# Musical pitch has multiple psychological geometries

Raja Marjieh [*] [1], Thomas L. Griffiths[1] [2] [†], Nori Jacoby [3] [†]

**Pitch perception is at the core of our experience of both speech and music[1,2,3]. Music theorists[4,5], psychologists[6,7], and neuroscientists[8,9,10] have sought to determine the psychological representation of musical pitch for centuries. The pitch helix, which jointly represents logarithmic scaling of the periodicity of a tone and the heightened similarity between tones separated by an octave, has been widely assumed to capture the psychological geometry of pitch[6,7,11]. However, empirical support for this structure is inconclusive, in part because it relies on studies with small sample sizes[6,12,13,14,15]. Here we revisit this problem using a series of comprehensive experiments involving musicians and non-musicians performing three established tasks based on similarity judgments and singing[16,17,18]. We show that a simple helical representation alone cannot explain the data. Rather, our results demonstrate that, depending on the task and musical experience[10,19], the geometry of pitch can exhibit linear, degenerate-helical, and double-helical structures, suggesting a new, broader understanding of how we perceive pitch.**

As we interact with the world around us, our minds constantly create internal representations to support perception, action and decision making[17,20]. The geometric structure of those representations determines how stimuli are internally organized and facilitates computation and generalization[17,21,22]. In audition, pitch is one of the most extensively studied psychological phenomena, being essential to both music and speech perception[2,3,10,16,23-26]. Western music organizes pitch linearly on a logarithmic scale, with tones with slow periodicity perceived as "low" and tones with fast periodicity perceived as "high"[16]. Western music additionally relies on octave equivalence, whereby tones with periodicities that differ by an octave (ratio of 2:1) are associated with the same note name, also known as chroma[6,16,27], and are perceived as similar. The conjunction of these two features, i.e., pitch height and octave equivalence (Figure 1A), led psychologists like Roger Shepard[6,7,27] to the hypothesis that the internal manifold of pitch representations can be captured by a helix (Figure 1B). Since then the pitch helix representation and its decomposition in terms of height and chroma have become canonical examples in many textbooks on perception[7,11,28].

Recent research, however, suggests that cultures differ in the type of verbal metaphors used to organize pitch[29] and in the way they respond to different acoustic pitch cues, including "pleasantness" or consonance of simultaneous tones[30,31] and octave equivalence[16]. Furthermore, research has shown that musical experience alters individuals' sensitivity to pitch, and possibly pitch representations as well[12,23,32-37]. Even within Western music theory, the helix representation fails to capture the role of other important musical intervals, like the perfect fifth (3:2 ratio), which play central roles in music-theoretic constructs. This observation motivated Shepard[6,12] and others[14,32], to propose alternative geometric constructs to support pitch perception and tonal perception like the double helix[12], the cone[14,38], and other higher-dimensional toroidal structures[6,39]. However, the empirical support for these alternative structures is weak, often relying on small sample sizes (e.g., three individual musicians[15]), unnatural tone timbres (e.g., sinusoidal pure tones[14,32], or artificial Shepard tones[39]), incomplete data[6,12], and very specific musical primes (e.g., diatonic C major scale[13,14]) that may induce relative saliency effects[12] (since they include only a subset of the tones being compared).

The recent success of large-scale online studies in shedding new light on classic problems in perception research[40-43], as well as the development of more ecologically valid paradigms to study pitch perception and production[16,18], provide a compelling opportunity for revisiting this classical question. In this work we probe the psychological geometry of musical pitch using three established psychoacoustic paradigms based on similarity judgments and singing which together provide a holistic view that spans both musical expertise (musicians vs. non-musicians) as well as different task modalities (perceptual-evaluative vs. production-based). Using these paradigms we construct detailed maps that capture the perceptual similarity between pitch pairs, which we then analyze using computational models to unravel the contributions of different geometric structures. We show that a simple helical representation is insufficient to explain the data. We find that pitch representations can exhibit an array of geometries, depending on task and experience, ranging from a strictly height-based linear representation in non-musicians to a complex double helix representation in musicians that accounts for heightened similarity at perfect intervals (i.e., fourth and fifth) as well as tritone aversion. This double helix representation had been suggested as a possibility by Shepard[6,12], but has never previously been observed in psychological data. Viewed together, these findings reveal a new, broader picture of pitch perception across tasks and musical experience.

[*] e-mail: raja.marjieh@princeton.edu
[1] Department of Psychology, Princeton University, USA
[2] Department of Computer Science, Princeton University, USA
[3] Max Planck Institute for Empirical Aesthetics, Germany
[†] These authors contributed equally to this work.

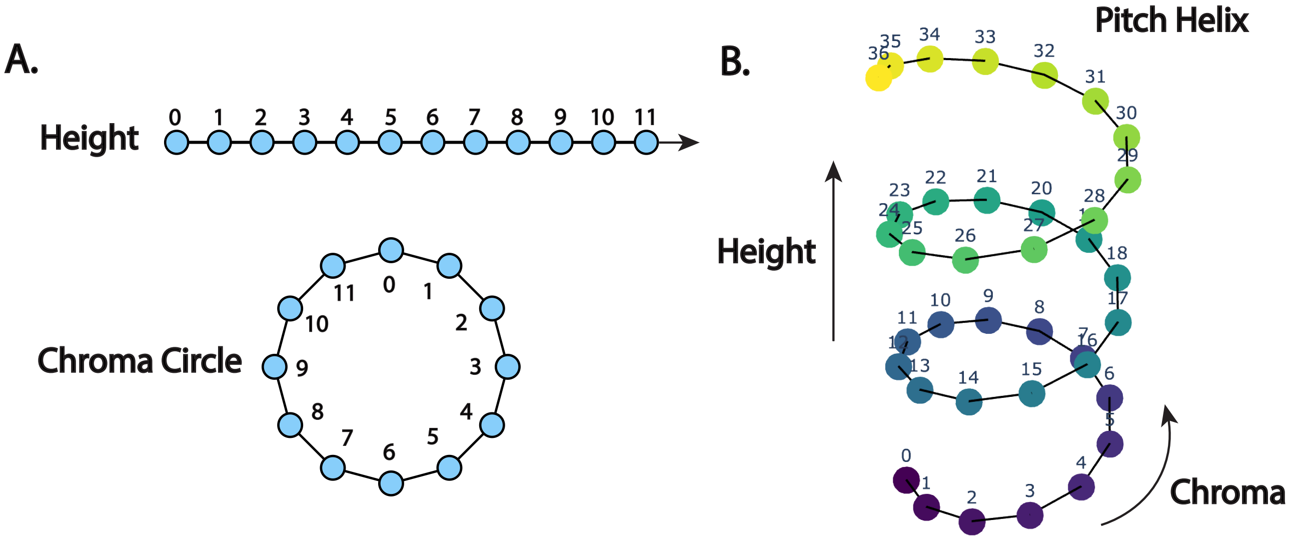


**Fig. 1** The pitch helix representation and its underlying components, namely, the pitch height line and the chroma circle.

## Results

To cover the different ways in which people process musical pitch, we considered three prominent behavioral paradigms (Figures 2A-C) that provide complementary forms of task engagement and applied them to both musician and non-musician participants. First, we wanted to probe pitch representations in the context of melody perception, which is one of the most prominent contexts that involve musical pitch perception across cultures[44]. To that end, we expanded the task of Demany & Armand[18] (also used in Jacoby *et al.*[16]) whereby people listened to pairs of randomly generated three-note melodies and rated their similarity on a Likert scale (Figure 2A; see Methods). Crucially, we generated the second melody from the first by applying a fixed transposition $t$ to the second and third tones across a wide two-octave range. Since these melodies were randomized across trials, we expected the average rating to be able to track the interplay between pitch height (how separated are the tones on a log-scale) and chroma (whether tone chroma is identical in both melodies, i.e., at octave transpositions). We recruited participants from three groups that differed in their musical experience. The first and second groups were recruited from Amazon Mechanical Turk (AMT). The first group comprised AMT participants who self-reported zero years of musical experience ($N = 102$; YME: $M = 0, SD = 0$), and the second comprised AMT participants who reported at least 3 years ($N = 60$; YME: $M = 10.5$, $SD = 9.6$). The third group was a cohort of professionally trained musicians ($N = 44$; YME: $M = 18.6$, $SD = 8.6$) recruited from music schools in Germany (see Methods). Figure 2D shows substantial group differences in the average profile as a function of the transposition interval. Non-musicians with no musical experience exhibited a predominantly linear profile (linear model explains 93.8% of the variance of the average profile, CI: $[91.7, 95.9]$) with a small but significant bump at the first octave (12 semitones, CI of the mean rating difference between 12 and the average rating of 11 and 13 semitones is $[0.028, 0.085]$ which does not include zero), and musicians exhibiting a highly non-linear profile (linear model explains only 16.9% of the variance, CI: $[7.1, 26.7]$) that is more flat with significant spikes at octave transpositions (12 and 24 semitones; CIs for mean rating difference for 12 and the average of 11 and 13, and 24 and 23 semitones were $[0.292, 0.410]$ and $[0.288, 0.427]$, respectively) suggesting both octave equivalence and chroma sensitivity. Non-musicians with 3+ years of experience, on the other hand, exhibited a profile that is somewhere in between those of the other two groups. Quantitatively, a linear model explained 85.3% of the variance of the average profile (CI: $[78.4, 92.3]$) and the CI of the mean difference in rating between 12 and the average of 11 and 13 semitones, and between 24 and 23 semitones were $[0.040, 0.129]$ and $[0.028, 0.111]$, respectively.

In order to provide a complementary perspective on pitch representations, we considered the paradigm of Jacoby *et al.*[16] whereby people were asked to sing back two-note melodies that extended outside their singing range (Figure 2B; see Methods). This task provides an ecologically-valid form of musical engagement with pitch as singing is present in virtually every human culture[44]. Moreover, singing also involves pitch production and as such complements the perceptual-evaluative task of melodic similarity presented above. Participants heard two-note melodies, with the first tone sampled from a frequency range of $45.5 - 105.5$ MIDI note corresponding to $113.2 - 3623.1$ Hz, and were asked to reproduce them by singing (the participants' singing range is approximately $80 - 1000$ Hz). Here we wanted to determine which pitch representation underlies the pattern of behavior observed in the data collected by Jacoby *et al.*[16]. We specifically focused on how people approximated tones outside their singing range and constructed similarity scores between pitch values based on

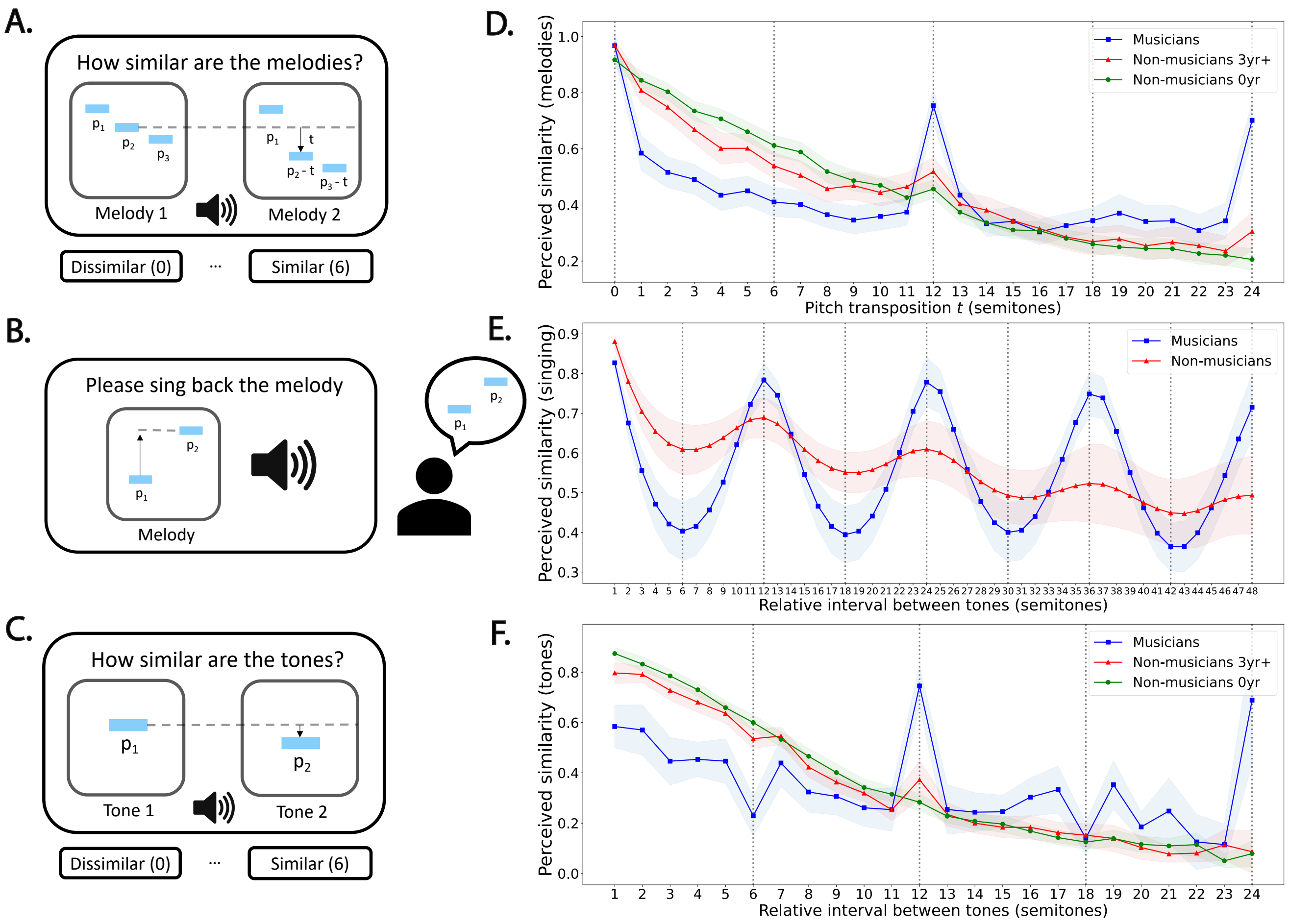


**Fig. 2** Probing pitch similarity across tasks and musical experience. Left panel: Schematics of the three paradigms **A.** similarity judgments over pairs of melodies that differ by a transposition, **B.** free imitation of two-note melodies through singing, and **C.** similarity judgments over pairs of isolated tones. Right panel (**D-F**): Corresponding normalized similarity profiles as a function of interval separations for the three behavioral paradigms and different participant groups. Shaded area, here and everywhere else, represents 95% confidence intervals bootstrapped over participants with 1,000 repetitions.

how similar their response distributions were (see Methods). This approach is analogous to confusion matrices[21] whereby two stimuli are similar in so far as they produce the same behavioral response. The data comprised two groups: participants with at most 3 years of self-reported years of musical experience (non-musicians, $N = 27$; YME: $M = 1.1$, $SD = 1.1$) and participants with at least 10 years of self-reported years of musical experience (musicians, $N = 28$; YME: $M = 19.2$, $SD = 6.8$). We estimated similarity scores between different target pitch values by computing the Jensen-Shannon distance (JSD) between their response distributions (see Methods; raw target-response distributions are provided in Supplementary Figure 1). Figure 2E shows the resulting average similarity profiles. Similar to the first paradigm, non-musicians and musicians exhibited qualitatively different profiles, with non-musicians showing a linear similarity trend with some residual periodicity at the octave (linear model explains 71.2% of the variance with CI $[42.8, 99.6]$, and a sinusoidal model with 12-semitone periodicity explains 20.1% of the variance with CI $[0, 43.8]$) and musicians exhibiting a highly periodic pattern with strong peaks at integer multiples of the octave (linear model explains 1.5% of the variance with CI $[0, 4.5]$ and a sinusoidal model with 12-semitone periodicity explains 93.6% of the variance with CI $[89.3, 98.0]$). These results provide further support for a structural linear-to-helical transition as a function of musical experience.

Finally, we returned to the classic setup of Shepard[17] and asked participants to directly rate the similarity between pairs of isolated tones with as little as possible additional context (Figure 2C; see Methods). We generated a high-powered similarity matrix by asking participants to rate pairs of harmonic complex tones taken from a two-octave range from C4 to C6. We collected data from three cohorts: participants recruited from AMT with zero reported years of musical experience ($N = 94$; YME: $M = 0$, $SD = 0$), participants from AMT with 3+ reported years ($N = 55$; YME: $M = 8.5$, $SD = 6.4$), and musicians from music schools in Germany ($N = 32$; YME: $M = 20.8$, $SD = 10.4$). Figure 2F shows the average similarity rating as a function of the interval between the two tones. Again, non-musicians with zero years of musical experience exhibited a near-perfectly flat profile (linear model explains 90.8% of the variance with CI $[89.1, 92.6]$), and musicians on

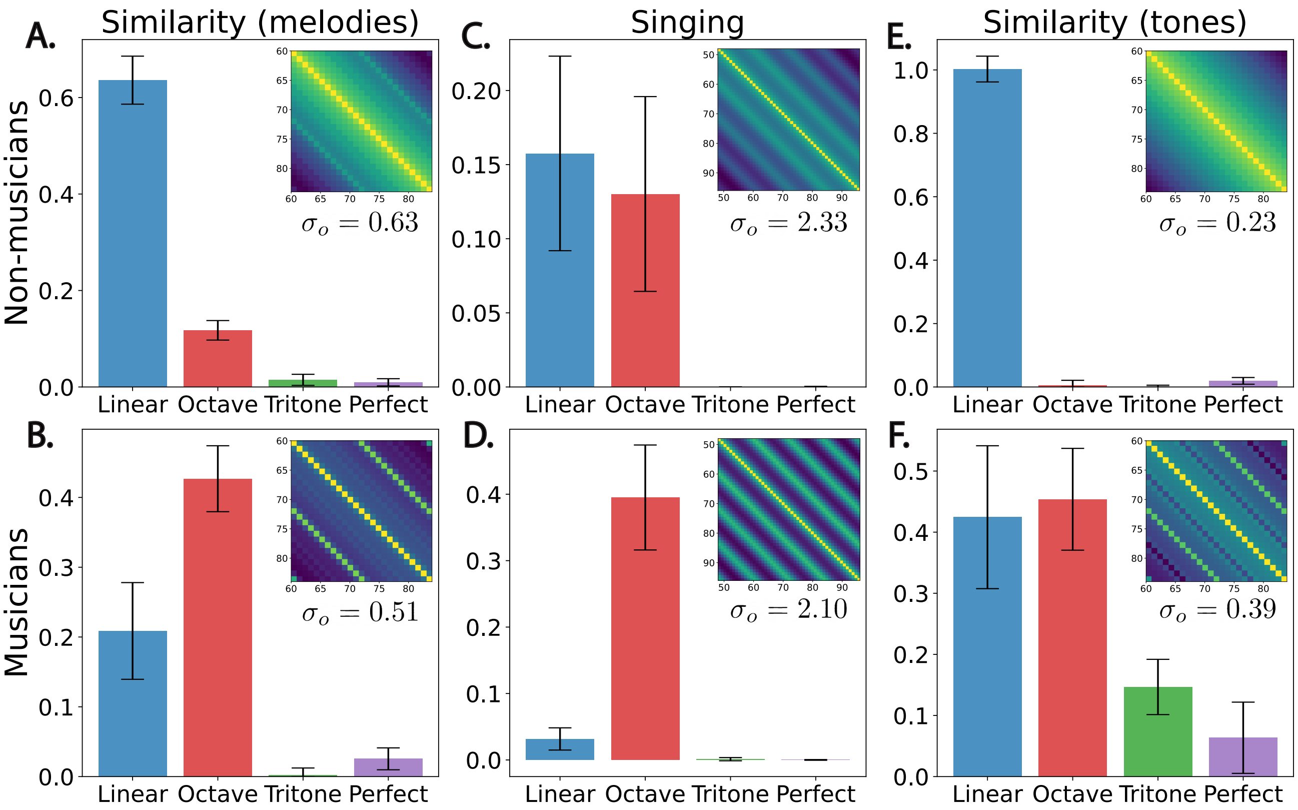


**Fig. 3** Geometric component analysis of similarity. Fitted geometric coefficients to data from non-musicians and musicians and their associated similarity matrices shown as insets. Error bars indicate 95% confidence intervals bootstrapped over participants with 1,000 repetitions.

the other hand exhibiting a highly non-linear pattern (linear model explains $17.2\%$ of the variance with CI $[1.5, 32.9]$) with spikes at the octaves (12 and 24 semitones; CIs of the mean difference in rating between 12 and average of 11 and 13 semitones, and between 24 and 23 semitones were $[0.401, 0.581]$ and $[0.279, 0.868]$, respectively) and sharp dips at the tritones (6 and 18 semitones; CIs of the mean difference in rating between 6 and average of 5 and 7 semitones, and between 18 and average of 17 and 19 semitones were $[-0.305, -0.122]$ and $[-0.285, -0.119]$, respectively) indicating strong octave equivalence but also interestingly strong tritone aversion and enhanced similarity at perfect intervals around it (i.e., 5, 7, 17 and 19 semitones). Non-musicians with 3+ years of musical experience interpolate between the two other groups (linear model explains $89.8\%$ of the variance with CI $[86.8, 92.8]$) and a small but significant peak at the first octave (CI of mean difference between 12 and the average of 11 and 13 semitones was $[0.045, 0.211]$) and an onset of a dip at the first tritone (CI of mean difference between 6 and the average of 5 and 7 semitones was $[-0.094, -0.017]$), but not around 18 semitones (CI of mean difference between 18 and the average of 17 and 19 semitones was $[-0.051, 0.054]$).

## Computational modeling

The results shown in Figure 2 indicate that different participant groups may rely on different perceptual cues to perform the tasks which in turn may translate into different representational geometries. To tease apart the underlying mechanisms, we used a computational modeling approach inspired by an analysis initially proposed by Shepard[12] and adapted here to capture the different possible sources of variance observed in our data. First, to enhance the contrast between the conditions we grouped the data into broader musician vs. non-musician groups (see Methods) and computed similarity matrices based on the average responses. We then determined which geometric structures best explained the similarity matrices. We did that by introducing a basis of geometric components $d_{ij}^{(c)}$ and estimating how the behavioral similarity $s_{ij}$ loaded on these components by fitting a metric solution of the form $1 - s_{ij} = b + \sum_c a_c d_{ij}^{(c)}$ via linear regression where $b, a_c \geq 0$ are non-negative coefficients (Figure 3; see also Methods and Supplementary Figures 2 and 3). The components were: a linear pitch height scale $d_{ij}^{l}$ and three additional components that represent the identification of particular intervals, namely, a component $d_{ij}^{o}$ that identifies octaves, a component $d_{ij}^{p}$ that identifies perfect intervals (fourth and fifth), and a component $d_{ij}^{t}$ that captures aversion to tritones. The octave component had an additional width hyperparameter $\sigma_o$ that interpolated between sharp octave recognition and smoother chroma matching (see Methods for explicit formulas).

The resulting geometric weights for each of the paradigms considered are summarized in Figure 3 (the numerical values

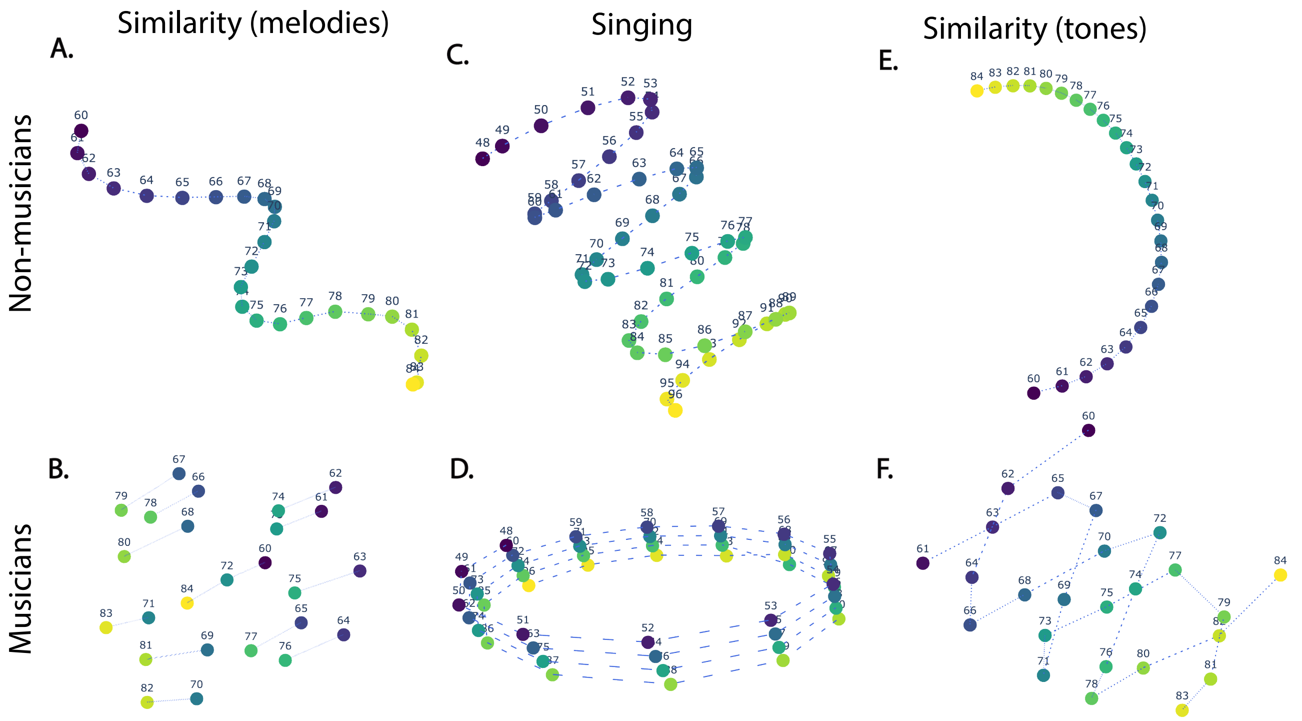


**Fig. 4** Multidimensional scaling solutions in three dimensions for the various behavioral similarity matrices. Left to right: similarity judgments over melodies that differ by a transposition (**A-B**), free imitation of two-note melodies via singing (**C-D**), and similarity judgments over pairs of isolated tones (**E-F**).

for all parameters can be found in Supplementary Table 1; enlarged model insets are provided in Supplementary Figure 3). We see that these weights vary widely depending on experience and task, and overall we found that the geometric component model provided excellent fit to the observed data with mean Pearson correlation of $r = .94$ ($r$ values ranged between $0.82 - 0.98$ depending on the experiment and corrected for attenuation; see Methods for details regarding model evaluation; full list of correlations and other evaluation metrics is provided in Supplementary Table 2). Starting from the melodic similarity paradigm (Figure 3A-B) we observe that most of the variability is driven by the linear and octave components in both musicians ($r = .98$, 95% CI: $[.97, .98]$) and non-musicians ($r = .98$, 95% CI: $[.98, .98]$) though the octave component was significantly more important for musicians compared with non-musicians. Moreover, the relatively diminished linear component in musicians as well as the narrow octave width ($\sigma_o = 0.51$) are suggestive of a degenerate octave recognition mechanism, whereby participants recognize the interval of an octave specifically as more similar without this enhanced similarity influencing other neighbouring intervals.

Turning next to the singing paradigm (Figure 3C-D), we see a similar pattern where the data is largely explained by the linear and octave components (musicians: $r = .97$, 95% CI: $[.96, .98]$; non-musicians: $r = .82$, 95% CI: $[.78, .86]$). In particular, we see that the linear component in musicians is nearly absent relative to that of the octave which is suggestive of a degenerate geometry whereby chroma dominates over pitch height. Unlike the previous paradigm, however, here we see that the optimal octave component width is much larger $\sigma_o = 2.10$ which is indicative of a representation in which octave similarity affects the similarity of other neighboring intervals and thus can be captured by distances on a chroma circle. Possible mechanisms for this result could be a perceptually broad representation (as a result of the higher task demands), and production noise. Finally, for the similarity over isolated tones paradigm (Figure 3E-F) we see a distinct pattern in which non-musicians ($r = .96$, 95% CI: $[.96, .97]$) exhibit a strictly linear representation, while musicians ($r = .93$, 95% CI: $[.90, .96]$), on the other hand, exhibit a highly complex pattern in which all components contribute, in particular, tritone aversion and enhanced similarity at perfect intervals, in addition to relatively balanced linear and octave components.

## The geometries of musical pitch

Our computational modeling suggests that different components are active in the different groups. What are the possible internal geometric representations that can support the observed profiles in Figure 3? To answer this question, we applied three dimensional multi-dimensional scaling (MDS[17]; see Methods). The resulting solutions, shown in Figure 4, exhibit a rich array of geometries, with a simple helical representation with a leading height component and a subleading chroma circle appearing in two out of the six regimes considered (Figure 4A: non-musicians similarity over melodies; Figure 4C: non-musicians singing). With increased musical experience, the simple helical representations become degenerate (i.e., chroma dominates over pitch height), though the nature of the degeneracy differs between the paradigms reflecting in one case a recognition-based representation with octave-equivalence captured as isolated strands (Figure 4B: musicians similarity over melodies) and in the other a degenerate chroma circle representation (Figure 4D: musicians singing). As for the third paradigm, here too we observe a complex structural transition as a function of experience whereby the linear representation observed in non-musicians (Figure 4E) factorizes into a double helix (Figure 4F) to account for tritone aversion and heightened similarity at perfect intervals, as hypothesized by Shepard[6,12].

## Discussion

Our results provide strong evidence for the conclusion that a simple helical representation is insufficient for explaining human pitch perception, exaggerating on the one hand the effect of octave equivalence in non-musicians, and missing, on the other, the implications of alternative sources of variance in the judgments of musicians, in particular, those pertaining to tritone aversion and preference towards perfect intervals. Moreover, our work highlights the dynamic nature of pitch representations, with production-based and perception-based tasks loading differently on the various components, and the way different representational mechanisms (octave recognition and chroma distance) are reflected in the psychological geometry.

## Multiple mechanisms contribute to pitch perception

The results of our study support the notion that pitch perception involves multiple separate mechanisms[24]. Across all experiments, we found four components that explained nearly all of the meaningful variance in pitch similarity (an average raw $R^2$ of .78 across datasets, and an average corrected Pearson correlation of $r = .94$). However, their relative importance varied greatly between populations and tasks. Finding population and task dependency in pitch perception supports the idea that pitch perception is processed in higher-order brain areas[9,45].

We found that a simple linear component dominated the responses of non-musicians, and also contributed to the responses of musicians (Figure 3). This is in line with the widely adopted definition of pitch by the American National Standards Institute[1] (ANSI; "pitch is the auditory attribute of sound that allows sounds to be ordered on a scale from low to high") which emphasizes the linear nature of pitch. This is also consistent with Jacoby *et al.*[16] who showed evidence that linear pitch perception is present in participants across cultures.

The other leading component we found was octave equivalence. Octave equivalence is fundamental to Western music theory[4,5,46] and octaves are common in songs from different cultures around the world[44]. However, its reliance on biological mechanisms and its prevalence in participants with and without musical training is debated in the literature[15,27,47-50]. For example, Hoeschele *et al.*[51,52] used an operant conditioning test that can be run in both humans and chickadees to show that only the former species exhibited octave equivalence. Moreover, Demany & Armand[18] observed octave equivalence in children but found that it was weaker in adults. Following this, Jacoby *et al.*[16] showed that octave equivalence is culturally-contingent (absent in the Tsimane', an indigenous population from Bolivia) and that it is manifested in different degrees even within Western participants[16]. In addition, these findings are consistent with work by Regev *et al.*[53] that used EEG signal for deviance detection and showed that octave equivalence was not detected automatically even by expert musicians. Our findings support the idea that octave equivalence varies across participants. More specifically, we show quantitatively how the strength of octave equivalence varies with musical experience (Figure 3). These findings are also consistent with a large body of evidence showing differences between musicians and non-musicians in auditory perception and production as well as their neural correlates[6,10,19,33-35,54,55].

Beyond these two components our results showed that musicians in the similarity judgement task also relied on two other components, specifically "aversion" to tritones and preference for perfect intervals. What can explain this behavior? One possible idea is that other aspects of music perception may influence musicians' responses. For example, the phenomenon of melodic consonance, or the perceived pleasantness of tone sequences. Recent research suggests that Western participants[41] exhibit a hierarchy of preferences when evaluating the pleasantness of two-tone melodies, with tones separated by an octave and perfect intervals being particularly pleasant, and those separated by a tritone being particularly unpleasant. This pattern overlaps to a certain extent also with the pattern of harmonic consonance (the pleasantness of simultaneous tones[31,40,56-58]). Moreover, these hierarchical patterns are also reflected in the distribution of melodic intervals in musical corpora, which may drive preference via a mechanism of familiarity[14,59-61]. Another possible and potentially related mechanism is tonality, namely, the hierarchical organization of tones within a scale. Previous work suggests that tonality may also be a shaping force in the structure of musical pitch representations[14,32,62]. While our method tried to minimize carry-on effects of tonality from previous trials (by using different roving of tones across trials), it is possible that some sense of tonality can be induced even within a single trial, at least when it comes to expert musicians (Figure 3F).

## Limitations and future work

We end by discussing limitations which point to important directions for future research. First, our participant cohort comes from the United States and Germany, which are Western countries. This limits the generalizability of the present findings as elements of pitch perception such as octave equivalence and melodic preferences vary cross-culturally[16]. A natural follow-up, therefore, could look at the way representations vary across cultures by applying the same methods deployed in this work which by design are suitable for cross-cultural research. Second, the observed linear to double-helical structural transition in pitch perception as a function of musical training resembles other structural-representational transitions in the literature such as that of numbers as a function of education[63]. In that case, a linear magnitude-based representation transforms into a non-linear representation that encapsulates different mathematical categories. It would be informative to investigate the parallels between those two developmental trajectories and to leverage large-scale online recruitment to construct detailed maps for the latter (number). Third, in the present work we exclusively focused on population level analysis of representations, however, since musical experience is very subjective one might expect to see individual differences[58] which are worth investigating. Finally, there are other established paradigms for studying pitch representations that exist in the literature which we did not consider[52]. Future work could apply geometric component analysis to these and see whether a similar picture emerges. We hope to explore these directions in future research.

To conclude, while pitch perception provides the fundamental backbone underlying both speech and music perception, it is neither simple nor static. Rather, our results reveal a highly complex and dynamical phenomenon that manifests multiple psychological geometries and bears parallels with mathematics and language. More broadly, our work showcases how combining large-scale behavioral studies with established psychophysical paradigms and computational modeling can provide new answers to fundamental questions in auditory research. We believe that scaling up psychological research in tandem with concurrent computational approaches will continue to provide exciting new hypotheses concerning the nature of human cognition.

# Methods

## Software implementation

The similarity judgment paradigms were designed and deployed using PsyNet[1], a modern framework for complex experiment design which builds on the Dallinger[2] platform for online experiment hosting and participant recruitment. Participants interact with the experiment in the browser using a front-end interface which, in turn, communicates with a back-end Python cluster that organizes the experiment timeline. All experiment and analysis code are available in the Supporting Information provided with this manuscript (see Code and Data availability below).

## Stimuli

We defined absolute pitch using the Musical Instrument Digital Interface (MIDI) semitonal scale which is based on a logarithmic pitch representation. This was motivated by previous research which showed that participants across many cultures use approximately logarithmic pitch representations[16]. Specifically, the MIDI scale was defined as follows: $f = 440 \times 2^{(p-69)/12}$ where $f$ is frequency in Hertz and $p$ is the corresponding pitch value. This means that a Concert A (A4 or 440 Hz) corresponds to a value of 69 on this scale.

**Similarity paradigms** The similarity judgment paradigms used complex harmonic tones that were synthesized using Tone.js[3], a Javascript library for audio synthesis in the browser. We used additive synthesis so that tones were given by $s(t) = \sum_{i=0}^{n_H-1} w_i \sin(2\pi f_i t)$ with $n_H = 10, f_i = f_0 \times (i+1)$ for some fundamental frequency $f_0$, $w_i = 10^{-\omega_i/20}, \omega_i = \rho \log_2(i+1)$ and $\rho = 3$ which corresponds to 3 dB/octave roll-off.

In the similarity judgment paradigm over melodies, random melodies were synthesized by first uniformly sampling a starting tone in the MIDI range 76-80, and then generating the second and third tones by subtracting a uniformly sampled integer interval in the range $5-7$ and $9-11$ semitones, respectively. This means that the resulting melodies could have a variety of melodic intervals per transposition so that pre-existing musical expectations would not prime the perception of the transposition interval. Transpositions were then generated by subtracting a fixed integer interval in the range $0-24$ semitones from the second and third tones. As for the similarity judgment paradigm over tones, tones covered the MIDI range $60 - 84$ corresponding to the notes C4-C6.

**Singing paradigm** The singing paradigm also involved additively synthesized harmonic complex tones with $n_H = 10$ harmonics, though with a roll-off of 12 dB/octave. Two-note melodies were produced by sampling the first tone from the range of $45.5 - 105.5$, and then creating the second tone by adding one of the intervals $0, \pm1, \pm2$, or $\pm3$ semitones. Each experiment consisted of a series of “blocks” presented in random order, each of which presented stimuli within a single frequency register. Within each block the $f_0$ of the first tone was constrained to particular register. Full information is provided in the section ”General Experiment Structure” of the paper by Jacoby *et al.*[16].

## Participants

**Similarity paradigms** Non-musician participants for the similarity judgment studies were recruited on Amazon Mechanical Turk (AMT) subject to the following recruitment criteria designed to ensure data quality: 1) participants must be 18 years of age or more, 2) they must reside in the United States, 3) they have a 99% approval rate or higher on prior AMT tasks, and 4) have successfully completed 5,000 tasks on AMT. These participants provided informed consent under a Princeton University Institutional Review Board (IRB) protocol (application 10859) and were paid a fair wage of 12 USD per hour. Overall, $N = 194$ participants completed the similarity over melodies study with a reported age range of $19 - 77$ ($M = 40.3, SD = 12.2$) and $0 - 53$ ($M = 3.5, SD = 7.1$) years of musical experience. As for the similarity over tones study, $N = 186$ participants completed the study with a reported age range of $20 - 78$ ($M = 40.2, SD = 10.3$) and $0 - 34$ ($M = 2.8, SD = 5.2$) years of musical experience.

Musicians for the similarity judgment studies were recruited from music schools through an internal participant pool at the Max Planck Institute for Empirical Aesthetics in Germany. Participants were recruited to this pool by research assistants who sent emails to local music conservatories and handing flyers at the entrance of the conservatory to possible participants. In order to participate, musicians were required to be at least 18 years of age and to preferably have at least

[1] https://www.psynet.dev

[2] https://dallinger.readthedocs.io/en/latest/

[3] https://tonejs.github.io/

10 years of active musical training. Participants provided consent under a Max Planck Ethics Council protocol (application 2021_42) and were paid at a rate of 15 USD per hour. Participant took the experiment remotely, and the web-interface was identical to the one used by the AMT participants. Overall, $N = 44$ musicians completed the similarity over melodies study with a self-reported age range of $20 - 50$ ($M = 30.4$, $SD = 8.6$) and $2 - 41$ years of musical experience ($M = 18.6$, $SD = 8.6$). As for the similarity over tones study, $N = 32$ musicians completed the study with a self-reported age range of $21 - 62$ ($M = 33.0$, $SD = 10.7$) and $2 - 46$ years of musical experience ($M = 20.8$, $SD = 10.4$). All but three individuals had more than 10 years of musical experience. We also note that the number of musicians varied between the similarity paradigms due to the nature of online recruitment whereby participants had the freedom to select which experiments they wanted to complete from a list of available studies.

**Singing paradigm** The singing data analyzed in the present work was reproduced from Experiment 4 in Jacoby *et al.*[16]. The experiment comprised $N = 28$ US musicians with an age range of $18 - 69$ ($M = 31.1$, $SD = 10.5$) and $10 - 38$ years of musical experience ($M = 19.2$, $SD = 6.8$), as well as, $N = 27$ US non-musicians with an age range of $20 - 49$ ($M = 32.1$, $SD = 8.4$) and $0 - 3$ years of musical experience ($M = 1.1$, $SD = 1.1$).

## Pre-screening

**Similarity paradigms** To enhance online data quality, in the similarity tasks participants were required to complete a headphone prescreening test[64] to ensure that they were wearing headphones. In each trial of this test, participants heard a series of three tones and were asked to judge which of these was the quietest. The tones were designed to induce a phase cancellation effect when played on speakers which would lead participants without headphones to answer incorrectly. Participants who failed the headphone test were not allowed to proceed to the main experiment but were nevertheless fully compensated for the time taken to perform the pre-screening test.

**Singing paradigm** In the singing task, participants performed a hearing task (see Jacoby *et al.*[16]) to make sure they had normal hearing. Participants who failed the task were excluded.

## Performance incentives

To motivate online participants to provide honest responses for the otherwise subjective tasks, participants in the similarity paradigms were informed that they could receive a performance bonus depending on the quality of their responses. Specifically, they received the following instructions: "The quality of your responses will be automatically monitored, and you will receive a bonus at the end of the experiment in proportion to your quality score. The best way to achieve a high score is to concentrate and give each trial your best attempt". While the tasks are subjective in nature, we used self-consistency as a measure of performance quality. This was done by repeating 5 random trials at the end of the experiment and then computing Spearman correlation between the original and repeated answers. The resulting score $s$ was then used to compute a small bonus of up to 10 cents using the formula $\min(\max(0, 0.1s), 0.1)$.

## Procedure

**Similarity paradigms** In the similarity over melodies task, the experiment proceeded as follows: upon completing the consent form participants received the following instructions: "In this experiment we are studying how people perceive melodies. In each round you will be presented with two three-note melodies and your task will be to simply judge how similar they are. You will have seven response options, ranging from 0 ('Completely Dissimilar') to 6 ('Completely Similar'). Choose the one you think is most appropriate. You will also have access to a replay button that will allow you to replay the sounds if needed. Note: no prior expertise is required to complete this task, just choose what you intuitively think is the right answer". Participants then completed two practice trials and then proceeded to the main experiment. The procedure for the similarity over isolated tones task was identical, except that we replaced "melodies" in the instructions with "sounds".

**Singing paradigm** For the singing task the procedure was as follows: in each session participants completed a series of trials whereby they listened to a melody and were instructed to replicate it as well as possible. Participants were seated in front of a microphone and facing the experimenter. All stimuli were presented through headphones and participants recorded their responses using the microphone (see Jacoby *et al.*[16] for full details).

## Data analysis

**Constructing similarity matrices** To generate an aggregate similarity matrix from the direct similarity judgment paradigms, we applied the following procedures. In the case of similarity over isolated tone pairs, we simply computed the average Likert score per pair of items and divided by a constant of 6 so that similarity scores would be normalized between 0 and 1 (rather than 0 and 6). As for similarity over melodies, since the data is one-dimensional in that case (a pitch transposition $t$ between random melodies) we constructed a two-dimensional matrix $a_{ij}$ from the one-dimensional data $s_t$ using the formula $a_{ij} = s_{|p_i - p_j|}$ where $p_i$ and $p_j$ are pitch values of interest. We again divided by a factor of 6 so that similarity scores would be normalized between 0 and 1. As for the singing data, the procedure is slightly more complicated and is described below.

**Singing response distributions** To construct similarity matrices based on the response distributions of human singers, we first applied a 2D Gaussian kernel density estimate (KDE) $\rho(p_t, p_r)$ to the target-response pitch pairs $(p_t, p_r)$ using the `KernelDensity` method of the `scikit-learn`[4] Python package with a bandwidth parameter of $\sigma = 1.2$ semitones and a resolution of $500 \times 500$ bins. We used this bandwidth as it provided a good tradeoff between reliability and resolution. Dissimilarity between two target pitches $p_{t_1}$ and $p_{t_2}$ was then computed by applying the Jensen-Shannon distance `jensenshannon` from the `scipy`[5] package to the KDE marginals $\rho(p_r|p_{t_1})$ and $\rho(p_r|p_{t_2})$.

**Multi-dimensional scaling** We constructed MDS embeddings using the `MDS` method of `scikit-learn`. This was done in two steps, first we applied a metric MDS to find an initial embedding which was

[4] https://scikit-learn.org/stable/

[5] https://scipy.org/

then fed as an initialization into a non-metric MDS. We used three components, a maximum iteration value of 10,000, and a tolerance parameter of 1e-100. MDS was applied to the fitted model similarity matrices to reveal their corresponding geometries. For completeness, we provide the MDS solutions associated with the raw unprocessed data in Supplementary Figure 4. While they are naturally noisy due to MDS overfitting to noise, they still exhibit the observed key features of the degenerate geometries and the interleaved strands of the double helix.

**Geometric components** The explicit formulas for the different geometric components were: a linear pitch scale $d_{ij}^{(l)} \propto |p_i - p_j|$, two positive interval recognition components $d_{ij}^{(c)} = 1 - \exp\left(-(\chi(p_i, p_j) - c)^2/2\sigma_c^2\right)$ where $\chi(p_i, p_j) = \min\{\psi_{ij}, 6 - \psi_{ij}\}$ and $\psi_{ij} = |p_i - p_j| \mod 12$ to capture chroma distance (i.e., intervallic distance irrespective of register) and $c = 0, 5$ for octave ($d_{ij}^{(o)}$) and perfect intervals (fourth and fifth; $d_{ij}^{(p)}$), respectively, and one negative recognition component $d_{ij}^{(c)} = \exp\left(-(\chi(p_i, p_j) - c)^2/2\sigma_c^2\right)$ with $c = 6$ to capture tritone aversion ($d_{ij}^{(t)}$). We fixed $\sigma_c = 0.25$ semitones for the tritone and perfect intervals as a narrow interval recognition threshold and optimized the octave width $\sigma_o$ as a hyperparameter since different values of $\sigma_o$ efficiently interpolate between sharp octave recognition and smooth distance on the chroma circle (since $c = 0$). As a sanity check for the chosen width parameter values, we refitted the model in the musician tone similarity condition (Figure 3F, the only condition in which the tritone and perfect consonance had significant contributions) but this time allowing all width parameters to vary. We found that the optimal width parameters did not differ much from our chosen values ($\sigma_o = 0.39$, $\sigma_t = 0.18$ and $\sigma_p = 0.50$). In addition, to ensure that the coefficients are properly normalized (i.e., each geometric component varied between 0 to 1), we rescaled the linear component such that its largest separation on a given range of interest equals to one (e.g. if pitch differences varied between 0 and 24 semitones, we defined $d_{ij}^{(l)} = |p_i - p_j|/24$).

**Model fitting and evaluation** We fitted geometric components to the data using the `LinearRegression` method of `scikit-learn`. This was done by first flattening the upper triangular part of each component's distance matrix $d_{ij}^{(c)}$ into a feature vector $v_i^{(c)}$ and then fitting a linear regression of the form $1 - s_i = b + \sum_c a_c v_i^{(c)}$ where $b, a_c \geq 0$ are non-negative coefficients and $s_i$ is the flattened upper-triangular matrix of the behavioral similarity data. We repeated this fitting process 1,000 times by bootstrapping over participants in a split-half fashion (i.e., generating random half-splits of the data based on participants and then fitting the model to one half and testing on the other). The optimal octave width hyperparameter $\sigma_o$ was fine-tuned using the blackbox optimizer `scipy.optimize.minimize` over the bootstrapped linear regression procedure described above. To quantify model performance, we computed three Pearson correlation coefficients for each of the 1,000 split-halves as follows: $r_{dd}$ the correlation between the corresponding human similarity matrices of each split, $r_{dm}$ the correlation between the model fit on one half with the human similarity of the other (there are two ways to compute this so we took the mean), and $r_{mm}$ the correlation between the fitted models on each half. We then used these values to compute the corrected correlation for attenuation $r = r_{md}/\sqrt{r_{dd} r_{mm}}$.

**Code and Data availability** All data and code used in this work can be accessed via the following link: `https://osf.io/nx586/?view_only=0baf1633d6f343d9a68f305763def28f`.

**Author contributions** All authors contributed to the design of the experiments and editing the manuscript. Author RM implemented the experiments, analyzed the results, and drafted the initial manuscript.

**Competing interests** The authors declare no competing interests.

**Acknowledgements** This work was supported by grant 61454 from the John Templeton Foundation.

## References

1. Association, A. S. *et al.* Acoustical terminology SI. *1-1960, American Standards Association* (1960).
2. Levitin, D. J. *This is your brain on music: The science of a human obsession* (Penguin, 2006).
3. Patel, A. D. *Music, language, and the brain* (Oxford university press, 2010).
4. d'Arezzo, G. Micrologus. ed. *Joseph Smits van Waesberghe. Rome: American Institute of Musicology* (1955).
5. Drabkin, W. *Octave (i)* 2001. doi:`10.1093/gmo/9781561592630.article.50054`.
6. Shepard, R. N. Geometrical approximations to the structure of musical pitch. *Psychological review* **89,** 305 (1982).
7. Krumhansl, C. L. *Cognitive foundations of musical pitch* (Oxford University Press, 2001).
8. Von Helmholtz, H. *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Longmans Green, 1912).
9. Schnupp, J., Nelken, I. & King, A. *Auditory neuroscience: Making sense of sound* (MIT press, 2011).
10. Zatorre, R. J., Chen, J. L. & Penhune, V. B. When the brain plays music: auditory–motor interactions in music perception and production. *Nature reviews neuroscience* **8,** 547–558 (2007).
11. Wolfe, J. M. *et al. Sensation & perception* (Sinauer Sunderland, MA, 2006).
12. Shepard, R. N. in *Psychology of Music* (ed Deutsch, D.) 343–390 (Academic Press, San Diego, 1982). ISBN: 978-0-12-213562-0.
13. Amano, S. Perception of tonal and microtonal structure over octaves. *Japanese Psychological Research* **34,** 89–95 (1993).
14. Krumhansl, C. L. The psychological representation of musical pitch in a tonal context. *Cognitive psychology* **11,** 346–374 (1979).
15. Kallman, H. J. Octave equivalence as measured by similarity ratings. *Perception & Psychophysics* **32,** 37–49 (1982).
16. Jacoby, N. *et al.* Universal and non-universal features of musical pitch perception revealed by singing. *Current Biology* **29,** 3229–3243 (2019).
17. Shepard, R. N. Multidimensional scaling, tree-fitting, and clustering. *Science* **210,** 390–398 (1980).
18. Demany, L. & Armand, F. The perceptual reality of tone chroma in early infancy. *The journal of the Acoustical Society of America* **76,** 57–66 (1984).

19. Gaser, C. & Schlaug, G. Brain structures differ between musicians and non-musicians. *Journal of neuroscience* **23,** 9240–9245 (2003).

20. Anderson, J. R. *The adaptive character of thought* (Psychology Press, 1990).

21. Shepard, R. N. Toward a universal law of generalization for psychological science. *Science* **237,** 1317–1323 (1987).

22. Shepard, R. N. & Metzler, J. Mental rotation of three-dimensional objects. *Science* **171,** 701–703 (1971).

23. Peretz, I. & Zatorre, R. J. Brain organization for music processing. *Annu. Rev. Psychol.* **56,** 89–114 (2005).

24. McPherson, M. J. & McDermott, J. H. Diversity in pitch perception revealed by task dependence. *Nature human behaviour* **2,** 52–66 (2018).

25. Wong, P. C. *et al.* Effects of culture on musical pitch perception. *PloS one* **7,** e33424 (2012).

26. Koelsch, S. & Siebel, W. A. Towards a neural basis of music perception. *Trends in cognitive sciences* **9,** 578–584 (2005).

27. Shepard, R. N. Circularity in judgments of relative pitch. *The journal of the acoustical society of America* **36,** 2346–2353 (1964).

28. Goldstein, E. B. & Cacciamani, L. *Sensation and perception* (Cengage Learning, 2021).

29. Dolscheid, S., Shayan, S., Majid, A. & Casasanto, D. The thickness of musical pitch: Psychophysical evidence for linguistic relativity. *Psychological science* **24,** 613–621 (2013).

30. McDermott, J. H., Schultz, A. F., Undurraga, E. A. & Godoy, R. A. Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* **535,** 547–550 (2016).

31. Lahdelma, I. & Eerola, T. Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Scientific reports* **10,** 8693 (2020).

32. Krumhansl, C. L. & Shepard, R. N. Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of experimental psychology: Human Perception and Performance* **5,** 579 (1979).

33. François, C., Chobert, J., Besson, M. & Schön, D. Music training for the development of speech segmentation. *Cerebral Cortex* **23,** 2038–2043 (2013).

34. Merzenich, M., Nahum, M. & van Vleet, T. *Changing brains: applying brain plasticity to advance and recover human ability* (Elsevier, 2013).

35. Jacoby, N. & Ahissar, M. What does it take to show that a cognitive training procedure is useful?: A critical evaluation. *Progress in brain research* **207,** 121–140 (2013).

36. Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R. & Pantev, C. Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of cognitive neuroscience* **16,** 1010–1021 (2004).

37. Trainor, L. J. & Corrigall, K. A. Music acquisition and effects of musical experience. *Music perception,* 89–127 (2010).

38. Patel, A. D. Language, music, syntax and the brain. *Nature neuroscience* **6,** 674–681 (2003).

39. Krumhansl, C. L. & Kessler, E. J. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological review* **89,** 334 (1982).

40. Marjieh, R., Harrison, P. M., Lee, H., Deligiannaki, F. & Jacoby, N. Reshaping musical consonance with timbral manipulations and massive online experiments. *bioRxiv* (2022).

41. Anglada-Tort, M., Harrison, P. M., Lee, H. & Jacoby, N. Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Current Biology* (2023).

42. Battleday, R. M., Peterson, J. C. & Griffiths, T. L. Capturing human categorization of natural images by combining deep networks and cognitive models. *Nature communications* **11,** 5418 (2020).

43. Hebart, M. N., Zheng, C. Y., Pereira, F. & Baker, C. I. Revealing the multidimensional mental representations of natural objects underlying human similarity judgements. *Nature human behaviour* **4,** 1173–1185 (2020).

44. Mehr, S. A. *et al.* Universality and diversity in human song. *Science* **366,** eaax0868 (2019).

45. Nelken, I. Music and the auditory brain: where is the connection? *Frontiers in Human Neuroscience* **5,** 106 (2011).

46. Aldwell, E., Schachter, C. & Cadwallader, A. *Harmony and voice leading* (Cengage Learning, 2018).

47. Humphreys, L. G. Generalization as a function of method of reinforcement. *Journal of Experimental Psychology* **25,** 361 (1939).

48. Deutsch, D. Octave generalization of specific interference effects in memory for tonal pitch. *Perception & Psychophysics* **13,** 271–275 (1973).

49. Idson, W. L. & Massaro, D. W. A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics* **24,** 551–565 (1978).

50. Kallman, H. J. & Massaro, D. W. Tone chroma is functional in melody recognition. *Perception & Psychophysics* **26,** 32–36 (1979).

51. Hoeschele, M., Weisman, R. G., Guillette, L. M., Hahn, A. H. & Sturdy, C. B. Chickadees fail standardized operant tests for octave equivalence. *Animal cognition* **16,** 599–609 (2013).

52. Hoeschele, M., Weisman, R. G. & Sturdy, C. B. Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, & Psychophysics* **74,** 1742–1760 (2012).

53. Regev, T. I., Nelken, I. & Deouell, L. Y. Evidence for linear but not helical automatic representation of pitch in the human auditory system. *Journal of cognitive neuroscience* **31,** 669–685 (2019).

54. Dellacherie, D., Roy, M., Hugueville, L., Peretz, I. & Samson, S. The effect of musical experience on emotional self-reports and psychophysiological responses to dissonance. *Psychophysiology* **48,** 337–349 (2011).

55. Koelsch, S., Schröger, E. & Tervaniemi, M. Superior preattentive auditory processing in musicians. *Neuroreport* **10,** 1309–1313 (1999).

56. Harrison, P. & Pearce, M. T. Simultaneous consonance in music perception and composition. *Psychological Review* **127,** 216 (2020).

57. Trainor, L. J. & Heinmiller, B. M. The development of evaluative responses to music:: Infants prefer to listen to consonance over dissonance. *Infant Behavior and Development* **21,** 77–88 (1998).

58. McDermott, J. H., Lehr, A. J. & Oxenham, A. J. Individual differences reveal the basis of consonance. *Current Biology* **20,** 1035–1041 (2010).

59. Vos, P. G. & Troost, J. M. Ascending and descending melodic intervals: Statistical findings and their perceptual relevance. *Music Perception* **6,** 383–396 (1989).

60. Zivic, P. H. R., Shifres, F. & Cecchi, G. A. Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences* **110,** 10034–10038. doi:10.1073/pnas.1222336110. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1222336110 (2013).

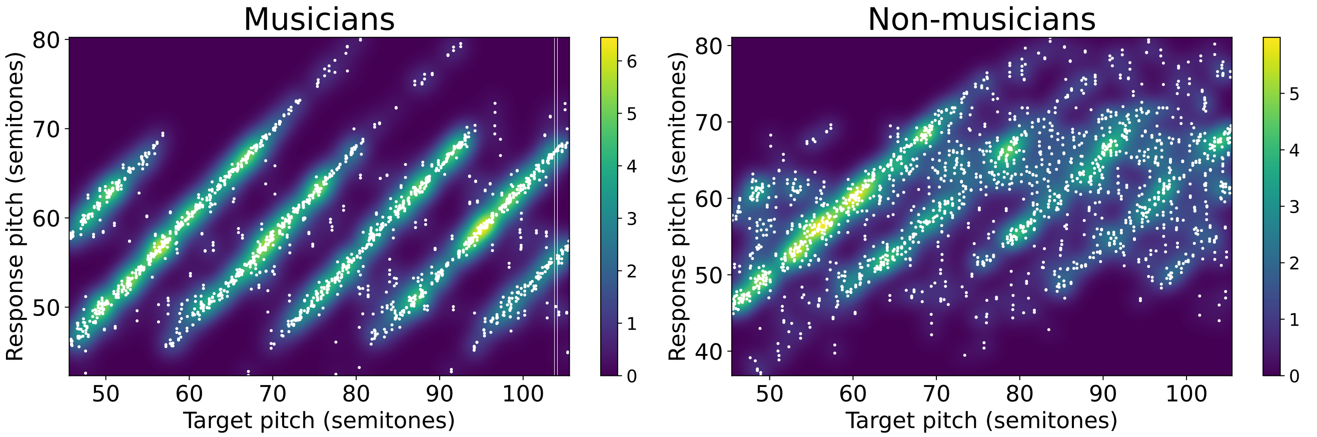
61. Eerola, T., Louhivuori, J. & Lebaka, E. Expectancy in Sami Yoiks revisited: The role of data-driven and schema-driven knowledge in the formation of melodic expectations. *Musicae Scientiae* **13,** 231–272 (2009).

62. Fogel, A. R., Rosenberg, J. C., Lehman, F. M., Kuperberg, G. R. & Patel, A. D. Studying musical and linguistic prediction in comparable ways: The melodic cloze probability method. *Frontiers in psychology* **6,** 1718 (2015).

63. Miller, K. & Gelman, R. The child's representation of number: A multidimensional scaling analysis. *Child development,* 1470–1479 (1983).

64. Woods, K. J., Siegel, M. H., Traer, J. & McDermott, J. H. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics* **79,** 2064–2072 (2017).

# Supplementary Information



**Supplementary Figure 1** Pitch imitation through singing. Target vs. response pitch distributions for musicians and non-musicians (using a Guassian kernel with $\sigma = 1.2$ semitones; density normalized relative to a uniform distribution).

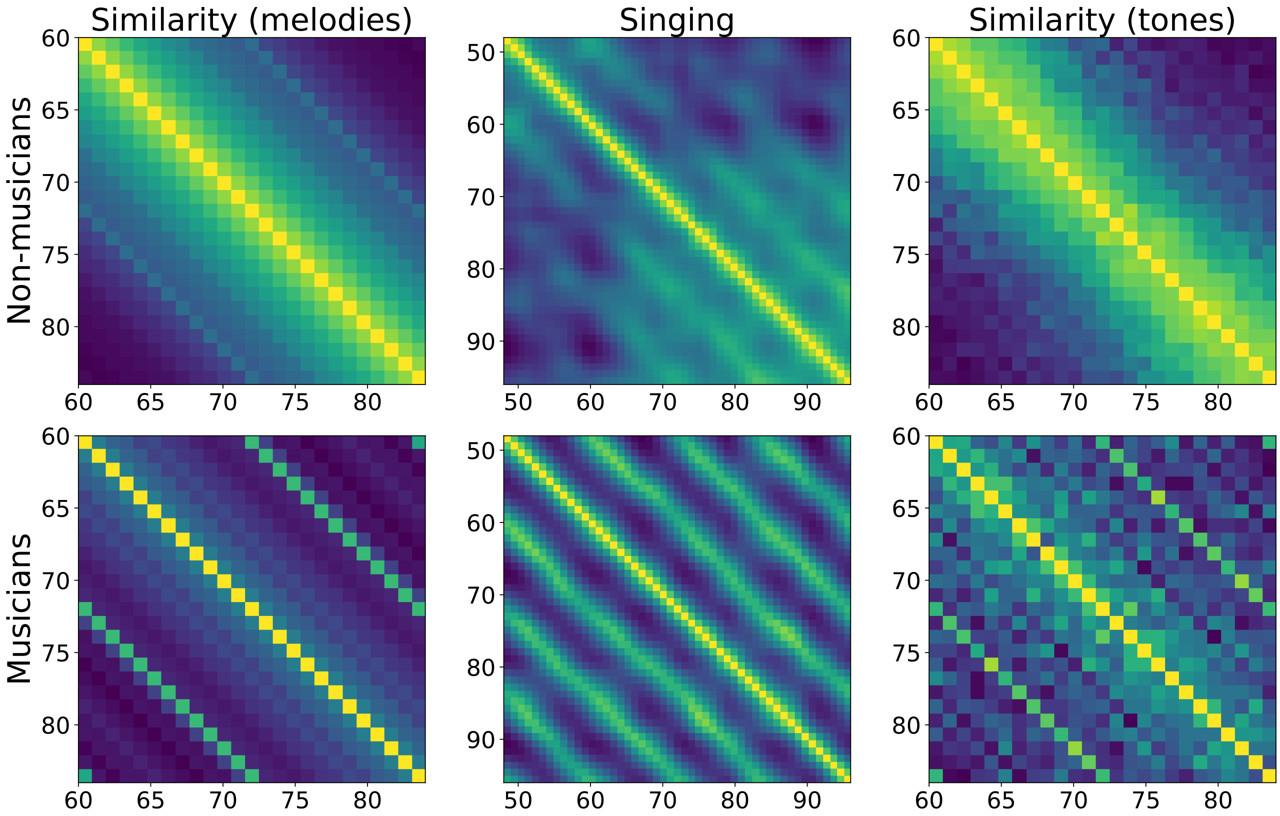
**Supplementary Table 1** Full list of model parameter values and their 95% confidence intervals

| Task | Group | Intercept | Linear | Octave | Tritone | Perfect | $\sigma_o$ |
|---|---|---|---|---|---|---|---|
| Similarity (melodies) | Non-musicians | $0.12 \pm 0.02$ | $0.64 \pm 0.05$ | $0.12 \pm 0.02$ | $0.01 \pm 0.01$ | $0.01 \pm 0.01$ | 0.63 |
| Similarity (melodies) | Musicians | $0.07 \pm 0.05$ | $0.21 \pm 0.07$ | $0.43 \pm 0.05$ | $0.00 \pm 0.01$ | $0.03 \pm 0.02$ | 0.51 |
| Singing | Non-musicians | $0.21 \pm 0.04$ | $0.16 \pm 0.07$ | $0.13 \pm 0.07$ | $0.00 \pm 0.00$ | $0.00 \pm 0.00$ | 2.33 |
| Singing | Musicians | $0.19 \pm 0.03$ | $0.03 \pm 0.02$ | $0.40 \pm 0.08$ | $0.00 \pm 0.00$ | $0.00 \pm 0.00$ | 2.10 |
| Similarity (tones) | Non-musicians | $0.13 \pm 0.03$ | $1.00 \pm 0.04$ | $0.01 \pm 0.02$ | $0.00 \pm 0.00$ | $0.02 \pm 0.01$ | 0.23 |
| Similarity (tones) | Musicians | $-0.03 \pm 0.11$ | $0.42 \pm 0.12$ | $0.45 \pm 0.08$ | $0.15 \pm 0.05$ | $0.02 \pm 0.06$ | 0.06 |

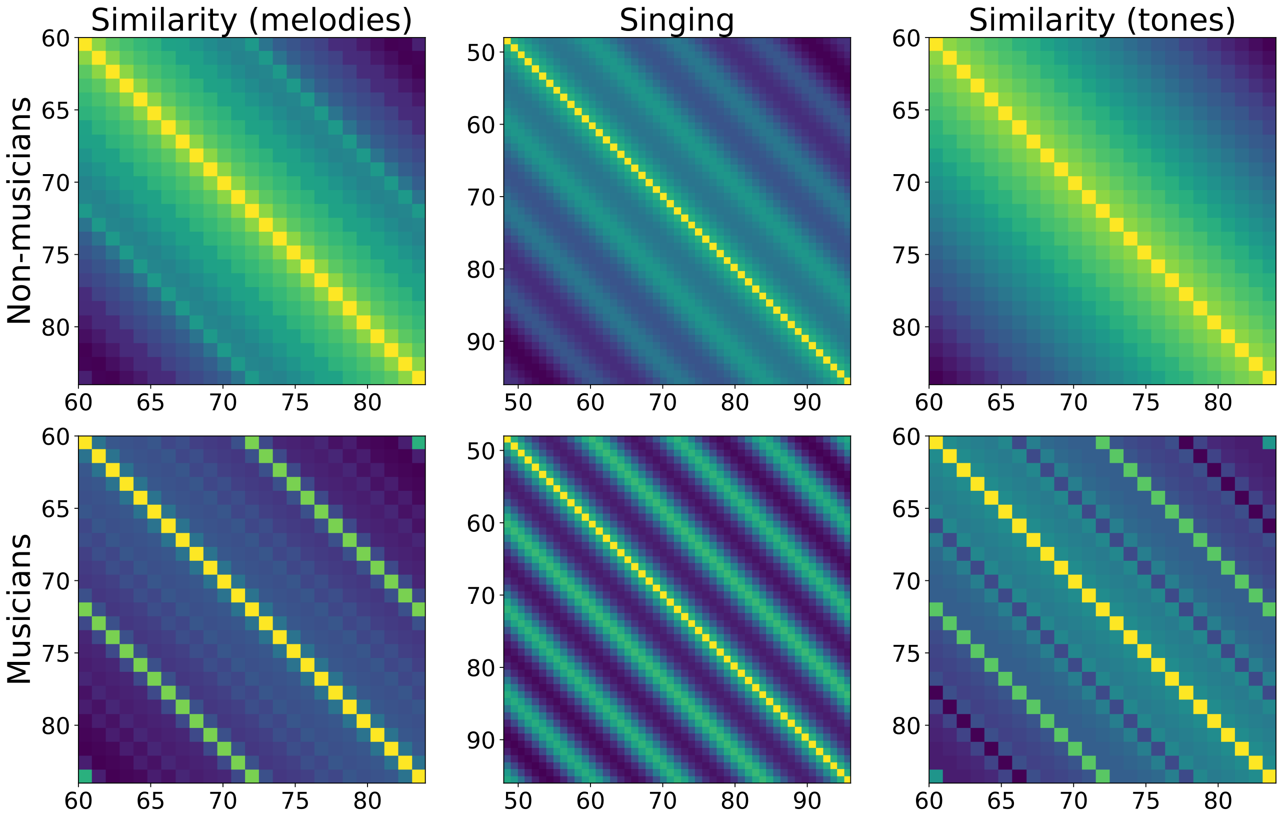
**Supplementary Table 2** Full list of evaluation metrics and their 95% confidence intervals using split-half bootstrap over participants with 1,000 repetitions.

| Task | Group | $R^2$ | $r_{dd}$ | $r_{mm}$ | $r_{dm}$ | $r$ |
|---|---|---|---|---|---|---|
| Similarity (melodies) | Non-musicians | $.95 \pm .01$ | $.99 \pm .00$ | $.99 \pm .00$ | $.98 \pm .00$ | $.98 \pm .00$ |
| Similarity (melodies) | Musicians | $.94 \pm .02$ | $.97 \pm .03$ | $.99 \pm .03$ | $.96 \pm .02$ | $.98 \pm .00$ |
| Singing | Non-musicians | $.54 \pm .11$ | $.74 \pm .12$ | $.94 \pm .15$ | $.69 \pm .10$ | $.82 \pm .04$ |
| Singing | Musicians | $.84 \pm .09$ | $.89 \pm .04$ | $.99 \pm .01$ | $.91 \pm .01$ | $.97 \pm .01$ |
| Similarity (tones) | Non-musicians | $.91 \pm .01$ | $.97 \pm .00$ | $.99 \pm .00$ | $.95 \pm .00$ | $.96 \pm .00$ |
| Similarity (tones) | Musicians | $.49 \pm .09$ | $.55 \pm .07$ | $.97 \pm .06$ | $.68 \pm .05$ | $.93 \pm .03$ |

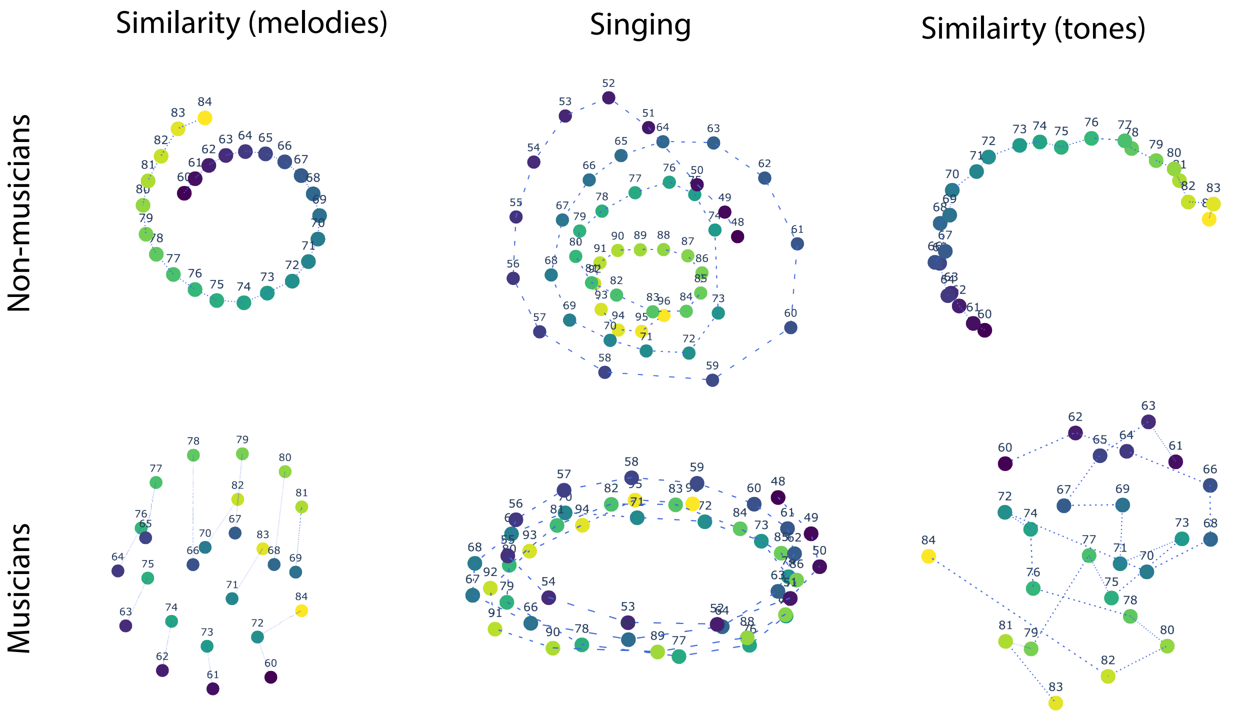
Note: The measures are: $R^2$ coefficient of determination, $r_{dd}$ data-data Pearson correlation, $r_{mm}$ model-model Pearson correlation, $r_{dm}$ data-model Pearson correlation, and $r$ data-model correlation corrected for attenuation (see Methods for full details).



**Supplementary Figure 2** Raw similarity matrices for the different behavioral paradigms considered in the paper.



**Supplementary Figure 3** Fitted model similarity matrices for the different behavioral paradigms considered.



**Supplementary Figure 4** Three-dimensional multidimensional scaling solutions for the raw unprocessed behavioral similarity matrices in Supplementary Figure 2. Left to right: similarity judgments over melodies that differ by a transposition, free imitation of two-note melodies via singing, and similarity judgments over pairs of isolated tones.