Automatic responses to musical intervals: Contrasts in acoustic roughness predict affective priming in Western listeners

James Armitage, Imre Lahdelma and Tuomas Eerola

Citation: The Journal of the Acoustical Society of America **150**, 551 (2021); doi: 10.1121/10.0005623
View online: https://doi.org/10.1121/10.0005623
View Table of Contents: https://asa.scitation.org/toc/jas/150/1
Published by the Acoustical Society of America

JASA ARTICLE

Automatic responses to musical intervals: Contrasts in acoustic roughness predict affective priming in Western listeners

James Armitage,[a] Imre Lahdelma,[b] and Tuomas Eerola[c]
Department of Music, Durham University, Durham, DH1 3RL, United Kingdom

ABSTRACT:
The aim of the present study is to determine which acoustic components of harmonic consonance and dissonance influence automatic responses in a simple cognitive task. In a series of affective priming experiments, eight pairs of musical intervals were used to measure the influence of acoustic roughness and harmonicity on response times in a word-classification task conducted online. Interval pairs that contrasted in roughness induced a greater degree of affective priming than pairs that did not contrast in terms of their roughness. Contrasts in harmonicity did not induce affective priming. A follow-up experiment used detuned intervals to create higher levels of roughness contrasts. However, the detuning did not lead to any further increase in the size of the priming effect. More detailed analysis suggests that the presence of priming in intervals is binary: in the negative primes that create congruency effects the intervals' fundamentals and overtones coincide within the same equivalent rectangular bandwidth (i.e., the minor and major seconds). Intervals that fall outside this equivalent rectangular bandwidth do not elicit priming effects, regardless of their dissonance or negative affect. The results are discussed in the context of recent developments in consonance/dissonance research and vocal similarity.
 https://doi.org/10.1121/10.0005623

(Received 4 February 2021; revised 19 June 2021; accepted 24 June 2021; published online 22 July 2021)
[Editor: James F. Lynch] Pages: 551–560

I. INTRODUCTION

The contrast between consonance and dissonance is a vital feature of Western music. Consonance is typically perceived as *agreeable* and *stable*, while dissonance, in turn, is perceived as *disagreeable* and *in need of resolution* (Tramo *et al.*, 2001). Consonance/dissonance has both a vertical and a horizontal aspect: single isolated *intervals* (two concurrent pitches) and *chords* (three or more concurrent pitches) represent *vertical consonance/dissonance*, while the sequential relationships between these in melodies and chord progressions represent *horizontal consonance/dissonance* (Parncutt and Hair, 2011). The current research refers exclusively to the vertical aspect.

Empirical research concerning consonance and dissonance frequently relies on self-report methods, which come with a well-documented set of limitations (see, e.g., Fazio and Olson, 2003, for review). Priming, on the other hand, captures participants' automatic responses to the stimuli, avoiding demand characteristics in a question that is underpinned by both physiology and culture (see, e.g., Herring *et al.*, 2013, for review). Previous studies using an affective priming paradigm have shown that valenced chords (e.g., consonant-positive, dissonant-negative) facilitate the evaluation of similarly valenced target words (see, e.g., Steinbeis and Koelsch, 2011). Recent research by Lahdelma *et al.* (2020) has found that this congruency effect is not present when intervals, as opposed to chords, are used as primes. Lahdelma *et al.* suggested tentatively that this finding was a consequence of the higher levels of roughness and/or harmonicity found in the four-note chords compared to the intervals rather than as a consequence of the number of notes in the chords *per se*. Roughness is often seen as prevalent in the perception of dissonance but not in the perception of consonance (see, e.g., Hutchinson and Knopoff, 1979), as dissonant intervals contain less overall roughness than dissonant chords. Apart from roughness, another major acoustic factor related to consonance and dissonance is *harmonicity*, which has been demonstrated to contribute to perception of consonance in a variety of settings (McDermott *et al.*, 2010). The study by Lahdelma *et al.* (2020), however, tested only four distinct intervals, which poses a possible limitation.

The present study aims to disentangle the role of specific acoustic properties of the intervals, namely *roughness* and *harmonicity*. It considers automatic responses to these acoustic components as indexed in an affective priming paradigm (see Sec. I B, below) rather than by self-report methods frequently employed in consonance-dissonance research.

A. Consonance and dissonance

1. Acoustic roughness

Historically, the acoustic and perceptual characteristics of consonance and dissonance were placed on a sound

[a]Electronic mail: james.e.armitage@durham.ac.uk, ORCID: 0000-0001-9802-7479.
[b]ORCID: 0000-0001-6193-928X.
[c]ORCID: 0000-0002-2896-929X.

https://doi.org/10.1121/10.0005623

theoretical and empirical footing by Helmholtz (1885) who proposed acoustic roughness as an explanation for why some musical intervals are considered dissonant and disagreeable. There is consensus that the sensation of roughness is caused by interference patterns between wave components of similar frequency that give rise to beating (see, e.g., Hutchinson and Knopoff, 1978), which in turn creates the sound qualities that listeners typically perceive as unpleasant. Based on his findings of the relationship between roughness and dissonance, von Helmholtz derived that consonance in turn is the absence of roughness and concluded that roughness (or lack thereof) is the cause of both consonance and dissonance in music. Coming to a similar conclusion later, Terhardt (1984) proposed that the evaluation of consonance in isolated intervals and chords is mostly governed by *sensory consonance*, i.e., a lack of unpleasant features of a sound such as sharpness (the presence of spectral energy at high frequencies) and roughness.

The perception of roughness has a biological substrate, as beating occurs at the level of the basilar membrane in the inner ear when the frequency components are too close together to separate (see, e.g., Tramo *et al.*, 2001). This range is known as the *critical bandwidth* (Fletcher and Munson, 1933). According to Smith and Abel (1999, p. 21) "a critical band is 100 Hz wide for center frequencies below 500 Hz, and 20% of the center frequency above 500 Hz." A more recent formulation of this concept, the equivalent rectangular bandwidth (ERB) (see Patterson, 1976) approximates the basilar membrane as being made up of rectangular bandpass filters. Fundamental frequencies processed within the same bandpass filter are perceived as acoustically rough. Compared to the critical bandwidth, the ERB typically encompasses a narrower range of frequencies (Smith and Abel, 1999). Although the ERB formulation is very much an approximation, it nevertheless provides a useful framework for considering aspects of vertical harmony.

The sensitivity to roughness seems to be present cross-culturally (see McDermott *et al.*, 2016), but its appraisal differs significantly across musical styles and cultures: while a typical Western listener hears roughness as disagreeable, it is deliberately harnessed in the vocal practice of *beat diaphony* (known as *Schwebungsdiaphon*ie in German literature) in for example the Baltic and Balkan regions of Europe (see, e.g., Ambrazevičius, 2017) and in Papua New Guinea (Messner, 1981). Several models of roughness exist such as those by Hutchinson and Knopoff (1978), Vassilakis (2001), and Wang *et al.* (2013). These models, which are based on emulations of the human auditory system perceiving the sensation of roughness, largely agree on the amount of roughness in different intervals. Figure 1 shows how roughness varies for intervals over the octave from C_4 to C_5 as the distance between the two fundamentals increases by one cent—equal to 1/1200 of an octave.

While roughness was long accepted as a sufficient explanation for consonance and dissonance (see Helmholtz, 1885; Terhardt, 1984), later counterarguments were made for why it is not all-encompassing in explaining its underlying cause. First, perceptions of consonance and dissonance seem to remain when the tones of a chord are presented independently to the ears (i.e., dichotically), precluding physical interaction at the input stage and thus greatly reducing the perception of roughness (McDermott *et al.*, 2010). However, it has been pointed out that beats could also occur centrally, within a binaural critical band rather than being based on cochlear interactions (Carcagno *et al.*, 2019). Second, the perceived consonance of a chord does not seem to increase when roughness is artificially reduced by removing partials from complex tones (e.g., Nordmark and Fahlén, 1988). Third, it has been shown that participants with congenital amusia (i.e., a neurogenetic disorder characterised by an inability to recognise or reproduce musical tones) exhibit abnormal consonance perception but normal roughness perception (Cousineau *et al.*, 2012). Recent research in roughness (Arnal *et al.*, 2015; Arnal *et al.*, 2019) highlights the presence of roughness in, for example, alarm signals and suggests that acoustic roughness is responsible for activating salience-related aversive responses, particularly in the amygdala. Interestingly, Arnal *et al.* (2015) found that this activation was common to screams, alarm signals and—crucially for the present study—acoustically rough musical intervals. Additionally, Koelsch *et al.* (2018) found that stimuli high in acoustic roughness were associated with increased activation in the left planum polare and the orbital sulcus of the orbitofrontal cortex, a region associated with negative reinforcement (Kringelbach and Rolls, 2004).

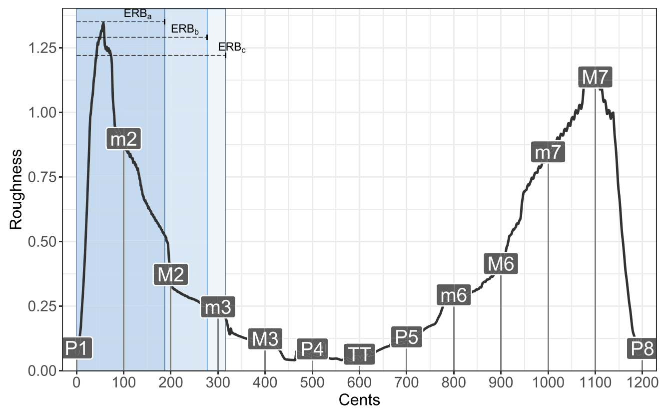

FIG. 1. (Color online) Roughness using the model by Wang *et al.* (2013) for intervals from unison to octave divided into 1200 cents. Frequency-dependent critical bandwidth boundaries are shown for the first three intervals ($ERB_a = m2$, $ERB_b = M2$, $ERB_c = m3$) reflecting the different mean frequencies of the intervals in our design (see Sec. II B) using the ERB bandwidths (Moore and Glasberg, 1983).

2. Harmonicity

Another possible explanation that has been put forward to explain consonance and dissonance in addition to roughness is the acoustic property of *harmonicity*, which denotes how closely the spectral frequencies of a sound correspond to a harmonic series. A single musical pitch is the combination of the fundamental frequency plus its overtones, which typically follow the harmonic series. If an interval is made of two tones that share a large proportion of their overtones (e.g., the perfect fifth) then the interval is high in harmonicity. Conversely, intervals whose overtones do not significantly overlap (e.g., the major seventh) are considered low in harmonicity. In other words, harmonicity posits that in consonant pitch combinations of the component frequencies produce an aggregate spectrum that is typically harmonic, i.e., it resembles the spectrum of a single sound. In contrast, dissonant intervals and chords produce an inharmonic spectrum (Cousineau *et al.*, 2012).

Harmonicity offers a conception of consonance that is defined constructively rather than by the absence of roughness. Like roughness, harmonicity may have a biological underpinning, albeit at the central rather than peripheral level (Tramo *et al.*, 2001), even if this is debated (Carcagno *et al.*, 2019). Preference for high levels of harmonicity has been found to correlate with preference for consonance (McDermott *et al.*, 2010), and it has been suggested that this link might be related to the advantages of recognising human vocalisation (Bowling and Purves, 2015). Another possibility is that, unless humans are born with a preference for harmonicity, it is acquired simply through exposure to natural sounds or music (McDermott *et al.*, 2010).

B. Affective priming

Affective priming is a common paradigm in both social and cognitive psychology and is employed in diverse domains such as memory, social psychology, psycholinguistics, and psychopathology as an indirect measure of attitudes (see, e.g., Herring *et al.*, 2013, for a review). In the affective priming paradigm, two stimuli are presented near-simultaneously. The extent to which the first (*prime*) stimulus influences responses to the second (*target*) stimulus in a classification task (for instance where the target is to be classified as positive or negative) is indexed by reaction time or accuracy rate. In particular, affective priming studies frequently consider *congruency effects*. Typically, in congruent conditions (i.e., when the prime and target stimuli share an affective feature such as valence), reaction times (RTs) are faster than in incongruent conditions (i.e., when the prime and target stimuli do not share the affective feature). Congruency effects have been associated with a range of auditory stimuli, for instance, affective sounds (Scherer and Larsen, 2011) and music (Goerlich *et al.*, 2011). Affective priming has been used in a small number of vertical harmony studies (Bakker and Martin, 2015; Costa, 2013; Lahdelma *et al.*, 2020; Sollberger *et al.*, 2003; Steinbeis and Koelsch, 2011). There is a consensus amongst these authors that chords are associated with congruency effects, for instance dissonant (or minor) chords paired with negative words and consonant (or major) chords paired with positive words are associated with faster RTs and/or higher accuracy rates than the converse. Despite the relative strength of agreement in the existence of congruency effects, there is relatively little discussion as to what drives these effects, either in terms of cognition or in terms of the specific acoustic properties of the chords. Steinbeis and Koelsch (2011) and Bakker and Martin (2015) suggest activation of affective concepts in the semantic memory and conflict-resolution in cross-modal integration respectively as possible causal factors on a cognitive level. The specific features of chords that may be responsible for congruency are not explored beyond modality, consonance-dissonance, register, and numerosity. However, although Steinbeis and Koelsch (2011) and Lahdelma *et al.* (2020) do tentatively suggest that acoustic roughness is responsible for presence of priming effects, the experimental manipulations varied consonance, modality, and numerosity rather than varying roughness directly. Moreover, the authors' conclusions in favour of roughness cannot be considered conclusive owing to the absence of consideration of other factors, such as harmonicity.

C. The present study

The extent to which acoustical aspects (roughness and harmonicity) contribute to the perception of consonance/dissonance in intervals has remained contentious. The present study sought to explore the extent to which roughness and harmonicity influence automatic responses on a simple word evaluation task. As affective priming is an indirect measure, it is expected to yield valuable information on the importance of these individual factors' contributions to the perception of consonance and dissonance. The present study considered whether intervals can influence behaviour on a word evaluation task. An important question in cognitive psychology that has remained unclear in previous research is what property of the prime is responsible for activating the nodes in the semantic-affective network that lead to affective priming – perceived valence (i.e., positive or negative) or consonance/dissonance (teasing apart the components of roughness and harmonicity). The present study used ten interval pairs that were chosen to contrast in terms of their roughness or harmonicity (see Sec. II B) to prise apart the differential contributions of these acoustic components of consonance and dissonance to affective priming. The intervals were chosen to maximise contrasts between the interval pairs in terms of roughness and harmonicity.

In a similar previous experiment (Lahdelma *et al.*, 2020), it was found that when positive or negative words were preceded by tetrachords (four concurrent pitches), there were significant congruency effects—i.e., positive words were classified more quickly when preceded by a consonant rather than a dissonant chord, and *vice versa* for negative words. Notably, these congruency effects were absent for intervals. Lahdelma *et al.* suggested that the context of exposure to tetrads might dampen responses to intervals and so the present experiments set out to examine whether diatonic intervals can drive priming effects without the confound of exposure to more complex chords. Ten separate within-subjects sub-experiments were conducted online, one for each pair of intervals (see Table II). There

TABLE I. Intervals, notation, and key descriptors.

Interval (Abbr.)	Roughness	Harmonicity	Partials/ERB
Minor second (m2)	10.00	1.00	23
Major second (M2)	4.79	5.42	25
Minor third (m3)	3.66	2.79	18
Major third (M3)	4.66	4.72	15
Perfect fourth (P4)	3.64	10.00	14
Tritone (TT)	2.70	1.02	15
Perfect fifth (P5)	1.00	10.00	13
Minor sixth (m6)	4.75	4.72	14
Major sixth (M6)	5.98	2.79	15
Minor seventh (m7)	5.70	5.42	14
Major seventh (M7)	2.58	1.00	12
Minor second (detuned piano)	14.44	-	
Minor second (detuned Shepard)	10.11	-	
Perfect Fifth (Shepard)	−5.56	-	

were separate participants for each sub-experiment. Participants were asked to classify a sequence of emotional words as either positive or negative. Each word was preceded by the brief sounding of a musical interval. To evaluate the extent to which manipulating the amount of roughness and harmonicity can influence results of the behavioural task, each sub-experiment used two intervals, chosen for their contrast in roughness, harmonicity, or both. We predicted that intervals where there were high contrasts in harmonicity or roughness (or both) would be associated with congruency effects in RT, whereas interval pairs that did not contrast greatly in roughness or harmonicity would not be associated with congruency effects.

Following the first ten sub-experiments, to probe the role of roughness further, we tested two interval pairs involving artificially detuned minor seconds (played with both the piano timbre and with the Shepard tone, see Sec. II B for details). It is speculated that response to acoustic roughness confers an evolutionary advantage. For instance, alarm signals, whether in nature or man-made, are frequently high in roughness (Arnal *et al.*, 2015). Indeed, specifically in the case of human vocalisation, roughness has been linked to perceived anger (Bänziger *et al.*, 2015), screams (Schwartz *et al.*, 2019), and infant cries (Koutseff *et al.*, 2018). Consequently, it was predicted that these interval pairs which exhibited a higher difference in roughness than the standard diatonic intervals would elicit even greater priming effects than the high roughness contrast intervals.

II. METHODS

A. Participants

Participants were recruited *via* Prolific Academic, a crowdsourcing platform targeted specifically for academic research. Following deletions (17 participants; see Sec. III A), 379 participants (197 female, 178 male, 4 other/prefer not to say, mean age = 36.2; standard deviation = 12.4) completed the study. All participants reported corrected-to-normal or normal vision and were right-handed native speakers of English. 295 participants identified as non-musicians. Ethical approval was granted by the Music Department Ethics Committee, University of Durham. Informed consent was provided via an online checkbox.

B. Materials and stimuli

Ten auditory stimuli were generated in total, eight diatonic intervals (equal tempered in common with Costa, 2013 and Sollberger *et al.*, 2003) plus two intervals manipulated in tuning to maximise roughness. The diatonic interval pairs were chosen so as to maximise the contrasts in harmonicity and roughness; the diatonic intervals were classified as being high or low in contrast in roughness (ΔRoughness) and high or low in contrast in harmonicity (ΔHarmonicity), so there were altogether four conditions spanning the contrasts in roughness and harmonicity (see Table II). The diatonic intervals were created in accordance with the procedure employed by Bowling *et al.* (2018), where the fundamental frequencies (f_0) of the pitches in each interval were adjusted so that the mean f_0 of both pitches in each interval pair was C_4 (261.63 Hz). Descriptors of roughness and harmonicity are given in Table I. Roughness calculations were carried out using the model developed by Wang *et al.* (2013). The analyses were duplicated using the Vassilakis (2001) and Hutchinson and Knopoff (1978) models, and with a composite model (mean roughness value of all three models). Harmonicity was calculated using the model by Harrison and Pearce (2018) which simulates the way listeners search the auditory spectrum for occurrences of harmonic spectra; harmonicity values under alternative harmonicity models operate similarly in this context and a composite model is detailed in the supplementary material. For the follow up sub-experiments, the two artificial intervals were combined with perfect fifths to create two additional interval pairs: a detuned minor second and a perfect fifth in Shepard Tones and piano timbre (hereafter abbreviated to s2/S5 and d2/P5, respecively). These interval pairs were classified as being "Extreme" in roughness contrast.

As an additional diagnostic measure of dissonance, the first 13 partials (fundamental plus twelve overtones) were extracted from each single tone from the intervals presented

TABLE II. Interval pairs tested for priming index (difference in roughness and harmonicity given in brackets) and number of participants.

Intervals	Δ Roughness	Δ Harmonicity	N
m3/M3	Low (1.00)	Low (1.97)	39
m6/M6	Low (1.23)	Low (1.93)	44
TT/P5	Low (1.70)	High (8.98)	37
M7/P5	Low (1.58)	High (9.00)	37
m2/P5	High (9.00)	High (9.00)	39
m2/TT	High (7.30)	Low (0.02)	38
M2/P5	High (3.79)	High (4.58)	33
m2/M3	High (5.34)	Low (3.72)	37
s2/S5	Extreme (12.61)	N/A	37
d2/P5	Extreme (15.67)	N/A	38

below. For each interval, we calculated the ERB about the mean of each pair of partials using the formula derived by Moore and Glasberg (1983), generating 169 (i.e., 13×13) ERBs and checked whether the frequencies fell within this band. Table I details how many partial pairs for each interval fell within ERBs. The ERB was defined as the frequency band between the boundary suggested by Moore and Glasberg (1983) and 10 Hz. Table I excludes those partial pairs that fell within 10 Hz because beating effects due to frequency differences of less than 10 Hz would be perceived as amplitude modulation or "beats" rather than roughness (see, e.g., Roederer, 2008, p. 38); a further limitation of this measure is that very close alignments of overtones would contribute to a sense of harmonicity, for instance in the case of P5.

For the primes in the initial experiment, there were eight different pairings of intervals: m3/M3, m6/M6, TT/P5, m2/P5, and M7/P5, m2/TT, M2/P5, and m2/M3. These interval pairings were generated with *Ableton Live 9* (a music sequencer software), using the *Synthogy Ivory Grand Pianos II* sample-based plug-in. For the piano interval pairs, the applied sound font was *Steinway D Concert Grand*. No reverb was used, and the intervals had a fixed velocity (65) in order to have a neutral and even sound. The artificial interval pairs comprised d2/P5 (detuned minor second) and s2/S5 (Shepard tones). In the d2/P5 interval pairing the minor second interval was created by taking a unison and detuning one pitch down by −90 cents; this procedure created a notably high amount of roughness when measured with the models by Vassilakis (2001) and Wang *et al.* (2013). The pairing s2/S5 was created using Shepard tones that have octave-spaced partials from 16 Hz to 20 kHz with cosine-curve-shaped spectral envelope (for the code used to generate the Shepard tones, see the supporting information)[4]. The detuning of the minor second interval using the Shepard tone was created by shifting the odd partials above the fundamental upward and partials below downward by a detuning constant d. The constant was determined to yield a maximal roughness at $d = 0.024$ using the roughness models by Vassilakis and Wang *et al.* Table II shows the pairs of primes and the differences in their acoustic parameters. We used a median split to classify the differences in roughness and harmonicity as High or Low. The classification as High or Low ΔRoughness and High or Low ΔHarmonicity remained unchanged when we calculated composite Roughness and Harmonicity measures (see supplementary material).

The loudness of the stimuli was equalised by setting them to the same peak sone level. The sound files were converted to stereo (same signal in both channels) as 44.1 kHz, 32 bits per sample waveform audio files. The length of each interval was exactly 800 ms including a 10 ms fadeout. Figure 2 shows the spectra of a selection of stimuli alongside details of the calculation of the differences in roughness.

The target words were chosen from the database of affective norms for English words compiled by Warriner *et al.* (2013). All the words consisted of one or two syllables and were controlled for arousal so that there were four low-arousal and four high-arousal words in the positive and negative categories Warriner *et al.* (2013). The negative words were *Flaccid (3.43), Hijack (1.84), Rabid (2.95), Coma (1.89), Saggy (2.62), Dismal (2.6), Arrest (2.33), Morgue (1.79)* (the Warriner *et al.*, 2013 valence ratings on a scale of 1–7 are given in brackets); the positive words were *Climax (7.53), Gentle (7.42), Lively (7.12), Rest (7.86), Excite (7.79), Payday (7.95), Relax (7.82), Comfy (7.25).* The target words were presented in white size 40 Arial font on a black background. The priming task was coded using PsyToolkit (Stoet, 2017). RT distributions collected *via* PsyToolkit and Prolific Academic have been found to be comparable to RT distributions collected in a controlled laboratory environment (Armitage and Eerola, 2019).[1]

C. Procedure

The experiment consisted of a standard word classification task with affective priming. Each item consisted of the prime (interval) presented simultaneously with a fixation cross for 250 ms. At 250 ms, the fixation cross was replaced with the target word. Participants were instructed to press the "z" key if the target word was negative and the "m" key if it was positive. The target word remained onscreen for 1500 ms; key presses greater than 2000 ms after the onset of the target word were classed as timeouts. Participants initially completed a ten-item familiarisation block, which was followed by the experimental block of 32 items. During the practice block, participants were informed whether or not their response was correct immediately after each item. No indication of accuracy was given during the experimental block.

III. RESULTS

A. Statistical analysis

Statistical analyses were carried out in *R* (R Core Team, 2018) with alpha = 5%. Each participant's RTs were fitted to a Gamma distribution using the R library *fitdistrplus* (Delignette-Muller and Dutang, 2015). RTs shorter than 250 ms or slower than the 95th percentile of each participant's Gamma distribution were deleted (the mean 95th percentile was 936 ms, so a typical participant's RTs would lie in the range from 200 to 936 ms). Similarly, incorrect answers and timeouts (i.e., RTs greater than 2000 ms) were also deleted prior to analysis. Data from participants who failed to reach a 75% accuracy rate was deleted from the analysis. Overall, 89.6% of data were retained. Data were analysed with a Generalized Linear Mixed model (GLMM) (model fitted using the glmer function from the R library *lme4*, Bates *et al.* (2015). GLMM analysis of RT data is discussed in Lo and Andrews (2015). Congruence, ΔHarmonicity and ΔRoughness were included as fixed factors; participants were treated as random factors. The model used a Gamma family with an inverse link function.[2] To test the appropriateness of the Gamma distribution, each

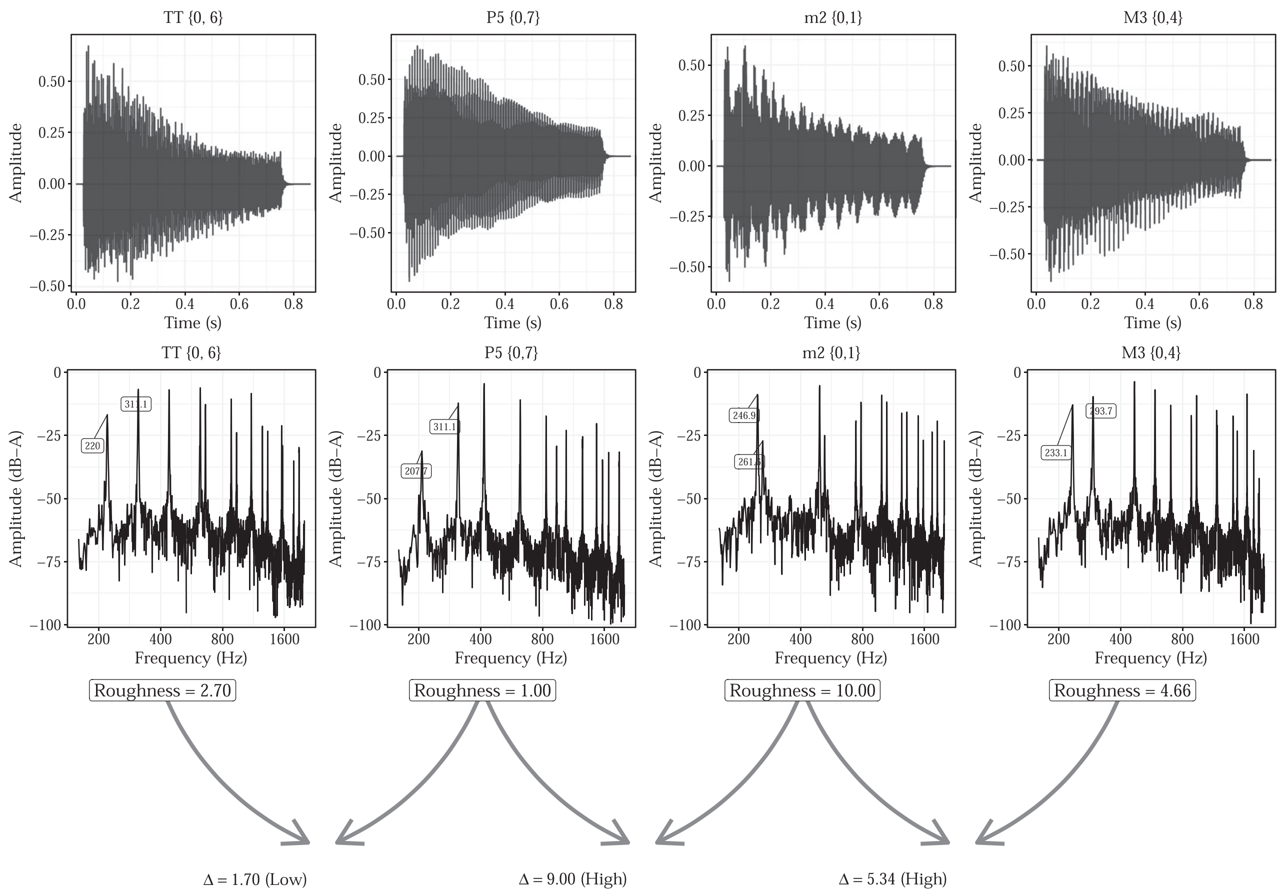

FIG. 2. Amplitudes and Spectra for sample Stimuli.

individual particpant's RT distribution was tested for goodness of fit against the Gamma distribution previously calculated (see above) *via* a Kolmogorov-Smirnov test. Of the 382 Kolmogorov-Smirnov tests, 11 returned significant results, $p < 0.05$ in each case. Finally, we carried out a binomial test ($r = 12$, $n = 379$, $p = 0.05$), which proved non-significant, $p = 0.98$, suggesting that significant Kolmogorov- Smirnov tests had occurred at no more frequently than chance levels and that the gamma function provides an adequate representation of the distributions. The structure of the GLMM was $RT \sim \Delta Harmonicity \times Congruence + \Delta Roughness \times Congruence + 1|Participant$. The effects of this model were then subject to a type III analysis of variance (ANOVA). The *p*-values in the subsequent planned contrast analysis were subject to Bonferroni corrections for multiple comparisons. Standardised effect sizes are not reported for the GLMM (Baguley, 2009).

B. Relationship between priming and stimulus features

Stimuli were grouped according to their acoustic properties: High vs Low difference in acoustic roughness (ΔRoughness), and High vs Low difference in harmonicity (ΔHarmonicity).[3] Response times in the High vs Low ΔRoughness difference conditions and the High vs Low ΔHarmonicity conditions were compared alongside the overall congruency effect (Congruent vs Incongruent conditions). As predicted, congruency effects were present in the High ΔRoughness condition. The key interaction ΔRoughness × Congruence proved significant, $\chi^2(1) = 5.67$, $p = 0.02$. The main effect of ΔHarmonicity was non-significant $\chi^2(1) = 0.54, p = 0.46$ as was the interaction ΔHarmonicity × Congruence, $\chi^2(1) = 0.35, p = 0.55$. Owing to the presence of the interaction, we do not report on the main effects of ΔRoughness or Congruence. Planned contrasts were carried out for the four ΔHarmonicity ×Δ Roughness conditions, with results as follows: The High ΔHarmonicity × High ΔRoughness condition showed significant congruency effects (Incongruent: mean = 671 ms, SD = 170; Congruent: mean = 664 ms, SD = 166), $z(\infty) = 2.72$, $p = 0.01$. Congruency effects were also present in the Low ΔHarmonicity × High ΔRoughness condition (Incongruent: mean = 684 ms, SD = 155; Congruent: mean = 670 ms, SD = 151), $z(\infty) = 3.11, p = 0.004$. However, the High ΔHarmonicity × Low ΔRoughness did not yield significant congruency effects, $z(\infty) = 0.492, p = 1.00$, (Incongruent: mean = 690 ms, SD = 184; Congruent: mean = 687 ms, SD = 169). Additionally, no congruency effects were present in the Low ΔHarmonicity × Low ΔRoughness condition, (Incongruent: mean = 683 ms, SD = 169; Congruent: mean = 682 ms, SD = 175), $z(\infty) = 0.920$, $p = 0.920$. Mean response times per Congruency, ΔRoughness and ΔHarmonicity are shown in Fig. 3. Within-subjects Wilcoxon tests on the congruency effects for individual prime pairs are reported in the supplementary material.

C. Roughness manipulation

To probe the role of roughness further, we introduced two artificial intervals that were designed to test the influence of more extreme differences in roughness, d2 and S2 (detuned minor seconds played with piano and Shepard tone timbres). Two further sub-experiments were carried out employing the same procedure but using these artificial stimuli as primes, with the expectation that the increase in ΔRoughness would, on average, increase the difference in response times between the congruent and incongruent conditions. As the key measure of congruency effects is the difference in response times, we carried out a simple linear correlation test on difference in roughness vs the difference in response time between the Congruent and Incongruent conditions (referred to from hereon as *priming index* for brevity) for the expanded data set of ten interval pairs. The correlation was statistically significant, $r = 0.12, t(377) = 2.63, p = 0.02$. However, the low r value and visual inspection of Fig. 3 suggested that, rather than a linear relationship, there is a step-like relationship—i.e., the increase in ΔRoughness is *not* associated with an additional increase in the priming index. To test this, the interval pairs were split into three categories by difference in roughness: Low (m3/M3, m6/M6, M7/P5, TT/P5), High (m2/P5, m2/TT, m2/M3, M2/P5), and Extreme (made up of the manipulated intervals d2/P5 and s2/S5). Owing to the clear *a priori* competing hypotheses (linear vs step-like dependence on ΔRoughness), we used a planned contrasts approach, where the dependent variable was the priming index. The planned comparisons were between the Low ΔRoughness and the combined High and Extreme ΔRoughness groups, and finally between the separate High and Extreme ΔRoughness groups.

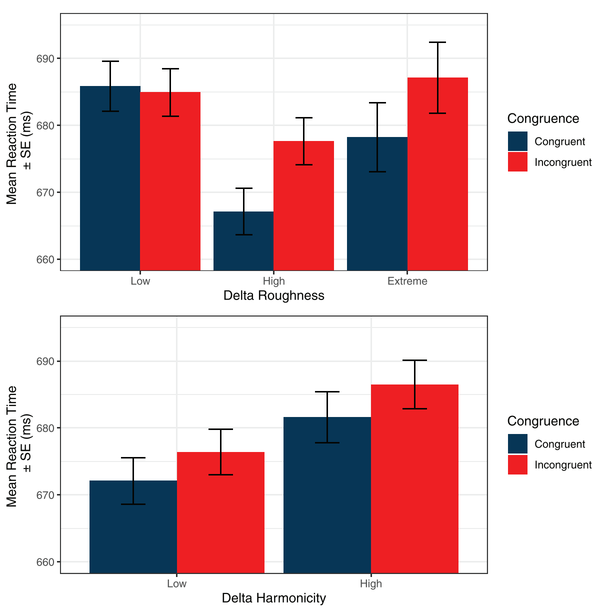

FIG. 3. (Color online) Mean RTs for Low, High, and Extreme levels of ΔRoughness and Low and High levels of ΔHarmonicity.

For the first contrast, there was a statistically significant difference in the size of the priming index between the Low (mean index = –2.21 ms, SD = 95.9 ms) and the combined High and Extreme groups (mean index = 22.4 ms, SD = 83.5), $t(376) = 2.38, p = 0.04$. However, there was no significant difference in priming index between the High (mean = 10.9 ms, SD = 42.3) and Extreme (mean = 11.3 ms, SD = 40.9) ΔRoughness groups, $t(376) = 0.07, p = 1.00$, suggesting that increasing the level of roughness does not increase the strength of the automatic response, supporting the hypothesis that the relationship between ΔRoughness and priming index is step-like rather than linear—that is, beyond a certain threshold, a further increase in the roughness contrast is not associated with a further increase in the priming index.

IV. DISCUSSION

We have demonstrated that interval pairs that differ by only a small amount in terms of acoustic roughness do not influence responses on a word classification task. On the other hand, intervals that differ more significantly in roughness do influence the responses. Notably, out of the two acoustic components that are seen as prevalent in the perception of consonance/dissonance, namely, roughness and harmonicity (see Harrison and Pearce, 2020; Parncutt and Hair, 2011), roughness was the only component that influenced the responses. Harmonicity failed to reach significance and, consequently, it seems likely that harmonicity contributes less to congruency effects in this setting than roughness—and the question of whether it contributes at all remains unresolved. Interestingly, the pairs m3/M3 and m6/M6 failed to influence responses to the classification task. This suggests that these interval pairs are not dissimilar enough in terms of roughness to elicit priming effects when presented by themselves, which is striking in the light that minor/major triads have been shown to be effective primes for negative/positive words in an affective priming setting (Steinbeis and Koelsch, 2011). Curiously, when used as the "negative" prime in the pairing tt/P5, the tritone was not congruent with negative words, and was, in fact, an effective prime for positive words when paired with the maximally rough minor second interval due to its relatively low roughness (see Table I). Given that the tritone carries conventionally a negative connotation in Western music (see, e.g., Costa *et al.*, 2000) and has historically been described as "diabolic" (Partch, 1974), it seems important that further research should probe the relationship between cultural convention and acoustic components and how they could differentially influence automatic and appraisal judgements about consonance and dissonance. See Fig. 4.

Although several contemporary models of consonance and dissonance attempt to integrate the concepts of roughness, harmonicity, and other non-acoustic factors (see, e.g.,

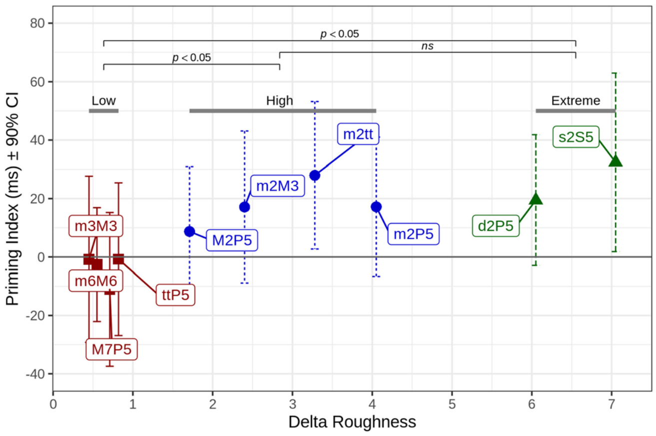

FIG. 4. (Color online) Difference in roughness vs priming index for diatonic intervals. Three groups of pairings (Low, High, and Extreme Δ Roughness) highlighting the contrasts.

Harrison and Pearce, 2020; Parncutt and Hair, 2011), the present study suggests that in this context of looking exclusively at musical intervals in an affective priming setting, roughness is the most important acoustic variable. Critically, the intervals that are associated with increased priming index, m2 and M2, fall within the critical bandwidth (see Patterson, 1976). The comparison of the minor and major thirds provides a particularly interesting case. The minor third falls theoretically just within the ERB. However, the priming index for the m3/M3 pairing did not differ significantly from zero. To probe this further, we considered whether the partials of the two notes in the interval co-occurred in the basilar membrane within the same ERB (see Table I). This suggests that although the fundamentals lie within the same ERB for m3, the number of overtones lying within the same ERBs (18) is considerably less than is the case for m2 (23) and M2 (25), creating a much less rough effect overall. It should be noted that this approach does not account for relative amplitudes of the partials and that the ERB formulation is an approximation—rather than a binary distinction, the degree of activation falls away more gradually. Moreover, it is not clear exactly how the summation of these interactions occurs. Nevertheless, it does provide a tentative explanation of why an automatic response was detected for other intervals in the ERB approximation (m2 and M2) but not for m3.

If we consider only the "rough" intervals (i.e., the diatonic intervals m2, M2, and the artificial intervals d2 and s2), we did not detect a difference in the size of the congruency effects between interval pairs where the difference in roughness was high compared to the extreme. This suggests that some qualitative difference exists between intervals that fall either within or outside this specific degree of ERB activation, but that beyond this threshold, the degree of roughness does not influence the priming index. Indeed, the results of the present experiment are consistent with the suggestion by Scharf (1970) that "listeners react one way when the stimuli are wider than the critical band and another way when the stimuli are narrower" (p. 196): intervals such as m2 and M2 create a significant overlap in critical bands in both fundamentals and higher partials.

An important question is why it is the contrast in roughness, rather than harmonicity, that has a greater influence on the size of the priming index? Compared to harmonicity, roughness is uniquely situated in being associated with, for example, alarm signals, and is thought to convey an advantage in enacting automatic responses (Arnal *et al.*, 2015). One speculative explanation is that it is not dissonance *per se* that is pre-activating negative concepts. The human auditory system is well attuned to human vocalisations as they carry acoustic information about for example bodily states (see Pouw *et al.*, 2020). Rough sounds are particularly salient and are associated with activation across large areas of the cortex (Arnal *et al.*, 2019). It may also be that the salience of roughness, rather than its contribution to consonance or dissonance, is responsible for the priming effects. However, both in producing and attending to rough sounds, for instance, cries of infants Koutseff *et al.* (2018) or angry voices (Bänziger *et al.*, 2015), sensitivity to roughness might confer an evolutionary advantage. This advantage explains why roughness in particular and not harmonicity is associated with an increase in the priming index, i.e., a stronger behavioural response to the valenced stimuli. It is plausible that the automatic response present in the priming index is driven by biological adaptation to acoustic factors: acoustically rough intervals activate responses associated with acoustically rough human vocalisations such as growls, screams, or cries. Indeed, recent research argues that responses to roughness in music have been exploited in the context of film music (Trevor *et al.*, 2020). Notably, in the case of the major second interval, this seems to happen quite subconsciously, if we go by the notion of composer/theorist Paul Hindemith who proposes that the major second sounds "almost consonant to our ears" (Hindemith, 1942, p. 85). On an empirical note, the major second interval has indeed been found to be perceived as more consonant than the minor seventh and major seventh intervals (Bowling *et al.*, 2018).

Although the present study provides a behavioural method for tapping the roughness construct, it should be noted that at this stage, the size of the effect is relatively small. Moreover, it is useful in indicating contrasts in roughness rather than as a direct measure of roughness. Nevertheless, roughness contrast is an objective method that taps into automatic perception as opposed to aesthetic judgements and it might mitigate semantic confounds which have been problematic for consonance/dissonance research in general as well as for specifically cross-cultural research endeavours into the question (see Bowling *et al.*, 2017; Lahdelma and Eerola, 2020). Such an objective new method can help to investigate the appreciation of dissonance across musical cultures.

The present result provides, to some extent, an explanation of the observation that, in the absence of previous exposure to Western diatonic harmony, there is no preference for consonant intervals over dissonant intervals, although there is a small preference for large over small intervals

(McDermott *et al.*, 2016). The results of the present study suggest that the binary division is not whether an interval is categorised as consonant or dissonant, but on how much overlap there is in ERBs between the various partials—which is potentially higher in the case of for instance m2 or M2 than M7. The present study offers an explanation of automatic responses to acoustically rough intervals based on biological imperatives to respond to alarm signals present in human vocalisation. However, it should be noted that the emphasis in the present study is on *automatic responses*. Consonance and dissonance more broadly (e.g., cognitive or aesthetic appraisal) may well be underpinned by harmonicity or factors not considered in the present study such as cultural conventions. Future research should consider whether this biological imperative underpins dissonance perception more broadly or whether this sort of priming paradigm presents a special case. The role of culture in this effect also warrants further investigation, in particular, whether it can be replicated with participants who have had frequent exposure to, for instance, *beat diaphony* in musical cultures that promote roughness for its aesthetic value (Ambrazevičius, 2017; Messner, 1981). Additionally, future research could test the viability of exploiting acoustic roughness in alarm signals, for instance, by using the major 2nd interval, which provokes an automatic response in this priming context yet is not as unpleasant as other more extreme sounds.

ACKNOWLEDGMENTS

Partial financial support was provided by the Faculty Pro-Vice Chancellor's award at Durham University. Additional funds were provided by the Society for Education, Music and Psychology Research (SEMPRE) in the form of a Reg and Molly Buck Award awarded to the second author.

[1]Owing to the Covid-19 pandemic, we were unable to collect a confirmatory lab sample. However, we collected a small more tightly controlled web sample of 12 participants - all using MacOS, Chrome/Firefox browser and all of whom passed a headphone check prior to completing the experiment. The results mirrored those of the main experiment and are reported in the SI.[4]

[2]The inverse link function was chosen as it offered the lowest AIC compared to logarithmic or identity functions.

[3]Wang *et al*. (2013) and Harrison and Pearce (2018).

[4]Supporting information (different roughness and harmonicity models, Wilcoxon tests for congruency effects per interval pair, and additional visualisations of spectra of stimuli) are publicly available, along with the RT data and analysis scripts at https://tuomaseerola.github.io/primingroughnessdata/. The PsyToolkit code and stimuli are available at https://osf.io/zmjpd/.

Ambrazevičius, R. (**2017**). "Dissonance/roughness and tonality perception in Lithuanian traditional Schwebungsdiaphonie," J. Interdiscipl. Music Stud. **8**, 39–53.

Armitage, J., and Eerola, T. (**2019**). "Reaction time data in music cognition: Comparison of pilot data from lab, crowdsourced, and convenience web samples," Front. Psychol. **10**, 2883.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., and Poeppel, D. (**2015**). "Human screams occupy a privileged niche in the communication soundscape," Curr. Biol. **25**(15), 2051–2056.

Arnal, L. H., Kleinschmidt, A., Spinelli, L., Giraud, A.-L., and Mégevand, P. (**2019**). "The rough sound of salience enhances aversion through neural synchronisation," Nat. Commun. **10**(1), 3671.

Baguley, T. (**2009**). "Standardized or simple effect size: What should be reported?," Br. J. Psychol. **100**(3), 603–617.

Bakker, D. R., and Martin, F. H. (**2015**). "Musical chords and emotion: Major and minor triads are processed for emotion," Cogn. Affective Behav. Neurosci. **15**(1), 15–31.

Bänziger, T., Hosoya, G., and Scherer, K. R. (**2015**). "Path models of vocal emotion communication," PLoS One **10**(9), e0136675.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Bowling, D. L., Hoeschele, M., Gill, K. Z., and Fitch, W. T. (**2017**). "The nature and nurture of musical consonance," Music Percept. **35**(1), 118–121.

Bowling, D. L., and Purves, D. (**2015**). "A biological rationale for musical consonance," Proc. Natl. Acad. Sci. **112**(36), 11155–11160.

Bowling, D. L., Purves, D., and Gill, K. Z. (**2018**). "Vocal similarity predicts the relative attraction of musical chords," Proc. Natl. Acad. Sci. **115**(1), 216–221.

Carcagno, S., Lakhani, S., and Plack, C. J. (**2019**). "Consonance perception beyond the traditional existence region of pitch," J. Acoust. Soc. Am. **146**(4), 2279–2290.

Costa, M. (**2013**). "Effects of mode, consonance, and register in visual and word-evaluation affective priming experiments," Psychol. Music **41**(6), 713–728.

Costa, M., Ricci Bitti, P. E., and Bonfiglioli, L. (**2000**). "Psychological connotations of harmonic musical intervals," Psychol. Music **28**(1), 4–22.

Cousineau, M., McDermott, J. H., and Peretz, I. (**2012**). "The basis of musical consonance as revealed by congenital amusia," Proc. Natl. Acad. Sci. **109**(48), 19858–19863.

Delignette-Muller, M. L., and Dutang, C. (**2015**). "fitdistrplus: An R package for fitting distributions," J. Stat. Softw. **64**(4), 1–34.

Fazio, R. H., and Olson, M. A. (**2003**). "Implicit measures in social cognition research: Their meaning and use," Ann. Rev. Psychol. **54**(1), 297–327.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," Bell Syst. Technol. J. **12**(4), 377–430.

Goerlich, K. S., Witteman, J., Aleman, A., and Martens, S. (**2011**). "Hearing feelings: Affective categorization of music and speech in alexithymia, an ERP study," PLoS One **6**(5), e19501.

Harrison, P., and Pearce, M. (**2018**). "An energy-based generative sequence model for testing sensory theories of Western harmony," in *Proceedings of the 19th International Society for Music Information Retrieval Conference*, September 23–27, Paris, France, pp. 160–167.

Harrison, P., and Pearce, M. (**2020**). "Simultaneous consonance in music perception and composition," Psychol. Rev. **127**(2), 216–244.

Helmholtz, H. (**1885**). *On the Sensations of Tone 1954*, translated by A. J. Ellis (Dover, Mineola, NY).

Herring, D. R., White, K. R., Jabeen, L. N., Hinojos, M., Terrazas, G., Reyes, S. M., Taylor, J. H., and Crites, S. L., Jr. (**2013**). "On the automatic activation of attitudes: A quarter century of evaluative priming research," Psychol. Bull. **139**(5), 1062–1089.

Hindemith, P. (**1942**). *The Craft of Musical Composition*, Vol. 1 (Belwin-Mills, New York).

Hutchinson, W., and Knopoff, L. (**1978**). "The acoustic component of Western consonance," Interface **7**(1), 1–29.

Hutchinson, W., and Knopoff, L. (**1979**). "The significance of the acoustic component of consonance in Western triads," J. Musicol. Res. **3**(1–2), 5–22.

Koelsch, S., Skouras, S., and Lohmann, G. (**2018**). "The auditory cortex hosts network nodes influential for emotion processing: An FMRI study on music-evoked fear and joy," PLoS One **13**(1), e0190057.

Koutseff, A., Reby, D., Martin, O., Levrero, F., Patural, H., and Mathevon, N. (**2018**). "The acoustic space of pain: Cries as indicators of distress recovering dynamics in pre-verbal infants," Bioacoustics **27**(4), 313–325.

Lahdelma, I., Armitage, J., and Eerola, T. (**2020**). "Affective priming with musical chords is influenced by pitch numerosity," Musicae Scientiae 102986492091112.

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

Lo, S., and Andrews, S. (**2015**). "To transform or not to transform: Using generalized linear mixed models to analyse reaction time data," Front. Psychol. **6**, 1171.
McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). "Individual differences reveal the basis of consonance," Curr. Biol. **20**(11), 1035–1041.
McDermott, J. H., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (**2016**). "Indifference to dissonance in native Amazonians reveals cultural variation in music perception," Nature **535**(7613), 547–550.
Messner, G. F. (**1981**). "The two-part vocal style on Baluan Island Manus province, Papua New Guinea," Ethnomusicology **25**(3), 433–446.
Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**(3), 750–753.
Nordmark, J., and Fahlén, L. E. (**1988**). "Beat theories of musical consonance," Q. Prog. Status Rep. **29**, 111–122.
Parncutt, R., and Hair, G. (**2011**). "Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies," J. Interdiscip. Music Stud. **5**(2), 119–166.
Partch, H. (**1974**). *Genesis of a Music* (Da Capo Press, New York).
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**(3), 640–654.
Pouw, W., Paxton, A., Harrison, S. J., and Dixon, J. A. (**2020**). "Acoustic information about upper limb movement in voicing," Proc. Natl. Acad. Sci. **117**(21), 11364–11367.
R Core Team (**2018**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
Roederer, J. G. (**2008**). *The Physics and Psychophysics of Music: An Introduction* (Springer Science & Business Media, New York).
Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Chap. 5, pp. 157–202.
Scherer, L. D., and Larsen, R. J. (**2011**). "Cross-modal evaluative priming: Emotional sounds influence the processing of emotion words," Emotion **11**(1), 203–208.
Schwartz, J. W., Engelberg, J. W., and Gouzoules, H. (**2019**). "Was that a scream? Listener agreement and major distinguishing acoustic features," J. Nonverbal Behav. **44**, 233–252.
Smith, J. O., and Abel, J. S. (**1999**). "Bark and ERB bilinear transforms," IEEE Trans. Speech Audio Process. **7**(6), 697–708.
Sollberger, B., Rebe, R., and Eckstein, D. (**2003**). "Musical chords as affective priming context in a word-evaluation task," Music Percept. **20**(3), 263–282.
Steinbeis, N., and Koelsch, S. (**2011**). "Affective priming effects of musical sounds on the processing of word meaning," J. Cogn. Neurosci. **23**(3), 604–621.
Stoet, G. (**2017**). "PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments," Teaching Psychol. **44**(1), 24–31.
Terhardt, E. (**1984**). "The concept of musical consonance: A link between music and psychoacoustics," Music Percept. **1**(3), 276–295.
Tramo, M. J., Cariani, P. A., Delgutte, B., and Braida, L. D. (**2001**). "Neurobiological foundations for the theory of harmony in Western tonal music," Ann. NY Acad. Sci. **930**(1), 92–116.
Trevor, C., Arnal, L. H., and Frühholz, S. (**2020**). "Terrifying film music mimics alarming acoustic feature of human screams," J. Acoust. Soc. Am. **147**(6), EL540–EL545.
Vassilakis, P. N. (**2001**). "Perceptual and physical properties of amplitude fluctuation and their musical significance," Ph.D. thesis, University of California, Los Angeles, Los Angeles, CA.
Wang, Y., Shen, G., Guo, H., Tang, X., and Hamade, T. (**2013**). "Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise," J. Sound Vib. **332**(16), 3893–3904.
Warriner, A. B., Kuperman, V., and Brysbaert, M. (**2013**). "Norms of valence, arousal, and dominance for 13,915 English lemmas," Behav. Res. Methods **45**(4), 1191–1207.