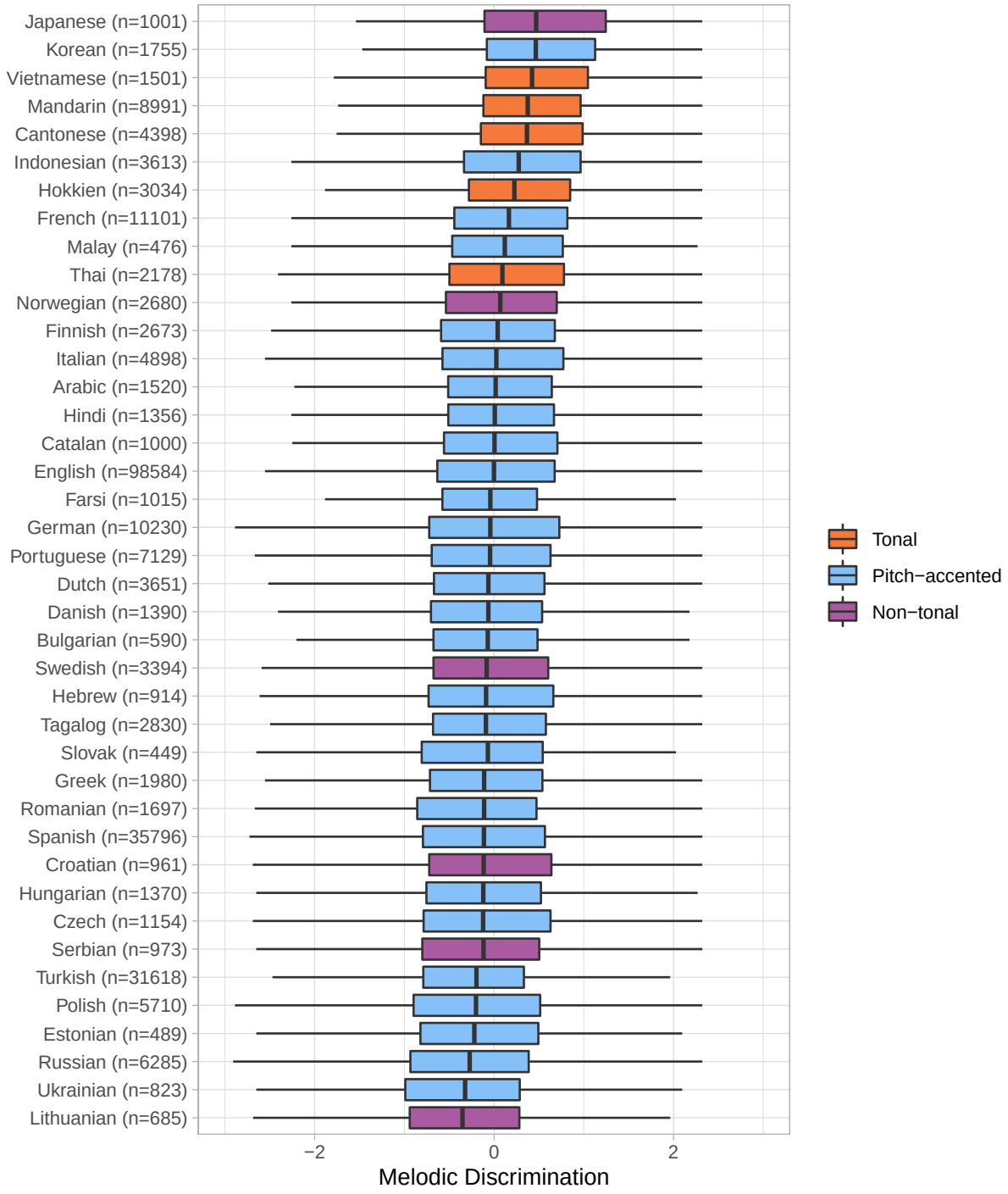

Figure S1a. Melodic discrimination scores ranked by language (ranked by median, detailed version of Figure 4 in the main text). Only participants who haven't taken music lessons are included.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

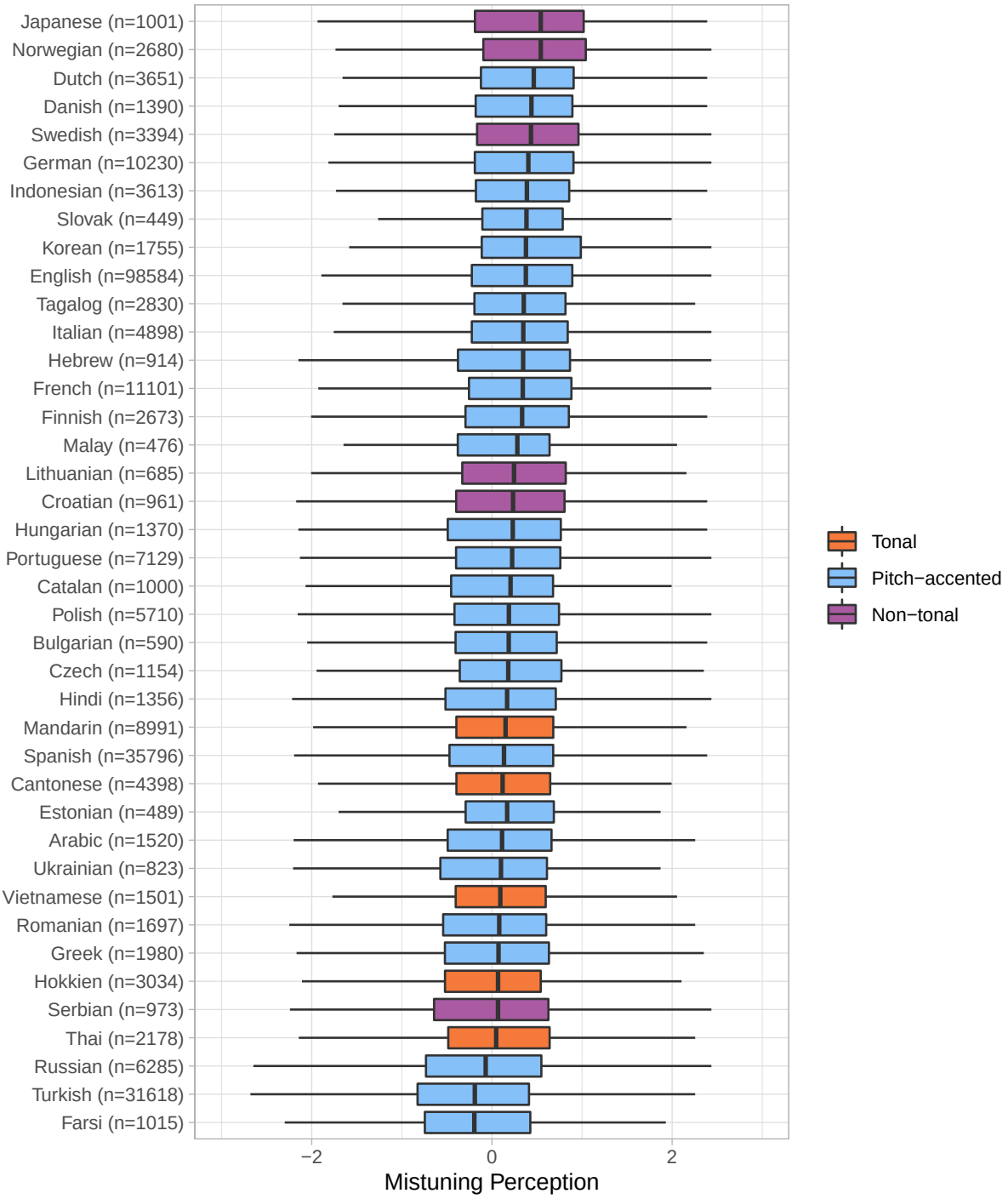

Figure S1b. Mistuning perception scores ranked by language (ranked by median, detailed version of Figure 4 in the main text). Only participants who haven't taken music lessons are included.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

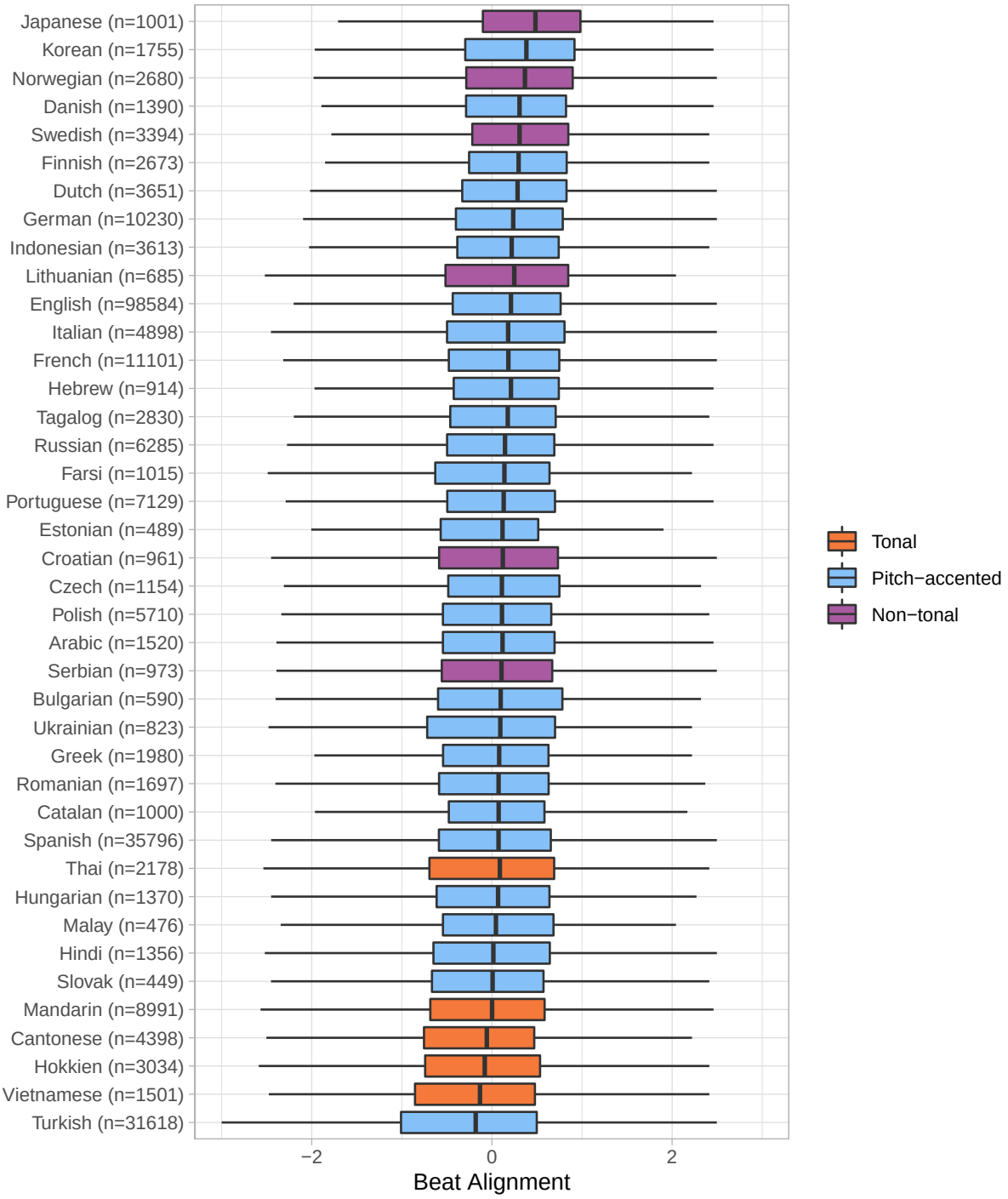

Figure S1c. Beat alignment scores ranked by language (ranked by median, detailed version of Figure 4 in the main text). Only participants who haven't taken music lessons are included.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

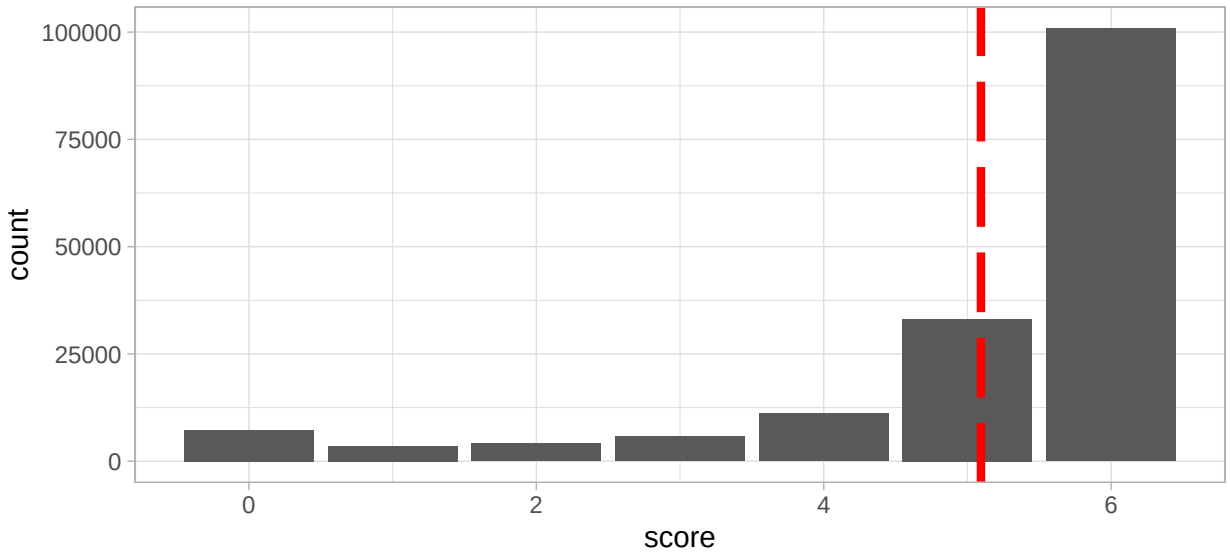

Figure S2. Scores on the headphone detection task, from participants who self-reported that they were wearing headphones. The maximum score was 6; the dashed red line indicates the mean score.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Alexander, J. A., Bradlow, A. R., Ashley, R. D., & Wong, P. C. M. (2008). Music melody perception in tone-language- and nontone-language speakers. *The Journal of the Acoustical Society of America*, *124*(4), 2495–2495. https://doi.org/10.1121/1.4782815

Asaridou, S. S., & McQueen, J. M. (2013). Speech and music shape the listening brain: Evidence for shared domain-general mechanisms. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00321

Austin, P. C., & 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. *Statistics in Medicine*, *34*(28), 3661–3679. https://doi.org/10.1002/sim.6607

Bailey, L. M. (1988). A non-linear analysis of pitch accent in Swedish. *Lingua*, *75*(2), 103–124. https://doi.org/10.1016/0024-3841(88)90028-9

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 97–103. https://doi.org/10.1037/0096-1523.32.1.97

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. https://doi.org/10.1111/desc.12715

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011a). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition*, *77*(1), 1–10. https://doi.org/10.1016/j.bandc.2011.07.006

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011b). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434.

Bidelman, G. M., Hutka, S., & 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*(4), e60676. https://doi.org/10.1371/journal.pone.0060676

Bidelman, G. M., & Lee, C.-C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *NeuroImage*, *120*, 191–200. https://doi.org/10.1016/j.neuroimage.2015.06.087

Bonneville-Roussy, A., Rentfrow, P. J., Xu, M. K., & Potter, J. (2013). Music through the ages: Trends in musical engagement and preferences from adolescence through middle adulthood. *Journal of Personality and Social Psychology*, *105*(4), 703–717. https://doi.org/10.1037/a0033770

Bradley, E. D. (2016). Phonetic dimensions of tone language effects on musical melody perception. *Psychomusicology: Music, Mind, and Brain*, *26*(4), 337–345. https://doi.org/10.1037/pmu0000162

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7-9), 1044–1098. https://doi.org/10.1080/01690965.2010.504378

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattanasone, N. X., Schwarz, I.-C., & Webster, D. (2015). Universality and language-specific experience in the perception of lexical tone and pitch. *Applied Psycholinguistics*, *36*(6), 1459–1491. https://doi.org/10.1017/S0142716414000496

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

Chang, D., Hedberg, N., & Wang, Y. (2016). Effects of musical and linguistic experience on categorization of lexical and melodic tones. *The Journal of the Acoustical Society of America*, *139*(5), 2432–2447. https://doi.org/10.1121/1.4947497

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience*, *31*(6), 751–760. https://doi.org/10.1080/23273798.2016.1156715

Cowen, A. S., Laukka, P., Elfenbein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, *3*(4), 369–382. https://doi.org/10.1038/s41562-019-0533-6

Creel, S. C., Weng, M., Fu, G., Heyman, G. D., & Lee, K. (2018). Speaking a tone language enhances musical pitch perception in 3-year-olds. *Developmental Science*, *21*(1), e12503. https://doi.org/10.1111/desc.12503

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*(1), 46–61. https://doi.org/10.1080/09541440802708136

Delogu, F., Lampis, G., & Olivetti Belardinelli, M. (2006). Music-to-language transfer effect: May melodic ability improve learning of tonal languages by native nontonal speakers? *Cognitive Processing*, *7*(3), 203–207. https://doi.org/10.1007/s10339-006-0146-7

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719–722. https://doi.org/10.1121/1.2151799

Eibl-Eibesfeldt, I. (1979). Human ethology: Concepts and implications for the sciences of man. *Behavioral and Brain Sciences*, *2*(01), 1–57. https://doi.org/10.1017/S0140525X00060416

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5). https://doi.org/10.1017/S0140525X09990720

Giuliano, R. J., Pfordresher, P. Q., Stanley, E. M., Narayana, S., & Wicha, N. Y. Y. (2011). Native Experience with a Tone Language Enhances Pitch Discrimination and the Timing of Neural Responses to Pitch Change. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00146

Gussenhoven, C., & Gussenhoven, P. of G. and E. P. C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Hannon, E. E., & Trehub, S. E. (2005). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639–12643.

Harrison, P. M. c. (2020). psychTestR: An R package for designing and conducting behavioural psychological experiments. *Journal of Open Source Software*, *5*(49), 2088. https://doi.org/10.21105/joss.02088

Harrison, P. M. C., Collins, T., & Müllensiefen, D. (2017). Applying modern psychometric techniques to melodic discrimination testing: Item response theory, computerised adaptive testing, and automatic item generation. *Scientific Reports*, *7*(1), 1–18. https://doi.org/10.1038/s41598-017-03586-z

Harrison, P. M. C., & Müllensiefen, D. (2018). Development and validation of the Computerised Adaptive Beat Alignment Test (CA-BAT). *Scientific Reports*, *8*(1), 1–19. https://doi.org/10.1038/s41598-018-30318-8

Hilton, C. B., & Goldwater, M. (2021). Linguistic syncopation: Meter-syntax alignment affects sentence comprehension and sensorimotor synchronization. *Cognition*. https://doi.org/10.31234/osf.io/hcngm

Hilton, C. B., & Mehr, S. A. (In Press). Citizen science can help to alleviate the generalizability crisis. *Behavioral and Brain Sciences*.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Bainbridge, C. M., Atwood, S., Simson, J., Knox, D., Glowacki, L., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., Schachner, A., … Mehr, S. A. (2021). Acoustic regularities in infant-directed speech and song across cultures. *PsyArXiv*. https://doi.org/10.1101/2020.04.09.032995

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hove, M. J., Sutherland, M. E., & Krumhansl, C. L. (2010). Ethnicity effects in relative pitch. *Psychonomic Bulletin & Review*, *17*(3), 310–316. https://doi.org/10.3758/PBR.17.3.310

Hutka, S., Bidelman, G. M., & 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. https://doi.org/10.1016/j.neuropsychologia.2015.03.019

Hyman, L. M. (2009). How (not) to do phonological typology: The case of pitch-accent. *Language Sciences*, *31*(2), 213–238. https://doi.org/10.1016/j.langsci.2008.12.007

Hyman, L. M. (2006). Word-Prosodic Typology. *Phonology*, *23*(2), 225–257.

Inkelas, S., & Zec, D. (1988). Serbo-Croatian Pitch Accent: The Interaction of Tone, Stress, and Intonation. *Language*, *64*(2), 227–248. https://doi.org/10.2307/415433

Jasmin, K., Dick, F., Stewart, L., & Tierney, A. T. (2020). Altered functional connectivity during speech perception in congenital amusia. *eLife*, *9*, e53539. https://doi.org/10.7554/eLife.53539

Jasmin, K., Sun, H., & Tierney, A. T. (2020). Effects of language experience on domain-general perceptual strategies. *bioRxiv*, 2020.01.02.892943. https://doi.org/10.1101/2020.01.02.892943

Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481. https://doi.org/10.1016/j.cognition.2020.104481

Konner, M. (2010). *The evolution of childhood: Relationships, emotion, mind*. Belknap Press of Harvard University Press.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, *109*(20), 7877–7881. https://doi.org/10.1073/pnas.1201575109

Krumhansl, C. L. (2004). The Cognition of Tonality as We Know it Today. *Journal of New Music Research*, *33*(3), 253–268. https://doi.org/10.1080/0929821042000317831

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Larrouy-Maestri, P., Harrison, P. M. C., & Müllensiefen, D. (2019). The mistuning perception test: A new measurement instrument. *Behavior Research Methods*, *51*(2), 663–675. https://doi.org/10.3758/s13428-019-01225-1

Li, X., & Gao, Z. (2018). The effect of language experience on lexical tone perception. *The Journal of the Acoustical Society of America*, *144*(3), 1865–1865. https://doi.org/10.1121/1.5068200

Liu, L., Chen, A., & Kager, R. (2020). *Simultaneous bilinguals who do not speak a tone language show enhancement in pitch sensitivity but not in executive function*. https://doi.org/10.1075/lab.19037.liu

Liu, L., & Kager, R. (2017). Enhanced music sensitivity in 9-month-old bilingual infants. *Cognitive Processing*, *18*(1), 55–65. https://doi.org/10.1007/s10339-016-0780-7

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, Experience, and Music Perception. *Psychological Science*, *1*(4), 272–276. https://doi.org/10.1111/j.1467-9280.1990.tb00213.x

Maddieson, I. (2013). Tone. In *The World Altas of Language Structures Online* (Dryer, Matthew S. & Haspelmath, Martin).

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

McMullen, E., & Saffran, J. R. (2004). Music and Language: A Developmental Comparison. *Music Perception*, *21*(3), 289–311. https://doi.org/10.1525/mp.2004.21.3.289

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mehr, S. A. (2013). Music and Success. *The New York Times*.

Mehr, S. A. (2014). Music in the home: New evidence for an intergenerational link. *Journal of Research in Music Education*, *62*(1), 78–88. https://doi.org/10.1177/0022429413520008

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2020). Origins of music in credible signaling. *Behavioral and Brain Sciences*, 1–41. https://doi.org/10.1017/S0140525X20000345

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., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), 957–970. https://doi.org/10.1126/science.aax0868

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*. https://doi.org/10.31234/osf.io/sqatb

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

Patel, A. D. (2008). Science & music: Talk of the tone. *Nature*, *453*(7196), 726–727.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology*, *2*.

Patel, A. D., & Iversen, J. R. (2014). The evolutionary neuroscience of musical beat perception: The Action Simulation for Auditory Prediction (ASAP) hypothesis. *Frontiers in Systems Neuroscience*, *8*, 57. https://doi.org/10.3389/fnsys.2014.00057

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. https://doi.org/10.1016/j.wocn.2010.09.003

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688.

Peretz, I., Nguyen, S., & Cummings, S. (2011). Tone Language Fluency Impairs Pitch Discrimination. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00145

Pfordresher, P. Q., & Brown, S. (2009). Enhanced production and perception of musical pitch in tone language speakers. *Attention, Perception, & Psychophysics*, *71*(6), 1385–1398. https://doi.org/10.3758/APP.71.6.1385

Pike, K. L. (1948). *Tone languages*. University of Michigan Press.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Sala, G., & Gobet, F. (2020). Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory & Cognition*, *48*(8), 1429–1441. https://doi.org/10.3758/s13421-020-01060-2

Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, *11*(2), e1521. https://doi.org/10.1002/wcs.1521

Stevens, C. J., Keller, P. E., & Tyler, M. D. (2013). Tonal language background and detecting pitch contour in spoken and musical items. *Psychology of Music*, *41*(1), 59–74. https://doi.org/10.1177/0305735611415749

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science : A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21. https://doi.org/10.1214/09-STS313

Swaminathan, S., Kragness, H. E., & Schellenberg, E. G. (2021). The Musical Ear Test: Norms and correlates from a large sample of Canadian undergraduates. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01528-8

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.18.464888; this version posted October 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tierney, A., Krizman, J., Skoe, E., Johnston, K., & Kraus, N. (2013). High school music classes enhance the neural processing of speech. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00855

Tong, X., Choi, W., & Man, Y. Y. (2018). Tone language experience modulates the effect of long-term musical training on musical pitch perception. *The Journal of the Acoustical Society of America*, *144*(2), 690–697. https://doi.org/10.1121/1.5049365

Tong, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage*, *28*(2), 417–428. https://doi.org/10.1016/j.neuroimage.2005.06.002

van der Hulst, H. (2011). Pitch Accent Systems. In *The Blackwell Companion to Phonology* (pp. 1–24). American Cancer Society. https://doi.org/10.1002/9781444335262.wbctp0042

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

Wagner, M., & McAuliffe, M. (2019). The effect of focus prominence on phrasing. *Journal of Phonetics*, *77*, 100930. https://doi.org/10.1016/j.wocn.2019.100930

Wang, Q. (2008). L2 stress perception: The reliance on different acoustic cues. *Proceedings of Speech Prosody*, 135–138. https://doi.org/https://www.isca-speech.org/archive/sp2008/papers/sp08_635.pdf

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L.-H., & Peretz, I. (2012). Effects of Culture on Musical Pitch Perception. *PLOS ONE*, *7*(4), e33424. https://doi.org/10.1371/journal.pone.0033424

Wong, P. C. M., Kang, X., Wong, K. H. Y., So, H.-C., Choy, K. W., & Geng, X. (2020). ASPM-lexical tone association in speakers of a tone language: Direct evidence for the genetic-biasing hypothesis of language evolution. *Science Advances*, *6*(22), eaba5090. https://doi.org/10.1126/sciadv.aba5090

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872

Woods, K. J. P., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 1–9. https://doi.org/10.3758/s13414-017-1361-2

Yan, R., Jessani, G., Spelke, E., Villiers, P. de, Villiers, J. de, & Mehr, S. (2021). *Across demographics and recent history, most parents sing to their infants and toddlers daily*. PsyArXiv. https://doi.org/10.31234/osf.io/fy5bh

Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*. https://doi.org/10.1017/S0140525X20001685

Yip, M. (2002). *Tone*. Cambridge University Press.

Yu, V. Y., & Andruski, J. E. (2010). A Cross-Language Study of Perception of Lexical Stress in English. *Journal of Psycholinguistic Research*, *39*(4), 323–344. https://doi.org/10.1007/s10936-009-9142-2

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7

Zhang, L., Xie, S., Li, Y., Shu, H., & Zhang, Y. (2020). Perception of musical melody and rhythm as influenced by native language experience. *The Journal of the Acoustical Society of America*, *147*(5), EL385–EL390. https://doi.org/10.1121/10.0001179

Zheng, Y., & Samuel, A. G. (2018). The effects of ethnicity, musicianship, and tone language experience on pitch perception. *Quarterly Journal of Experimental Psychology*, *71*(12), 2627–2642. https://doi.org/10.1177/1747021818757435