Sex Differences in Human Music Perception are Negligible

Mila Bertolo (mila.bertolo@mail.mcgill.ca)
Integrated Program in Neuroscience, McGill University
Montreal H2X 3P8 Canada

Daniel Müllensiefen (d.mullensiefen@gold.ac.uk)
Department of Psychology, Goldsmiths, University of London
London SE14 6NW

Isabelle Peretz (isabelle.peretz@umontreal.ca)
Department of Psychology, University of Montreal
Montreal H3T 1J4 Canada

Sarah C. Woolley (sarah.woolley@mcgill.ca)
Department of Biology, McGill University
Montreal H2X 3P8 Canada

†Jon T. Sakata (jon.sakata@mcgill.ca)
Department of Biology, McGill University
Montreal H2X 3P8 Canada

†Samuel A. Mehr (sam@auckland.ac.nz)
Department of Psychology, University of Auckland
Auckland 1010 New Zealand

†These authors contributed equally.

Abstract

Since Darwin (1871), researchers have proposed that musicality evolved in a reproductive context in which males produce music to signal their mate quality. The extent to which evidence supports this contention, however, remains unclear. Related traits in many non-human animals are sexually differentiated, and while some sex differences in human auditory perception have been documented, the pattern of results is murky. Here, we study melodic discrimination, mistuning perception, and beat alignment perception in 360,009 men and 194,291 women from 208 countries. We find that, in contrast to other non-music human traits, and in contrast to non-human traits, there was no overall advantage for either sex, and the observed sex differences were minuscule (Cohen's d: 0.009 - 0.11) and of inconsistent direction. These results do not provide compelling support for human music perception being a sexually dimorphic trait, and therefore it is unlikely to have been shaped by sexual selection.

Keywords: music perception; sex differences; sexual selection; musical listening ability

Music is ubiquitous across human societies, and there are numerous similarities in the form (Jacoby & McDermott, 2017; Jacoby et al., 2019; Savage et al., 2015) and function of musical styles across societies and cultures (Mehr et al., 2019; Mehr et al., 2018). Consequently, the biological and evolutionary roots of musicality (i.e., the ability to produce and perceive music) have been intensely debated (Fitch, 2006b; Honing, 2018; Ravignani, 2018). Even before the appreciation of universals in musical form and function, researchers have proposed that human musicality may have been shaped by selective forces, in particular sexual selection. For example, Darwin (1871) and others proposed that music could function in a reproductive context in which males produce music to signal their quality to females and in which females assess potential mates based on their musical performance (Miller, 2000). The extent to which evidence supports this contention, however, remains unclear (Bowling, Hoeschele, & Dunn, 2021; Fitch, 2006a; Marcus, 2012; Mehr et al., 2021; Ravignani, 2018; Zentner, 2021).

Proponents of a sexual selection account of music evolution have relied on the observation of sex differences in musicality to support the framework. This is because sex differences in musicality could be attributable to a plausible sexual selection force: sex differences in reproductive investment. In species in which females incur the larger cost for reproduction (e.g., larger gametes, internal fertilization, pregnancy, and lactation), females are proposed to be more discriminating with social signals like courtship vocalizations because the cost of reproducing with a "low quality" male is higher for females (Dabelsteen & Pedersen, 1993; Green & Swets, 1966; Otter & Ratcliffe, 1997; Searcy & Brenowitz, 1988; Trivers, 1972; Wiley, 2006).

Correspondingly, in these species, males are subject to selection pressures to perform more complex signals.

Examples of dramatic sexual dimorphisms in polygynous non-human species abound, with ornate plumage, elaborate vocalizations, enlarged musculature and brain circuitry for the production of courtship signals, and other visual and behavioral cues of health and vigor observed in males (Ryan, 2018; Sakata et al., 2020). Correspondingly, sexual selection has also led to sex differences in sensory processing in non-human animals, with a conspicuous female advantage in the sensory detection and discrimination of courtship signals in bird (Hahn et al., 2017; Hoeschele, Guillette, & Sturdy, 2012), amphibian (Bernal, Stanley & Ryan, 2007), and mammalian species (Krizman et al., 2021).

If sexual selection was a primary force underlying the evolution of human musicality, one might expect comparable sexual dimorphisms to be consistently observed in music-related measures. Some studies report such findings: men score higher on certain scales of musical training assessment (Wolf & Kopiez, 2018), and anatomical reflections of in-utero testosterone exposure (the ratio of second- to fourth-finger lengths) positively correlate with high status within orchestras (Sluming & Manning, 2000; for conflicting evidence see Borniger, Chaudhry, & Muehlenbein, 2013). Conversely, women have been found to be better at recognizing familiar melodies (Miles, Miranda, & Ullman, 2016) and to produce more otoacoustic emissions, a manifestation of heightened auditory sensitivity (Wisniewski et al., 2014).

However, a number of empirical findings and scientific discussions challenge whether such findings are sufficient to claim that sexual selection has shaped musicality. For example, a large-scale twin study found that correlations between musical aptitude and mating success were either null or in the opposite direction (e.g., individuals who are more musically accomplished demonstrated lower reproductive success; Mosing et al., 2015). It is also ambiguous whether observed sex differences in low-level motor or sensory processes (such as otoacoustic emissions) could in principle be acted upon in sexual selection. And as has been noted by Marcus (2012), even contemporarily found sex differences are not necessarily indicative of biologically-based sexual dimorphisms and may reflect sociocultural forces.

Further, some papers that describe findings consistent with a role of sexual selection in musicality have been retracted (e.g., women are more attracted to men with musical ability; Guéguen, Meineri, & Fischer-Lokou, 2014, retracted in 2020) or not replicated (e.g., the idea that women prefer more complex music around ovulation (Charlton, 2014; Charlton, Filippi, & Fitch, 2012), and given the systemic barrier for publishing null results, it is likely that many studies that do not observe sex differences in musicality remain unpublished. Discrepancies across studies could also be due to investigations of relatively few individuals within a restricted population, thereby limiting the scope of generalizability to humans as a species, as they may be artifacts of the particular settings of the studies.

Taken together, there is growing skepticism concerning whether human musicality does in fact display any biologically-based sexual dimorphism, and by extension, skepticism towards the idea that sexual selection had a role to play in human musicality (Marcus, 2012; Mehr et al., 2021).

In the present study, we aim to address sex differences in music perception using well-validated tests of musical perception, and employing a large-scale, citizen-science approach. To this end, we recruited a large global sample of participants to complete tests that utilize validated measures of music perception acuity (Harrison, Collins, & Müllensiefen, 2017; Harrison & Müllensiefen, 2018; Larrouy-Maestri, Harrison, & Müllensiefen, 2019). To further contextualize our study, we compare observed effect sizes to those reported for other sexually dimorphic traits, behaviors, and cognitive processes in humans as well as to those found in non-human animals for which we have a clearer understanding of sexual selection pressures.

Methods

Participants

Our publicly accessible online test (found at themusiclab.org/quizzes/miq) reached 562,853 participants from 208 countries (at time of writing, we considered data gathered between 22 Nov 2019 and 14 Dec 2020) (for details, see Liu & Hilton, 2023). Because sexual selection theory does not make predictions with regards to how non-binary participants may differ from men and women, for all analyses we considered data only from participants who self-identified as either male or female. Many participants were from North America (179,368, or 28.6%) and Europe (189,478, or 30.2%). Additional participant information is detailed in Table 1 (and in SI[1]).

Measures of Music Perceptual Abilities

Participants completed three tasks that tested their abilities in the following domains:

i) Melodic discrimination (Harrison et al., 2017): Participants listened to three different versions of the same melody, played at different absolute pitch levels, where one of the three differed in one note. The participant's task was to identify the odd-one-out. This task is based on an explanatory item response theory (IRT) model and is adaptive; the procedure selects melody items dynamically to match item difficulty to the estimated skill of the participant. Participant scores were computed according to the underlying IRT model and have a theoretical range from -4 to 4, where higher scores represent greater melodic discrimination ability. All three music perception tests were adaptive in this way, with participant scores computed as described above.

ii) Mistuning perception (Larrouy-Maestri et al., 2019): Participants were given two nearly identical excerpts of

vocal music. In one excerpt the vocal track was pitch-shifted against the instrumental track. The participant's task was to indicate in which of the two excerpts the singer sounded more out of tune.

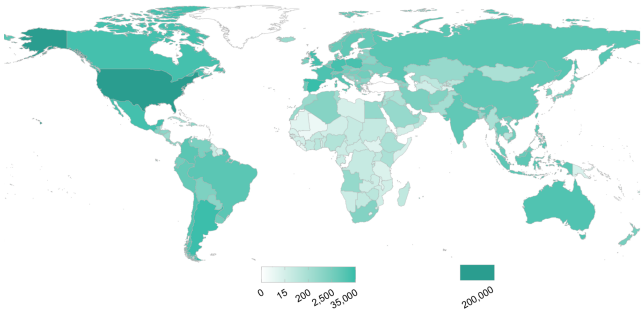

Figure 1: Global spread of participants

Table 1: Participants by gender, age, musical lessons, and age of start of lessons (if any). (See SI[1]).

Gender	N	mean age (sd)	music lessons	age start lessons
Female	194,291	23.6 (9.6)	72%	9.1
Male	360,009	23.5 (9.5)	63%	11.0

iii) Beat alignment (Harrison & Müllensiefen, 2018): Participants listened to the same clip of music twice, with an overlaid beep track. In one version, the beep track was aligned with the musical beat of the music, and in the other version the beep track was slightly shifted in time. The participant's task was to identify in which of the two clips the beep track is best aligned with the beat of the music.

In each of these tests, the excerpts of music were all excerpts of Western popular music - music in a style that would be familiar to listeners acculturated to Western music, but not songs that they were likely to have heard before (all stimulus materials were sourced from stock libraries containing pop songs that had not been commercially released). The significance of using Western music for global participants is addressed in the Discussion.

Before starting these tests, participants were asked to voluntarily report their age, country, self-identified gender, native language, whether they had a hearing impairment, and whether they had completed this game before. They were also asked about the noisiness of their immediate environment, and if they were wearing headphones. After the tasks, they were asked questions concerning their musical experience (see SI[1]).

Procedure

Participants completed the experiment on computers, smartphones, or tablets. They first answered demographic questions (see above) and then completed the three music perception tests. Each test had 15 trials and the order of the tests was randomized.

Analysis

Data Pre-processing Participants were filtered in order to exclude those who reported an age younger than 7 or older than 100. Of the >2 million people who had started the online experiment, we considered only those that had completed all three musical perception tests, who self-reported no hearing deficits, and who self-reportedly had not completed this game before (n= 562,853). Questions that had written response labels were re-coded into numeric variables. For example, for "On an average day, how much time do you spend making music?" the response option 'No time at all' was scored "1"; '1-5 minutes"' was scored "2"; and so on. Answers to these questions were grouped and used as covariates in analyses (see SI file[1]).

Factor Analysis to Extract Composite Scores For each of the above groups of measures, information from their constituent sources was aggregated by factor analysis. For each of these factor analyses, the sample was randomly split into two equal sized subsamples. First, an exploratory minimum residual factor analysis was run on one sub-sample with the correlation matrix of the variables belonging to the same conceptual group. Questions differed in their response options and, thus, the measurement level of corresponding variables. In these cases, a mixed-type correlation matrix was computed employing polychoric, tetrachoric, biserial, point-biserial or Pearson correlations as appropriate for each pair of questions. The resulting correlation matrix is then subjected to a factor analysis. In all cases only a single factor is specified and only variables that loaded with a minimum of 0.3 were retained in the factor model. The second subsample was used for a confirmatory factor analysis and the factor structure was confirmed by inspecting measures of absolute model fit. On the condition that fit measures were satisfactory, the confirmatory factor model was then computed a second time, this time using the entire sample of participants, and factor scores were computed for all participants with complete data on all variables included in the factor model. We used the Empirical Bayes Modal approach for computing factor scores from the confirmatory model. The SI file[1] details each factor's individual items, their loadings on the factor, and how much variance among the indicators is explained by the single factor. For all factors, the confirmatory fit measures indicated a good model fit (General musical ability, the aggregate of the three listening test scores RMSEA < .001, SRMR, < .001; self-reported musical ability RMSEA = 0.029, SRMR = 0.037; self-reported musical training

1 Anonymized supplementary material can be found at: https://osf.io/3v27y/

RMSEA < .001, SRMR, < .001; self-reported musical listening RMSEA < .001, SRMR, < .001).

Testing Whether Musical Abilities Differed as a Function of Participant Variables We analyzed sex differences in music perceptual abilities in several ways to confirm the potential existence and magnitude of sex differences. First, a general additive model (GAM) was fit to each dependent variable (i.e., melodic discrimination, mistuning perception, and beat alignment, and the aggregate factor score for general musical ability). The GAM allows for the simultaneous estimation of means and variances for men and women, and the coefficients for the difference between men and women's mean scores and variances were tested for significance.

Second, the overlap of distributions of performance scores was plotted to visualize the extent of differences and overlap. Third, effect sizes were computed. We calculated Cohen's d as a parametric effect size, as well as the percentage overlap of the two distributions as a non-parametric effect size, and the probability that one group scored higher than the other. For all models, we also computed the R^2, and the η^2 and partial η^2 for the factor sex.

Results

Negligible Sex Differences in Musical Perception

Statistically significant but minuscule differences were found in general musical ability, and in the three individual tests ($p < 0.05$ for each; Figure 2, Table 2). Women, on average, obtained a mean score of 0.002 (SD = 0.331) for general musical ability, which was only minimally higher than the average general musical ability for men (m = -0.001, SD = 0.34). With regard to individual tests, women outperformed men in the mistuning perception test (m = 0.486, SD = 0.857; men: m = 0.402, SD = 0.900), whereas men, on average, scored higher than women on the beat alignment (women: m = 0.126, SD = 0.979; men: m = 0.234, SD = 0.962) and melodic discrimination tests (women: m = 0.309, SD = 1.05; men: m = 0.330, SD = 1.06). Effect sizes ranged from 0.009-0.112: all of these values would be considered as null or very small effects (Hyde, 2014; Archer, 2019; see Discussion).

Sex differences in variances were also observed ($p < 0.05$ for each). Variances were significantly larger for men in the mistuning perception (men:women variance ratio: 1.104) and melodic discrimination tests (men:women variance ratio: 1.019), whereas variances were significantly smaller

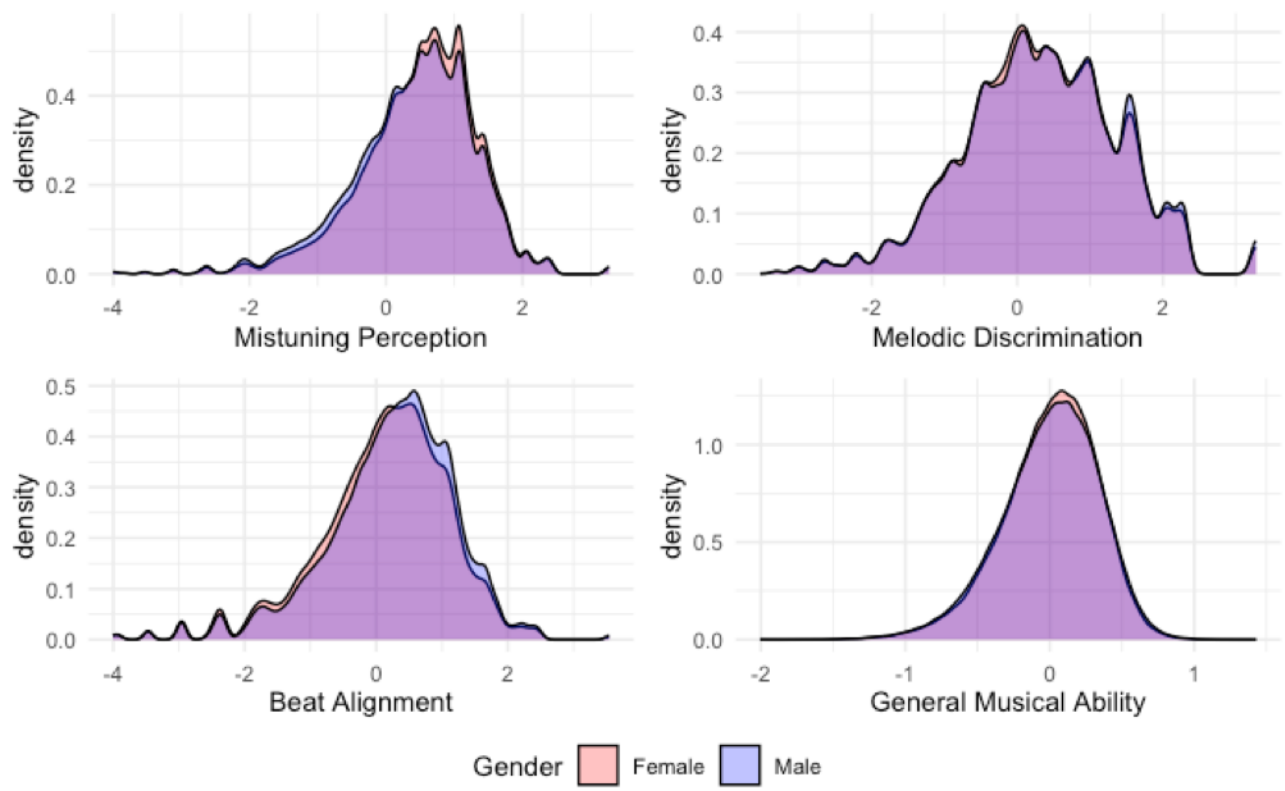

Figure 2: Distributions of musical listening ability scores for men and women. Statistically significant but minuscule differences were found between men and women. All effect sizes ranged from 0.009 - 0.112; these effect sizes are typically considered negligible or very small.

for men in the beat alignment test (men:women variance ratio: 0.966; Table 3). Overall, these differences in variance are considered small (e.g., Hyde, 2014; Archer 2019; see Discussion).

Table 2: Effect sizes of sex differences. Negative values indicate men's scores are higher, and vice versa.

Task	Cohen's d	% overlap distributions	R^2
Mistuning Perception	0.096	95.4	0.002
Melodic Discrimination	-0.020	93.4	<0.001
Beat Alignment	-0.111	95.0	0.003
General Musical Ability	0.009	98.0	<0.001

Magnitude of Sex Difference not Altered by Covariates

Performance on music perception tests can be affected by various demographic variables. In order to control for these covariates and help isolate an effect of sex, we ran further linear models that included potentially important covariates. In particular, we assessed how the inclusion of aggregate scores for musical training, musical listening, age, education, and start of music lessons affected the effect size for sex. The inclusion of these covariates did not have any substantial effect on effect sizes. Partial η^2 values for sex from the covariate-adjusted models are of the same magnitude as the η^2 values from the models containing only sex (Table 3).

Table 3: Explained variance of sex and of full model, in models that include additional participant data.

Task	partial η^2 of sex	R^2 of full model
Mistuning Perception	0.001	0.074
Melodic Discrimination	0.001	0.076
Beat Alignment	0.004	0.037
General Musical Ability	<0.001	0.116

Discussion

In the present study, we assessed sex differences in music perception in a large global sample. In contrast to prior studies on this subject, we employed a citizen-science approach to gather a large sample (>550,000) of men and women from across the globe to complete three music perception tests. The three tests assessed mistuning perception, melodic discrimination, and beat misalignment, and have previously been validated to relate to music perception (Harrison et al., 2017; Harrison & Müllensiefen, 2018; Larrouy-Maestri et al., 2019; Müllensiefen, Elvers, & Frieler, 2022). We observed statistically significant albeit miniscule sex differences in each measure of music perception as well as in the general factor for music perception, with the direction of sex difference varying across tests.

The sample size of the present dataset is unconventionally large, and therefore unconventionally well-powered to detect minuscule differences, and the magnitude of observed differences indicate that they should be interpreted with caution. Based on convention laid out by Cohen (1988), the observed effect sizes (Cohen's |d|: 0.009 - 0.111) should be considered as "small" (i.e., d<0.2), though what constitutes a "small" effect is context dependent, as a function of what competing theories predict. Here, we find that our observed effect sizes are small compared to effect sizes for sexually selected traits, for other types of sex differences in humans, and for sex differences in auditory perception in non-human animals.

Sex Differences are Small Compared to Traits Under Strong Sexual Selection

Various degrees of sex differences in morphology, behavior, and cognition are observed in humans. For example, there are consistent and robust sex differences in height (d = 1.6; Lippa, 2009), vocal acoustic features that signal threat potential (2.7 < d < 5.7; Puts, Apicella, & Cárdenas, 2011), physical aggression (d = 0.33 - 0.84; Archer, 2004), and violent crime (d = 1.11; Archer, 2019). Body size and aggressiveness have long been argued to be under sexual selection pressure throughout human evolution; observed large sex differences in these domains are consistent with these theories. In addition to these large differences in putatively sexually selected traits, meta-analyses of sex differences in human traits and behaviors (Archer, 2019; Hyde, 2005, 2014; Zell et al., 2015) equally reveal functionally negligible sex differences in many domains that might not be shaped by sexual selection; these include mathematical performance (range of Cohen's d: -0.05 to 0.16; Archer, 2019; Hyde, 2014), verbal performance (range of Cohen's d: -0.33 to 0.25; Hyde, 2014), and reward sensitivity (d = 0.01; Archer, 2019). The magnitude of sex differences observed here for music perception are within the range or smaller than those observed for these latter measures.

Sex differences in auditory perception have been observed in a variety of non-human animals in which acoustic signals are important for courtship and sexual behavior. Because song is an important signal that shapes reproductive opportunities in songbirds, studies of song perceptual discrimination in songbirds are particularly relevant here. Song learning and production in the zebra finch, for example, are highly sexually dimorphic, with only males learning to produce complex song and using these songs to court females. Likely as a consequence of male zebra finches having greater forebrain song regions, males are ~9% more accurate than female zebra finches at discriminating between rhythmic and arrhythmic stimuli (Rouse et al., 2023). In chickadees, the consistency of the ratio of frequencies between two consecutive notes (frequency ratio) in a male chickadee's "fee-bee" song is correlated with dominance and reproductive success (Christie, Mennill, & Ratcliffe, 2004), and female chickadees outperform males on the discrimination of pitch ratios in fee-bee songs (partial eta-squared: 0.163) (Hoeschele et al., 2012). Black-capped chickedee females are also significantly quicker than males at learning to discriminate between socially dominant and subordinate male songs (Hahn et al., 2017). While the direction of sex differences varies across studies and species, the effect sizes in these non-human animals are large compared to those observed in our global sample of humans.

No Evidence of Greater Male Variability in Music Perception

With regards to what evidence would support a sexual selection account, sex dimorphisms may manifest as overlapping distributions with similar means, but with differences in variance, typically with greater variability among males than among females (resulting in men being overrepresented at the very high and very low ranges of a distribution: the Greater Male Variability Hypothesis; Arden & Plomin, 2006; Deary, 2003; Feingold, 1992; Hedges & Nowell, 1995). Such a model is not a parsimonious account of our data: not only are all observed differences in variance very small (see Table 2), but our data indicate alternately larger variances for men and women in a pattern that appears unrelated to the small differences in means (mistuning perception - higher m for women, larger SD for men; melodic detection - higher m for men, larger SD for men; beat alignment perception - higher m for men, larger SD for women). This pattern of results is inconsistent with a canonical assessment strategy for sex differences.

Music Perception vs Production

While theories about asymmetric reproductive investment predict more acute female sensory processing of male reproductive signals, the degree to which sex differences in sensory processing would be observed for music remains unclear. Many of the traits that are commonly proposed to be sexually selected (e.g., plumage, ornamentation) are not behaviors or traits that require experiential input. By contrast, achieving a high level of music performance requires intense sensory processing and sensorimotor integration. In other words, in the sexual selection model where it is advantageous for men to produce highly complex musical signals, they must have acute musical perception in order to do so. Therefore, even if music or musicality is under sexual selection pressure, sensory processing abilities might be similar across the sexes, as observed in our study. Other weaknesses of the sexual selection model of music evolution notwithstanding (Mehr et al., 2021), a lack of observable sex differences in perception, in isolation, is not strictly incompatible with the possibility of sexual selection having some other role in musicality (Verpooten & Eens, 2021).

Limitations of Our Dataset

Our tests consisted of musical excerpts in the Western musical tradition, thus raising the question of whether it is interpretable to recruit participants with varying familiarity with this style (ie. variation in performance across cultures likely simply reflects variation in familiarity to Western musical styles). For the scope of this paper, because men and women from all geographic regions participated in the study, we argue that our test remains useful for quantifying sex variation around the world; any culture-driven differences in performance would apply equally to both men and women, and therefore would not confound the presently reported findings.

Our participant pool included more men than women. The reason for this remains unclear, but we argue that, given the total number of participants, this sex imbalance should not impact the nature of our results. It is possible that the sex imbalance in the dataset contributed to our estimates of variance on mistuning perception and melodic discrimination tests (because more men were sampled, it might be more likely to observe extremes for men than for women), but given that even in this imbalanced pool we still observed larger variance for women on the beat alignment test, it appears unlikely that the sex imbalance in sample size would produce a global skew on our estimates of sex variances.

Conclusions

Overall, given the scale and global reach of our dataset and our ability to statistically account for demographic variables relevant to musicality (years of musical training, age, etc.: see Table 1 and SI), we argue that our study has appreciable advantages over existing studies of sex differences in music perception with relatively small sample sizes in limited populations. In summary, while we report statistically significant sex differences in music perception, differences were small compared to other types of dimorphisms in humans and in traits under sexual selection in non-human animals. A number of theories stand as potential explanations of the origins of music (Mehr et al., 2021; Savage et al., 2021); of them, our data do not provide compelling support for sexual selection playing a substantive role in music processing.

References

Archer, J. (2004). Sex Differences in Aggression in Real World Settings: A Meta-Analytic Review. *Review of General Psychology*, 8(4), 291–322. https://doi.org 10.1037/1089-2680.8.4.291

Archer, J. (2019). The reality and evolutionary significance of human psychological sex differences. *Biological Reviews*, 94(4), 1381–1415. https://doi.org/10.1111/brv.12507

Arden, R., & Plomin, R. (2006). Sex differences in variance of intelligence across childhood. *Personality and Individual Differences*, 41(1), 39–48. https://doi.org/10.1016/j.paid.2005.11.027

Bernal, X. E., Stanley Rand, A., & Ryan, M. J. (2007). Sexual Differences in the Behavioral Response of Túngara Frogs, Physalaemus pustulosus, to Cues Associated with Increased Predation Risk. *Ethology*, 113(8), 755–763.https://doi.org/10.1111j.1439-0310.2007.01374.x

Borniger, J. C., Chaudhry, A., & Muehlenbein, M. P. (2013). Relationships among Musical Aptitude, Digit Ratio and Testosterone in Men and Women. *PLOS ONE*, 8(3), e57637. https://doi.org/10.1371/journal.pone.0057637

Bowling, D. L., Hoeschele, M., & Dunn, J. C. (2021/ed). Progress without exclusion in the search for an evolutionary basis of music. *Behavioral and Brain Sciences*, 44, e97. https://doi.org/10.1017/S0140525X2 0001466

Charlton, B. D. (2014). Menstrual cycle phase alters women's sexual preferences for composers of more complex music. *Proceedings of the Royal Society B: Biological Sciences*, 281(1784), 20140403. https://doi.org/10.1098/rspb.2014.0403

Charlton, Benjamin D., Filippi, P., & Fitch, W. T. (2012). Do women prefer more complex music around ovulation? *PLoS ONE*, 7(4), e35626. https://doi.org/10.1371/journal.pone.0035626

Christie, P. J., Mennill, D. J., & Ratcliffe, L. M. (2004). Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology*, 55(4), 341–348. https://doi.org/10.1007/s00265-003-0711-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates.

Dabelsteen, T., & Pedersen, S. B. (1993). Song-based species discrimination and behaviour assessment by female blackbirds, Turdus merula. *Animal Behaviour*, 45 (4), 759–771. https://doi.org/10.1006/anbe.1993.1090

Darwin, C. (1871). *The descent of man*. London: Watts & Co.

Deary, I. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence*, 31(6), 533–542. https://doi.org/10.1016/S0160-2896(03)00053-910

Feingold, A. (1992). Sex Differences in Variability in Intellectual Abilities: A New Look at an Old Controversy. *Review of Educational Research*, 62(1), 61–84. https://doi.org/10.3102/00346543062001061

Fitch, W. T. (2006a). On the Biology and Evolution of Music. *Music Perception*, 24(1), 85–88. https://doi.org/10.1525/mp.2006.24.1.85

Fitch, W. T. (2006b). The biology and evolution of music: A comparative perspective. *Cognition*, 100(1), 173–215.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Guéguen, N., Meineri, S., & Fischer-Lokou, J. (2014). Men's music ability and attractiveness to women in a real-life courtship context. *Psychology of Music*, 42(4), 545–549. https://doi.org/10.1177/0305735613482025

Hahn, A. H., Guillette, L. M., Hoeschele, M., Otter, K. A., Ratcliffe, L. M., & Sturdy, C. B. (2017). Discrimination of male black-capped chickadee songs: Relationship between acoustic preference and performance accuracy. *Animal Behaviour*, 126, 107–121. https://doi.org/10.1016/j.anbehav.2017.02.001

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

Hedges, L. V., & Nowell, A. (1995). Sex Differences in Mental Test Scores, Variability, and Numbers of High-Scoring Individuals. *Science*, 269 (5220), 41–45. https://doi.org/10.1126/science.7604277

Hoeschele, M., Guillette, L. M., & Sturdy, C. B. (2012). Biological relevance of acoustic signal affects discrimination performance in a songbird. *Animal Cognition*, 15(4), 677–688. https://doi.org/10.1007/s10071-012-0496-8

Honing, H. (2018). On the biological basis of musicality. *Annals of the New York Academy of Sciences*, 1423(1), 51–56.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Hyde, J. S. (2014). Gender Similarities and Differences. *Annual Review of Psychology*, 65(1), 373–398. https://doi.org/10.1146/annurev-psych-010213-115057

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*, 27(3), 359–370.

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology*, 29 (19), 3229–3243. https://doi.org/10.1016/j.cub.2019.08.020

Krizman, J., Rotondo, E. K., Nicol, T., Kraus, N., & Bieszczad, K. M. (2021). Sex differences in auditory processing vary across estrous cycle. *Scientific Reports*, 11(1), 22898. https://doi.org/10.1038/s41598-021-02272-5

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

Levitin, D. J. (2006). *This Is Your Brain on Music: The Science of a Human Obsession*. Penguin.

Lippa, R. A. (2009). Sex Differences in Sex Drive, Sociosexuality, and Height across 53 Nations: Testing Evolutionary and Social Structural Theories. *Archives of Sexual Behavior*, 38(5), 631–651.

Liu, J., Hilton, C. B., Bergelson, E., & Mehr, S. A. (2023). Language experience predicts music processing in a half-million speakers of fifty-four languages. *Current Biology* 33, 1–10.e1–e4

Marcus, G. F. (2012). Musicality: Instinct or acquired skill? *Topics in Cognitive Science*, 4(4), 498–512. https://doi.org/10.1111/j.1756-8765.2012.01220.x

Mehr, Samuel A., Krasnow, M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *Behavioral and Brain Sciences*. https://doi.org/10.31234/osf.io/nrqb3

Mehr, Samuel A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., . . . Glowacki, L. (2019). Universality and diversity in human song. *Science*, 366 (6468), 957–970. https://doi.org/10.1126/science.aax0868

Mehr, Samuel A., Singh, M., York, H., Glowacki, L., & Krasnow, M. M. (2018). Form and function in human song. *Current Biology*, 28 (3), 356–368. https://doi.org/10.1016/j.cub.2017.12.042

Miles, S. A., Miranda, R. A., & Ullman, M. T. (2016). Sex Differences in Music: A Female Advantage at Recognizing Familiar Melodies. *Frontiers in Psychology*, 7.

Miller, G. (2000). Evolution of human music through sexual selection. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music*. Cambridge, Mass.: MIT Press.

Mosing, M. A., Verweij, K. J. H., Madison, G., Pedersen, N. L., Zietsch, B. P., & Ullén, F. (2015). Did sexual selection shape human music? Testing predictions from the sexual selection hypothesis of music evolution using a large genetically informative sample of over 10,000 twins. *Evolution and Human Behavior*, 36 (5),359–366. https://doi.org/10.1016/j.evolhumbehav.2015.02.004

Müllensiefen, D., Elvers, P., & Frieler, K. (2022). Musical development during adolescence: Perceptual skills, cognitive resources, and musical training. *Annals of the New York Academy of Sciences*, 1518(1), 264–281.https://doi.org/10.1111/nyas.14911

Otter, K., & Ratcliffe, L. (1997). Female initiated divorce in a monogamous songbird: Abandoning mates for males of higher quality. *Proceedings of the Royal Society of London*. Series B: Biological Sciences, 263(1368), 351–355. https://doi.org/10.1098/rspb.1996.0054

Puts, D. A., Apicella, C. L., & Cárdenas, R. A. (2011). Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B: Biological Sciences*, 279 (1728), 601–609. https://doi.org/10.1098/rspb.2011.0829

Ravignani, A. (2018). Darwin, sexual selection, and the origins of music. *Trends in Ecology & Evolution*, 33(10), 716–719.

Rouse, A. A., Patel, A. D., Wainapel, S., & Kao, M. H. (2023). Within-species differences in vocal production learning in a songbird are associated with differences in flexible rhythm pattern perception. *bioRxiv*.
https://doi.org/10.1101/2022.07.13.499954

Ryan, M. J. (2018). A Taste for the Beautiful: The Evolution of Attraction. In *A Taste for the Beautiful*. Princeton University Press. https://doi.org/10.1515/9781400889150

Sakata, J. T., Woolley, S. C., Fay, R. R., & Popper, A. N. (Eds.). (2020). *The Neuroethology of Birdsong*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-34683-6

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, 112 (29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences*, *44*, e59.

Searcy, W. A., & Brenowitz, E. A. (1988). Sexual differences in species recognition of avian song. *Nature*, 332 (6160), 152–154. https://doi.org/10.1038/332152a0

Sluming, V. A., & Manning, J. T. (2000). Second to fourth digit ratio in elite musicians: Evidence for musical ability as an honest signal of male fitness. *Evolution and Human Behavior*, 21 (1), 1–9.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. G. Campbell (Ed.), *Sexual selection and the descent of man*. Chicago: Aldine.

Van Den Broek, E. M. F., & Todd, P. M. (2009). Evolution of rhythm as an indicator of mate quality. *Musicae Scientiae*, 13 (2_suppl), 369–386. https://doi.org/10.1177/1029864909013002161

Verpooten, J., & Eens, M. (2021). Singing is not associated with social complexity across species. *The Behavioral and Brain Sciences*, 44, e92. https://doi.org/10.1017/S0140525X20001120

Werner, G. M., & Todd, P. M. (1997). Too many love songs: Sexual selection and the evolution of communication. In *Fourth European Conference on Artificial Life* (pp. 434-443). Cambridge, MA: MIT Press/Bradford Books.

Wiley, R. H. (2006). Signal Detection and Animal Communication. In *Advances in the Study of Behavior*

(Vol. 36). Academic Press. https://doi.org/10.1016/S0065-3454(06)36005-6

Wisniewski, A. B., Espinoza-Varas, B., Aston, C. E., Edmundson, S., Champlin, C. A., Pasanen, E. G., & McFadden, D. (2014). Otoacoustic Emissions, Auditory Evoked Potentials and Self-Reported Gender in People Affected by Disorders of Sex Development (DSD). *Hormones and Behavior*, 66(3), 467–474. https://doi.org/10.1016/j.yhbeh.2014.07.004

Wolf, A., & Kopiez, R. (2018). Development and Validation of the Musical Ear Training Assessment (META). *Journal of Research in Music Education*, 66(1), 53–70. https://doi.org/10.1177/0022429418754845

Zell, E., Krizan, Z., & Teeter, S. (2015). Global Gender Differences Can Be Operationalized and Tested. *The American Psychologist*, 70, 664–665. https://doi.org/10.1037/a0039590

Zentner, M. (2021). Social bonding and credible signaling hypotheses largely disregard the gap between animal vocalizations and human music. *The Behavioral and Brain Sciences*, 44, e120. https://doi.org/10.1017/S0140525X2000165X12