Effect of Timbre on Goodness-of-Fit Ratings of Short Chord Sequences

Ivan Jimenez, Tuire Kuusi, & Juha Ojala
University of the Arts Helsinki, Helsinki, Finland

Isabella Czedik-Eysenberg
University of Vienna, Vienna, Austria

Peter Harrison
University of Cambridge, Cambridge, United Kingdom

Aditya Chander
Yale University

Previous research has found that instrument-like timbres (hereafter, timbres) can affect the Goodness-of-Fit (GoF) evaluations of cadences (Vuvan & Hughes, 2019). Here, we expand these findings by testing more timbres and chord sequences and analyzing a wide array of chordal and timbral variables. One hundred and thirty-five participants with varying levels of music training provided GoF ratings for 15 C-C-X chord sequences constructed with piano, clean electric guitar, and choir timbres. The third chord was a major, minor, major-minor seventh, or minor seventh chord. The ratings of choir stimuli were higher and their range narrower than the ratings for the other two timbres, regardless of music training. Additionally, the profile formed by the GoF ratings of the 15 choir stimuli was different from the profiles of the other two timbres. Further analyses provided detailed information about the effect of timbre as well as harmonic and melodic factors on the ratings. Attack was identified as a likely contributor to GoF ratings of choir stimuli being higher than the ratings of the other stimuli. Tonal contextuality (Leman, 2000), affected by diffuse partials, and the importance given to the soprano are discussed as two plausible explanations for the narrow range and other idiosyncrasies of the GoF ratings of the choir stimuli.

Received: January 12, 2024, accepted August 21, 2024.

Key words: chord sequences, harmony, timbre, goodness of fit, sensory dissonance

Timbre may affect listeners' perceptions of how well certain chords go together. There is evidence that instrument-like timbres (hereafter, timbres)[1] affect Goodness-of-Fit (hereafter, GoF) ratings of chord sequences. Vuvan and Hughes (2019) found that GoF ratings for two cadences, bVII-I and V-I, were less differentiated from each other when played with a distorted electric guitar timbre than when played with a piano timbre. Since bVII-I is common in popular music but rare in classical music[2] while V-I is common in both styles of music, results were attributed to the two timbres activating different style-specific harmonic schemata. However, Vuvan and Hughes (2021) later acknowledged that purely sensory mechanisms related to the distortion of the electric guitar (e.g., greater presence of higher partials affecting the clarity of pitch structures) could provide an alternative explanation for the effect of timbre on GoF ratings in their experiments. The results from Vuvan and Hughes are intriguing and invite further inquiry. The current study further explores the effect of timbre on GoF ratings by expanding the number of chord sequences and timbres investigated as well as the number of chordal and timbral variables considered.

[1] The expression "instrument-like timbre" refers to the fact that the block chords in Vuvan and Hughes (2019) and the current experiment were constructed using sampled tones as opposed to using a recording of human performers playing or singing the chords. Although the tones we used are sampled from recordings of real instruments and real singers, it is possible that some listeners can tell the difference between chords created with sampled tones and recordings of humans performing the chords. Throughout the article, we use the word "timbre" as an abbreviation of "instrument-like timbres." The term "timbral" will be used with a more comprehensive meaning as in the expression "timbral features." The term "instrument" will be reserved for discussing research that uses recordings of human performances, for referring to our manipulation of the instrument libraries within the context of a digital audio station, and for referring to one of the components of the PCA in our statistical analysis.

[2] In this article we use the term "classical music" to refer to "common-practice music" which is tonal music from the Western European tradition, ca. 1750–1900. We use the term "popular music" to refer to English-language pop and rock music. Although Vuvan and Hughes (2019) used the specific term "rock music" instead of "popular music," we will use the term "popular music" instead since all the main currently available harmonic corpora include songs from both pop and rock repertoires.

Music Perception, pp. 1–24, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2025.2397481

The Effect of Timbre on Other Experiments Using Block Chords

The effect of timbre on GoF ratings found by Vuvan and Hughes (2019) is part of a growing body of evidence that extra-harmonic features such as duration and timbre can affect the perception of harmony. For instance, Jimenez et al. (2024) found that the duration of the chords can make chord-type changes more salient in comparison to voicing changes. In the specific case of timbre, its effect on chord perception has already been found in experiments using a wide variety of approaches including behavioral and brain responses. For instance, it has been shown that when asked whether two chords had the exact same pitches, participants are slower and less accurate when the timbre of the two chords is different (Beal, 1985; Cho et al., 1991, 1993). Virtala et al. (2018) investigated event-related EEG potentials in an oddball paradigm using major and power chords (i.e., a major chord without its third) played with clean and distorted electric guitar tones. The effects of timbre and harmony changes were additive, with the strongest mismatch negativity (MMN) responses obtained when both timbre and harmony were changed at the same time. Although it could be expected that harmonic contrast will elicit greater MMNs with clean than distorted electric guitar because distorted electric guitar power chords tend to sound major (Juchniewicz & Silverman, 2013), this pattern failed to reach statistical significance. Jimenez et al. (2023) asked participants to rate the similarity of chords in pairs of same-root piano chords taken from the openings of commercial recordings of well-known pop songs. They found that timbral brightness and pitch register were the two most important factors affecting the ratings, while the ratings were not affected by chord type differences.

There is also some evidence that timbre can affect the perception of chord type, one of the factors that has influenced GoF ratings in previous studies (Krumhansl, 1990). Rodrigues et al. (2017) found that participants with no previous music training were more successful at learning to identify major, minor, diminished, and augmented chords when played on acoustic guitar than when played on piano. Skorik et al. (2018) found that participants with no music training were consistently able to differentiate between major and minor chords when played on piano, but the differentiation was hindered when chords were constructed using sine waves with a long fade-out, and no differentiation was possible when the chords were constructed using sine waves without any amplitude ramps.

Finally, other studies on the perception of isolated chords provide some insights that familiarity of timbre can affect chord perception (Lahdelma & Eerola, 2020). Results from similar experiments provide support for the notion that timbral features have an effect on the perception of stability and consonance (Arthurs et al., 2018), pleasantness (Baltes et al., 2023; Marjieh et al., 2022) and some emotions that have been traditionally linked to the major-minor dichotomy (Lahdelma & Eerola, 2016a).

However, there are some studies showing no effect of timbre on the harmonic perception of chords. For instance, Virtala et al. (2014) tested musicians and nonmusicians by introducing occasional oddball chords (minor chord or the second inversion of the major chord) in a series of randomly transposed major chords. All chords were played with piano tones or sinusoidal tones. They found an effect of timbre for the behavioral detection of the inversion oddball but not the chord-type oddball. Parncutt et al. (2023) asked musicians and nonmusicians to rate dissonance of trichords played using piano and octave-complex tones (whose timbre resembles that of an electric organ) and found no effect of timbre.

GoF Ratings and Harmonic Perception

Studying the effect of timbre on chord perception using GoF ratings can be particularly informative because it can reveal the extent to which timbre affects the perceived tonal relationship between chords (e.g., harmonic syntax), arguably one of the most important functions of chords in tonal music. Although the great majority of GoF studies have focused on the perception of single tones (for a review, see Krumhansl & Cuddy, 2010), there have been several GoF studies investigating the perception of chords (Bharucha & Krumhansl, 1983; Chander & Aslin, 2023; Craton et al., 2021; Creel, 2011; Hughes, 2011; Krumhansl, Bharucha, & Castellano, 1982; Krumhansl, Bharucha, & Kessler, 1982; Sears et al., 2019; Vuvan & Hughes, 2019). While most of the recent GoF studies on chords have focused on investigating whether GoF can be influenced by the occurrence frequency of chords and chord transitions in classical and popular music, only Vuvan and Hughes (2019) have used timbre alone as a cue for trying to activate style-specific harmonic schemata. However, taking into consideration the potential spectral overlap between harmonic and timbral information, it seems reasonable to expect that the effect of timbre on harmonic perception may extend beyond the activation of style-specific schemata.

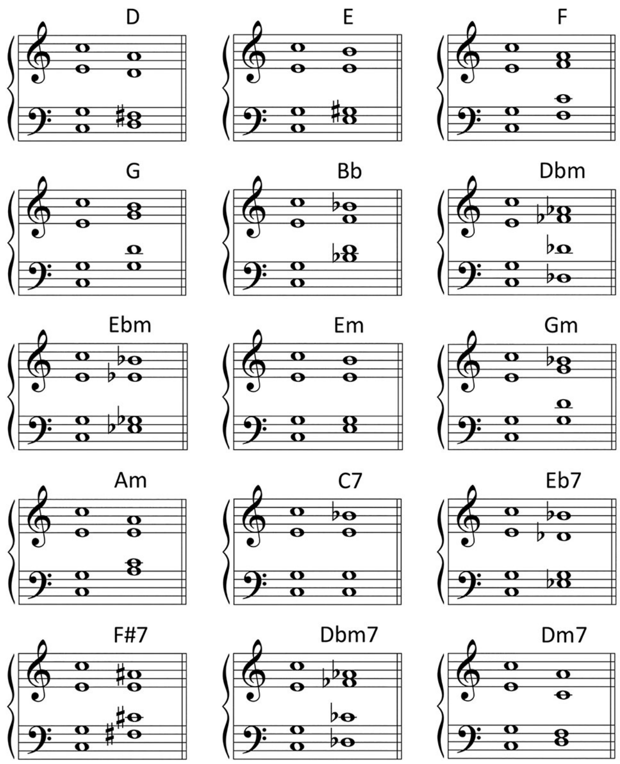

FIGURE 1. Transcriptions of the second and third chord of the fifteen chord sequences used in the experiment.

Aims

The aim of the current study was to analyze the effect of 15 chord sequences and 3 timbres as well as of more fine-grained chordal and timbral variables and participants' music training on GoF ratings. Three general questions guided our analyses: 1) Can an effect of timbre on GoF ratings be found using different timbres and chord sequences than those used by Vuvan and Hughes (2019)? 2) Is that effect modulated by music training? 3) What can an analysis of fine-grained chordal and timbral variables tell us about the effect of timbre on GoF ratings?

All chord sequences in the experiment followed the pattern C-major – C-major – test chord, that is, C-C-X (see Stimuli and Figure 1) and the timbres used in the study were piano, electric guitar, and choir. These chord sequences and timbres were selected based on pilots (see *Stimuli*). The fine-grained variables were selected based on their effect on chord and timbre perception in previous experiments. Table A1 (Appendix) provides detailed descriptions of all the chordal and timbral variables used in the analysis. Most of the chordal variables described harmonic or melodic attributes of the chord sequences. The harmonic variables included the theoretical tonal distance between chords C and X in Lerdahl's Tonal Pitch Space (hereafter, TPS distance, Lerdahl, 1988), and Tonal Contextuality (hereafter, TC, Leman, 2000) of the X chord in the context of the C chord, which is a measure of harmonic similarity related to the concept of pitch commonality (Bigand et al., 1996) but calculated from raw audio. Harmonic variables also included two variables related to style-specific harmonic schemata, the frequency of occurrence of chord transitions from C to X in a classical music corpus (Harrison & Pearce, 2018) and a popular music corpus (Chander & Quinn, 2023). Only one chordal variable, Roughness, calculated using the model by Vassilakis (2001), described a harmonic attribute of the X chord as an isolated event. Although the roughness measure was calculated from audio in this study, we considered this variable to be a harmonic chordal variable because it was more influenced by pitch structure (chord type and voicing) than the timbres used in this experiment. Among the chordal variables were also the soprano and bass motions between chords C and X. The timbral variables included temporal descriptors (e.g., attack time, temporal centroid), spectral variables (e.g., spectral centroid, spectral flatness), and one spectro-temporal descriptor (spectral flux). The values for the timbral variables were calculated by analyzing the raw audio of the X chord.

Although timbre tends to be more perceptually salient than harmonic structure (Jimenez et al., 2023), music training has sometimes been found to attenuate the tendency to focus on timbre over harmony (Beal, 1985). Music training has also been found to facilitate the mental processing of harmony in a variety of tasks (e.g., tension ratings in Bigand et al., 1996; stability and consonance ratings in Arthurs et al., 2018; noticing differences between chord sequences in Eitel et al., 2024; identifying songs from chord sequences in Jimenez & Kuusi, 2018; assessing the harmonization of vocals from well-known songs in Jimenez et al., 2022). Music training has also been found to affect GoF ratings of piano chords (Craton et al., 2021, Experiment 2). On the other hand, there are studies that have found no statistically significant effect of music training in tasks rating pleasantness of chords in isolation (Arthurs et al., 2018; Johnson-Laird et al., 2012) or in context (Roberts, 1986), liking of piano chords (Craton et al., 2021, Experiment 1), in identifying songs from single chords (Jimenez et al., 2023), nor in the general level of GoF ratings of cadences in studies using multiple timbres (Vuvan & Hughes, 2019). In our case, music training was assessed using the Gold-MSI musical training scale (see, *Procedure*).

Method

PARTICIPANTS

The project was approved by the Research Ethics Committee of the University of the Arts Helsinki (Reference number 055111). The online experiment, hosted in PsyToolkit (Stoet, 2010, 2017), was visited 210 times between July 7 and August 5, 2023. Since we knew that the number of non-serious visitors and survey bots is large in crowdsourcing platforms (Ahler et al., 2019; Dennis et al., 2020), we used a relatively difficult loudness pre-test to screen participants and to minimize the influence of the quality of participants' headphones, the environmental noise, and participants' hearing deficiencies like hearing loss. Participants' task in the loudness pre-test was to choose the loudest tone of a series of five otherwise identical piano tones. The loudness pre-test included three separate trials, and participants were allowed to listen to the series of five tones from each trial as many times as they wanted before moving on to the next trial. Altogether, 6 visitors abandoned the survey before taking the pre-test, and 50 visitors failed to answer the pre-test correctly. Additionally, we excluded 19 participants for various reasons; for example, leaving parts of the experiment unanswered, or being inconsistent (see description of consistency checks below). Participants' data were excluded if their responses for the 9 consistency checks (see Procedure) were too inconsistent with their own previous responses to the same exact stimuli, and the correlation between their ratings averaged across timbres and the average ratings of all the other participants was lower than .20.

The total number of participants whose responses were included in our main analysis was 135 (67 male, 63 female, 5 other; age $M = 30.78$, $SD = 10.76$). We grouped the 135 participants according to the Gold-MSI factor 3 (Musical Training). The degrees of training were defined as follows: 1–2.99 = low training ($n = 87$); 3–4.99 = medium training ($n = 29$); 5–7.0 = high training ($n = 19$). All the participants in our group with high music training have lived most of their lives in Anglophone and European countries, and most of the participants in our low music training group were from Spanish-speaking countries (e.g., Mexico, Chile). However, since our online participants in Spanish-speaking countries are likely to have had at least some exposure to Anglophone popular music, we considered that any differences found between the music training groups would be primarily driven by music training rather than major differences in listening exposure across countries.

STIMULI

Stimuli are available at https://osf.io/tajyz/. Each stimulus consisted of three chords: two C major chords and the test chord. Fifteen test chords were included: Major triads (D, E, F, G, Bb), minor triads (Dbm, Ebm, Em, Gm Am), major-minor seventh chords (C7, Eb7, Gb7), or minor seventh chords (Dbm7, Dm7). Figure 1 provides transcriptions of the last two chords from each stimulus.

The chord sequences and their voice leading followed those used in Bigand et al. (1996) and were convenient for our study for several reasons:

1. Chord sequences in Bigand et al. (1996) are short.
2. All chord sequences in Bigand et al. (1996) are voiced using the same contrapuntal pattern in which the outer voices moved inwards by contrary motion, systematizing the direction of the soprano and bass voice in the stimuli and simplifying their analysis. We did not change the pattern even though it caused a "parallel" voice-leading with some chord combinations.
3. Chord sequences in Bigand et al. (1996) use the four most common chord types found in Western tonal music according to numerous corpora (e.g., Arthurs et al., 2018; Kolchinsky et al., 2017; Nadar et al., 2019).
4. The chord sequences used in Bigand et al. (1996) are diverse and relatively balanced in terms of: a) tonal distances according to Lerdahl's Tonal Pitch Space (1988), and b) frequency of occurrence in corpora of Western tonal music.

Since our goal was not to test chord tension or cadences but to test GoF ratings for the last chord of short chord sequences regardless of whether listeners could hear the sequences as cadential or not, we followed a chord pattern C-C-X (instead of the C-X-C used by Bigand et al., 1996). That is, we increased the salience of the context chord C by repeating it, and omitted the last C chord, making it possible for the participants to evaluate chord X in the context of the C chords. Choosing only 15 chord sequences allowed us to test chord sequences with three timbres without creating an experiment that was too long. Furthermore, the repetition of the C major chord: a) increased the chances that all the stimuli were heard in the key of C major (potentially increasing inter-rater agreement), b) gave the participants time within each stimulus to concentrate on the C major chord regardless of the timbre (that constantly changed from one trial to another), and c) decreased the temporal uncertainty related to the three timbres having different attack envelopes. Finally, the repetition of the C major chord could facilitate harmonic perception by doubling the time of the C major chord (Jimenez et al., 2024).

Specific Chord Sequences

The main goal when selecting the 15 chord sequences for this study was to achieve a diverse and relatively balanced set in terms of: a) tonal distances according to Lerdahl's Tonal Pitch Space (1988), b) frequency of occurrence in corpora of Western tonal music, c) chord types, d) roots, e) melodic intervals in the top voice, and f) the extent to which their ratings of GoF were affected by timbre in pilots. When referring to the chord sequences in the main text of this article, the first two chords will be described with a single "C" symbol as opposed to repeating the symbol (e.g., C-D instead of C-C-D). In the figures, however, each chord sequence will be referred to by only specifying the X chord (e.g., D instead of C-C-D) for brevity.

Timbre

Three versions of each of the 15 chord sequences were created using three different timbres: piano, electric guitar, and choir. These specific timbres were selected based on three criteria:

1. *Ecological validity:* Creating block-chord stimuli with piano, electric guitar, and choir is naturalistic because piano, electric guitar, and choir often play/sing block chords in real music and because those instruments are commonly used in Western tonal music. Additionally, the authors believed that the selected libraries sounded particularly realistic and most nonmusicians were able to identify the instruments from the timbres in pilots.
2. *Liking*: Pilots with nonmusicians and 30 different timbres showed similar moderate levels of liking for chords played with the three timbres that were ultimately chosen for the main experiment. The goal of using timbres with similar levels of liking was to reduce the potential effect of preference on GoF ratings. Additionally, we avoided timbres with low levels of liking because these types of timbres could have increased fatigue, reducing the quality of the responses and the likelihood that participants finish the experiment.
3. *Effect of timbre on GoF ratings:* Pilots indicated that GoF ratings for chord sequences played with the choir timbre that was ultimately selected for the main experiment were different from the ratings for chords sequences played with piano and electric guitar timbres. These differences were larger than the differences we found between the GoF ratings of other timbres.

The following are the specific details of how the timbres were generated:

Piano tones were generated using the "Bösendorfer Grand Piano" from Logic Pro. All the plugins for that instrument were turned off and the MIDI key velocity for all tones was set to 75.

Electric guitar tones were generated using the "Session Guitarist - Electric Sunburst" from Kontakt, Native Instruments and using the following settings: Preset = Clean Groove (Toolbox A), Retro Echo = off, and Expression = 75. These settings produced clean (low level of distortion) but slightly twangy electric guitar tones. Clean tones were used to avoid seventh chords becoming too dissonant.

Choir tones were generated using "Labs Choir" from Spitfire Audio and using the following settings: Expression = 100%, Dynamics = 100%, Reverb = 0%, Attack = 0 ms, Decay = 0 ms, Sustain = 50%, Release = 600 ms. The samples from this library consist of choir humming long tones with a smooth attack and a subtle gradual crescendo in intensity (see Figure 2).

Monophonic audio stimuli were used instead of stereo audio stimuli to reduce the number of variables that could potentially affect participants' ratings.

Loudness

The loudness of the tones used to create the chords was equalized following a two-step procedure: We first performed an initial round of loudness equalization using pyloudnorm (Steinmetz & Reiss, 2021). One of the authors then listened to all stimuli and adjusted the loudness of any tones that sounded anomalously loud or quiet. These manual changes reached no more than +/- 4 dB. Two other authors verified that the equalization with the additional subjective changes was better than the automated equalization alone.

Duration

The interonset interval between chords was 1,500 ms. The MIDI duration of the chords was also set to 1,500 ms but the actual duration of the tones was slightly longer because, even when the reverb setting of both the instrument and the DAW were turned off, the generated sounds included the natural resonance that tends to briefly linger in real-world situations due to physical properties of the instruments and the acoustical spaces in which they are usually played. This short resonance was kept in our stimuli to prevent timbres from sounding unnatural.

An interonset interval of approximately 5 ms was added between each of the four tones of all the

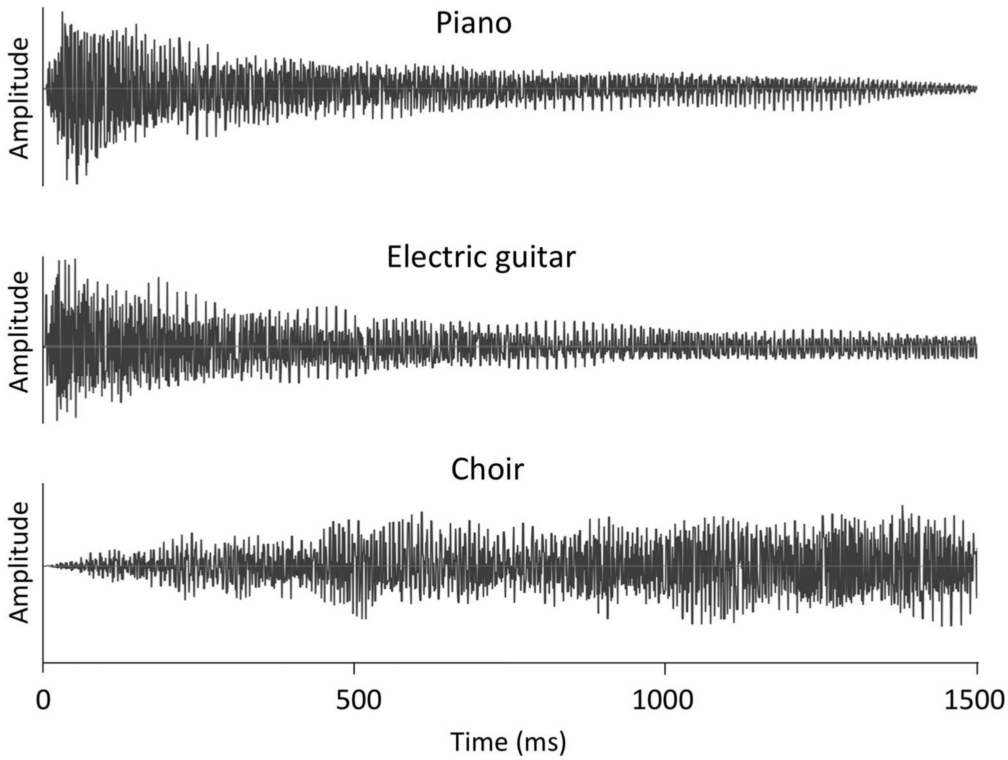

FIGURE 2. Waveforms of the C chord played on the three timbres.

electric-guitar chords. The onsets of the tones were ordered according to register from low to high tones (i.e., bass, tenor, alto, soprano). This interonset interval was added to mimic the sound of a quick down strum (lower strings first) with a pick. Pilots indicated that this procedure facilitated the identification of the instrument. Longer interonset intervals were avoided to prevent participants from hearing the individual tones as separate events (Simon et al., 2019), which would have introduced additional melodic and rhythmic factors that could have influenced participants' ratings of the chords. Recordings of real strumming were not used because that would have prevented us from fully controlling the loudness of the individual tones of the chords.

PROCEDURE

Participants were recruited online by word of mouth and using the crowdsourcing platform Prolific. We used PsyToolkit software (Stoet, 2010, 2017) for data collection.

Participants that successfully completed the loudness pre-test were presented with five training trials and the following instructions:

> *In this page, we will show some examples of the type of main questions we will ask you in this survey. Please listen to all the five excerpts before providing any ratings. Then listen to the excerpts again and rate them in terms of how well the third chord fits with the previous two chords in the excerpt. This task is comparable with the question "How well do two colors go with each other?"*[3] *The excerpts in this page exemplify different degrees of goodness-of-fit you will find in the main section of the experiment (i.e., "very good fit," "very poor fit," or intermediate degrees of fit). Try to provide a wide range of ratings accordingly, but try to also respond intuitively without thinking too much about all the different specific ways in which the third chord may be different from the previous two chords.* ***<u>Important</u>****: We will sometimes omit the third chord to verify that you are paying attention to the task. When the third chord is omitted, please select the option "the third chord is missing" instead of providing a goodness-of-fit rating.*

[3] The instruction *"How well do two colors go with each other?"* was adopted from Parncutt et al. (2019).

Participants rated the GoF of the third chord in each sequence using a 7-point scale (1 = *very poor fit*, 7 = *very good fit*). The five training stimuli had the same C-C-X pattern as the main stimuli, but the X chord in these chord sequences were not used in the main experiment. Additionally, chords for the training stimuli did not use the timbres from the main experiment but were instead constructed using sinusoidal tones in a way that resembled elements of piano and guitar tones (amplitude envelope) and choir tones (absence of broadband inharmonicity related to percussive attacks).

The experiment also included six attention checks that were interspersed among the main trials and included two C chords but no X chord. If the participants missed an attention check for the first time, they were notified of the miss, and after a second miss they were not allowed to complete the experiment. The experiment also included nine consistency checks that were identical to main trials except that they were presented to participants at the end of main section, after the 45 main trials and six attention checks, and were not considered for any analysis other than for calculating a self-consistency score.

The main section of the study consisted of 60 trials: 45 main trials (15 chord sequences x 3 timbres), 9 consistency checks, and 6 attention checks. The stimuli were presented in one of 18 pseudo-randomized orders in which the timbre always changed from one trial to the next and stimuli with the same X chord (but different timbre) were separated by 10 or more trials. To reduce participants' fatigue, the 60 trials were presented in three blocks of 20 trials. Participants were asked the questions from Gold-MSI (Müllensiefen et al., 2014) between blocks. After the 60 trials, participants heard audio clips in which the chord pattern C-C-C was played with each of the 3 timbres used in the experiment and were asked to use a 7-point scale to rate their liking and familiarity with the timbres. Most participants completed the entire session in less than 30 minutes.

DATA ANALYSIS

To respond to our three general research questions (see Aims), we analyzed GoF ratings in two different ways. We used a three-way repeated-measures mixed ANOVA with factors of *training* (low, medium, high), *chord* (each of the 15 chord sequences described in the *Stimuli* subsection of the *Method*) and *timbre* (piano, electric guitar, and choir). Since four participants had incomplete rating data, their responses were removed before running the ANOVA, resulting in a total of 5,895 observations across 131 participants.

To explore potential mechanisms for the results we observed, we used principal components analysis (PCA) for grouping the 7 chordal and 9 timbral fine-grained numerical variables, followed by a linear regression to analyze whether the PCA components predicted the participants' GoF ratings.

Results

THREE-WAY ANALYSIS OF VARIANCE

Figure 3 shows the goodness-of-fit ratings for three timbres and fifteen chord sequences for all participants together and for the three groups of music training separately. Table 1 shows the results of the corresponding three-way ANOVA. There were significant main effects of *timbre* and *chord* (but not *training*), as well as significant two-way interactions of *timbre* × *chord* and *training* × *chord* (but not *training* × *timbre*). The three-way interaction between *timbre*, *chord,* and *training* was not significant.

For the main effect of *timbre*, post hoc paired *t*-tests (with timbre ratings averaged across chords and Bonferroni corrections for multiple comparisons) showed that GoF ratings in the choir timbre (M_{choir} = 5.57, SD_{choir} = 0.80) were clearly higher than GoF ratings in the piano timbre (M_{piano} = 4.45, SD_{piano} = 0.81), $t(130) = 13.25$, $p < .001$, Cohen's $d = 1.39$. On the other hand, the GoF ratings in the piano timbre were also higher than the GoF ratings in the guitar timbre, but the effect was not as strong (M_{guitar} = 4.17, SD_{guitar} = 0.86), $t(130) = 4.18$, $p < .001$, Cohen's $d = 0.33$. For the main effect of *chord*, reporting all paired *t*-test outcomes was unfeasible, but there were clear differences between chords as shown in Figure 3.

The two-way interaction between timbre and chord was most likely driven by the fact that, as shown in Figure 3, there was less variation in the GoF ratings for chords in the choir timbre than with those in the piano and guitar timbres. Indeed, the range of ratings was narrowest for the choir stimuli (2.21 for choir, 2.63 for piano, and 2.86 for guitar). Moreover, since the highest ratings for choir are close to the upper end of the ratings scale, it is likely that the narrowing of the range for the chord ratings is, at least partially, a consequence of a ceiling effect.

There were also other, more specific differences that may also have driven the interaction between chord and timbre. For instance, in the guitar timbre, the Am chord was rated clearly lower than the G and F chords, while in the piano and choir timbres they were not significantly different. Additionally, the Dbm and Dbm7 chords were rated clearly lower than every other chord

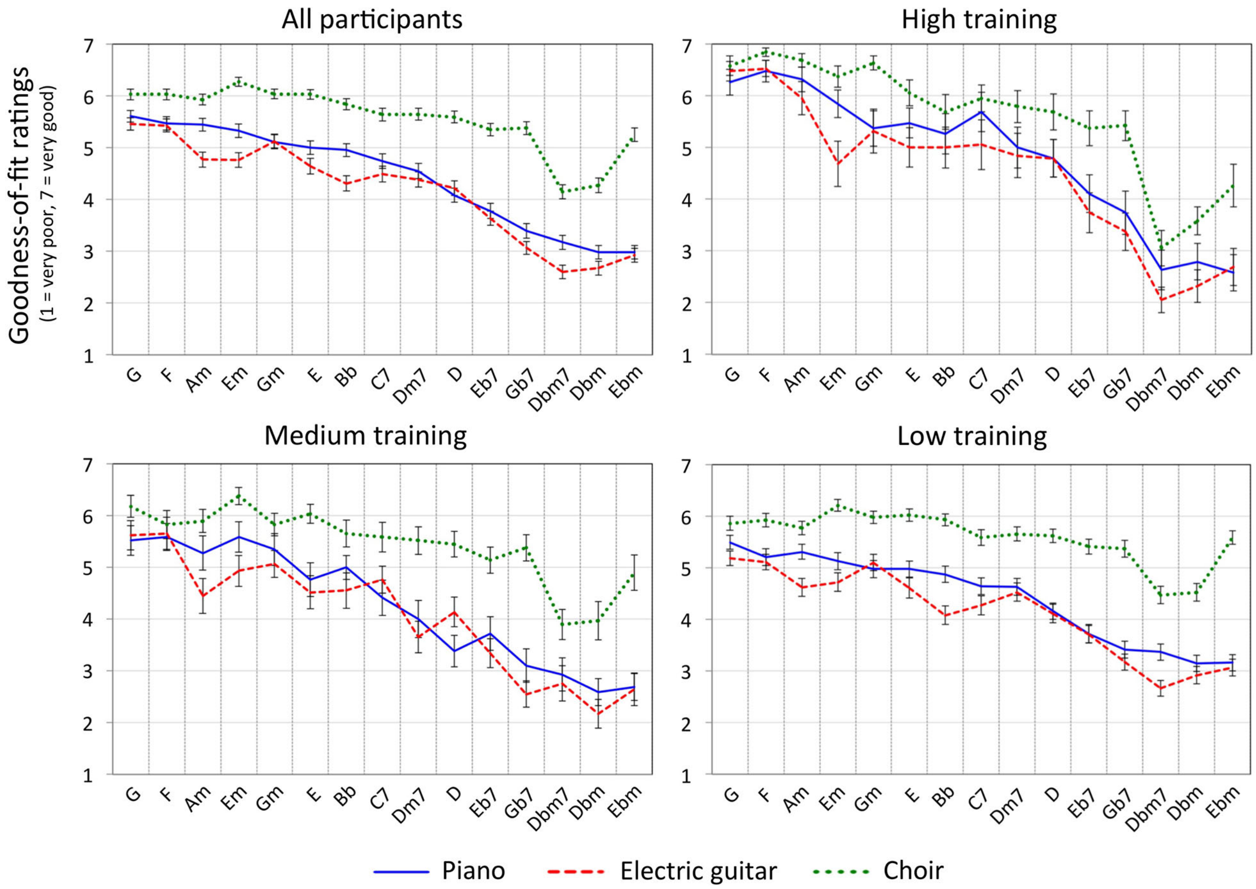

FIGURE 3. Goodness-of-fit ratings for three timbres and fifteen chord sequences. All participants together (upper left plot), participants with Gold-MSI musical training 5-7 (upper right plot), 3-4.99 (lower left plot), and 1-2.99 (lower right plot). The chord names on the *x*-axis describe the X chords from the 15 C-C-X patterns. To facilitate comparison, chord sequences are arranged according to the GoF ratings for piano by all participants together. Error bars indicate standard errors.

TABLE 1. *Results of Three-way ANOVA with Factors of Training, Chord, and Timbre*

Predictor	df_{Num}	df_{Den}	*Epsilon*	SS_{Num}	SS_{Den}	*F*	*p*	η^2_p
(intercept)	1.00	128.00		89403.67	2305.25	4964.19	< .001	.97
training	2.00	128.00		77.49	2305.25	2.15	.121	.03
timbre	1.86	238.70	0.93	1258.01	1553.88	103.63	< .001	.45
chord	9.77	1251.18	0.70	3416.32	4079.45	107.19	< .001	.46
training × timbre	3.73	238.70	0.93	25.04	1553.88	1.03	.389	.02
training × chord	19.55	1251.18	0.70	321.68	4079.45	5.05	< .001	.07
timbre × chord	21.41	2740.72	0.76	261.08	4624.48	7.23	< .001	.05
training × timbre × chord	42.82	2740.72	0.76	70.37	4624.48	0.97	.520	.01

Note. df_{Num} = degrees of freedom numerator. df_{Den} = degrees of freedom denominator. Epsilon = Greenhouse-Geisser multiplier for degrees of freedom, *p* values and degrees of freedom in the table incorporate this correction. SS_{Num} = sum of squares numerator. SS_{Den} = sum of squares denominator. η^2_p = partial eta-squared. Training = music training factor from Gold-MSI; timbre = piano, electric guitar, and choir; and chord = 15 C-C-X chord sequences.

in the choir timbre, but not in the guitar and piano timbres. A ceiling effect cannot explain these specific differences between the ratings of the three timbres. Alternative explanations for the narrowing of range and a potential explanation for the more specific differences will be explored later in this section.

The two-way interaction between training and chord was most likely driven by the differences between the

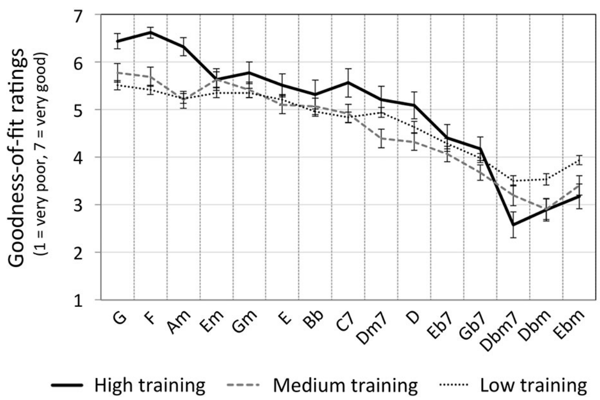

FIGURE 4. Goodness-of-fit ratings for the three levels of music training.

training groups in terms of the range of GoF ratings (see, Figure 4). Overall, the range of GoF ratings was greater for the high training group than the other two groups of participants. However, since the groups were highly imbalanced, it is unclear how robust this effect is.

PRINCIPAL COMPONENT ANALYSIS AND LINEAR REGRESSION

To provide a more fine-grained understanding of the effect of timbre on the GoF ratings, we continued the analysis with a two-phase procedure. The first phase of the analysis was a PCA in which we had 7 chordal variables and 9 timbral variables (as listed and explained in Appendix Table A1). The main goal of the PCA was to minimize any issues related to multicollinearity (Appendix Table A2 shows the correlations between all the variables). For the PCA with the 16 variables, the Kaiser–Meyer–Olkin (KMO) index (.560) and Bartlett's test ($p < .001$) showed that the data were suitable for factor analysis. The analysis revealed five components that together explained 87.604% of the variance. The component matrix was varimax rotated to make the five components easier to interpret. The rotated component matrix was interpretable (see the loadings in Table 2), and the components were interpreted as follows:

Component 1 (explaining 34.7% of the total variance) was interpreted as the "*Instrument*" component because it grouped variables that consistently differentiate the choir stimuli from the piano and electric guitar stimuli regardless of chord sequence. This component included three temporal descriptors (attack time, decay time, and temporal centroid), two spectral descriptors (spectral spread and spectral entropy), and one spectro-temporal descriptor (spectral flux). According to these variables and compared to other stimuli, choir stimuli had smoother and more sustained amplitude envelopes, and spectra that were more complex, less spread, and that changed more over time.

Component 2 (explaining 17.7% of the variance) was interpreted as the "*Harmonic motion*" component because it consisted of variables that pertained to the chord sequence (TPS, frequency of occurrence of chord transitions) and of one variable that was predominantly driven by chord sequence with very little influence of timbre (TC).

Component 3 (explaining 14.1% of the variance) was interpreted as the "*Brightness*" component because one of its two variables was spectral centroid, which is commonly associated with perceived timbral brightness. The other variable in this component, spectral flatness, is connected to spectral centroid not only in our set of stimuli but also in a large set of recordings of isolated tones from orchestral instruments (Reymore et al., 2022). Both spectral centroid and spectral flatness were also connected to average pitch of chords, thus describing the average register of the chord.

Component 4 (explaining 11.6% of the variance) was interpreted as the "*Dissonance*" component. This component included two main variables: Roughness and inharmonicity, which were both influenced by the chord-type, voicing, and timbre. Although frequency of occurrence in popular music primarily loaded on component 2, its loading on component 4 was also high. Frequency of occurrence in popular music was the only variable to have loadings higher than .500 on more than one component.

Finally, Component 5 (explaining 9.5% of the variance) was interpreted as "*Melodic motion*" component since it included two melody-related variables: the bass motion and the soprano motion.

As the second phase, we ran a linear regression analysis. In this analysis we used the five PCA components as model 1 to explain the participants' GoF ratings. Since two of the three training groups were small, we used all participants as one group.

The Durbin-Watson statistic (1.651) and the residuals showed that the data were suitable for a regression analysis. The model summary showed that the model explained 90.6% of the GoF ratings (F Change = 85.790; $p < .001$). We also tried a model with the three timbres and the four chord types added as dummy variables, since it was possible that the repetition of stimuli had an effect on the GoF ratings. However, they did not affect the model fit, so we do not report that more complex model here.

Further analysis of the five components showed that four of them were statistically significant predictors of the GoF ratings. The most important predictor was C2

TABLE 2. *Rotated[a] Component Matrix for PCA of Chordal and Timbral Variables.*

Type	Variable	C1 Instrument	C2 Harmonic motion	C3 Brightness	C4 Dissonance	C5 Melodic motion
Chordal	1. TPS distance (distance in Tonal Pitch Space)	.009	**-.945**	−.009	.103	−.206
	2. Frequency of occurrence classical music	−.007	**.943**	.014	−.074	−.033
	3. Frequency of occurrence popular music	.057	**.655**	.126	**-.586**	.224
	4. TC (harmonic similarity according to Leman)	.061	**.729**	−.022	.381	.189
	5. Roughness	.367	−.001	.061	**.735**	−.391
	6. Soprano motion	.043	−.290	.063	−.091	**-.798**
	7. Bass motion	.012	.076	.374	−.329	**.721**
Timbral	1. Attack time ms (linear)	**.989**	.023	−.074	.044	.017
	2. Decay time ms (linear)	**.985**	.034	.032	.068	.064
	3. Temporal centroid ms	**.892**	.021	.277	.177	−.112
	4. Inharmonicity	.301	.010	−.007	**.699**	.115
	5. Spectral centroid	−.157	.022	**.969**	.050	.019
	6. Spectral entropy	**.911**	−.002	.297	.187	.013
	7. Spectral flatness	.240	.017	**.945**	−.068	.114
	8. Spectral flux	**.976**	−.005	−.125	.111	.008
	9. Spectral spread	**-.845**	−.011	.256	−.351	.169

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. [a]Rotation converged in 6 iterations.

"*Harmonic motion*" (*Beta* = .646, *t* = 13.981, *p* < .001). The effect can be explained by the variables loading on C2: stimuli with smaller TPS distance (i.e., the chords were tonally closer), greater TC (i.e., the chords were spectrally closer), and greater frequency of occurrence in classical and popular music had higher GoF ratings. Another important predictor was C1 "*Instrument*" (*Beta* = .581, *t* = 12.564, *p* < .001), showing that stimuli with smoother and more sustained amplitude envelopes, and with more complex, less spread, and more strongly changing spectra had higher GoF ratings. This component differentiated the choir timbre from the other two timbres. Component C5 "*Melodic motion*" (*Beta* = .374, *t* = 8.094, *p* < .001) was explained by the two variables loading on it: the smaller the soprano motion and the wider the bass motion, the higher the GoF rating. Component C4 "*Dissonance*" (*Beta* = -.147, *t* = -3.177, *p* = .003) showed that the more dissonant the chords were, the lower the GoF ratings were. This indicates that consonant chords had generally higher GoF ratings than dissonant chords. This was the case even though our stimuli only included relatively consonant chord types. Finally, component C3 "Brightness" (*Beta* = .008, *t* = .177, *p* = .861) was not a statistically significant predictor in the regression.

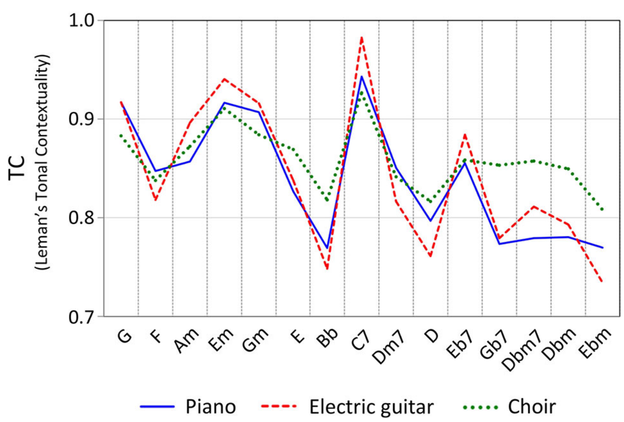

FIGURE 5. TC for the 3 timbres and 15 chord sequences. Greater TC values can be understood as lower spectral contrast between chords.

POTENTIAL EFFECT OF PARTIALS DIFFUSENESS

Earlier, a ceiling effect was identified as a potential explanation for the choir GoF ratings to be narrower in range than the ratings for the other two timbres. However, factors other than a ceiling effect could have contributed to the narrowing of GoF ratings. Figure 5 shows that the range of TC is narrowest for the choir timbre.

This pattern may be related to the partials of the choir being more diffuse than the piano or electric guitar partials. Figure 6 shows spectrograms for an isolated A3 pitch and the C chord from the experiment using the three timbres. We used the term "partials diffuseness" to describe how each partial in the choir timbre tends to be spread over a wider frequency range (i.e., the

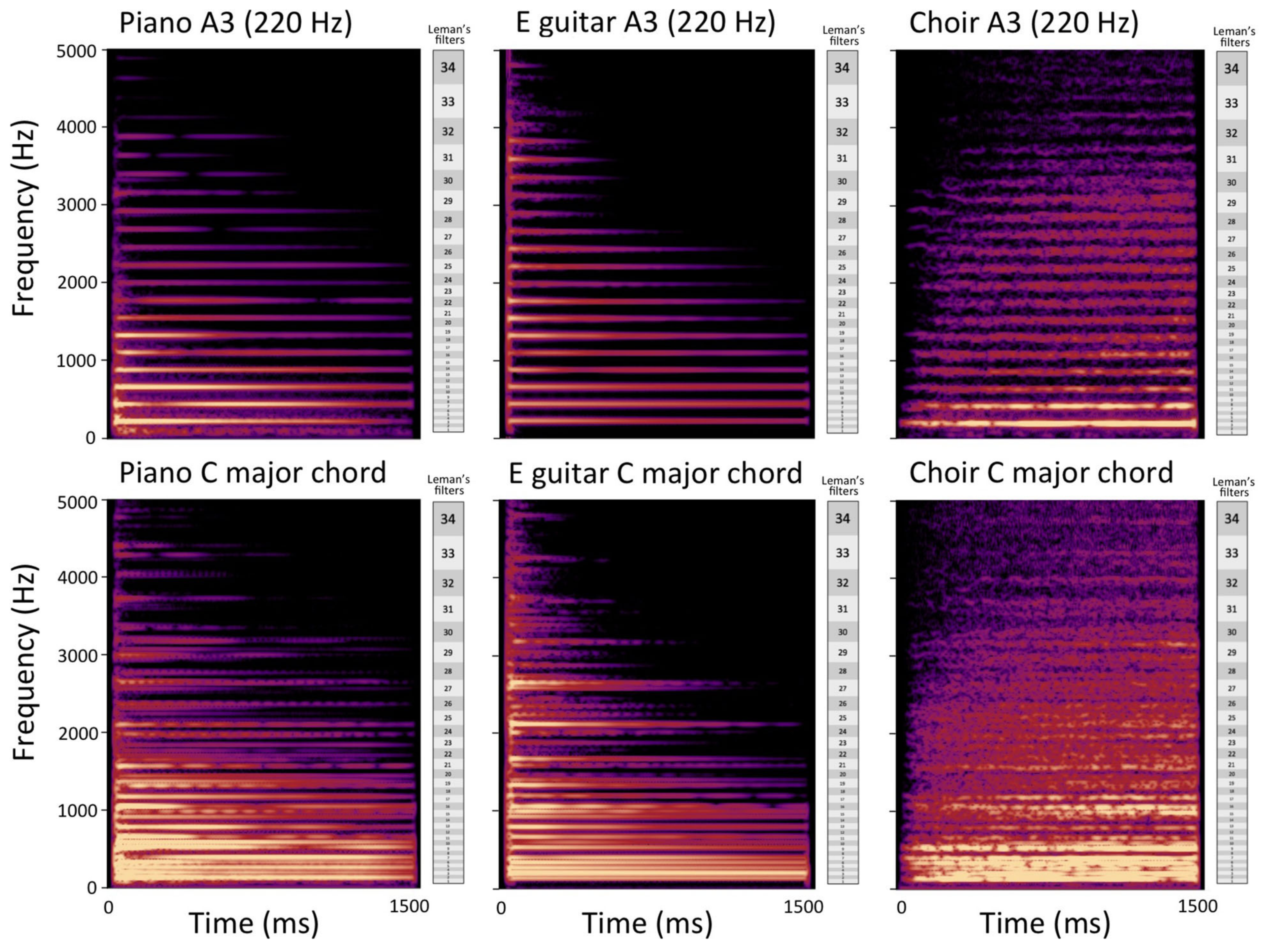

FIGURE 6. Spectrograms of an isolated A3 pitch and the C chord using the three timbres. The height of the grey cells to the right of each spectrogram indicates the bandwidth of the first 34 of the 40 filters that Leman's model uses. The full range covered by the center frequencies of the auditory filters in Leman's model is from 140 Hz to 8,877 Hz.

horizontal line appears thicker in the spectrogram) than the partials for the two other timbres.

One possible reason for the partials of the choir being diffuse is the fact that each note is sung by more than one person, and in those circumstances, there is likely to be a slight offset of fundamental frequencies (Cuesta et al., 2018; Jers & Ternström, 2005). Other reasons for the diffuse partials may include vibrato (Cuesta et al., 2018; Jers & Ternström, 2005) and timbral differences between the voices that are singing the same note. However, in this study it is not possible to compare the multiple singers of individual tones since we do not have access to separate recordings of the individual singers (i.e., the multiple singers are already mixed in each of the tones used to create the choir chords).

One of the steps in Leman's model (2000) is to transform the fine-grained spectra into 40 frequency bins that Leman refers to as filters. In Figure 6, grey cells to the right of each spectrogram indicate the bandwidths of 34 of the 40 filters that Leman's model uses. Since the bandwidths of these filters correspond to critical bands (Bark scale), the filters do not exactly correspond to partials. The more diffuse a partial is, the more likely it is that the partial contributes to more than one filter. This situation is likely to make the chord images within Leman's model blurrier, ultimately affecting the resulting TC values.

The clarity of the pitch content of the chords can be measured via chroma distributions. That type of audio analysis tells a similar story as the TC values. For instance, the range of Euclidean distances between the chroma distribution of the C chord and the chroma distribution of the 15 target chords in the current study is narrower for choir (1.26) than for piano (1.56) or electric guitar (1.52).

TABLE 3. *Correlations Between GoF Ratings, Melodic Motion of Soprano, and TPS Distance for All Participants and Participants Grouped by Music Training*

	Soprano motion			TPS distance		
	Piano	E guitar	Choir	Piano	E guitar	Choir
All participants	−.48	−.48	−.72	−.90	−.92	−.81
High music training	−.45	−.41	−.56	−.94	−.91	−.85
Medium music training	−.51	−.49	−.75	−.84	−.91	−.81
Low music training	−.46	−.49	−.75	−.90	−.90	−.75

POTENTIAL EFFECT OF THE IMPORTANCE GIVEN TO THE SOPRANO

Another factor that could explain the narrow range of the choir ratings is the relative importance given to the soprano. Giving more importance to melodic motion and less importance to harmony can narrow the range of GoF ratings by, for instance, reducing the impact of the most chromatic chord sequences. Although our stimuli included four voices, the soprano was the voice to most likely attract attention because of the high-voice superiority effect (Trainor et al., 2014). The importance given to the soprano can be influenced by its musical characteristics (e.g., how it is played) and by listeners' attention or sensitivity to melody. Correlations between GoF ratings and the absolute interval size of the melodic motion of the soprano in semitones (Table 3), are consistent with the notion that the soprano was most salient in the choir stimuli and that music training modulated that tendency. Correlations between GoF ratings and the melodic motion of the soprano suggest that the influence of the soprano is greatest for the choir stimuli.

The correlations in Table 3 also suggest that the tendency for GoF ratings to be more influenced by the soprano motion, although present for all participants groups, is stronger for participants with medium and low music training. Although there are only 19 participants with high music training in the current study, this relatively small number may be sufficient to obtain a reliable pattern in the context of this analysis since intersubject agreement is frequently higher among musicians than nonmusicians in experiments where participants are asked to rate block-chord sequences (Bigand & Parncutt, 1999) or short passages of music (Krumhansl, 1996).

Table 3 also includes the correlation between the GoF ratings and TPS distance, the variable with the highest loading in the *Harmonic motion* component (C2). The correlations between GoF and TPS distance show patterns that are often opposite to the correlations between GoF and soprano.

The relative importance given to the soprano also offers a plausible explanation for somewhat idiosyncratic patterns such as the GoF ratings of choir C-Dbm and C-Dbm7, which are considerably lower than the ratings of the other chromatic chord sequences (see Figure 3). The soprano motion of both C-Dbm and C-Dbm7 is C5-Ab4, which is the largest interval in the soprano of all the 15 chord sequences (for the transcriptions of the stimuli, see Figure 1). The other chromatic chord sequences (C-Ebm, C-Eb7, and C-Gb7) use the soprano C5-Bb4, which is the second smallest soprano motion in the stimuli set. The GoF ratings for these five chromatic stimuli are consistent with participants providing lower GoF ratings to chord sequences with larger melodic intervals in the soprano when the stimuli were played with the choir timbre.

Discussion

The current study investigated the effect of timbre on GoF ratings of short chord sequences played using piano, electric guitar, and choir tones. Timbre and chord sequence affected the GoF ratings. Choir stimuli received higher GoF than piano and electric guitar stimuli, and this general effect was observed regardless of the participants' music training. Interactions were found between timbre and chord sequences as well as between chord sequences and music training. The analysis identified fine-grained timbral variables that differentiated the choir timbre from the other two and could have therefore contributed to the generally higher GoF ratings for choir stimuli. The analyses also identified a variety of chordal variables that can explain the differences between chord sequences in terms of GoF. Finally, two potential explanations for the interaction between timbre and chord sequences were proposed. These explanations were analyzed using some of the timbral and chordal fine-grained variables.

THE EFFECT OF THE THREE TIMBRES AND NINE TIMBRAL VARIABLES

The clearest differences between the three timbres in terms of GoF were found between choir and the other two timbres. Our analyses and discussions on the differences between timbres focused on the three main differences found between the GoF ratings of the choir stimuli and the other stimuli: 1) GoF ratings were higher for choir for all chord sequences; 2) the range of GoF rating was narrowest for choir; and 3) choir was different from the other two timbres in terms of the profile of GoF ratings for the 15 chord sequences. In the following section we discuss the potential explanations for those three differences.

1. Why were GoF ratings higher for choir?
One of the simplest possible explanations for the fact that the GoF ratings were higher for choir is that participants simply liked the choir timbre better than the other two timbres used in the experiment. However, participants reported liking the piano and choir timbres used in the experiment more than the electric guitar timbre (5.30, 5.39, and 4.29 respectively on a 7-point scale). The extent to which this pattern of preference differs from the pattern of GoF ratings (4.44, 5.56, and 4.17 for piano, choir, and electric guitar, respectively) indicates that preference is not sufficient to explain the effect of timbre on the GoF ratings. Similarly, familiarity cannot explain the GoF ratings for choir since participants reported being least familiar with the choir timbre (3.67, 4.71, and 5.78, for choir, electric guitar, and piano on a 7-point scale).

Another factor that reduces the likelihood that higher GoF ratings for choir were simply a consequence of participants' preference for the choir timbre is the way the three timbres were selected. One of the criteria to select the three timbres was the liking ratings of participants in pilots. Liking ratings for chords played using 30 contrasting timbres were collected in pilots with 69 participants. In that context, all three timbres ultimately selected for this experiment were liked and the differences between their liking ratings were small (less than 0.18 on a 7-point scale). Differences in liking ratings collected in the main experiment may be larger than the liking ratings in the pilots because the main experiment only uses 3 timbres and because liking ratings in the main experiment were always collected in the same order (piano, electric guitar, and choir).

Factors other than preference are likely to have driven the main effect of timbre in the GoF ratings. We found temporal variables, spectral variables, and one spectro-temporal variable that distinguished choir stimuli from piano and electric guitar stimuli. Compared to piano and electric guitar, the choir stimuli had smoother and more sustained amplitude envelopes, more complex and less spread spectra, and a spectrum that changed more over time. It is possible that gradual transition between chords created by the smooth attack of the choir stimuli made chord changes sound less abrupt (having higher GoF) than when played with the sharper attacks of piano and electric guitar.

There is evidence that a smoother attack can reduce perceived tension. For instance, participants in Paraskeva and McAdams (1997) rated piano excerpts as being tenser than the orchestral version of the same excerpts and attributed that result to the piano having a sharper attack than the orchestral instruments. Similarly, ratings of nostalgia and tenderness of isolated chords played with either piano or string tones in Lahdelma and Eerola (2016a) were positively correlated with attack. However, it is possible that the effect of attack on chord perception depends on the specific timbres being tested. For instance, Arthurs et al. (2018) showed that isolated chords were perceived as being more consonant, pleasant, stable, and relaxed when played on piano than when played with organ, despite organ tones typically having slower attacks than piano tones. It is also possible that the acoustic characteristics of the performance spaces (real or simulated) such as reverb can further soften the transition between chords complementing the effect of attack. Future research would be needed to separately test the effect of attack.

The other timbral features that differentiated the choir chords from the piano and electric guitar chords (component C1) could have also contributed to the higher GoF ratings for choir chords. For instance, it is possible that some of the timbral characteristics of the choir chords (e.g., sustained amplitude envelope, complex spectra, and a spectrum that changes over time), made the timbre more salient, taking attention away from the harmonic contrast between chords. Hearing the chord changes as being less contrasting in the choir stimuli could have led to higher GoF ratings for that timbre.

The only PCA component that did not predict the GoF ratings in our experiment was C3 "brightness," constituted by spectral centroid and spectral flatness. The absence of an effect of brightness on GoF ratings is somewhat intriguing, taking into consideration that spectral centroid has consistently been one of the most perceptually salient features of timbre (McAdams & Siedenburg, 2019) and high spectral centroid is associated with lower consonance, pleasantness, stability, and relaxation ratings of isolated chords (Arthurs et al., 2018). Because the number of GoF studies with chords and concentrating on timbre is small, our study calls for future research on the connection between brightness and GoF ratings.

2. Why was the range of the GoF ratings narrowest for choir?
Although the fact that GoF ratings were narrowest for choir could be a by-product of a ceiling effect, the differences between the three profiles formed by the GoF ratings of the 15 chord sequences for the three timbres (see Figure 3) suggest that the ceiling effect is, at the very least, not the only factor affecting the narrowing of the range.

TC and a visual inspection of spectrograms suggest an alternative, compelling explanation for the narrowing of the range. Since TC can be understood as the sensory component of harmonic contrast between two musical events, a narrower TC suggests that the chord sequences sound less harmonically contrasting when the choir timbre is used, thus explaining the narrower range of GoF for the choir stimuli. In turn, the narrower range of TC can be explained by the choir timbre having more diffuse partials, presumably due to the choir tones being sung by more than one singer.

The association just described between TC and GoF can be understood as an effect of pitch clarity on the clarity or salience of harmonic relationships between chords. To further explore the association between the range of TC and GoF, we calculated the TC values of the stimuli used in the second experiment of Vuvan and Hughes (2019). We found that the range of TC (.09 for piano and .05 for distorted electric guitar) was consistent with the range of GoF ratings (.78 for piano and .37 for distorted electric guitar in a 6-point scale). This relationship between GoF and TC provides further support for the notion that pitch clarity in general and TC specifically, can affect the clarity of harmonic relationships between successive chords.

3. Why was the choir different in terms of GoF profiles?
The choir timbre is different from the other two timbres in terms of the GoF profiles (see Figure 3). One potential explanation for these differences is the importance of the soprano motion in the ratings. The correlation between GoF ratings and the melodic motion of the soprano is highest for the choir stimuli whereas correlations between GoF ratings and TPS, the variable with the highest loading in the *Harmonic motion* component (C2), is lowest for the choir stimuli.

The strong relationship between melodic motion of the soprano and the GoF ratings of the choir stimuli could be the consequence of the choir timbre prompting participants to pay attention to the melodic aspects of the stimuli. Since attentional resources are finite, allocating attentional resources to melodic aspects can result in paying less attention to harmonic aspects (Williams, 2005, 2008) and the correlation between TPS distance and GoF ratings being lower for choir than for piano and electric guitar is consistent with that shift of attention.

The tendency for GoF ratings to be more correlated with the motion of the soprano for choir than other stimuli was stronger for participants with medium and low music training. This effect of music training is consistent with previous research that has found that participants with less music training tend to pay more attention to melody and less attention to harmony compared to trained participants (e.g., Williams, 2005).

It is also possible that participants did not consciously attend to either melody or harmony but simply are more sensitive to one or the other parameter regardless of whether they consciously paid attention to them during the experiment (Bigand et al., 1996; Sears et al., 2014). However, the choir timbre is more likely to prompt listeners to attend to melody than the piano or electric guitar timbre because listeners' tendency to make sense of music by unconsciously imitating the music via subvocalization is likely to be stronger when the music includes singing (Cox, 2001, 2016). Additionally, the choir timbre in the current experiment differed from the other two timbres in terms of sustain and spectral flux. The fact that the choir tones were more sustained and their spectra changed over time more than the piano and electric guitar tones could have further shifted the focus towards (horizontal) melodic motions instead of (vertical) individual chords when rating the choir stimuli.

Importantly, a shift towards paying more attention to melody can also explain the other differences between the rating profiles of choir and the two other timbres. Paying more attention to melody and less attention to harmony in the choir stimuli can make the harshness of the chromatic chord sequences less prominent and consequently reduce the range of GoF ratings for the choir stimuli. Alternatively, paying more attention to melody and less attention to harmony in the choir stimuli can reduce the prominence of harmony not only of the chromatic chord sequences but for all the choir stimuli, and this can make the GoF ratings more susceptible to the effect of other more salient attributes such as the smoothness of attack. It should be clarified, however, that in the later scenario, the narrowing of the range of GoF ratings would need to be explained by a different factor such as a ceiling effect.

The potential effects of TC and attention to melody on GoF ratings are not mutually exclusive. If pitch is slightly less clear in the choir chords as suggested by TC values and chroma distributions, this attribute can promote attention to the soprano in choir chords since it may be easier to attend to the most salient voice when pitch is less clear in general. However, since the current experiment does not allow us to confirm that both TC and attention to melody affected GoF ratings, additional research will be needed to test the effects of TC and attention to melody on harmonic perception.

THE EFFECT OF THE SEVEN CHORDAL VARIABLES

Our factor analysis grouped chordal variables into a *Harmonic motion* component and a *Dissonance* component, both of which had an effect on GoF ratings. These components were predominantly harmonic. In addition, one component, *Melodic motion* (C5), was clearly melodic in nature, including the soprano and bass motion.

The effect of the *Harmonic motion* and *Dissonance* components on GoF ratings resembled the results from Bigand et al. (1996) despite the differences between the two studies: ratings in both studies were influenced by TPS, pitch commonality (defined by TC in our study), melodic motion, and roughness. Our study, further, suggested that frequency of occurrence and inharmonicity can affect GoF ratings via an association between those variables, TPS, and roughness. As shown, inharmonicity, a variable that has traditionally been considered a timbral descriptor in the context of analyzing single tones, loaded on the same PCA component as chordal variable roughness. Roughness and harmonicity has been linked to perceived dissonance of harmonic intervals and chords in many studies (for a review, see Harrison & Pearce, 2020). Additionally, roughness and harmonicity were also highly correlated to each other in a large set of chords played using different transpositions and voicings (Parncutt et al., 2023).

The loadings in the PCA also showed that frequency of occurrence in classical music loaded on component C2 while the frequency of occurrence in popular music loaded both on components C2 and C4. One important variable loading on C2 was TPS, whereas C4 was characterized by dissonance variables. The differences between frequency of occurrence in classical and popular music in terms of their relationship to TPS distance and dissonance is possibly influenced by two factors. First, the development of TPS was inspired in part by traditional music theory (Lerdahl, 1988), which in turn was developed to describe the harmonic patterns of classical music (Christensen, 2006). Second, the prevalence of major chords over less consonant chords such as major-minor seventh chords is far greater in popular than in classical music in which the major-minor seventh chord plays a very important harmonic role. In the particular case of the two corpora used in the current experiment, the frequency of occurrence of major and major-minor seventh chords is 18% and 10% respectively in the classical corpus and 56% and 1% in the popular music corpus. Indeed, the presence of major-minor seventh chords has decreased quite dramatically over time in popular music. Léveillé Gauvin (2016) found that in the McGill Billboard corpus, containing Anglophone popular music from 1958–1991, the proportion of songs containing a major-minor seventh chord decreased from close to 100% in the late 1950s to around 40% in the late 1980s/early 1990s. In the Yale-Geerdes corpus, only 12% of songs featured a major-minor seventh chord. It thus stands to reason that chord frequency and dissonance would be more clearly associated in the popular corpus than in the classical corpus.

Several studies have shown that frequency of occurrence of chord transitions in classical and popular music can predict harmonic surprise as measured by surprise ratings (Cheung et al., 2023), tonal priming (Sears et al., 2019), and neural responses (Goldman et al., 2021). However, there are only two studies about chords investigating the activation of both classical and popular music schemata within the same experiment (Hughes, 2011; Vuvan & Hughes, 2019). Both studies used GoF ratings but only Vuvan and Hughes (2019, Experiment 2) used timbre to test the activation of harmonic schemata. Yet their results are not necessarily only due to style-specific harmonic schemata since they could be explained by timbre affecting the clarity of the pitches.

In the current experiment, we found no evidence of timbre activating style-specific harmonic schemata. For instance, GoF ratings for both piano and clean electric guitar were associated to a similar extent to the frequency of occurrence of the chord transitions in both classical ($r = .78$ for piano and $r = .80$ for electric guitar) and popular music ($r = .84$ for piano and $r = .83$ for electric guitar). This result is not too surprising since the frequency of occurrence of chord transitions in classical and popular music tend to also be correlated (e.g., $r = .56$ for 48 chord sequences used in Bigand et al., 1996, and $r = .59$ for the 15 chord sequences used in the current experiment). Focusing on chord sequences that are typical in one style but atypical in the other and vice versa (e.g., Sears & Forrest, 2021) will be required in future studies to properly test the extent to which timbre can activate style-specific harmonic schemata.

An important difference between most research on harmonic expectation and the current study is the length of the chord sequences. Although early research on tonal priming, a phenomenon believed to be influenced by harmonic expectation, used two-chord sequences (e.g., Bharucha & Stoeckig, 1986), most research on harmonic expectation uses chord sequences of seven or more chords. Future experiments can test the effect of timbre on the GoF ratings of longer chord sequences, but there are some points worth considering.

Despite its strong relationship to GoF ratings in this study, TPS distance may not be an adequate variable to test with longer chord sequences. Although TPS distance has been applied to long chord sequences (Bigand & Parncutt, 1999; Lerdahl & Krumhansl, 2007), Lerdahl's model does not calculate tonal distances but tonal tension when applied to long chord sequences, and in that context, the model involves complex and somewhat subjective decisions that are hard to generalize and automatize. In that sense, TC and models based on frequency of occurrence such as IDyOM (Goldman et al., 2021), can be easier to operationalize and test. Ultimately, since TPS distance loaded on the same *Harmonic motion* component as the other harmonic variables (being highly correlated with them), we suggest that it could be excluded from explorations of longer chord sequences.

This said, testing short chord sequences in future experiments can still be beneficial. Short chord sequences are not only practical in terms of reducing time demands on participants, and a convenient way to control for harmonic variables (whose complexity grows exponentially in longer chord sequences), but are also arguably more ecologically valid now than when the participants of Bigand et al. (1996) were tested. The use of looped short chord sequences in popular music has become more prevalent in recent decades (Chander & Quinn, 2023), and two-chord sequences, commonly referred as two-chord shuttles (Tagg, 2014), are part of that repertoire of loops. Additionally, two-chord sequences are also very common in film music of recent decades where they often carry specific narrative associations (Murphy, 2006, 2014).

Conclusions

The current study provides further evidence that timbre can affect the GoF ratings of chords. The effect of timbre in Vuvan and Hughes (2019) and most of the effect of timbre in the current study was observed regardless of music training and despite instructions and training trials encouraging participants to focus on harmony. This finding suggests that the effect of timbre on GoF ratings is not an artifact of participants' inability to focus on harmonic information but a more integral aspect of tonal perception. There seems to be a complex, and most likely unconscious, connection between timbre, style, harmonic expectations, and GoF. Participants without specific music training learn timbres specific to music and gain knowledge about harmony by exposure to music.

Further research is needed to identify the specific mechanisms by which timbre affects GoF ratings of chords. For instance, future work can investigate the extent to which GoF ratings of chords are affected by specific timbral parameters (e.g., attack) or by timbres that activate harmonic schemata of particular styles of music. These types of investigation will likely involve some challenges regarding the selection of timbres. On the one hand, although specific timbral parameters can be controlled rigorously, the idea of creating ecologically valid timbres that do not evoke any specific musical styles is contradictory. On the other hand, investigating the role of timbre on activating style-specific harmonic schemata would involve using timbres that necessarily differ in one or more timbral parameters. A potential solution to these challenges could be to study both types of effects simultaneously.

Investigating the effect of timbre on the activation of style-specific harmonic schemata involves some additional challenges. Different styles of Western tonal music can be similar in terms of the most common chord types (e.g., major and minor) and the types of chord transitions (e.g., diatonic as opposed to chromatic). In addition to diatonic chord sequences, the current study included some chromatic chord sequences (e.g., C-Ebm) that are infrequent in most styles of Western tonal music. By including the salient contrast between these two types of chord sequences, we aimed at making it easier for participants who had no music training to focus on harmony as opposed to timbre. However, that type of harmonic difference between stimuli may need to be avoided in experiments focusing on distinctions between chord sequences that are common in one style of Western tonal music but uncommon in another.

An additional challenge involves the fact that, in our increasingly globalized world, it is difficult to find styles of Western tonal music that are different from each other in terms of harmonic patterns, and for which it is reasonably easy to find a group of participants that is very familiar with one style but not the other. Dealing with this challenge most likely would involve some type of compromise in which smaller differences between participants in terms of their familiarity with different styles of music are considered instead of aiming for an all-or-none categorization of participants' familiarity with musical styles. Another possibility for future exploration of this topic would be to teach novel timbre-harmony combinations to test whether participants can learn timbre-specific harmonic expectations within one or several experimental sessions. In that case, particular attention should be given to the balance between novelty (minimizing the effect of familiar timbres and harmonies on the perception of novel timbres and harmonies) and ecological validity.

Author Note

Stimuli and data are available at https://osf.io/tajyz/.

Correspondence concerning this article should be addressed to Ivan Jimenez (ivan.jimenez.rodriguez@uniarts.fi).

References

Ahler, D. J., Roush, C. E., & Sood, G. (2019). The micro-task market for lemons: Data quality on Amazon's Mechanical Turk. *Political Science Research and Methods*, 1–20. https://doi.org/10.1017/psrm.2021.57

Ajayakumar, R., & Rajan, R. (2020). *Predominant instrument recognition in polyphonic music using gmm-dnn framework* [Conference session]. 2020 International Conference on Signal Processing and Communications (SPCOM), IEEE. https://doi.org/10.1109/SPCOM50965.2020.9179626

Arthurs, Y., Beeston, A. V., & Timmers, R. (2018). Perception of isolated chords: Examining frequency of occurrence, instrumental timbre, acoustic descriptors and musical training. *Psychology of Music*, *46*(5), 662–681. https://doi.org/10.1177/0305735617720834

Baltes, F., Chemnitz, L., & Lange, E. B. (2023). Pleasantness of nonlinear distortion in isolated triads of synthetic timbre. *Journal of the Acoustical Society of America*, *154*(2), 1028–1040. https://doi.org/10.1121/10.0020667

Beal, A. L. (1985). The skill of recognizing musical structures. *Memory and Cognition*, *13*, 405–412. https://doi.org/10.3758/BF03198453

Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*(1), 63–102. https://doi.org/10.1016/0010-0277(83)90003-3

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(4), 403–410. https://doi.org/10.1037/0096-1523.12.4.403

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, *62*(4), 237–254. https://doi.org/10.1007/s004260050053

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience*, *8*(94), 1–27. https://doi.org/10.3389/fnsys.2014.00094

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception and Psychophysics*, *58*, 125–141. https://doi.org/10.3758/BF03205482

Chander, A., & Aslin, R. N. (2023). Expectation adaptation for rare cadences in music: Item order matters in repetition priming. *Cognition*, *240*, 105601. https://doi.org/10.1016/j.cognition.2023.105601

Chander, A., & Quinn, I. (2023). *The decline of harmonic schemata in popular music chord loops* [Conference session]. 17th International Conference on Music Perception and Cognition and the 7th Conference of the Asia-Pacific Society for the Cognitive Sciences of Music. https://easychair.org/publications/preprint/vRwsT

Cheung, V. K., Harrison, P. M., Koelsch, S., Pearce, M. T., Friederici, A. D., & Meyer, L. (2023). Cognitive and sensory expectations independently shape musical expectancy and pleasure. *Philosophical Transactions of the Royal Society B*, *379*(1895), 20220420. https://doi.org/10.1098/rstb.2022.0420

Cho, J. L., Hall, M. D., & Pastore, R. E. (1991). Normalization process in the human auditory system. *Journal of the Acoustical Society of America*, *89*(4B), 1988–1988. https://doi.org/10.1121/1.2029785

Cho, J. L., Hall, M. D., & Pastore, R. E. (1993). Stimulus properties critical to normalization of instrument timbre. *Journal of the Acoustical Society of America*, *93*(4 Supplement), 2402–2402. https://doi.org/10.1121/1.406008

Christensen, T. (Ed.). (2006). *The Cambridge history of Western music theory*. Cambridge University Press.

Cox, A. (2001). The mimetic hypothesis and embodied musical meaning. *Musicae Scientiae*, *5*(2), 195–212. https://doi.org/10.1177/102986490100500

Cox, A. (2016). *Music and embodied cognition: Listening, moving, feeling, and thinking*. Indiana University Press. https://iupress.org/9780253032317/music-and-embodied-cognition/

Craton, L. G., Lee, J. H. J., & Krahe, P. M. (2021). It's only rock 'n roll (but I like it): Chord perception and rock's liberal harmonic palette. *Musicae Scientiae*, *25*(1), 92–110. https://doi.org/10.1177/1029864919845023

Creel, S. C. (2011). Specific previous experience affects perception of harmony and meter. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1512–1526. https://doi.org/10.1037/a0023507

Cuesta, H., Gómez Gutiérrez, E., Martorell Domínguez, A., & Loáiciga, F. (2018). *Analysis of intonation in unison choir singing* [Conference session]. 15th International Conference on Music Perception and Cognition and the 10th Triennial Conference of the European Society for the Cognitive Sciences of Music.

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. *Behavioral Research in Accounting*, *32*(1), 119–134. https://doi.org/10.2308/bria-18-044

Eitel, M., Ruth, N., Harrison, P., Frieler, K., & Müllensiefen, D. (2024). Perception of chord sequences modeled with prediction by partial matching, voice-leading distance, and spectral pitch-class similarity: A new approach for testing individual differences in harmony perception. *Music and Science*, *7*. https://doi.org/10.1177/20592043241257654

Farbood, M. M., & Price, K. C. (2017). The contribution of timbre attributes to musical tension. *Journal of the Acoustical Society of America*, *141*(1), 419–427. https://doi.org/10.1121/1.4973568

Goldman, A., Harrison, P. M., Jackson, T., & Pearce, M. T. (2021). Reassessing syntax-related ERP components using popular music chord sequences: A model-based approach. *Music Perception*, *39*(2), 118–144. https://doi.org/10.1525/mp.2021.39.2.118

Harrison, P., & Pearce, M. T. (2018). *An energy-based generative sequence model for testing sensory theories of Western harmony* [Conference session]. 19th International Society for Music Information Retrieval Conference. https://doi.org/10.48550/arXiv.1807.00790

Harrison, P., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition. *Psychological Review*, *127*(2), 216–244. https://doi.org/10.1037/rev0000169

Hughes, B. (2011). *Harmonic expectation in twelve-bar blues progressions*. (Publication No. 3483565) [Doctoral Dissertation, The Florida State University]. ProQuest Dissertations Publishing. https://www.proquest.com/openview/27f5385b5c1bd981fb878cf9da7c47fc/1?pq-origsite=gscholar&cbl=18750

Jers, H., & Ternström, S. (2005). Intonation analysis of a multi-channel choir recording. *TMHQPSR Speech, Music and Hearing: Quarterly Progress and Status Report*, *47*(1), 1–6.

Jimenez, I., & Kuusi, T. (2018). Connecting chord progressions with specific pieces of music. *Psychology of Music*, *46*(5), 716–733. https://doi.org/10.1177/0305735617721638

Jimenez, I., Kuusi, T., Czedik-Eysenberg, I., & Reuter, C. (2023). Identifying songs from their piano-driven opening chords. *Musicae Scientiae*, *27*(1), 27–52. https://doi.org/10.1177/10298649211003631

Jimenez, I., Kuusi, T., & Ojala, J. (2022). Veridical and schematic memory for harmony in melody-and-accompaniment textures. *Music Perception*, *40*(2), 89–111. https://doi.org/10.1525/mp.2022.40.2.89

Jimenez, I., Kuusi, T., Ojala, J., & Harrison, P. (2024). The effect of chord duration on the relative salience of chord-type and voicing changes. *Psychology of Music* (OnlineFirst). https://doi.org/10.1177/03057356241247533

Johnson-Laird, P. N., Kang, O. E., & Leong, Y. C. (2012). On musical dissonance. *Music Perception*, *30*(1), 19–35. https://doi.org/10.1525/mp.2012.30.1.19

Juchniewicz, J., & Silverman, M. J. (2013). The influences of progression type and distortion on the perception of terminal power chords. *Psychology of Music*, *41*(1), 119–130. https://doi.org/10.1177/0305735611422506

Kolchinsky, A., Dhande, N., Park, K., & Ahn, Y. Y. (2017). The minor fall, the major lift: inferring emotional valence of musical chords through lyrics. *Royal Society Open Science*, *4*, 170952. https://doi.org/10.1098/rsos.170952

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Krumhansl, C. L. (1996). A perceptual analysis of Mozart's Piano Sonata K. 282: Segmentation, tension, and musical ideas. *Music Perception*, *13*(3), 401–432. https://doi.org/10.2307/40286177

Krumhansl, C. L., Bharucha, J. J., & Castellano, M. A. (1982). Key distance effects on perceived harmonic structure in music. *Perception and Psychophysics*, *32*(2), 96–108. https://doi.org/10.3758/BF03204269

Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(1), 24–36. https://doi.org/10.1037/0096-1523.8.1.24

Krumhansl, C. L., & Cuddy, L. L. (2010). A theory of tonal hierarchies in music. In M. R. Jones, R. Fay, & A. Popper (Eds.), *Music perception* (Vol. 36, pp. 51–87). Springer. https://doi.org/10.1007/978-1-4419-6114-33

Kurzom, N., Lorenzi, I., & Mendelsohn, A. (2023). Increasing the complexity of isolated musical chords benefits concurrent associative memory formation. *Scientific Reports*, *13*(1), 1–12. https://doi.org/10.1038/s41598-023-34345-y

Lahdelma, I., & Eerola, T. (2016a). Single chords convey distinct emotional qualities to both naïve and expert listeners. *Psychology of Music*, *44*(1), 37–54. https://doi.org/10.1177/0305735614552006

Lahdelma, I., & Eerola, T. (2016b). Mild dissonance preferred over consonance in single chord perception. *i-Perception*, *7*(3), 2041669516655812. https://doi.org/10.1177/2041669516655812

Lahdelma, I., & Eerola, T. (2020). Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Scientific Reports*, *10*(1), 8693. https://doi.org/10.1038/s41598-020-65615-8

Lartillot, O., & Toiviainen, P. (2007). *A Matlab toolbox for musical feature extraction from audio* [Conference session]. 10th International Conference on Digital Audio Effects. https://www.dafx.de/paper-archive/2007/Papers/p237.pdf

Leman, M. (2000). An auditory model of the role of short-term memory in probe-tone ratings. *Music Perception*, *17*(4), 481–509. https://doi.org/10.2307/40285830

Lerdahl, F. (1988). Tonal pitch space. *Music Perception*, *5*(3), 315–349. https://doi.org/10.2307/40285402

Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception*, *24*(4), 329–366. https://doi.org/10.1525/mp.2007.24.4.329

Léveillé Gauvin, H. (2016). Changing use of seventh chords: A replication of Mauch et al. (2015). *Empirical Musicology Review*, *11*(1), 103–107. https://emusicology.org/index.php/EMR/article/view/4949/4393

Marjieh, R., Harrison, P. M., Lee, H., Deligiannaki, F., & Jacoby, N. (2022). Timbral effects on consonance illuminate psychoacoustics of music evolution. *bioRxiv*. https://doi.org/10.1101/2022.06.14.496070

McAdams, S., & Siedenburg, K. (2019). Perception and cognition of musical timbre. In P. J. Rentfrow & D. J. Levitin (Eds.), *Foundations in music psychology: Theory and research* (pp. 71–120). MIT Press.

Milne, A. J., & Holland, S. (2016). Empirically testing Tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, *10*(1), 59–85. https://doi.org/10.1080/17459737.2016.1152517

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). Measuring the facets of musicality: The Goldsmiths Musical Sophistication Index (Gold-MSI). Personality and Individual Differences, *60*, S35. https://doi.org/10.1016/j.paid.2013.07.081

Murphy, S. (2006). The major tritone progression in recent hollywood science fiction films. *Music Theory Online*, *12*(2). http://www.mtosmt.org/issues/mto.06.12.2/mto.06.12.2.murphy.html

Murphy, S. (2014). Transformational theory and the analysis of film music. In D. Neumeyer (Ed.), *The Oxford handbook of film music studies* (pp. 471–499). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195328493.013.019

Nadar, C. R., Abeßer, J., & Grollmisch, S. (2019). *Towards CNN-based acoustic modeling of seventh chords for automatic chord recognition*. 16th Sound & Music Computing Conference. https://smc2019.uma.es/proceedings.html

Paraskeva, S., & McAdams, S. (1997). Influence of timbre, presence/absence of tonal hierarchy and musical training on the perception of musical tension and relaxation schemas. *Proceedings of the 1997 International Computer Music Conference*. http://hdl.handle.net/2027/spo.bbp2372.1997.116

Parncutt, R., Engel, I., & Radovanovic, L. (2023). Consonance and dissonance of simultaneous trichords in Western music: A listening experiment to test models of harmonicity and roughness. *Psihologijske teme*, *32*(1), 13–34. https://hrcak.srce.hr/300435

Parncutt, R., Sattmann, S., Gaich, A., & Seither-Preisler, A. (2019). Tone profiles of isolated musical chords: Psychoacoustic versus cognitive models. *Music Perception*, *36*(4), 406–430. https://doi.org/10.1525/mp.2019.36.4.406

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The timbre toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America*, *130*(5), 2902–2916. https://doi.org/10.1121/1.3642604

Reymore, L., Beauvais-Lacasse, E., Smith, B. K., & McAdams, S. (2022). Modeling noise-related timbre semantic categories of orchestral instrument sounds with audio features, pitch register, and instrument family. *Frontiers in Psychology*, *13*, 796422. https://doi.org/10.3389/fpsyg.2022.796422

Roberts, L. A. (1986). Consonance judgements of musical chords by musicians and untrained listeners. *Acta Acustica United with Acustica*, *62*(2), 163–171. https://www.ingentaconnect.com/content/dav/aaua/1986/00000062/00000002/art00009

Rodrigues, R. M., Cedro, Á. M., Fonseca, R. M., Friedlaender, C. V., Torres, P. H. R., Oliveira, V. F. D., et al. (2017). The influence of timbre in the acquisition of rudimentary musical skills. *Revista Psicologia e Saúde*, *9*(3), 77–93. http://dx.doi.org/10.20435/pssa.v9i3.543

Sears, D., Caplin, W. E., & McAdams, S. (2014). Perceiving the classical cadence. *Music Perception*, *31*(5), 397–417. https://doi.org/10.1525/mp.2014.31.5.397

Sears, D., & Forrest, D. (2021). Triadic patterns across classical and popular music corpora: stylistic conventions, or characteristic idioms? *Journal of Mathematics and Music*, *15*(2), 140–153.

Sears, D., Pearce, M., Spitzer, J., Caplin, W., & McAdams, S. (2019). Expectations for tonal cadences: Sensory and cognitive priming effects. *Quarterly Journal of Experimental Psychology*, *72*(6), 1422–1438. https://doi.org/10.1177/1747021818814472

Simon, J., Balla, V., & Winkler, I. (2019). Temporal boundary of auditory event formation: An electrophysiological marker. *International Journal of Psychophysiology*, *140*, 53–61. https://doi.org/10.1016/j.ijpsycho.2019.04.006

Skorik, S. O., Almayev, N. A., & Bessonova, Y. V. (2018). Towards psychophysiological mechanisms of major and minor chords perception. *International Journal of Psychophysiology*, *131*, S163. https://doi.org/10.1016/j.ijpsycho.2018.07.432

Steinmetz, C. J., & Reiss, J. (2021). *pyloudnorm: A simple yet flexible loudness meter in Python*. Audio Engineering Society Convention 150. http://www.aes.org/e-lib/browse.cfm?elib=21076

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Tagg, P. (2014). *Everyday tonality II: Towards a tonal theory of what most people hear*. The Mass Media Music Scholars' Press.

Thiesen, F. C., Kopiez, R., Müllensiefen, D., Reuter, C., & Czedik-Eysenberg, I. (2020). Duration, song section, entropy: Suggestions for a model of rapid music recognition processes. *Journal of New Music Research*, *49*(4), 334–348. https://doi.org/10.1080/09298215.2020.1784955

Trainor, L. J., Marie, C., Bruce, I. C., & Bidelman, G. M. (2014). Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hearing Research*, *308*, 60–70. https://doi.org/10.1016/j.heares.2013.07.014

Vassilakis, P. N. (2001). *Perceptual and physical properties of amplitude fluctuation and their musical significance*. (Publication No. 3032914) [Doctoral Dissertation, University of California, Los Angeles]. ProQuest Dissertations Publishing. https://www.proquest.com/openview/74fc5469e73c7b6feed81b6598febfe8/1?pq-origsite=gscholar&cbl=18750&diss=y

Virtala, P., Huotilainen, M., Lilja, E., Ojala, J., & Tervaniemi, M. (2018). Distortion and Western music chord processing: An ERP study of musicians and nonmusicians. *Music Perception*, *35*(3), 315–331. https://doi.org/10.1525/mp.2018.35.3.315

Virtala, P., Huotilainen, M., Partanen, E., & Tervaniemi, M. (2014). Musicianship facilitates the processing of Western music chords—An ERP and behavioral study. *Neuropsychologia*, *61*, 247–258. https://doi.org/10.1016/j.neuropsychologia.2014.06.028

Vuvan, D. T., & Hughes, B. (2019). Musical style affects the strength of harmonic expectancy. *Music and Science*, *2*, 2059204318816066. https://doi.org/10.1177/2059204318816066

Vuvan, D. T., & Hughes, B. (2021). Probe tone paradigm reveals less differentiated tonal hierarchy in rock music. *Music Perception*, *38*(5), 425–434. https://doi.org/10.1525/mp.2021.38.5.425

Williams, L. R. (2005). Effect of music training and musical complexity on focus of attention to melody or harmony. *Journal of Research in Music Education*, *53*(3), 210–221. https://doi.org/10.1177/002242940505300303

Williams, L. R. (2008). Effect of music complexity on nonmusicians' focus of attention to melody or harmony. *Update: Applications of Research in Music Education*, *26*(2), 27–32. https://doi.org/10.1177/8755123308317952

Appendix

TABLE A1. *Description of Variables*

Type of variable*	Variable	Description	Pertinent findings from previous studies
Harmonic	TPS distance (2, 21)**	Distance between chords C and X in Lerdahl's Tonal Pitch Space (Bigand et al., 1996).	Highly correlated with tension ratings of the middle chord in C-X-C patterns (Bigand et al., 1996).
	TC (.73, .98)	Leman's Tonal Contextuality (Leman, 2000) of the test (X) chord in the context of C chord. TC is closely related to the concept of pitch commonality (Bigand et al., 1996) and can be understood as a measure of the sensory component of harmonic contrast between two musical events (the lower the TC, the less pitch commonality, and the higher the sensory harmonic contrast). The TC values used in the current study were average TC values obtained from using 75 combinations of global and local decays. The local decay varied from 100 to 500 ms by steps of 100 ms and the global decay, from 0.5 to 4 sec by steps of 0.25 sec. Average of TC values from multiple combinations of global and local decays were used because Leman's model is sensitive to the decay parameters (Bigand et al., 2014; Jimenez et al., 2024) and because the global decay is intended to represent the duration of echoic memory, but there is no current consensus on the duration of echoic memory.	Highly correlated with GoF ratings of individual major chords played after an ascending and descending major scale and two tonic chords (Craton et al., 2021). Predicts the sensory component of harmonic surprise in chord sequences (Cheung et al., 2023). In conjunction with frequency-of-occurrence data predicts brain responses to harmonic surprise in chord sequences (Goldman et al., 2021). Predicts many instances of tonal priming (Bigand et al., 2014). Pitch commonality, a predecessor of TC, is highly correlated with tension ratings of the middle chord in C-X-C patterns (Bigand et al., 1996). Pitch-class spectrum distance (Milne & Holland, 2016), a simplified version of TC, is associated with the salience of harmonic differences between short chord sequences (Eitel et al., 2024). Note: Craton et al. (2021) and the current study differ in terms of the general level of and ranking according to TC values for the chord sequences C-D, C-E, C-F, C-G, and C-Bb. These differences are likely influenced by voice-leading, use of scales, and the global and local decay parameters. Further information about these differences can be found at https://osf.io/tajyz/
	Frequency of occurrence in classical music (−12.0, −3.8)	Logarithmic version of the frequency of occurrence of the C-X transitions in a harmony corpus of 1,022 keyboard and ensemble pieces from Bach, Haydn, Mozart, Beethoven, and Chopin (Harrison & Pearce, 2018).	Frequency of occurrence of chord types in a large corpus of classical music was strongly associated with roughness and inharmonicity, two features that are known to affect perceived dissonance (Harrison & Pearce, 2020).
	Frequency of occurrence in popular music (−15.0, −2.3)	Logarithmic version of the frequency of occurrence of the C-X transitions in a harmony corpus of 537 pop songs that appeared on the Billboard Year-End Hot 100 Songs chart from 2002 to 2021 (Chander & Quinn, 2023).	Frequency of occurrence of chord sequences in a corpus of popular music contributed to perceived surprise in chord sequences (Cheung et al., 2023), salience of harmonic differences (Eitel et al., 2024), and brain responses to harmonic surprise in chord sequences (Goldman et al., 2021). The analysis in the current study used a harmonic corpus of more recent popular music.
Melodic	Soprano motion (1, 4)	Melodic motion of the soprano in semitones (motion was always descending in the stimuli used in the experiment).	Highly correlated with tension ratings of the middle chord in C-X-C patterns (Bigand et al., 1996).
	Bass motion (0, 10)	Melodic motion of the bass in semitones (motion was always ascending in the stimuli used in the experiment)	Correlated with tension ratings of the middle chord in C-X-C patterns (Bigand et al., 1996).

(continued)

TABLE A1. *(continued)*

Type of variable*	Variable	Description	Pertinent findings from previous studies
Temporal***	Attack time ms (linear) (34, 116)	Duration of the attack portion of the sound.	Important feature in similarity estimation between instruments as well as for instrument identification (McAdams,& Siedenburg, 2019). Salient difference between the amplitude envelopes of our choir stimuli and the piano and guitar stimuli. Can facilitate the perceptual distinction between major and minor chords (Skorik et al., 2018). Attack was positively correlated with nostalgia and tenderness ratings of isolated chords played with either piano or strings tones (Lahdelma & Eerola, 2016a). Slow attack was associated with greater perceived tension of isolated chords played with either piano or organ tones (Arthurs et al., 2018) but lower tension of excerpts played on either piano or orchestra (Paraskeva & McAdams, 1997).
	Decay time ms (linear) (125, 1135)	Duration of the decay portion of the sound.	Salient difference between the amplitude envelopes of our choir stimuli and the piano and guitar stimuli.
	Temporal centroid ms (linear) (360, 912)	Center of gravity of the temporal envelope.	Distinguishes percussive from sustained sounds (Peeters et al., 2011). The distinction between percussive and sustained sounds can affect consonance, pleasantness, stability, and relaxation ratings of isolated chords (Arthurs et al., 2018).
Spectral****	Spectral centroid (967.4, 1746.9)	Geometric center of gravity of the spectrum. Related to perceived brightness.	Important feature in similarity estimation between instruments (McAdams, 2019). Perceptually salient in similarity estimation of chords (Jimenez et al., 2023). Higher spectral centroid is associated with lower consonance, pleasantness, stability and relaxation ratings of isolated chords (Arthurs et al., 2018). Related to perceived noisiness and perceived physical exertion (Reymore et al., 2022). Timbral brightness was positively correlated with nostalgia and tenderness ratings of isolated chords played with either piano or strings tones (Lahdelma & Eerola, 2016a).
	Spectral entropy (.599, .734)	Shannon entropy of the spectral distribution.	Related to instrument identification (Ajayakumar & Rajan, 2020) and memorability of brief polyphonic excerpts (Thiesen et al., 2020).
	Spectral flatness (.042, .073)	Ratio between geometric and arithmetic means of the spectrum. Indicates whether the spectral distribution is smooth or spiky and can be understood as measuring how similar the spectrum is to white noise.	Related to perceived tension (Farbood & Price, 2017) and noisiness (Reymore et al., 2022). Greater spectral flatness is associated with lower consonance, pleasantness, stability and relaxation ratings of isolated chords (Arthurs et al., 2018).
	Spectral spread (1923.6, 2488.7)	Standard deviation of the spectrum around the mean. Typically, the more upper partials and/or broadband noise, the larger the spread.	Related to perceived noisiness, harshness, and perceived physical exertion (Reymore et al., 2022).

(continued)

TABLE A1. *(continued)*

Type of variable*	Variable	Description	Pertinent findings from previous studies
	Roughness (.144, .793)	Interference between partials due to beating. In this study, roughness is calculated using Vassilakis (2001) model and influenced by chord type, voicing, and timbre.	Roughness (in general) is related to dissonance (Harrison & Pearce, 2020; Parncutt et al., 2023), tension (Farbood & Price, 2017) and highly correlated with musicians' tension ratings of the middle chord in C-X-C patterns (Bigand et al., 1996). Roughness, as calculated via the Vassilakis model, was positively correlated with nostalgia and tenderness ratings of isolated chords played with either piano or strings tones (Lahdelma & Eerola, 2016a) and positively correlated with perceived tension, energy, and attention and negatively correlated with consonance and valence of isolated chords played with piano tones (Lahdelma & Eerola, 2016b; Kurzom et al., 2023).
	Inharmonicity (.168, .468)	Degree to which frequencies of the partials depart from multiples of the fundamental frequency.	Related to dissonance (Harrison & Pearce, 2020), tension (Farbood & Price, 2017), harshness and perceived physical exertion (Reymore et al., 2022). Highly correlated with roughness in a large set of chords played using different transpositions and voicings (Parncutt et al., 2023).
Spectro-temporal****	Spectral flux (7.0, 67.7)	A measure of variability of the spectrum over time.	Greater spectral flux was associated with lower consonance, pleasantness, stability and relaxation ratings of isolated chords (Arthurs et al., 2018). Spectral flux was positively correlated with nostalgia and tenderness ratings of isolated chords played with either piano or strings tones (Lahdelma & Eerola, 2016a).

* In this paper, harmonic and melodic variables are referred to as chordal variables, whereas temporal, spectral, and spectro-temporal variables are referred to as timbral variables. The only exception in this categorization is the variable roughness which we referred to as being a chordal variable because of its connection to chord type and voicing.

** Numbers in parentheses under the name of the variable indicate the minimum and maximum values for the variable in our set of stimuli.

*** Temporal variables were calculated using the Timbre toolbox (Peeters et al., 2011).

**** Spectral and spectro-temporal variables were calculated using the MIRtoolbox (Lartillot & Toiviainen, 2007).

TABLE A2. *Intercorrelation Between Variables*

		Chordal							Timbral							
Type	Variable	1	2	3	4	5	6	7	1	2	3	4	5	6	7	8
Chordal	1. TPS distance (distance in Tonal Pitch Space)															
	2. Frequency of occurrence classical music	−.60**														
	3. Frequency of occurrence popular music	−.89**	.59**													
	4. TC (harmonic similarity according to Leman)	−.75**	.21	.59**												
	5. Roughness Vassilakis	.15	.12	−.03	−.44**											
	6. Soprano motion	.42**	−.39**	−.20	−.27	.20										
	7. Bass motion	−.27	−.02	.16	.47**	−.45**	−.35*									
Timbral	1. Attack time ms (linear)	−.01	.12	.01	.02	.37*	.02	−.02								
	2. Decay time ms (linear)	−.02	.14	.01	.05	.39**	−.04	.03	.98**							
	3. Temporal centroid ms	.03	.14	−.01	−.06	.48**	.06	−.09	.86**	.90**						
	4. Inharmonicity	−.02	.16	−.02	−.23	.51**	.11	.02	.33*	.33*	.32*					
	5. Spectral centroid	−.04	−.02	.01	.12	.04	−.02	.32*	−.23	−.12	.14	−.04				
	6. Spectral entropy	.02	.13	−.01	−.02	.49**	.06	.09	.89**	.91**	.88**	.40**	.15			
	7. Spectral flatness	−.04	.06	.02	.17	.02	−.06	.40**	.18	.28	.48**	−.03	.87**	.48**		
	8. Spectral flux	.04	.12	.00	−.05	.43**	.02	−.06	.98**	.96**	.85**	.36*	−.28	.88**	.12	
	9. Spectral spread	−.05	−.12	.04	.18	−.64**	−.08	.33*	−.85**	−.82**	−.77**	−.48**	.33*	−.75**	.11	−.88**

Note. $^{*}p < .05$ $^{**}p < .01$