

ORIGINAL ARTICLE OPEN ACCESS

Consonance Shapes the Multisensory and Emotional Mappings of Musical Intervals Across English- and Mandarin-Speakers

Nicola Di Stefano¹ | Alessandro Ansani² | Andrea Schiavio³ | Suvi Saarikallio² | Petri Toiviainen² | Elvira Brattico^{4,5,6} | Charles Spence⁷

¹Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy | ²Centre of Excellence in Music, Mind, Body and Brain—Department of Music, Art and Culture Studies, University of Jyväskylä, Jyväskylä, Finland | ³School of Arts and Creative Technologies, University of York, York, UK | ⁴Center for Music in the Brain, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark | ⁵The Royal Academy of Music, Aarhus/Aalborg, Aarhus, Denmark | ⁶Department of Education, Psychology, Communication, University of Bari Aldo Moro, Bari, Italy | ⁷Department of Experimental Psychology, University of Oxford, Oxford, UK

Correspondence: Nicola Di Stefano (nicola.distefano@istc.cnr.it)

Keywords: consonance | crossmodal associations | dissonance | gradus suavitatis | multisensory

ABSTRACT

The exploration of sound's multisensory attributes has become increasingly central to the cognitive science of music, yet research has largely focused on a limited set of sensory dimensions, often within a single modality. Moreover, the extent to which these associations are consistently perceived by listeners from different countries and linguistic backgrounds remains an open question. To address this issue, we conducted a comprehensive investigation of how two groups of listeners—146 English-speakers from North America and the United Kingdom and 64 Mandarin-speakers from China—map harmonic intervals onto 23 sensory and affective dimensions spanning vision, touch, taste, and emotion. Our results revealed consistent associations across groups between musical intervals and gustatory, tactile, visual, and emotional attributes. Importantly, these multisensory mappings clustered according to the consonance/dissonance dimension: consonant intervals were rated as more sweet, smooth, warm, round, symmetric, and bright, whereas dissonant intervals were rated as more rough, cold, spiky, asymmetric, and dark. Crucially, country of origin and language had minimal influence on these mappings, as statistical analysis based on model comparisons overwhelmingly favored simpler models without group interaction terms. These findings suggest that consonance–dissonance may serve as a core organizing principle underlying multisensory mappings of musical intervals.

1 | Introduction

The past few decades have seen a surge of interest in the multiple sensory dimensions of sound across cognitive science. Studies have extensively explored cross-sensory associations between auditory percepts and non-auditory features in typical (i.e., non-synesthetic) individuals to better understand the

underlying mechanisms of perception, the integration of sensory information, and the cognitive and affective processes that shape how we experience and interpret the world. The available evidence demonstrates that these perceptual phenomena—commonly referred to as “crossmodal associations” or “crossmodal correspondences” [1]—are consistently observed across individuals [2, 3].

Nicola Di Stefano and Alessandro Ansani contributed equally

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Annals of the New York Academy of Sciences* published by Wiley Periodicals LLC on behalf of The New York Academy of Sciences.

Studies have primarily investigated how various musical features—such as pitch, timbre, mode, loudness, rhythm, and harmony—are associated with visuospatial properties like brightness, lightness, and elevation [4–11]. For instance, increases in pitch are linked to brighter visual stimuli [7] and higher spatial elevation [4]. Notably, many of these correspondences have been observed across remote cultures and languages. For instance, the correspondence between pitch–height has been observed in a tribe from Cambodia [12] and the bouba/kiki effect in a population from Namibia [13].

Expanding beyond basic auditory features, Maimon et al. [6] found that tonal stability—a concept related to the degree to which pitches and chords are perceived as anchored, resolved, or central within a tonal framework—is also associated with visual brightness (see also the work by Maimon et al. [14], for links between tonal and spatial relationships). Additionally, more consonant intervals have been linked to cooler, softer, more transparent, and less arousing color patches [15].

Researchers have explored auditory-based associations beyond the spatial senses, documenting a variety of mappings between auditory stimuli and gustatory/flavor stimuli [16, 17]. Studies consistently show that higher pitched sounds are associated with sweet and sour tastes, whereas lower pitched sounds are linked to bitter tastes [18–20]. Harmonic properties have also been examined, with consonant chords typically associated with sweet tastes and dissonant chords with sour tastes [21, 22].

Mesz et al. [23] conducted two experiments investigating taste–music associations. In the first one, musicians were invited to improvise using a MIDI keyboard with piano timbre on the basis of four taste words: sweet, sour, bitter, and salty. The analysis revealed consistent musical patterns—“bitter” improvisations were low-pitched and legato, “salty” improvisations were staccato, “sour” improvisations were high-pitched and dissonant, and “sweet” improvisations were consonant, slow, and soft. In a second experiment, non-musicians listened to excerpts from these improvisations and, in a forced-choice task, tried to identify which taste word each piece represented. The results revealed that participants identified the correct taste word at rates significantly above chance. Further research has explored auditory–olfactory crossmodal correspondences, with findings showing that higher pitched sounds (~G4) are matched with scents like candied orange and iris flower, whereas lower pitched sounds (~G3) are associated with odors such as musk and roasted coffee [24].

Although most studies have used verbal stimuli to evoke sensory features, Murari et al. [25] used physical materials such as wood and polystyrene foam to represent hardness and softness, sandpaper to convey smoothness and roughness, bitter and sweet liquids for gustatory contrasts, and cold and hot water for temperature variations. Additionally, they selected excerpts from the Western classical repertoire (e.g., Vivaldi, Bach, Mozart, and Wagner) as auditory stimuli. The results of their study revealed several consistent patterns of crossmodal associations among participants, with certain sensory qualities—such as smooth/rough and sweet/bitter—distinguishing musical excerpts more effectively than others, such as soft/hard (see Di Stefano and Spence [26] for a multisensory review on roughness).

Another strand of research has explored the association between musical intervals and emotional descriptors [5, 27, 28]. For example, Costa et al. [29] invited participants to rate harmonic intervals (i.e., combinations of two tones simultaneously played defined in terms of the number of scale steps separating the tones) using 30 adjectives commonly associated with musical expressiveness, such as cheerful/cheerless, serene/gloomy, and quiet/restless. Their findings revealed significant effects of both register and consonance/dissonance. Participants tended to rate low-pitched and dissonant intervals more negatively, whereas higher pitched and consonant intervals were rated more positively.

Despite a growing body of research on crossmodal associations involving musical stimuli, most studies have focused on a narrow range of sensory features—typically within one or two domains—and have used diverse methodologies, limiting the comparability and generalizability of findings. Additionally, the extent to which these associations remain consistent across participants from different countries, or continents, and with different linguistic backgrounds is largely unexplored. To advance current understanding, the present study systematically examines how musical intervals are associated with a broad set of sensory and affective dimensions in a group of English-speakers from North America and the United Kingdom and another group of Mandarin-speakers from China. In doing so, we aim to contribute to ongoing debates in music cognition, multisensory perception, and the foundations of crossmodal correspondences.

2 | Materials and Methods

2.1 | Participants

We recruited 210 participants through Prolific.com ($M_{\text{age}} = 37.40$ years, $SD = 13.66$, 61.42% females, 2 did not disclose their gender). One group was composed of 146 English-speakers mainly from the United Kingdom (87%), United States (5%), and Canada (5%) ($M_{\text{age}} = 40.26$ years, $SD = 14.52$, 57.30% females, one did not disclose their gender). Another group was composed of 64 Mandarin-speakers from China ($M_{\text{age}} = 30.89$ years, $SD = 8.50$, 68.75% females, one did not disclose their gender).

Information about musical background was collected using both subjective and objective measures. With regard to self-reported musical expertise, as assessed by the single-item measure proposed by Zhang and Schubert [30], 47.6% defined themselves as music-loving non-musicians, 29.0% non-musicians, 15.7% amateur musicians, 4.3% serious amateur musicians, 1.9% semi-professional musicians, and 1.4% professional musicians. The distribution of expertise categories did not differ significantly between the two samples, as indicated by a chi-square test, $\chi^2(5, N = 210) = 7.04, p = 0.218$.

To obtain a more quantitative assessment of musical expertise, we administered two subscales of the Goldsmith Musical Sophistication Index [31], namely, Musical Training (i.e., amount of formal musical training received) and Emotional Engagement with Music (i.e., ability to talk about emotions that music expresses). Both subscales had good reliability in our samples ($\alpha_{\text{MT-West}} = 0.92$, 95% CI [0.90, 0.94]; $\alpha_{\text{Emo-West}} = 0.74$, 95% CI [0.66, 0.80]; $\alpha_{\text{MT-East}} = 0.84$, 95% CI [0.78, 0.88]; $\alpha_{\text{Emo-East}} = 0.79$, 95%

CI [0.64, 0.87]). The samples were not statistically different in terms of emotional engagement with music ($U = 4243$, $p = 0.289$, $r_{rb} = 0.09$, 95% CI [-0.07, 0.25]), but the Eastern sample showed higher musical training ($U = 6209$, $p < 0.001$, $r_{rb} = 0.33$, 95% CI [0.17, 0.47]).

2.2 | Musical Stimuli

Using the piano Noire pure VST instrument (see https://www.native-instruments.com/fileadmin/ni_media/downloads/manuals/NOIRE_Manual.pdf for further information about the sampling), 144 auditory stimuli (wav, 24-bit, 44.1 kHz) were generated consisting of two piano notes played simultaneously, that is, harmonic intervals. Starting from the G below middle C (G_3 , 195.9 Hz) up to $F\#/Gb_4$ (369.9 Hz), we recorded all of the dyads from the minor second to the octave. In this process, each note in the range has been used as the root note, and all the above-mentioned intervals were generated from each root note. Therefore, the full pitch range was $G_3-F\#/Gb_5$. The stimuli were identical in terms of dynamics (velocity MIDI value = 100, corresponding to a *forte*) and duration (2 s).

2.3 | Sensory and Emotional Features

We selected 23 sensory and emotional/affective features. The sensory features spanned three domains: visuospatial, tactile, and gustatory. Visuospatial features included far/close, left/right, low/high, small/big, dark/bright, *maluma/takete* (i.e., curved and angular shapes, respectively), colorful/black and white, warm/cold hue, and symmetric/asymmetric. Tactile features included heavy/light, soft/hard, rough/smooth, and cold/hot (note that roughness has been conceived of as a sensory feature that can be associated with visual, auditory, and, to a lesser extent, gustatory stimuli [26]). Gustatory features included spicy, sweet, salty, bitter, and umami. Additionally, we incorporated five pairs of opposite emotional/affective adjectives: sad/happy, pleasant/unpleasant, tender/tough, dominant/submissive, and arousing/relaxing.

2.4 | Procedure

In an online procedure administered via Qualtrics, participants were presented with all 12 harmonic intervals, randomly selected from the sample of 144 stimuli. Each harmonic interval was presented once to each participant, and randomization ensured that all of the intervals were presented in a random order. Western participants completed the procedure in English, whereas Eastern participants completed it in Mandarin Chinese.

Participants were invited to rate each audio stimulus for all of the 23 sensory and affective dimensions using a 100-point visual analog scale (VAS). For bipolar features, the scale presented the opposite adjectives at the extreme of the scale (e.g., far/near). For unipolar features (e.g., bitter), only one term was presented close to one end of the VAS. During the rating task, participants were allowed to replay each interval as many times as needed. The median duration of the procedure was 22.49 min ($SD = 12.08$). The protocol was approved by the Research Ethics

and Integrity Committee of the National Research Council of Italy (no. 0323801).

2.5 | Statistical Analysis

All analyses were performed on RStudio (2024.04.2). To model the differences of the intervals across all dimensions, a generalized linear mixed model (GLMM) approach was carried out via the *lme4* package [32]. The participants' and the stimuli's keys were modeled as random intercepts to account for their variability [33]. The following formula was used:

$$\text{Dimension value} = \text{interval} + (1|\text{ID}) + (1|\text{Key})$$

In greater detail, as the distribution of several of our dependent variables diverted from normality, two instances of every model were built: a general (i.e., Gaussian) version and a generalized version, wherein a gamma distribution was used with a logarithmic link function [34]. The two instances were compared in terms of Akaike and Bayesian information criteria (AIC and BIC) and Bayes factors (BFs) computed via BIC approximation (see next paragraph). Except for the gustatory dimensions, which showed a better fit with the gamma-distributed generalized models, all the other dimensions showed a significantly better fit with the Gaussian model. The reported significance levels derive from the instance with the best fit.

To inspect the role of the group (i.e., Western and Eastern), a model comparison approach was used [35]. Each model (one per dimension) was fit into two versions: simple and complex. In the simple model, the interval was the only predictor; in the complex model, an interaction term *interval* $\times$ *group* was added. Subsequently, the models were compared based on BIC and AIC. Moreover, BFs were computed with BIC approximation. (Given the high number of observations ($N = 2520$), likelihood ratio tests (LRTs) were avoided because they do not penalize model complexity. Furthermore, LRT is inherently biased toward more complex models, as the likelihood always increases with additional parameters. Moreover, LRT does not account for sample size, making it unsuitable for evaluating model parsimony). Such BFs represent a ratio of the relative evidence for the simpler model over the complex and answer the question "Under which model are the observed data more probable?". They are computed as follows: $BF_{10} = e^{(BIC_0 - BIC_1)/2}$ where 1 and 0 represent the complex and simple models, respectively [36].

This procedure was conducted to assess whether the addition of the interaction term improved the fit significantly, thus confirming the importance of the group. Notably, the two metrics behave differently when it comes to penalization. Although AIC penalization is $2k$ (k being the number of parameters being estimated), BIC penalization is larger because it explicitly takes into account the number of observations, namely, $\ln(n)k$. The former is therefore more liberal and suitable for prediction purposes, whereas the other's protection against overfitting makes it more conservative and optimal for inference purposes [37].

Subsequently, through the *emmeans* package [38], we computed the estimated marginal means (EMMs) of all intervals and performed deviation contrasts (also known as sum contrasts

[39]) for the best-performing model configuration. Thus, we compared the EMM of each interval against the grand mean of the dimension at hand. To control for multiple comparisons, the contrasts were adjusted through Bonferroni correction (i.e., for 12 tests).

Lastly, a *K*-means algorithm was used to generate five alternative cluster solutions (range: 2–6). Several indices were computed with *NbClust* [40] to select the correct number of clusters: gap statistic [41], silhouette [42], KL7 [43], CH [44], and SD validity index [45]. Multidimensional scaling (MDS) was performed using the *factoextra* package [46].

As an attempt to quantify the degree of invariance between the two groups, we computed the cosine similarity between the two matrices of mean ratings (12 intervals $\times$ 23 dimensions). Cosine similarity assesses the alignment (angular similarity) of multivariate profiles rather than absolute differences in magnitude, making it appropriate for evaluating pattern-level invariance across groups [47]. As not all variables were measured on the same scale (e.g., hue ranges from 0 to 2800 pixels), all dimensions were *z*-transformed prior to the computation, thus ensuring that each variable contributed equally to the similarity metric. The cosine similarity was then computed on the vectorized standardized matrices.

In addition, to explore the geometric organization of the sensory and affective dimensions in a two-dimensional space, a principal component analysis (PCA) was conducted. Prior to the PCA, all the ratings were *z*-transformed within participants, thus ensuring an equal contribution to the analysis by all the participants. The PCA was conducted separately for the English- and Mandarin-speakers' groups, as well as for the full sample. Our focus was not on component retention or variance explanation, but rather on visualizing the relative positioning of the dimensions in the space defined by the first two principal components. For visualization, we plotted the first two components and their squared cosine values ($\cos^2$), which indicate the quality of representation of each score on the selected components.

3 | Results

3.1 | Model Comparison

According to the BIC, the simpler model is always preferable. In other words, the fit improvement of the models, which consider the group (i.e., Western/Eastern), is not sufficient to justify the group's role in predicting participants' ratings. When considering the BFs, the evidence for the simpler models is overwhelming, the smallest BF being 6957, and the largest being 3.60×10^{18} (see Table S1).

When considering the AIC metric, however, in 9 out of 23 cases the complex model is preferred. This means that, in such models, the ratings of some of the intervals are different in the two samples. It is worth stressing that a statistically significant interaction does not necessarily imply substantive or theoretical relevance, particularly in the context of model comparison. Although the inclusion of the group interaction term leads to a better fit according to the AIC, this does not inherently mean that the

interaction effect is noteworthy. The discrepancy between AIC and BIC in model selection underscores the importance of considering model complexity alongside interpretability. Given that BIC applies a stricter penalty for additional parameters, its lack of support for the more complex model suggests that the interaction effect may be an artifact of model flexibility rather than a meaningful effect. Thus, statistical significance alone should not be conflated with practical or theoretical importance, particularly in the presence of multiple comparisons and high-dimensional estimation. Although many of the observed differences appear minor and not readily interpretable, the variation in ratings of "perfect" consonances (e.g., see Figures S1 and S4) may plausibly relate to their distinct functional roles in Western tonal and traditional Chinese music. Consistent with the model comparison results, we will describe the estimates stemming from the simple version of all models. However, according to the guidelines provided by Burnham and Anderson [48, p. 71], it was deemed necessary to report in the Supporting Information the graphs of the features whose evidence for the complex model was defined as substantial (i.e., $\Delta AIC > 10$, see Figures S1–S7).

3.2 | Modeling Results

The results of the GLMMs indicate significant relationships between many sensory and affective features and musical intervals (see Figure 1). In the following sections, we explore these dimensions in greater detail, focusing on gustatory, spatial and physical, emotional and affective, tactile, and color and shape associations.

3.2.1 | Gustatory

The intervals did not differ in the umami dimension, except for the minor second, which was rated significantly less umami than the average. In terms of spiciness, minor third intervals were rated as not spicy, whereas the major seventh is associated with higher values of spiciness. The pattern of the salty dimension presents high values for seconds, tritone, and major sevenths and low values for major thirds, perfect fifths, and sixths. The minor second is perceived as the most bitter interval, followed by major seventh, tritone, and major seconds. Major third, fifth, and minor sixth were the least bitter intervals. The dimension of sweetness showed an opposite pattern, with major sixth being the sweetest interval, followed by fifth, minor sixth, major third, and perfect fourth. The least sweet were minor second, tritone, major second, and major seventh.

3.2.2 | Spatial and Physical Attributes

No differences were found in the far/close and left/right dimensions. Low/high ratings seem to be affected by the size of the interval, with the seconds and minor third intervals being associated with the low end of the spectrum, whereas the fifth, sixths, and major seventh are associated with the high end of the spectrum. The same intervals were also rated as the lightest ones (except for the major seventh). The minor seconds were also rated as bigger and heavier. Concerning size, the major second appears

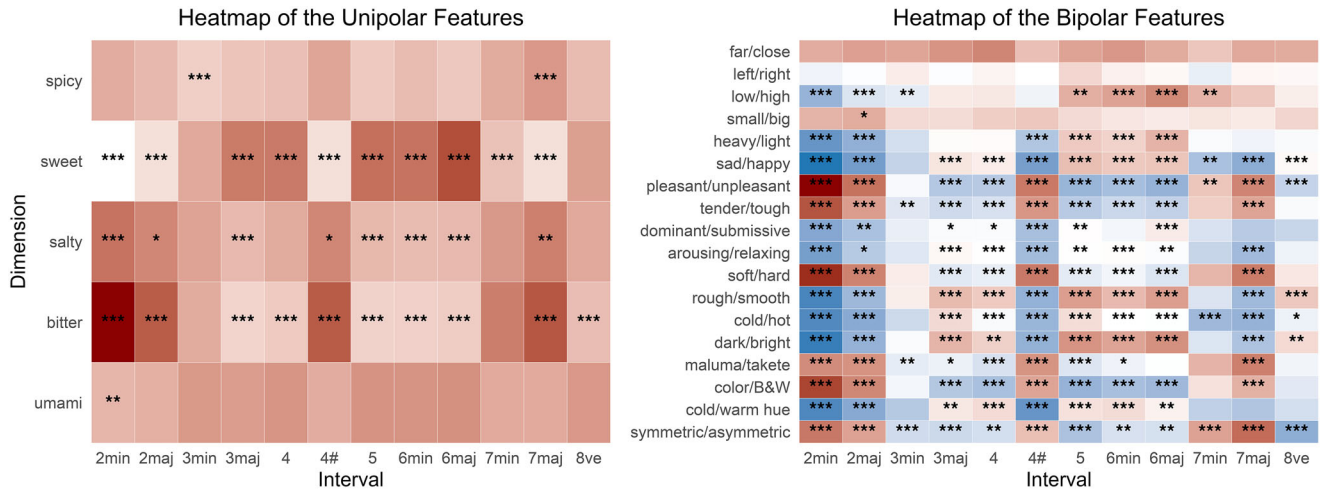


FIGURE 1 | Left. Heatmap of the unipolar features. Hues vary according to estimated marginal mean (EMM) values, with higher EMM values being represented by more saturated colors. Right. Heatmap of the bipolar features. Hues vary according to EMM values, with higher EMM values being represented by more saturated colors. Blue/red colors represent the feature at the left/right side of the bipolar continuum (e.g., far is blue and close is red). Significant relationships between intervals and unipolar/bipolar features are marked with “*,” that is, the EMM of the interval is significantly higher/lower than the grand mean of the same dimension. To provide details on the significance, we added *** for $p < 0.001$, ** for $0.001 < p < 0.009$, and * for $0.010 < p < 0.049$. Empty cells indicate the absence of the deviation contrast significance.

to be the only “big” interval. Finally, the tritone was perceived as heavy.

3.2.3 | Emotional/Affective

All emotional and affective features polarize intervals. Negatively valenced features (e.g., sad, unpleasant, tough, arousing, dominant) tend to be associated with seconds, tritones, and sevenths, while positively valenced ones (e.g., happy, pleasant, tender, relaxing, submissive) with major thirds, perfect fourths and fifths, sixths, and octaves.

3.2.4 | Tactile

The dimensions in the tactile domain showed great consistency in terms of the association patterns across the eight intervals. Minor second, tritone, and major seventh were judged as harder, rougher, and colder. Conversely, major third, fourth, fifth, and major sixth were deemed softer, smoother (see Figure 2), and hotter. Octaves and minor sixths were also judged as smooth (see Figure 2), the latter being also perceived as soft. Finally, minor sevenths were perceived as cold.

3.2.5 | Color and Shape

A high degree of consistency across participants was found in the color (see Figure 3) and shape domain. The seconds and tritone were consistently perceived as dark, spiky (*takete*), B&W, asymmetrical, and with a cold hue. The major seventh had the same associations except for the cold hue. In contrast, major thirds, fourths, fifths, and sixths were judged as bright, round (*maluma*), colorful, symmetrical, and with a warm hue. The

major sixth lacked significance in the roundness dimension, whereas the minor third was perceived as round and symmetric. The minor seventh was asymmetric, whereas the octave was symmetric. Interestingly, dissonances tend to be associated more likely with blue hues (e.g., sevenths, tritone, and seconds), whereas consonances with yellowish ones (octaves, sixths, perfect fourths and fifths, and thirds).

3.3 | Cluster Analysis, MDS, and Cosine Similarity

We noticed that the patterns of participants’ ratings seemed to differentiate between consonant and dissonant intervals. To further inspect such a claim, we ran a cluster analysis. As an additional confirmation of the non-determinant role of the group, three cluster analyses were performed on the English-speakers, Mandarin-speakers, and complete samples, including all of the 23 sensory/emotional features. In all cases, all of the indices converged on a clear two-cluster solution. In Figure 4, we present the MDS representation for all samples.

For the sample who ran the experiment in English, seven intervals fell into Cluster 1 (within cluster sum of squares—WCSS = 38.10, in blue in Figure 4): minor third, major third, perfect fourth, perfect fifth, minor sixth, major sixth, and octave. Cluster 2 (WCSS = 43.67, in red in Figure 4) was composed of the remaining intervals: minor second, major second, tritone, minor seventh, and major seventh. The ratio of the between-cluster sum of squares to the total sum of squares was 67.70%, thus indicating a strong separation between the clusters.

In the Mandarin-speakers sample, the solution was overlapping, in that the same intervals fell into Cluster 1 (WCSS = 47.44) and Cluster 2 (WCSS = 59.23). The ratio of the between-cluster sum of squares to the total sum of squares was 57.81%. When considering the whole sample, Cluster 1 (WCSS = 28.90, in blue in Figure 4)

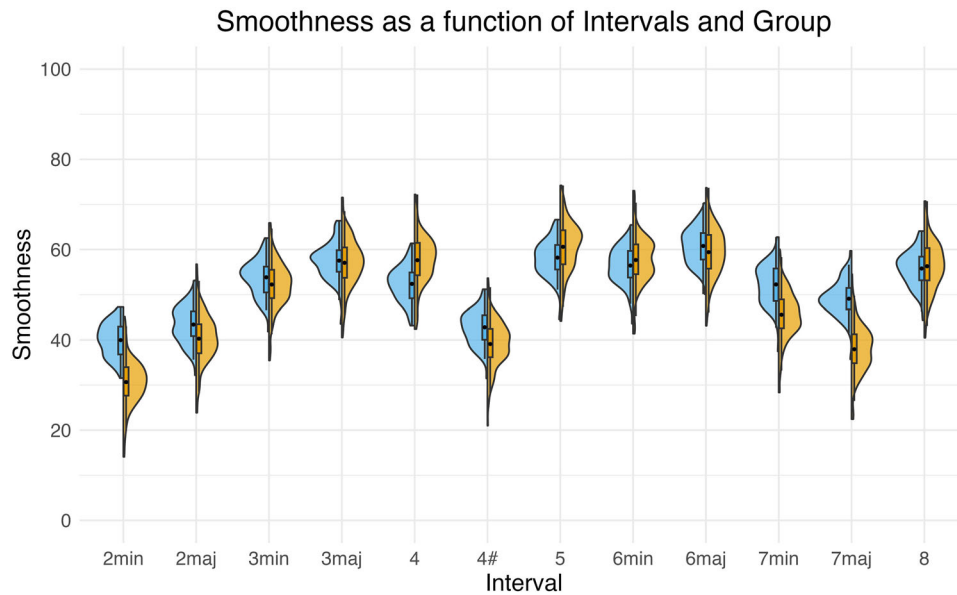


FIGURE 2 | Smoothness as a function of intervals and group. On the y-axis, a value of 100 indicates that participants rated the interval as smooth, whereas 0 indicates that the interval is perceived as rough. The form of the violin plots indicates the distribution curve. The boxplots within each violin represent interquartile ranges (IQRs). Black dots within the boxplots indicate mean values. Colors represent groups, namely, Mandarin-speakers (blue) and English-speakers (orange).

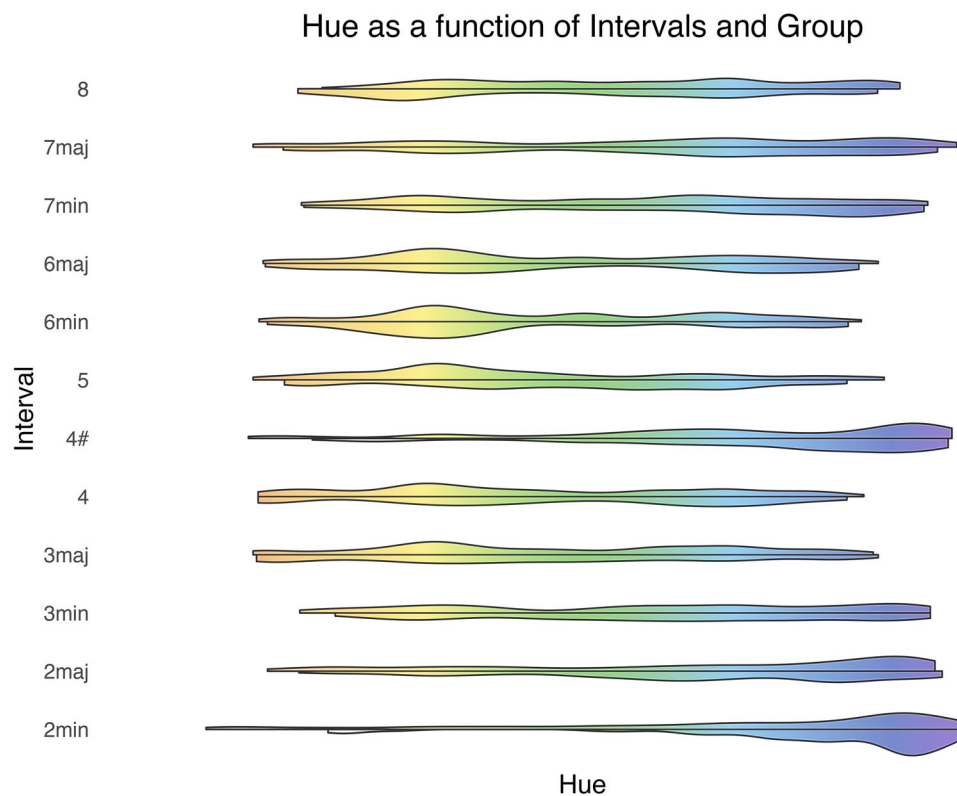


FIGURE 3 | Hue as a function of intervals and group. Split violin plots represent the hue distribution across intervals. The density estimation was performed using kernel density estimation (KDE) with a bandwidth adjustment factor of 0.5 to provide a more localized representation of the data distribution. The violin plots were trimmed and scaled to ensure equal area across conditions. Different from the other figures, in this one, the observed (not predicted) values are plotted. However, to reduce the influence of extreme values, the dataset was filtered by removing observations below the 10th percentile and above the 90th percentile within each interval separately, preserving the relative distribution within each group. The upper split violin represents the English-speakers' sample; the lower one pertains to the Mandarin-speakers' sample.

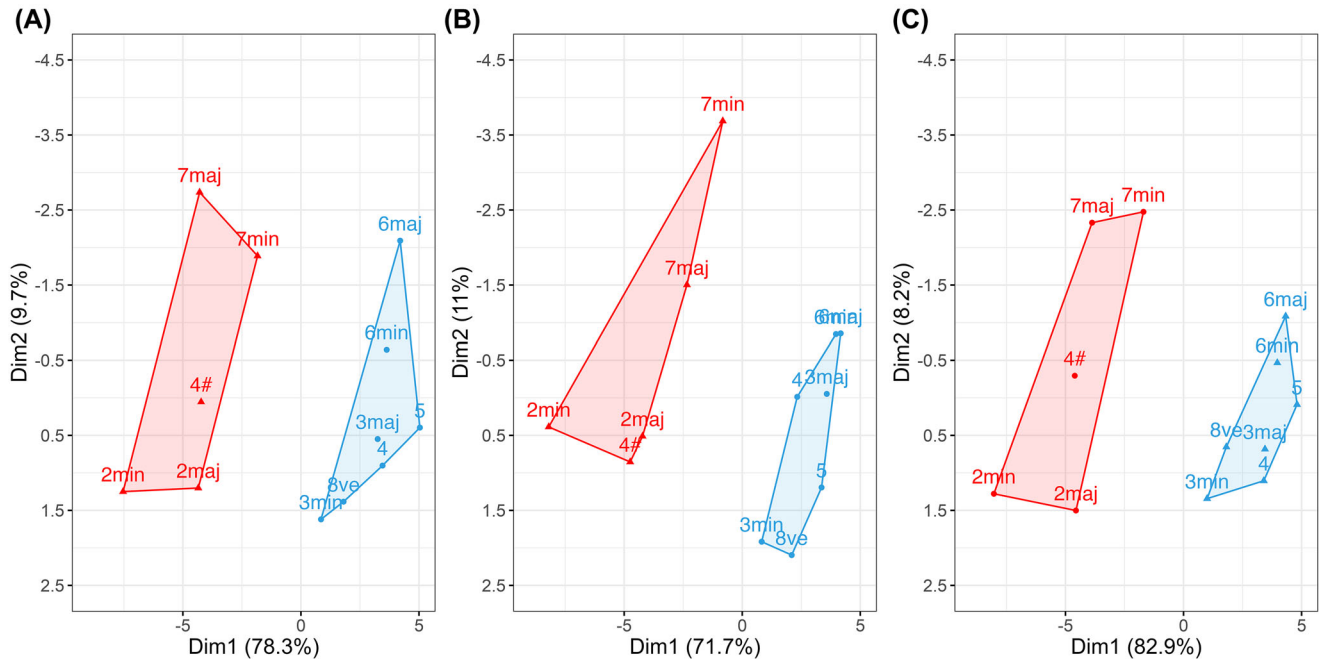


FIGURE 4 | Multidimensional scaling (MDS) solutions for English-speakers (A), Mandarin-speakers (B), and overall samples (C). In all solutions, all the dissonant intervals appear to be in Cluster 1 (red), whereas the consonant intervals group to Cluster 2 (blue). Note that in the Mandarin-speakers sample, the major and minor sixth intervals overlap.

and Cluster 2 (WCSS = 44.42) appear well differentiated. The ratio of the between-cluster sum of squares to the total sum of squares was 71.03%.

The cosine similarity computed between the two matrices of mean ratings (12 intervals $\times$ 23 dimensions) was 0.71. Although no universal thresholds exist, values higher than 0.70 are commonly interpreted as moderate-to-high degree of structural similarity in high-dimensional settings, thus supporting the similarity between the patterns of the two samples. It is worth noting that, due to the z -transformation of the scores, cosine similarity is mathematically equivalent to Pearson's correlation ($r(274) = 0.71$, 95% CI [0.65, 0.77], $p < 0.001$, see Figure S8).

Interestingly, in all samples, the cluster differentiation appears to overlap with the differentiation between consonant and dissonant intervals. Such a result is of particular interest if we consider that nowhere in the experimental procedure was the concept of consonance/dissonance mentioned, nor was it implied. Furthermore, when inspecting the x and y coordinates of the MDS solutions, it is quite straightforward to interpret their intrinsic sense. The intervals seem to be ordered on the x -axis from the most dissonant to the most consonant. For this reason, we correlated the x Cartesian coordinates of the intervals with two indices of consonance/dissonance, namely, roughness and *gradus suavitatis*.

Roughness was calculated using two different metrics, one audio-based and the other symbolic. For the audio-based, we used Sethares' model [49] as implemented in the MIR toolbox [50]. The process begins with extracting spectral peaks from the audio signal to identify prominent frequency components. Roughness is then computed for each frequency pair using the Plomp–Levelt curve [51], which models dissonance based on frequency

separation. The total roughness is obtained by summing these values, weighted by the amplitudes of the peaks.

For the symbolic metric, we used the Hutchinson and Knopoff [52] model as implemented in the *incon* R package [53]. This model computes sensory dissonance from symbolic pitch representations by estimating interference between partials of the component tones according to critical-band interactions derived from the Plomp–Levelt framework. Unlike the Sethares implementation, which operates directly on the acoustic signal and therefore incorporates timbral and spectral properties of the stimuli, the Hutchinson and Knopoff model relies on abstract pitch relationships and harmonic structures.

Gradus suavitatis (degree of agreeableness), proposed by Leonhard Euler [54], quantifies the consonance of an interval based on the simplicity of its frequency ratio, assuming just intonation [55]. For an interval ratio p/q , it is defined as $1 + \sum_i e_i (r_i - 1)$, where r_i are the prime factors with multiplicity e_i of the least common multiple of p and q . Lower values indicate greater consonance, as simpler frequency ratios are associated with more agreeable auditory perception.

The x coordinates did not correlate with audio-based roughness' metric in the complete sample ($\rho = -0.34$, $p = 0.287$) nor in the English/Mandarin-speakers' groups separately ($\rho_W = -0.42$, $p = 0.168$; $\rho_E = -0.23$, $p = 0.457$). However, they were highly correlated with the symbolic metric of roughness ($\rho = -0.77$, $p = 0.005$, see Figure 5) and with *gradus suavitatis* ($\rho = -0.77$, $p = 0.003$, see Figure 5) in the overall sample (*gradus suavitatis* resulted in being highly correlated with the symbolic metric of roughness, $\rho = 0.83$, $p < 0.001$). As for the y coordinates of the MDS, they seem to imply a measure of the intervals' size. Thus, we correlated them with a measure of the interval size in semitones

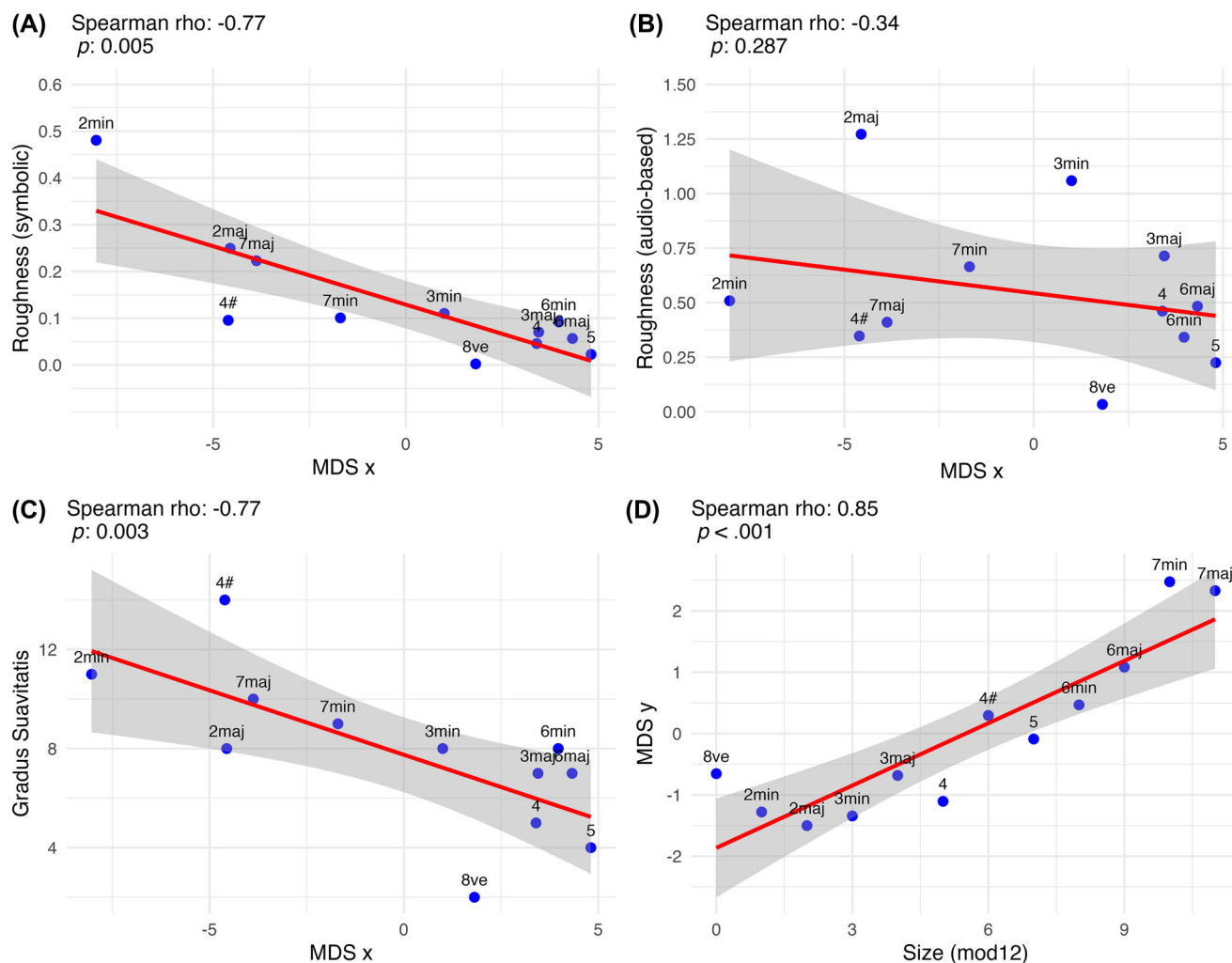


FIGURE 5 | Correlation between the x-axis of multidimensional scaling (MDS) and roughness as computed by the Hutchinson and Knopoff (1978) model (A) and by the Sethares [104] model (B). Correlation between the x-axis and *grados suavitatis* (C). Correlation between the y-axis and interval size (D).

assuming octave equivalence ($\text{Distance}_{\text{mod}} = \text{Semitone count mod } 12$), finding a very sound correlation ($\rho = 0.85$, $p < 0.001$, see Figure 5). This correlation was stronger in the Western sample as opposed to the Eastern ($\rho_W = 0.93$, $p < 0.001$; $\rho_E = 0.75$, $p = 0.006$).

3.4 | Principal Component Analysis

Figure 6 shows the PCA of the sensory and affective dimensions for the English-speaking (A), Mandarin-speaking (B), and overall (C) samples. Across samples, the first two components accounted for a comparable proportion of variance ($\text{Dim1} \approx 32\%–35\%$; $\text{Dim2} \approx 9\%–11\%$) and revealed a highly similar dimensional organization. A closer visual inspection suggests that affective and hedonic dimensions (e.g., happy, relaxing) contribute prominently to the first component (x-axis) but do not fully determine the spatial structure. Several perceptual and sensory descriptors (e.g., height, size) project along partially independent directions, suggesting additional dimensions beyond a purely hedonic axis. In particular, the second component (y-axis) seems to capture variations in perceived intensity or sensory salience, partially orthogonal to hedonic evaluation. The cosine-squared values

indicate that not all ratings are equally well represented by the first two components, consistent with a multidimensional organization of responses rather than a single evaluative dimension. Notably, this overall pattern is qualitatively consistent across English- and Mandarin-speakers.

4 | Discussion

This study reports an investigation of the multisensory mappings between musical intervals and various sensory and affective qualities in two samples, one comprising participants from the United Kingdom and North America who completed the study in English, and another comprising participants from China who completed the study in Mandarin. Notably, despite differences between the two samples in terms of musical training, the findings reveal consistent cross-group mappings between musical intervals and a range of gustatory, visual, tactile, and emotional attributes. Furthermore, MDS analysis demonstrated that these associations led to a clear partitioning of the harmonic space into two distinct regions, that is, consonance and dissonance. These

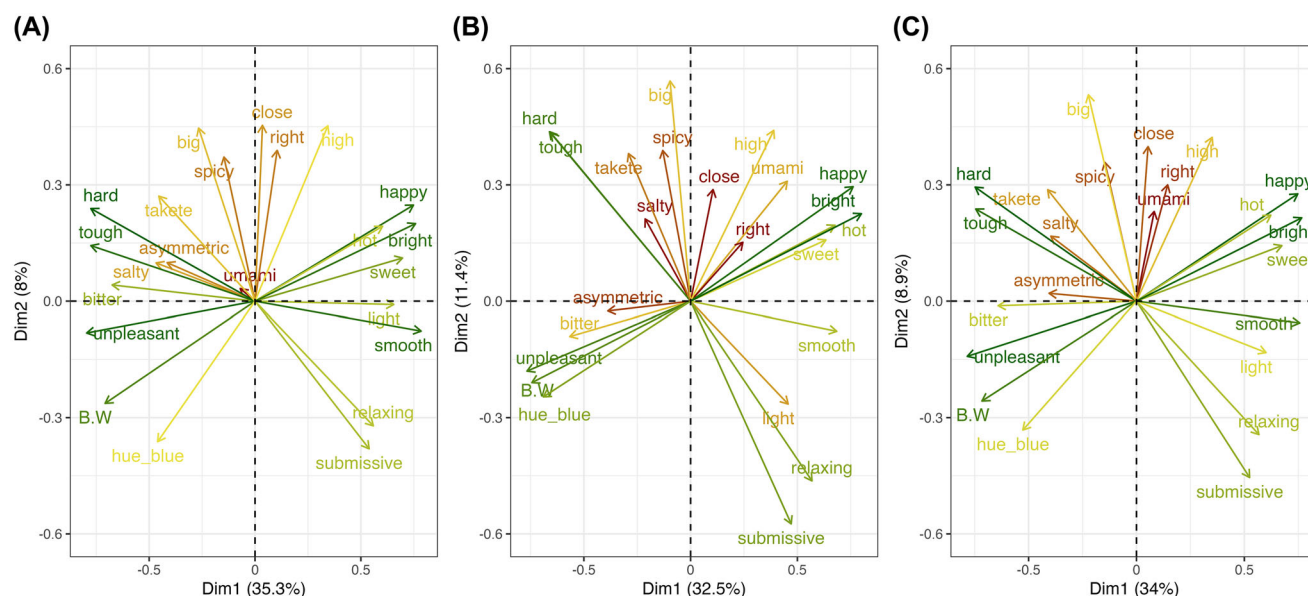


FIGURE 6 | The English-speakers (A) and Mandarin-speakers (B) samples show very similar principal component analysis (PCA) solutions, with comparable orientations of ratings across the two dimensions, thus indicating a consistent perceptual structure across groups. The similarity across samples is reflected in the overall PCA solution (C). It is worth stressing that the PCA configuration displays a star-like structure, suggesting that the overall pattern is not driven by a single dominant dimension. Colors correspond to the $\cos^2$ values, which express the quality of representation of each variable on the PCA axes. The gradient ranges from red (low values) to green (high values), with yellow indicating intermediate values.

categories emerged as the primary underlying factors mediating the observed mappings in both groups.

The strength and prevalence of significant relationships between intervals and multisensory/emotional attributes varied across domains. Emotional attributes showed the largest proportion of significant associations, followed by tactile and visual (color/shape) domains, with gustatory mappings being somewhat less pronounced. In contrast, only a few spatial and physical features exhibited significant relationships with musical intervals. This pattern likely reflects the well-established salience of emotional responses in music perception [56]. It also mirrors the predominance of crossmodal associations involving spatially grounded senses such as vision and touch, compared to the less consistent or weaker links with chemical senses, particularly taste [1, 57].

When the participants were presented with contrasting dimensions (e.g., low/high) or unipolar dimensions (e.g., bitter), most ratings were polarized on the basis of the consonant or dissonant nature of the musical stimuli (see Costa et al. [29], for similar findings using emotional descriptors). Specifically, consonant intervals tend to be perceived as lighter, happier, more pleasant, softer, more tender, more relaxing, smoother, more *maluma*, more colorful, hotter, and more symmetric. In contrast, dissonant intervals were perceived as bigger, heavier, sadder, more unpleasant, tougher, more dominant and arousing, harder, rougher, colder, darker, more *takete*, more black-and-white, and more asymmetric.

Nearly all variance in the data can be accounted for by consonance and dissonance alone. The two dimensions of the MDS analysis likely represent distinct perceptual qualities of musical intervals. Specifically, in Figure 4, the horizontal axis appears to

reflect the degree of consonance, with more consonant intervals positioned on the right and more dissonant ones on the left, as suggested by the strong correlation between the horizontal axis and consonance indices such as roughness and Euler's *gradus suavitatis* [58, 59]. Meanwhile, the vertical axis corresponds to interval size, with intervals tending to increase in size from bottom to top, except for the octave.

The prominence of consonance–dissonance as the principal axis of multisensory and emotional mappings suggests that this dimension may serve as a core organizing principle in the cognitive representation of musical structure. This interpretation aligns with a substantial body of research indicating that the perception of consonance and dissonance is grounded in neurobiological mechanisms [60–65]. These results demonstrate that a single, musically intrinsic feature (i.e., harmonic stability) can predict cross-sensory mappings across multiple modalities beyond the auditory domain.

Consonance and dissonance are commonly defined in terms of the perceptual effects they generate in listeners, with consonances typically perceived as smooth, stable, and harmonious, and dissonances as rough, tense, or unstable [62]. Considerable debate remains regarding the primary mechanisms underlying consonance perception, as well as the respective roles of biological predispositions and culturally situated musical practices. Recent work has increasingly suggested that consonance and dissonance emerge from interactions between perceptual constraints and cultural exposure [66–68]. Studies by Lahdelma and Eerola [69, 70] showed that judgments of consonance and dissonance do not merely reflect or mirror the commonly assumed related constructs of pleasantness or preference, indicating that they likely index partially distinct perceptual dimensions or factors. In this line, the comprehensive review by Harrison

and Pearce [53] proposed that consonance judgments are best explained as a combination of chord harmonic, roughness, and familiarity. Subsequent studies have largely supported and further refined these conclusions [71–73]. That being said, for the purposes of the present paper, consonance and dissonance are primarily used as descriptive categories for identifying groups of musical intervals that emerged from MDS, with no implications regarding the ultimate origins or mechanisms underlying consonance perception.

When explaining the data based on the hypotheses that are currently discussed in the literature to date [1, 74], the affective mediation hypothesis appears to be the strongest candidate to account for many of the consistent mappings that emerged in our study [75]. According to this account, crossmodal associations might be explained by the similar emotional profile, or meaning, of the stimuli that happen to be associated [76, 77]. For instance, colors and musical excerpts might be consensually matched by observers due to the shared emotional meaning [67]. This hypothesis would be supported by the emotional attributes ascribed to or perceived in musical intervals [27, 29, 61, 78] and the clear emotional connotations of many investigated features, such as rough/smooth, *maluma/takete*, and sweet/salty [26, 79, 80]. For instance, our finding indicating that consonances (i.e., major thirds, perfect fourths and fifths, major and minor sixths, and octaves) were more frequently associated with yellow hues may align with the studies demonstrating that such warm hues are typically rated as pleasant [81, 82, 10]. However, the color-emotion literature also indicates that short-wavelength hues (e.g., blue–purple) often receive positive valence ratings, which complicates a straightforward interpretation. Adding further nuance, evidence suggests that warm–cool judgements correlate with like–dislike ratings in British observers but not in Chinese observers [83]. Furthermore, the literature on the aesthetic/hedonic appreciation of other sensory qualities, such as “salty” and “B&W,” reports mixed and often context-dependent findings [84–87], thereby precluding a unidirectional pleasant/unpleasant interpretation of our results.

A PCA was conducted to further refine the interpretation of the results, particularly with respect to the role of pleasantness and unpleasantness. The first PCA dimension accounted for approximately one-third of the total variance, suggesting the presence of a meaningful hedonic or affective component in the data. This is perhaps unsurprising, given that many of the sensory descriptors used in the study can themselves be broadly organized along a positive–negative evaluative dimension.

Importantly, however, the starred configuration observed in the PCA plots indicates that the structure of the data cannot be straightforwardly reduced to a single pleasant–unpleasant axis. Several descriptors—including close, big, spicy, high, submissive, relaxing, and hard—appear nearly orthogonal to the hedonic dimension in at least one of the PCA solutions (English-speaking, Mandarin-speaking, or overall). Moreover, the central finding of the present study is that consonance and dissonance emerge as major organizing principles of the observed multisensory associations, and these dimensions do not simply reduce to or align with pleasantness and unpleasantness. We, therefore, adopt a more cautious interpretation, suggesting that the observed struc-

ture likely reflects the interaction of multiple factors, including hedonic, auditory, and acoustic properties.

Another possible explanation draws on the literature highlighting the role of conceptual metaphors in structuring listeners’ multisensory associations. Within this framework, crossmodal correspondences are grounded in cross-domain metaphors that reflect broader conceptual structures, rather than solely through affective mediation [88]. Previous research has demonstrated systematic crossmodal mappings between musical parameters and non-auditory sensory domains. In particular, Eitan et al. have shown that listeners consistently associate musical pitch with spatial, visual, and tactile dimensions, such as height, brightness, and sharpness, suggesting that musical perception is structured by networks of cross-domain correspondences [77, 89]. These mappings have been interpreted as reflecting underlying conceptual metaphors linking auditory features to spatial and embodied schemas [90]. The present study extends this line of research by investigating whether such multisensory mappings are systematically organized by the consonance–dissonance dimension across a broader range of sensory and affective attributes. However, because language plays a central role in shaping how sensory experiences are conceptualized [91], this account must also consider cross-linguistic differences in how such metaphors are expressed (e.g., in English and Mandarin), and whether comparable conceptual mappings underpin the associations in each language.

Our study might contribute also to the ongoing debate on the origins of foundational constructs in harmony—that is, consonance and dissonance—where conflicting perspectives have often emerged [92, 93, 66]. The present results align with findings on shared traits in both music perception and crossmodal correspondences involving auditory stimuli [69, 94–96]. However, although the difference in terms of musical training did not affect the multisensory mappings, the data collected on musical background do not allow us to clearly determine the extent to which the two groups differ in their exposure to Western music and harmony.

One limitation of the present study is the use of linguistic labels for tastes and odors rather than actual tastants or odors. Consequently, participants’ associations may have been influenced by the linguistic properties of the food names themselves (e.g., the roundedness or angularity of the graphemes [97]). Future research should therefore incorporate more ecologically valid stimuli to directly assess sensory mappings and eliminate potential confounds related to the presentation of verbalized sensory features [98, 25]. Another potential limitation concerns the absence of detailed data on bilingualism and English proficiency among the Chinese participants, and specific non-linguistic cultural practices, which could have provided additional informative variables for examining cross-group differences more closely. Furthermore, all musical stimuli were presented using a piano timbre. Given the role of timbre in shaping crossmodal associations with complex musical stimuli [88], future research could explore how different timbres influence multisensory mappings of simple musical stimuli, such as harmonic intervals. Finally, given the relatively narrow pitch range used in the present study, the interaction between interval structure, pitch register,

and crossmodal mappings was not directly addressed here and therefore remains an important question for future research.

The present findings may offer several potential avenues for application. First, the multisensory mappings could be leveraged in crossmodally augmented artistic practices, where sound is systematically paired with visual, tactile, or gustatory elements to enhance aesthetic experience [99]. Relatedly, the mappings between intervals and dimensions of the chemical senses could provide a reliable basis for the sonification of olfactory and gustatory experiences. For instance, mapping fragrances, food textures, or wine profiles onto harmonic structures in ways that are perceptually meaningful and systematically grounded [100, 101]. Second, they offer a principled basis for the design of sensory substitution systems, in which auditory signals could convey information typically accessed through other modalities (e.g., translating visual features into structured harmonic patterns) [102]. Third, these correspondences may inform sensory augmentation technologies, enabling more robust mapping across senses in a systematic manner [103].

5 | Conclusion

This study shows that the crossmodal mapping of musical intervals onto sensory and affective dimensions is predominantly structured by a musical factor, namely consonance and dissonance. Unlike previous research, which has largely focused on isolated sensory correspondences (e.g., pitch-brightness or pitch-taste), these findings demonstrate a robust, domain-general principle in which consonance and dissonance structure multiple sensory mappings. The fact that no explicit mention of consonance/dissonance was made in the experiment, yet participants' responses consistently aligned with this dimension, suggests that these associations might well be perceptually grounded and shared across the two groups of participants, namely, English-speakers from North America and the United Kingdom and Mandarin-speakers from China. This provides a new perspective on the multisensory dimension of music perception, supporting the idea that musical consonance and dissonance reflect deeper cognitive and perceptual regularities that extend across sensory systems.

Author Contributions

Nicola Di Stefano: conceptualization, methodology, writing – original draft, funding acquisition. Alessandro Ansani: methodology, writing – original draft, formal analysis, funding acquisition. Andrea Schiavio: methodology, writing – original draft. Suvi Saarikallio: methodology, writing – original draft, funding acquisition. Petri Toiviainen: methodology, formal analysis, writing – original draft, funding acquisition. Elvira Brattico: methodology, writing – original draft. Charles Spence: conceptualization, methodology, supervision. All authors read and approved the final version of the manuscript.

Acknowledgments

The authors are grateful to Erica Volta for her assistance in preparing the auditory stimuli. This research received financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, call for tender No. 104 published on 2.2.2022 by

the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—project title “The multisensory and interactional nature of music experience. Merging aesthetics with bioengineering to investigate the multidimensional structure of musical sound.” Grant Assignment Decree No. 1079 was adopted on 19/07/2023 by the Italian Ministry of University and Research (MUR) and the project FIS-2023-03371 “X-SONANCE: Unraveling consonance perception in audition and beyond: A crossmodal, cross-cultural, and cross-species investigation.” This work was funded in part by the Research Council of Finland through grants to the Centre of Excellence in Music, Mind, Body and Brain (Grant n. 346210) and to the FRAMES project (Grant n. 368151).

Open access publishing facilitated by Consiglio Nazionale delle Ricerche, as part of the Wiley – CRUI-CARE agreement.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data and musical stimuli that support the findings of this study are openly available in OSF at <https://osf.io/8nrx7>.

Peer Review

For transparency, the peer review documents associated with this article are available at <https://doi.org/10.1111/nyas.70337>.

References

1. C. Spence, “Crossmodal Correspondences: A Tutorial Review,” *Attention, Perception, & Psychophysics* 73 (2011): 971–995, <https://doi.org/10.3758/s13414-010-0073-7>.
2. O. Deroy, I. Fernandez-Prieto, J. Navarra, and C. Spence, “Unravelling the Paradox of Spatial Pitch,” in *Spatial Biases in Perception and Cognition*, ed. T. L. Hubbard (Cambridge University Press, 2018), 77–93, <https://doi.org/10.1017/9781316651247>.
3. C. Spence, “On the Relative Nature of (Pitch-Based) Crossmodal Correspondences,” *Multisensory Research* 32, no. 3 (2019): 235–265, <https://doi.org/10.1163/22134808-20191407>.
4. L. Bonetti and M. Costa, “Pitch-Verticality and Pitch-Size Cross-Modal Interactions,” *Psychology of Music* 46, no. 3 (2018): 340–356, <https://doi.org/10.1177/0305735617710734>.
5. L. Bonetti and M. Costa, “Musical Mode and Visual-Spatial Cross-Modal Associations in Infants and Adults,” *Musicae Scientiae* 23, no. 1 (2019): 50–68, <https://doi.org/10.1177/1029864917705001>.
6. N. B. Maimon, D. Lamy, and Z. Eitan, “Crossmodal Correspondence Between Tonal Hierarchy and Visual Brightness: Associating Syntactic Structure and Perceptual Dimensions Across Modalities,” *Multisensory Research* 33, no. 8 (2020): 805–836, <https://doi.org/10.1163/22134808-bja10006>.
7. L. E. Marks, “On Cross-Modal Similarity: Auditory–Visual Interactions in Speeded Discrimination,” *Journal of Experimental Psychology: Human Perception & Performance* 13 (1987): 384–394.
8. C. Møller, A. Højlund, K. B. Bærentsen, N. C. Hansen, J. C. Skewes, and P. Vuust, “Visually Induced Gains in Pitch Discrimination: Linking Audio-Visual Processing With Auditory Abilities,” *Attention, Perception, & Psychophysics* 80 (2018): 999–1010.
9. M. Tiihonen, T. Jacobsen, N. Trusbak Haumann, S. Saarikallio, and E. Brattico, “I Know What I Like When I See It: Likability Is Distinct From Pleasantness Since Early Stages of Multimodal Emotion Evaluation,” *PLoS ONE* 17, no. 9 (2022): e0274556, <https://doi.org/10.1371/journal.pone.0274556>.
10. C. Spence and N. Di Stefano, “Coloured Hearing, Colour Music, Colour Organs, and the Search for Perceptually Meaningful

- Correspondences Between Colour and Sound,” *I-Perception* 13, no. 3 (2022): 1–42, <https://doi.org/10.1177/20416695221092802>.
11. C. Spence and N. Di Stefano, “Sensory Translation Between Audition and Vision,” *Psychonomic Bulletin & Review* 31, no. 2 (2024): 599–626, <https://doi.org/10.3758/s13423-023-02343-w>.
 12. C. Parkinson, P. J. Kohler, B. Sievers, and T. Wheatley, “Associations Between Auditory Pitch and Visual Elevation Do Not Depend on Language: Evidence From a Remote Population,” *Perception* 41, no. 7 (2012): 854–861, <https://doi.org/10.1068/p7225>.
 13. A. J. Bremner, S. Caparos, J. Davidoff, J. de Fockert, K. J. Linnell, and C. Spence, “‘Bouba’ and ‘Kiki’ in Namibia? A Remote Culture Make Similar Shape–Sound Matches, but Different Shape–Taste Matches to Westerners,” *Cognition* 126, no. 2 (2013): 165–172, <https://doi.org/10.1016/j.cognition.2012.09.007>.
 14. N. B. Maimon, D. Lamy, and Z. Eitan, “Space Oddity: Musical Syntax Is Mapped Onto Visual Space,” *Scientific Reports* 11, no. 1 (2021): 22343, <https://doi.org/10.1038/s41598-021-01393-1>.
 15. S. Wang, J. Liu, X. Lan, Q. Hu, J. Jiang, and J. Zhang, “Cross-Modal Association Analysis and Matching Model Construction of Perceptual Attributes of Multiple Colors and Combined Tones,” *Frontiers in Psychology* 13 (2022): 970219, <https://doi.org/10.3389/fpsyg.2022.970219>.
 16. D. Guedes, M. Vaz Garrido, E. Lamy, B. Pereira Cavalheiro, and M. Prada, “Crossmodal Interactions Between Audition and Taste: A Systematic Review and Narrative Synthesis,” *Food Quality and Preference* 107 (2023): 104856, <https://doi.org/10.1016/j.foodqual.2023.104856>.
 17. K. Knöferle and C. Spence, “Crossmodal Correspondences Between Sounds and Tastes,” *Psychonomic Bulletin & Review* 19 (2012): 992–1006, <https://doi.org/10.3758/s13423-012-0321-z>.
 18. A.-S. Crisinel and C. Spence, “Implicit Association Between Basic Tastes and Pitch,” *Neuroscience Letters* 464, no. 1 (2009): 39–42, <https://doi.org/10.1016/j.neulet.2009.08.016>.
 19. A.-S. Crisinel and C. Spence, “As Bitter as a Trombone: Synesthetic Correspondences in Nonsynesthetes Between Tastes/Flavors and Musical Notes,” *Attention, Perception & Psychophysics* 72, no. 7 (2010): 1994–2002, <https://doi.org/10.3758/APP.72.7.1994>.
 20. A.-S. Crisinel and C. Spence, “A Sweet Sound? Food Names Reveal Implicit Associations Between Taste and Pitch,” *Perception* 39, no. 3 (2010): 417–425, <https://doi.org/10.1068/p6574>.
 21. Q. J. Wang and C. Spence, “‘Striking a Sour Note’: Assessing the Influence of Consonant and Dissonant Music on Taste Perception,” *Multisensory Research* 29, no. 1–3 (2016): 195–208, <https://doi.org/10.1163/22134808-00002505>.
 22. K. Bronner, K. Frieler, H. Bruhn, R. Hirt, and D. Piper, “What Is the Sound of Citrus? Research on the Correspondences Between the Perception of Sound and Flavour,” in *Proceedings of the 12th International Conference of Music Perception and Cognition (ICMPC) and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music (ESCOM) (ICMPC, 2012)*.
 23. B. Mesz, M. A. Trevisan, and M. Sigman, “The Taste of Music,” *Perception* 40, no. 2 (2011): 209–219, <https://doi.org/10.1068/p6801>.
 24. A.-S. Crisinel, C. Jacquier, O. Deroy, and C. Spence, “Composing With Cross-Modal Correspondences: Music and Odors in Concert,” *Chemosensory Perception* 6 (2013): 45–52, <https://doi.org/10.1007/s12078-012-9138-4>.
 25. M. Murari, A. Rodà, S. Canazza, G. De Poli, and O. Da Pos, “Is Vivaldi Smooth and Takete? Non-Verbal Sensory Scales for Describing Music Qualities,” *Journal of New Music Research* 44, no. 4 (2015): 359–372, <https://doi.org/10.1080/09298215.2015.1101475>.
 26. N. Di Stefano and C. Spence, “Roughness Perception: A Multisensory/Crossmodal Perspective,” *Attention, Perception, & Psychophysics* 84, no. 7 (2022): 2087–2114.
 27. D. Cooke, *The Language of Music* (Oxford University Press, 1959).
 28. E. Willems, *L’oreille Musicale*, [The Musical Ear] (Éditions Pro Musica, 1977).
 29. M. Costa, P. E. Ricci Bitti, and L. Bonfiglioli, “Psychological Connotations of Harmonic Musical Intervals,” *Psychology of Music* 28, no. 1 (2000): 4–22, <https://doi.org/10.1177/0305735600281002>.
 30. J. D. Zhang and E. Schubert, “A Single Item Measure for Identifying Musician and Nonmusician Categories Based on Measures of Musical Sophistication,” *Music Perception* 36, no. 5 (2019): 457–467, <https://doi.org/10.1525/mp.2019.36.5.457>.
 31. D. Müllensiefen, B. Gingras, J. Musil, and L. Stewart, “The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population,” *PLoS ONE* 9, no. 2 (2014): e89642, <https://doi.org/10.1371/journal.pone.0089642>.
 32. D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software* 67, no. 1 (2015): 1–48, <https://doi.org/10.18637/jss.v067.i01>.
 33. C. M. Judd, J. Westfall, and D. A. Kenny, “Treating Stimuli as a Random Factor in Social Psychology: A New and Comprehensive Solution to a Pervasive but Largely Ignored Problem,” *Journal of Personality and Social Psychology* 103, no. 1 (2012): 54–69, <https://doi.org/10.1037/a0028347>.
 34. V. K. Y. Ng and R. A. Cribbie, “Using the Gamma Generalized Linear Model for Modeling Continuous, Skewed and Heteroscedastic Outcomes in Psychology,” *Current Psychology* 36, no. 2 (2017): 225–235, <https://doi.org/10.1007/s12144-015-9404-0>.
 35. J. L. Rodgers, “The Epistemology of Mathematical and Statistical Modeling: A Quiet Methodological Revolution,” *American Psychologist* 65, no. 1 (2010): 1–12, <https://doi.org/10.1037/a0018326>.
 36. E.-J. Wagenmakers, “A Practical Solution to the Pervasive Problems of P Values,” *Psychonomic Bulletin & Review* 14, no. 5 (2007): 779–804, <https://doi.org/10.3758/BF03194105>.
 37. J. Ding, V. Tarokh, and Y. Yang, “Model Selection Techniques: An Overview,” *IEEE Signal Processing Magazine* 35, no. 6 (2018): 16–34, <https://doi.org/10.1109/MSP.2018.2867638>.
 38. R. V. Lenth, *Emmeans: Estimated Marginal Means, Aka Least-Squares Means (P. 1.10.4)* (CRAN, 2017), <https://doi.org/10.32614/CRAN.package.emmeans>.
 39. D. J. Schad, S. Vasishth, S. Hohenstein, and R. Kliegl, “How to Capitalize on a Priori Contrasts in Linear (Mixed) Models: A Tutorial,” *Journal of Memory and Language* 110 (2020): 104038, <https://doi.org/10.1016/j.jml.2019.104038>.
 40. M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, “NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set,” *Journal of Statistical Software* 61, no. 6 (2014): 1–36, <https://doi.org/10.18637/jss.v061.i06>.
 41. R. Tibshirani, G. Walther, and T. Hastie, “Estimating the Number of Clusters in a Data Set via the Gap Statistic,” *Journal of the Royal Statistical Society Series B: Statistical Methodology* 63, no. 2 (2001): 411–423, <https://doi.org/10.1111/1467-9868.00293>.
 42. P. J. Rousseeuw, “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis,” *Journal of Computational and Applied Mathematics* 20 (1987): 53–65, [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
 43. W. J. Krzanowski and Y. T. Lai, “A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering,” *Biometrics* 44, no. 1 (1988): 23–34, <https://doi.org/10.2307/2531893>.
 44. T. Calinski and J. Harabasz, “A Dendrite Method for Cluster Analysis,” *Communications in Statistics—Theory and Methods* 3, no. 1 (1974): 1–27, <https://doi.org/10.1080/03610927408827101>.
 45. M. Halkidi, M. Vazirgiannis, and Y. Batistakis, “Quality Scheme Assessment in the Clustering Process,” in *Principles of Data Mining and Knowledge Discovery: 4th European Conference, PKDD 2000 Lyon, France*,

- September 13–16, 2000 Proceedings, ed. D. A. Zighed, J. Komorowski, and J. Żytkow (Springer Berlin Heidelberg, 2000), 265–276.
46. A. Kassambara and F. Mundt, *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses (P. 1.0.7)* (CRAN, 2016), [https://doi.org/10.32614/CRAN.package.factoextra/\(Dataset\)](https://doi.org/10.32614/CRAN.package.factoextra/(Dataset)).
47. H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval* (Cambridge University Press, 2008).
48. K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. (Springer, 2010).
49. W. A. Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer, 1998), <https://doi.org/10.1007/978-1-4471-4177-8>.
50. O. Lartillot and P. Toivainen, “A Matlab Toolbox for Musical Feature Extraction From Audio,” in *Proceedings of International Conference on Digital Audio Effects (DAFx)*. (Osterreichische Computer Gesellschaft, Vienna: Austria).
51. R. Plomp and W. J. M. Levelt, “Tonal Consonance and Critical Bandwidth,” *Journal of the Acoustical Society of America* 38, no. 4 (1965): 548–560, <https://doi.org/10.1121/1.1909741>.
52. W. Hutchinson and L. Knopoff, “The Acoustic Component of Western Consonance,” *Journal of New Music Research* 7, no. 1 (1978): 1–29.
53. P. M. C. Harrison and M. T. Pearce, “Simultaneous Consonance in Music Perception and Composition,” *Psychological Review* 127, no. 2 (2020): 216–244, <https://doi.org/10.1037/rev0000169>.
54. L. Euler, *Tentamen Novae Theoriae Musicae [An Attempt at a New Theory of Music]* (Academiae Scientiarum, 1739).
55. H. L. F. Helmholtz, *Die Lehre von den Tonempfindungen Als Physiologische Grundlage für die Theorie der Musik [On the Sensations of Tone as a Physiological Basis for the Theory of Music]*, trans. A. J. Ellis (Vieweg, 1863).
56. P. N. Juslin and J. A. Sloboda, *Music and Emotion: Theory and Research* (Oxford University Press, 2001), <https://doi.org/10.1093/oso/9780192631886.001.0001>.
57. C. Spence, “Simple and Complex Crossmodal Correspondences Involving Audition,” *Acoustical Science and Technology* 41, no. 1 (2020): 6–12, <https://doi.org/10.1250/ast.41.6>.
58. F. Stolzenburg, “Harmony Perception by Periodicity Detection,” *Journal of Mathematics and Music* 9, no. 3 (2015): 215–238, <https://doi.org/10.1080/17459737.2015.1033024>.
59. A. Wand, “On the Conception and Measure of Consonance,” *Leonardo Music Journal* 22 (2012): 73–78, https://doi.org/10.1162/LMJ_a_00102.
60. G. M. Bidelman and A. Krishnan, “Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem,” *Journal of Neuroscience* 29, no. 42 (2009): 13165–13171, <https://doi.org/10.1523/JNEUROSCI.3900-09.2009>.
61. D. L. Bowling and D. Purves, “A Biological Rationale for Musical Consonance,” *Proceedings of the National Academy of Sciences* 112, no. 36 (2015): 11155–11160, <https://doi.org/10.1073/pnas.1505768112>.
62. I. Shapira Lots and L. Stone, “Perception of Musical Consonance and Dissonance: An Outcome of Neural Synchronization,” *Journal of the Royal Society Interface* 5, no. 29 (2008): 1429–1434, <https://doi.org/10.1098/rsif.2008.0143>.
63. D. Perani, M. C. Saccuman, P. Scifo, et al., “Functional Specializations for Music Processing in the Human Newborn Brain,” *Proceedings of the National Academy of Sciences* 107, no. 10 (2010): 4758–4763, <https://doi.org/10.1073/pnas.0909074107>.
64. N. di Stefano, P. Vuust, and E. Brattico, “Consonance and Dissonance Perception. A Critical Review of the Historical Sources, Multidisciplinary Findings, and Main Hypotheses,” *Physics of Life Reviews* 43 (2022): 273–304.
65. P. Virtala and M. Tervaniemi, “Neurocognition of Major-Minor and Consonance-Dissonance,” *Music Perception* 34, no. 4 (2017): 387–404, <https://doi.org/10.1525/mp.2017.34.4.387>.
66. J. H. McDermott, A. F. Schultz, E. A. Undurraga, and R. A. Godoy, “Indifference to Dissonance in Native Amazonians Reveals Cultural Variation in Music Perception,” *Nature* 535, no. 7613 (2016): 547–550, <https://doi.org/10.1038/nature18635>.
67. R. Parncutt, D. Reisinger, A. Fuchs, and F. Kaiser, “Consonance and Prevalence of Sonorities in Western Polyphony: Roughness, Harmonicity, Familiarity, Evenness, Diatonicity,” *Journal of New Music Research* 48, no. 1 (2019): 1–20, <https://doi.org/10.1080/09298215.2018.1477804>.
68. R. Parncutt and G. Hair, “A Psychocultural Theory of Musical Interval,” *Music Perception* 35, no. 4 (2018): 475–501, <https://doi.org/10.1525/mp.2018.35.4.475>.
69. I. Lahdelma and T. Eerola, “Mild Dissonance Preferred Over Consonance in Single Chord Perception,” *I-Perception* 7, no. 3 (2016): 2041669516655812, <https://doi.org/10.1177/2041669516655812>.
70. I. Lahdelma and T. Eerola, “Cultural Familiarity and Musical Expertise Impact the Pleasantness of Consonance/Dissonance but Not Its Perceived Tension,” *Scientific Reports* 10, no. 1 (2020): 8693, <https://doi.org/10.1038/s41598-020-65615-8>.
71. R. S. Friedman, D. A. Kowalewski, D. T. Vuvan, and W. T. Neill, “Consonance Preferences Within an Unconventional Tuning System,” *Music Perception* 38, no. 3 (2021): 313–330, <https://doi.org/10.1525/mp.2021.38.3.313>.
72. I. Engel, L. Radovanovic, and R. Parncutt, “Consonance and Dissonance of Simultaneous Trichords in Western Music,” *Psiholgijske Teme* 32, no. 1 (2023): 13–34, <https://doi.org/10.31820/pt.32.1.2>.
73. R. Marjeh, P. M. Harrison, H. Lee, F. Deligiannaki, and N. Jacoby, “Timbral Effects on Consonance Disentangle Psychoacoustic Mechanisms and Suggest Perceptual Origins for Musical Scales,” *Nature Communications* 15, no. 1 (2024): 1482, <https://doi.org/10.1038/s41467-024-45812-z>.
74. K. Motoki, L. E. Marks, and C. Velasco, “Reflections on Cross-Modal Correspondences: Current Understanding and Issues for Future Research,” *Multisensory Research* 37, no. 1 (2023): 1–23, <https://doi.org/10.1163/22134808-bja10114>.
75. C. Spence, “Assessing the Role of Emotional Mediation in Explaining Crossmodal Correspondences Involving Musical Stimuli,” *Multisensory Research* 33, no. 1 (2020): 1–29, <https://doi.org/10.1163/22134808-20191469>.
76. S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, “Music-Color Associations Are Mediated by Emotion,” *Proceedings of the National Academy of Sciences* 110, no. 22 (2013): 8836–8841, <https://doi.org/10.1073/pnas.1212562110>.
77. N. Di Stefano, A. Ansani, A. Schiavio, and C. Spence, “Prokofiev Was (Almost) Right: A Cross-Cultural Investigation of Auditory-Conceptual Associations in *Peter and the Wolf*,” *Psychonomic Bulletin & Review* 31 (2024): 1735–1744, <https://doi.org/10.3758/s13423-023-02435-7>.
78. G. Carraturo, V. Pando-Naude, M. Costa, P. Vuust, L. Bonetti, and E. Brattico, “The Major-Minor Mode Dichotomy in Music Perception,” *Physics of Life Reviews* 52 (2025): 80–106, <https://doi.org/10.1016/j.plrev.2024.11.017>.
79. R. Etzi, C. Spence, M. Zampini, and A. Gallace, “When Sandpaper Is ‘Kiki’ and Satin Is ‘Bouba’: An Exploration of the Associations Between Words, Emotional States, and the Tactile Attributes of Everyday Materials,” *Multisensory Research* 29, no. 1–3 (2016): 133–155, <https://doi.org/10.1163/22134808-00002497>.
80. Y. Zhou and C. S. Tse, “The Taste of Emotion: Metaphoric Association Between Taste Words and Emotion/Emotion-Laden Words,” *Frontiers in Psychology* 11 (2020): 986, <https://doi.org/10.3389/fpsyg.2020.00986>.

81. P. Valdez and A. Mehrabian, "Effects of Color on Emotions," *Journal of Experimental Psychology: General* 123, no. 4 (1994): 394–409, <https://doi.org/10.1037/0096-3445.123.4.394>.
82. N. Kaya and H. H. Epps, "Relationship Between Color and Emotion: A Study of College Students," *College Student Journal* 38, no. 3 (2004): 396–405.
83. L. C. Ou, M. R. Luo, A. Woodcock, and A. Wright, "A Study of Colour Emotion and Colour Preference. Part I: Colour Emotions for Single Colours," *Color Research & Application* 29, no. 3 (2004): 232–240, <https://doi.org/10.1002/col.20010>.
84. Y. Wang, T. Wang, W. Mu, and Y. Sun, "What is the Glamor of Black-and-White? The Effect of Color Design on Evaluations of Luxury Brand Ads," *Journal of Consumer Behaviour* 21, no. 5 (2022): 973–986, <https://doi.org/10.1002/cb.2030>.
85. S. Pastoor, "Legibility and Subjective Preference for Color Combinations in Text," *Human Factors: The Journal of the Human Factors and Ergonomics Society* 32, no. 2 (1990): 157–171, <https://doi.org/10.1177/001872089003200204>.
86. L. Yang, B. Qi, and Q. Guo, "The Effect of Icon Color Combinations in Information Interfaces on Task Performance Under Varying Levels of Cognitive Load," *Applied Sciences* 14, no. 10 (2024): 4212, <https://doi.org/10.3390/app14104212>.
87. C. Cattaneo, S. Spinelli, C. Dinnella, et al., "Patterns of Sensory and Hedonic Responses for Salty and Umami Tastes and Their Impact on Food Familiarity, Consumption, and Nutritional Status: A Gender-Based Analysis From a Large Population Sample," *Current Research in Food Science* 10 (2025): 100970, <https://doi.org/10.1016/j.crfs.2025.100970>.
88. Z. Eitan and I. Rothschild, "How Music Touches: Musical Parameters and Listeners' Audio-Tactile Metaphorical Mappings," *Psychology of Music* 39, no. 4 (2011): 449–467, <https://doi.org/10.1177/0305735610377592>.
89. Z. Eitan and R. Timmers, "Beethoven's Last Piano Sonata and Those Who Follow Crocodiles: Cross-Domain Mappings of Auditory Pitch in a Musical Context," *Cognition* 114, no. 3 (2010): 405–422, <https://doi.org/10.1016/j.cognition.2009.10.013>.
90. Z. Eitan, *The Routledge Companion to Music Cognition* (Routledge, 2017), <https://doi.org/10.4324/9781315194738>.
91. S. Dolscheid, S. Shayan, A. Majid, and D. Casasanto, "The Thickness of Musical Pitch," *Psychological Science* 24, no. 5 (2013): 613–621, <https://doi.org/10.1177/0956797612457374>.
92. D. L. Bowling, M. Hoeschele, K. Z. Gill, and W. T. Fitch, "The Nature and Nurture of Musical Consonance," *Music Perception* 35, no. 1 (2017): 118–121, <https://doi.org/10.1525/mp.2017.35.1.118>.
93. N. Di Stefano, "Cultural Accounts of Consonance Perception. A Lakatosian Approach to Save Pythagoras," *Music Perception* 42, no. 2 (2024): 165–176.
94. G. Athanasopoulos, T. Eerola, I. Lahdelma, and M. Kaliakatsos-Papakostas, "Harmonic Organisation Conveys Both Universal and Culture-Specific Cues for Emotional Expression in Music," *PLoS ONE* 16, no. 1 (2021): e0244964, <https://doi.org/10.1371/journal.pone.0244964>.
95. N. Di Stefano, A. Ansani, A. Schiavio, S. Saarikallio, and C. Spence, "Audiovisual Associations in Saint-Saëns' Carnival of the Animals: A Cross-Cultural Investigation on the Role of Timbre," *Empirical Studies of the Arts* 43, no. 2 (2025): 1162–1180.
96. X. Wan, A. T. Woods, J. J. van den Bosch, K. J. McKenzie, C. Velasco, and C. Spence, "Cross-Cultural Differences in Crossmodal Correspondences Between Basic Tastes and Visual Features," *Frontiers in Psychology* 5 (2014): 1365, <https://doi.org/10.3389/fpsyg.2014.01365>.
97. J. Simner, C. Cuskley, and S. Kirby, "What Sound Does That Taste? Cross-Modal Mappings Across Gustation and Audition," *Perception* 39, no. 4 (2010): 553–569, <https://doi.org/10.1068/p6591>.
98. V. Alluri and P. Toiviainen, "The Naturalistic Paradigm: An Approach to Studying Individual Variability in Neural Underpinnings of Music Perception," *Annals of the New York Academy of Sciences* 1530, no. 1 (2023): 18–22, <https://doi.org/10.1111/nyas.15075>.
99. C. Spence and N. Di Stefano, "Augmenting Art Crossmodally: Possibilities and Pitfalls," *Frontiers in Psychology* 16 (2025): 1605110, <https://doi.org/10.3389/fpsyg.2025.1605110>.
100. C. Spence, "Multisensory Flavour Perception: Blending, Mixing, Fusion, and Pairing Within and Between the Senses," *Foods* 9, no. 4 (2020): 407, <https://doi.org/10.3390/foods9040407>.
101. C. Spence, N. Di Stefano, F. Reinoso-Carvalho, B. Mesz, and A. Zacharakis, "Play the Fragrance': Designing Musical Soundscapes to Match Fragrances Based on Olfactory-Auditory Crossmodal Correspondences," *I-Perception* 17, no. 1 (2026): 1–33, <https://doi.org/10.1177/20416695251409265>.
102. G. Hamilton-Fletcher, T. D. Wright, and J. Ward, "Cross-Modal Correspondences Enhance Performance on a Colour-to-Sound Sensory Substitution Device," *Multisensory Research* 29, no. 4–5 (2016): 337–363, <https://doi.org/10.1163/22134808-00002519>.
103. M. Pinardi, N. Di Stefano, G. Di Pino, and C. Spence, "Exploring Crossmodal Correspondences for Future Research in Human Movement Augmentation," *Frontiers in Psychology* 14 (2023): 1190103, <https://doi.org/10.3389/fpsyg.2023.1190103>.
104. W. A. Sethares, "Local Consonance and the Relationship Between Timbre and Scale," *The Journal of the Acoustical Society of America* 94, no. 3 (1993): 1218–1228.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information: nyas70337-sup-0001-SuppMat.docx