The Journal of the Acoustical Society of America

Consonance perception beyond the traditional existence region of pitch

--Manuscript Draft--

Manuscript Number:	JASA-04437R1
Full Title:	Consonance perception beyond the traditional existence region of pitch
Short Title:	Consonance at high frequencies
Article Type:	Regular Article
Corresponding Author:	Samuele Carcagno, Ph.D. Lancaster University Lancaster, Lancashire UNITED KINGDOM
First Author:	Samuele Carcagno, Ph.D.
Order of Authors:	Samuele Carcagno, Ph.D.
	Saday Lakhani
	Christopher John Plack, PhD
Section/Category:	Psychological and Physiological Acoustics
Keywords:	consonance perception; temporal coding; rate-place coding; neural phase locking
Abstract:	Some theories posit that the perception of consonance is based on neural periodicity detection, which is dependent on accurate phase locking of auditory nerve fibers to features of the stimulus waveform. In the current study, 15 listeners were asked to rate the pleasantness of complex tone dyads (two note chords) forming various harmonic intervals, and bandpass filtered in a high frequency region (all components > 5.8 kHz), where phase locking to the rapid stimulus fine structure is thought to be severely degraded or absent. The two notes were presented to opposite ears. Consonant intervals (minor third, and perfect fifth) received higher ratings than dissonant intervals (minor second, and tritone). The results could not be explained in terms of phase locking to the slower waveform envelope, because the preference for consonant intervals was higher when the stimuli were harmonic, compared to a condition in which they were made inharmonic by shifting their component frequencies by a constant offset, so as to preserve their envelope periodicity. Overall the results indicate that, if phase locking is indeed absent at frequencies greater than ~5 kHz, neural periodicity detection is not necessary for the perception of consonance.

Consonance perception beyond the traditional existence region of pitch

Samuele Carcagno,[1, a)] Saday Lakhani,[1] and Christopher J. Plack[1, b]

Department of Psychology, Lancaster University, Lancaster, LA1 4YF, United Kingdom

Some theories posit that the perception of consonance is based on neural periodicity detection, which is dependent on accurate phase locking of auditory nerve fibers to features of the stimulus waveform. In the current study, 15 listeners were asked to rate the pleasantness of complex tone dyads (two note chords) forming various harmonic intervals, and bandpass filtered in a high frequency region (all components > 5.8 kHz), where phase locking to the rapid stimulus fine structure is thought to be severely degraded or absent. The two notes were presented to opposite ears. Consonant intervals (minor third, and perfect fifth) received higher ratings than dissonant intervals (minor second, and tritone). The results could not be explained in terms of phase locking to the slower waveform envelope, because the preference for consonant intervals was higher when the stimuli were harmonic, compared to a condition in which they were made inharmonic by shifting their component frequencies by a constant offset, so as to preserve their envelope periodicity. Overall the results indicate that, if phase locking is indeed absent at frequencies greater than ~ 5 kHz, neural periodicity detection is not necessary for the perception of consonance.

[a)] sam.carcagno@gmail.com

[b] Also at: Manchester Centre for Audiology and Deafness, University of Manchester, Manchester Academic Health Science Centre, M13 9PL, United Kingdom

I. INTRODUCTION

In Western music certain harmonic intervals such as the perfect fifth and the perfect fourth are regarded as consonant, and are described as producing a pleasant and stable sound sensation; other harmonic intervals, such as the minor second and the tritone, are regarded as dissonant, and are described as producing an unpleasant and tense sound sensation. The origins of this distinction between consonant and dissonant intervals have been debated for centuries. The earliest theory of consonance is often attributed to the Greek mathematician Pythagoras (Bowling and Purves, 2015), who considered as consonant those musical intervals whose frequencies formed "simple" ratios between small integers (e.g. 2:1, 3:2, 4:3).

In the last two centuries the debate has focused on the possible physiological mechanisms leading to the sensation of consonance. One of the major psychoacoustical theories of consonance posits that the sensation of dissonance is directly related to the sensation of "roughness" caused by the amplitude fluctuations, also known as "beats", produced when the frequencies of two tones are close enough as to interact within the same cochlear filter (Helmholtz, 1954; Kameoka and Kuriyagawa, 1969a; Plomp and Levelt, 1965). There are several pieces of evidence against this theory. Cochlear interactions cannot explain the fact that the sensation of dissonance persists when tones are presented dichotically to prevent cochlear interactions (Bidelman and Krishnan, 2009; Terhardt, 1974b). However, beats could also occur centrally, within a binaural critical band (Feeney, 1997), rather than being based on cochlear interactions. Additionally, it could be that the negative affect of dissonance is transferred by associative learning from naturally occurring conditions in which cochlear interactions are present, to artificial conditions in which these are eliminated by dichotic presentation.

Additional evidence against the idea that dissonance is caused by amplitude beats comes from the fact that a) intervals such as the perfect fifth are still considered consonant even when presented at low fundamental frequencies (F0s), where the notes produce considerable roughness (Terhardt, 1974b); b) dissonance does not grow with increasing number of harmonics in chords, which increases the sources of amplitude beats (McLachlan *et al.*, 2013); c) interindividual differences in preference for consonant over dissonant musical intervals correlate with preference for harmonic stimuli, rather than with preference for stimuli lacking beats (McDermott *et al.*, 2010); d) people with amusia show an aversion to amplitude beats similar to that of controls, but do not show a preference for consonant over dissonant intervals (Cousineau *et al.*, 2012).

Another major psychoacoustical theory of consonance holds that consonance is based on harmonicity: Simultaneously presented tones are perceived as more or less consonant depending on how well their frequency components match a single harmonic series (Terhardt, 1974b). For example, the frequency components of tones forming musical intervals such as the perfect fifth and the perfect fourth, that are generally perceived as highly consonant, fall closely into a single harmonic series, while the frequency components of tones such as the minor second and the tritone, which are generally perceived as dissonant, do not.

Harmonicity plays a key role in the perception of pitch (Plack and Oxenham, 2005), as well as in the segregation of concurrent sounds (Darwin, 2005), but the way harmonicity is encoded in the auditory system is unclear. One major theory holds that harmonicity is encoded by the detection of neural periodicities arising as a result of the phase locking of auditory nerve fibers to periodicities present in the stimuli (Meddis and O'Mard, 2006). Mathematical (Ebeling, 2008), functional (Patterson, 1986) and physiological (Bidelman and Heinz, 2011) models of consonance have been proposed on the basis of this theory. These models are supported by single-fiber recordings in non-human animals showing that temporal information for the perception of consonance is available at the level of the auditory nerve (Tramo *et al.*, 2001). Additionally, several studies in humans have shown that consonance ratings are related to measures of harmonicity derived from the scalp-recorded frequency following response (FFR) (Bidelman and Krishnan, 2009; Bones *et al.*, 2014; Bones and Plack, 2015a,b), a response that reflects neural phase locking in the brainstem.

Phase locking in the auditory nerve, which is the basis of the "temporal" models of consonance mentioned above, declines progressively with increasing frequency, and for most species studied becomes undetectable above about 5 kHz (Johnson, 1980; Palmer and Russell, 1986; Winter, 2005). The upper limit of phase locking in humans has been estimated from recordings of the auditory nerve compound action potentials to be at best similar to, and likely worse than, this 5 kHz limit (Verschooten *et al.*, 2018). Therefore, "temporal" models of consonance predict that the perception of consonance should break down for stimuli presented above 5 kHz. This frequency was long held to be the upper limit for the perception of pitch, another sensory attribute which is crucial for music and which has also been explained on the basis of neural temporal models. The evidence for such an upper limit came from several pieces of data, including the fact that a) the ability to perceive pure-tone melodies (Attneave and Olson, 1971), or melodic intervals (Semal and Demany,

1990) breaks down above 5 kHz, b) pure-tone frequency discrimination declines dramatically above 5 kHz (Moore, 1973), and c) the upper note of most musical instruments lies below 5 kHz. However, Oxenham *et al.* (2011), while confirming that the ability to perceive melodies is severely degraded for pure tones above 6 kHz, found that for complex tones whose harmonics all fell above 6 kHz performance in a melody discrimination task was comparable to that obtained with low-frequency pure tones. A follow-up study found that F0 difference limens (F0DLs) for complex tones with all harmonics falling above 8 kHz, although worse than F0DLs for complex tones presented in a low-frequency region, were considerably better than frequency difference limens (FDLs) for pure tones above 8 kHz (Lau *et al.*, 2017). F0DLs for the high-frequency complex tones were also considerably lower than predicted by the optimal integration of information based on FDL performance, suggesting that the poor FDLs for high-frequency pure tones are not due to peripheral coding constraints related to decreased phase locking at high frequencies.

It has also been argued that consonance does not directly depend on low-level physiological or psychoacoustical invariants such has cochlear beats or regularity of neural firing, but is instead the result of learned cultural conventions. Some support for this "cultural" theory of consonance comes from the observation that in music theory categorizations of consonant and dissonant intervals have changed over the centuries (Tenney, 1988). Further evidence for the cultural theory of consonance comes from the finding of McDermott *et al.* (2016) that an isolated Amazonian tribe with limited exposure to Western music did not show preference for consonant over dissonant musical intervals. There are, however, arguments against a purely cultural origin of consonance preference (reviewed in Bowling *et al.*, 2017), such as the striking similarity of tonal structures across musical cultures from different geographical regions, and different epochs, which suggests that these structures are partly shaped by biological constraints. The debate on the relative role of biological vs cultural factors in the determination of consonance preference remains open.

Models of consonance based on harmonicity are closely linked to models of pitch perception (Terhardt, 1974b) because both are often based on the concept of an F0 to explain the relations between the frequency components present in a stimulus. Cultural theories of consonance are also linked to pitch perception because they posit that consonance is a culturally learned preference for certain pitch combinations. If pitch perception is possible for stimuli with harmonics above 5 kHz, as suggested by the results of Oxenham *et al.* (2011), and Lau *et al.* (2017), it is reasonable to hypothesize that the perception of consonance may also be possible above this frequency limit. The present study sought to test this hypothesis by measuring the pleasantness ratings of 15 listeners for dyads (two note chords) with frequency components falling entirely above 5 kHz, and forming musical intervals traditionally considered consonant, or dissonant. To ensure that pleasantness ratings were not based on low-frequency envelope periodicities we used a manipulation similar to the one employed by Oxenham *et al.* (2011): Ratings for harmonic dyads were compared to those for dyads whose components had been shifted by a fixed frequency offset, so as to preserve their envelope periodicity while disrupting their harmonicity. The results were also compared to pleasantness ratings for stimuli with the same root note, but with frequency components below 5 kHz.

An additional experiment measured the ability of the same listeners to discriminate musical melodies composed of complex tones with frequency components above 5 kHz. The main purpose of this experiment was to rule out the possibility that, if consonance perception were to be found poor or absent in the high frequency region, this was due to an inability of our listeners to perceive melodic pitch above 5 kHz.

It seems reasonable to hypothesize that if melodic pitch perception is present at high frequencies, consonance perception should be present too. However, the results of a study by Gockel and Carlyon (2018) suggest that different aspects of pitch processing may show unexpected dissociations at high frequencies. In a series of experiments they found that while F0 discrimination performance at high frequencies was good, and could not be accounted for by residual envelope cues, in line with previous results (Lau *et al.*, 2017; Oxenham *et al.*, 2011), mistuning detection at high frequencies was unexpectedly poor. Detecting a mistuning of the 8^{th} harmonic of a 1400 Hz F0 complex tone was only slightly above chance level even for a mistuning of $\sim 6\%$, and listeners did not report hearing the mistuned component as perceptually segregated from the complex. Gockel and Carlyon (2018) concluded that either harmonic templates at high frequencies have wider tolerances than those at low frequencies, or even though they have comparable tolerances the mechanism that leads to the perceptual segregation of a mistuned component is absent at high frequencies. In either case, these results suggest that it cannot be assumed that consonance perception at high frequencies will be present simply because melodic pitch perception for complex tones is present at these frequencies.

II. METHODS

A. Audiometric screening

Participants were screened for hearing loss by measuring their thresholds for the detection of a 200-ms pure tone in quiet at octave frequencies ranging from 0.25 to 8 kHz. Only participants with thresholds below 20 dB HL for both ears were included in the study. Additionally, participants were screened for their ability to hear a 300-ms (including 10-ms onset and offset raised-cosine ramps) 12-kHz pure tone in a background of 45 dB SPL/ERB threshold-equalizing noise (TEN) (Moore *et al.*, 2000) bandpass filtered between 0.02 and 16 kHz. Only participants with thresholds $\leq$ 50 dB SPL for both ears in this task were included in the study. Both the audiometric thresholds in quiet and the tone-in-noise detection at 12 kHz were measured using a two-interval two-alternative forced-choice task with a two-down one-up adaptive rule tracking the 70.7% correct point on the psychometric function (Levitt, 1971). The step size was 4 dB for the first four reversals, and 2 dB thereafter. For the audiometric thresholds in quiet the adaptive track terminated after eight reversals, and thresholds were estimated by averaging the values of the adaptive track at the last four reversals. For the tone-in-noise detection at 12 kHz the adaptive track was stopped after 14 reversals, and thresholds were estimated by averaging the values of the adaptive track at the last 10 reversals.

B. Participants

Twenty-five listeners in their 20s took part in the study. Fifteen listeners (eight males) passed the audiometric screening and proceeded to run the main experiments, while the 10 listeners who failed the audiometric screening were excluded from the study. Nine out of the 15 listeners who passed the audiometric screening were musicians with more than five years of practice with a musical instrument. All participants gave written informed consent for participation in the study, and the study protocols were approved by the Lancaster University Psychology Department Ethics Committee.

C. Pleasantness ratings

In the rating experiment, listeners were asked to rate the pleasantness of dyads consisting of a low ("root") note, and a high ("interval") note. Participants rated each dyad on a scale ranging from -3 to +3 in 0.1 steps by moving, through a computer mouse, a slider presented on a computer monitor (Bones *et al.*, 2014; Bones and Plack, 2015a,b; McDermott *et al.*, 2010). The notes composing the dyads were equal-amplitude complex tones and were presented each to a different ear to eliminate the possibility of cochlear interactions between components of the root and interval notes, which can lead to amplitude fluctuations and perceived "roughness" (Terhardt, 1984). The dyads were bandpass filtered so that their components would fall either in a "low" frequency region, or in a "high" frequency region above the traditional existence region of pitch. The complex tones composing the dyads were either harmonic, or were made inharmonic by shifting all their components by a fixed frequency offset in hertz. In the "harmonic" conditions the root note of each dyad had an F0 of 1174.659 Hz (D6 note in the equal temperament scale). The F0s of the interval notes were 100 cents (minor second), 300 cents (minor third), 600 cents (tritone), or 700 cents (perfect fifth) above the root note (100 cents = 1 semitone), so as to form musical intervals of the equal-tempered scale. The F0s of the interval notes are shown in Table I.

Two dyads formed musical intervals which are traditionally considered consonant: the minor third, and the perfect fifth. The other two dyads formed musical intervals which are traditionally considered dissonant: the minor second, and the tritone. Music theory classifications of intervals in terms of consonance and dissonance have evolved and changed in the course of the centuries (Tenney, 1988), and often distinctions are made in terms of their degree of consonance for intervals within each category. The perfect fifth is typically considered a "perfect consonance", while the minor third is often classified as an "imperfect consonance".

The stimuli for the inharmonic conditions were obtained by shifting each component of the complex tones forming the dyads in the harmonic conditions by 234.9318 Hz (20% of the root note F0). The spectra for the harmonic stimuli are shown in Fig. 1, those of the inharmonic stimuli are shown in Fig. S1 of the supplementary materials[1]. The frequency components of the dyads in the harmonic conditions are listed in Table S1, and those of the dyads in the inharmonic conditions in Table S2 of the supplementary materials.

In all experimental conditions the dyads had a 2-sec duration, including 10-ms raised-cosine onset and offset ramps. The complex tones forming the dyads had a level of 55 dB SPL per component; each component had a random starting phase. In the low-frequency conditions the dyads were bandpass filtered between 1 and 6 kHz, while in the high-frequency conditions they were bandpass filtered between 7 and 12 kHz, using a 256-taps finite-impulse-response filter (90 dB/octave slope). Keeping the F0 of the root note constant while filtering the stimuli within two different frequency regions with the same bandwidth leads to differences both in the harmonic rank and in the total number of harmonics present in each dyad. These differences are particularly marked between dyads presented in the low and in the high frequency region (harmonic ranks, and total number of harmonics are higher in the high-frequency region). Although these differences could presumably affect pleasantness ratings, they are unlikely to affect greatly the difference in pleasantness ratings between the consonant and dissonant dyads *within* each frequency region, which was the main variable of interest in the experiment.

All the dyads were presented in a background TEN bandpass filtered between 0.02 and 16 kHz, with a level of 45 dB SPL/ERB. An additional band of 55-dB SPL/ERB TEN bandpass filtered between 0.02 and 5 kHz was added to the dyads in the high-frequency conditions to ensure that low-frequency distortion products would be masked (Oxenham *et al.*, 2011). These noises were gated on and off simultaneously with the dyads with 10-ms raised-cosine onset and offset ramps. On each trial a 2-sec, 45-dB SPL/ERB TEN bandpass filtered between 0.02 and 16 kHz was presented before the presentation of the dyad to "weaken" the sensory memory trace of the dyad presented in the previous trial, so as to minimize any effect it might have on the judgment of the dyad in the current trial (Bones *et al.*, 2014; McDermott *et al.*, 2010). There was a 500-ms silent interval between the presentation of this noise and the onset of the dyad. All noise samples, including those presented together with the dyad and those presented before the dyad, were independent between the left and right ear.

There were 16 experimental conditions overall, resulting from the combination of four musical intervals (minor second, minor third, tritone, and perfect fifth), two frequency regions (low, and high), and two harmonicity conditions (harmonic, and inharmonic). Participants first rated two dyads for each condition. These practice trials were discarded from subsequent analyses. After the practice trials participants rated eight dyads for each experimental condition. For both the practice phase and the main phase the trials were organized in four blocks, corresponding to the four combinations of frequency region and harmonicity: only stimuli of a given frequency region and harmonicity were presented in a block of trials, and all of the different intervals were presented within each block. The presentation order of the blocks was randomized. Within each block the presentation order of the intervals was also randomized. For each interval, the root note was presented to the right ear on half the trials in each block, and to the left ear on the other half, in random order. The interval note was always presented to the opposite ear.

D. Melody discrimination task

This task was similar to the melody discrimination task of Oxenham *et al.* (2011). On each trial participants were presented with two four-note melodies. The first melody consisted of 45 dB SPL pure tones drawn from a set of notes from the diatonic scale (C6=1046.502 Hz, D6, E6, F6, G6, A6, B6, C7). On each trial the notes were drawn sequentially at random, with the constraint that if the first and second note, or if the second and third note were the same, that note could not be drawn again for that trial. This constraint implied that no three consecutive notes could be the same. The second melody consisted of harmonic complex tones bandpass filtered between 7 and 12 kHz. The harmonics of the complex tones were added in sine phase[2] and had a level of 55 dB SPL. Two bands of TEN were added to each note of the second melody to mask low-frequency combination tones and promote harmonic fusion. The first noise was bandpass filtered between 0.02 and 5 kHz, and had a level of 55 dB SPL/ERB. The second noise was bandpass filtered between 0.02 and 16 kHz, and had a level of 45 dB SPL/ERB. On "same" trials the notes of the second melody had the same F0s as the notes of the first melody. On "different" trials the F0 of the third note of the second melody was changed by a step up, or a step down in the diatonic scale with respect to the third note of the first melody, while the F0s of the other three notes were the same in the two melodies. Each note had a duration of 300 ms, including 10-ms raised-cosine ramps. The noise bands for the second melody were gated on and off simultaneously with each note with 10-ms raised-cosine onset and offset ramps. Within each melody the notes were separated by 200-ms silent intervals. A 500-ms silent interval separated the two melodies. Each melody was marked by a flashing light on a computer screen. Listeners had to indicate by means of a key press on a computer keyboard whether the two melodies were the same or different. Feedback was provided by means of a colored light at the end of each trial. Listeners first completed a block of 10 practice trials in which both the first and the second melody consisted of pure tones. They then completed two 100-trial blocks in which the first melody consisted of pure tones, and the second melody consisted of complex tones filtered in the 7–12 kHz frequency region.

E. Equipment

Testing took place in a double-walled, sound-insulated booth (IAC Acoustics, UK). The stimuli were generated digitally with a 32-bit resolution and a 48-kHz sampling rate in Python, on a GNU/Linux workstation housed outside the booth. The stimuli were sent to a 24-bit digital-to analog converter (E-MU 0204 USB), and played via Sennheiser HDA300 headphones. These headphones were chosen both because of their extended high-frequency response, and because being closed-cup headphones they minimize acoustic cross-talk which may have otherwise re-introduced cochlear interactions effects in spite of the dichotic presentation of the root and interval notes in the pleasantness rating task.

F. Statistical analyses

Analyses were performed using Bayesian models implemented by Markov Chain Monte Carlo (MCMC) simulations using JAGS (Plummer, 2003) and R (R Core Team, 2019). Bayesian analysis methods have several strengths, including the ability to seamlessly fit complex models without having to rely on assumptions of normally distributed residuals, the ability to quantify the uncertainty over parameters of interest without relying on sampling distributions (Kruschke, 2014), and the ability to keep false alarms at bay in multiple comparison settings without sapping statistical power, by means of hierarchical modeling (Gelman *et al.*, 2012). For all MCMC simulations the chains were monitored for convergence using trace plots. All chains were also monitored for autocorrelation to ensure an effective sample size of at least around 10,000 samples for the main parameters of interest.

The pleasantness ratings of each listener were converted to z scores by subtracting the mean and scaling by the standard deviation of the scores given by that listener across all stimulus conditions (McDermott *et al.*, 2010). These standardized pleasantness ratings were then modeled using a hierarchical Bayesian linear model that estimated the effect of interval, frequency region, harmonicity, and the two- and three-way interactions between these factors, at the level of individual listeners, as well as at the group level. The model is based on a model proposed by Kruschke (2010, chap. 19, p. 532) for the analysis of within-subject designs in which subjects provide more than one datum per condition.

The hit and false alarm rates obtained by each listener in the melody-discrimination task were modeled using a Bayesian hierarchical model based on the equations of Macmillan and Creelman (2004) to calculate d' in the same-different task for an observer using the differencing strategy. The model estimated d' both at the individual and at the group level. The model has the advantage of taking into account the uncertainty around the d' estimate for each listener when computing across-listener statistics, rather than relying on point estimates of d'. Another advantage of the model is that it does not require corrections for extreme sampled proportions (i.e. hits or false alarm rates of 0, or 1) that can bias d' estimates (Hautus and Lee, 1998).

Details of the models are given in the supplementary material. Effects were summarized by 95% credibility intervals (CIs) of the posterior distribution of the parameter of interest. These indicate that, according to the model, the parameter has a 95% probability of being enclosed between the bounds of the interval. For inferential purposes parameters were deemed as credibly different from zero when the bounds of their 95% CIs did not enclose zero.

III. RESULTS

A. Pleasantness ratings

The mean standardized pleasantness ratings are shown in Fig. 2. The ratings for the harmonic dyads in the low frequency region follow the pattern expected from the literature (Bones and Plack, 2015a; Kameoka and Kuriyagawa, 1969b; Malmberg, 1918; McDermott *et al.*, 2010; Schwartz *et al.*, 2003) with higher average ratings for the consonant over the dissonant intervals. In the inharmonic condition, while the average ratings of the minor third, tritone, and perfect fifth, appear relatively well matched, those of the minor second are lower than those of the other intervals. The dyads in the high-frequency region received generally lower ratings than in the low frequency region, but the pattern with respect to interval type and harmonicity is similar to that observed in the low frequency region.

Figure 3 shows the mean consonance preference scores, which were calculated by subtracting the standardized scores given to the two dissonant intervals (minor second and tritone) from the standardized scores given to the consonant ones (minor third and perfect fifth). Posterior distributions and 95% CIs for consonance preference scores estimated by the Bayesian model are shown in Fig. 4. The 95% CIs indicate that for harmonic stimuli, consonant intervals are rated higher than dissonant intervals in both the low (CI: 0.53 – 1.17), and the high (CI: 0.22–0.8) frequency region. There was a tendency for consonant intervals to be rated higher than dissonant ones also in the inharmonic condition for both the low (CI: -0.03–0.589), and the high (CI: -0.1–0.44) frequency region. This largely reflects the fact that the minor second was rated lower than consonant intervals also in the inharmonic condition, as shown in Fig. 5 which displays 95% CIs for contrasts between each consonant and dissonant interval by frequency region and harmonicity. Importantly, the posterior distributions for consonance preference shown in Fig. 4 indicate that consonance preference was higher for the harmonic than for the inharmonic conditions not only in the low frequency region (CI: 0.23–0.91), but also in the high frequency region (CI: 0.02–0.66). Therefore, the consonance preference scores obtained in the high frequency region cannot be explained solely on the basis of envelope periodicity cues, which would have been the same for the harmonic as for the inharmonic stimuli.

B. Same-different melody task

The results of the melody discrimination experiment are shown in Fig. 6. Performance in this task was very good, with an average d' close to 3. The 95% CIs of d' estimates for individual listeners generated by the Bayesian model indicate that performance was well above chance level for every listener. The group-level 95% CI for d' ranged from 2.27 to 3.46.

IV. HARMONIC SIEVE

It is unclear why the minor second dyad tended to be given lower ratings than the "consonant" dyads in the inharmonic conditions. One possible reason is that the degree of inharmonicity of the minor second in the inharmonic conditions may have been higher than that of the other intervals. To investigate this possibility we passed the dyads through harmonic sieves.

As noted by McDermott *et al.* (2010, supplementary materials) there is no standard method for the measurement of the degree of harmonicity of a sound. While it is trivial to distinguish a perfectly harmonic from an inharmonic sound, quantifying the degree of harmonicity of a sound that is not perfectly harmonic, or consists of both harmonic and inharmonic components is not straightforward. A method often employed involves passing the power spectrum of a sound through a harmonic sieve with meshes centered at harmonic frequencies of a given fundamental frequency (F0). The width of the meshes defines the tolerance for slight degrees of inharmonicity. The ratio of the power of the sound passing through the meshes to the power of the sound rejected by the sieve [Harmonic to noise ratio (HNR)] provides a measure of how well the sound fits the harmonic template of a given F0. The sound can be passed through a range of sieves with different F0s to find the best matching template. The HNR for the best matching template provides a measure of the harmonicity of the sound. This measure is affected by several parameters, which can be partly constrained by perceptual considerations but are otherwise to a large extent arbitrary. These include the width and the shape of the meshes, the range of F0s used for finding the best fitting template, and the number of harmonics used in the template.

The lowest F0 for a harmonic sieve has been generally chosen to be 30 Hz, which corresponds to the lower limit for pitch perception (Krumbholz *et al.*, 2000). The width of the meshes of the sieve has been sometimes chosen to have a fixed value of a few Hz ($\sim 2-8$), in order to tolerate small deviations from perfect harmonicity expected for intervals defined with the equal temperament scale (Bones *et al.*, 2014; Bones and Plack, 2015a,b). While meshes with widths of a few Hz in this range work well at relatively low frequencies, they cannot accommodate small deviations from harmonicity at high frequencies. The largest difference between intervals of the just intonation and of the equal temperament scale occurs for the tritone, and has a value of ± 17.49 cents (corresponding to $\sim 1\%$). At 500 Hz this corresponds to a deviation of ~ 5 Hz, while at 5000 Hz the deviation becomes ~ 51 Hz. A possible solution to this issue is to define the width of the meshes proportionally to their center frequencies. Duifhuis *et al.* (1982) for example used meshes with a width of $\sim 5\%$ of their center frequency in a model of pitch estimation in speech. Templates with a low F0, however, will generate sieves with progressively larger meshes relative to the harmonic spacing as the center frequency increases, even with a relatively small tolerance. Thus they will increasingly pass more components of a sound at high center frequencies and will eventually pass all components above a certain frequency when the meshes become so large relative to the harmonic spacing that they start overlapping. For example, a template with an F0 of 30 Hz and a tolerance of ± 17.49 cents will have overlapping meshes above ~ 1500 Hz, which will effectively pass through all components above that frequency. This issue is largely avoided by pitch models that use only templates with low-numbered harmonics ($\lesssim 10$ Duifhuis *et al.*, 1982; Terhardt *et al.*, 1982), which perceptually are the most important for the determination of the pitch of a sound.

Given the considerations above, we passed the spectra of the dyads used in the experiment through harmonic sieves with meshes ± 17.5 cent wide, and F0s ranging from 30 to

1174.569 Hz in 0.1 Hz steps. Only the portions of the spectra between 0.8–7.2 kHz for the low-frequency dyads, and 5.6–14.4 for the high-frequency dyads, were passed through the harmonic sieves. The HNRs for the best fitting template are shown in Fig. 7 for sieves with harmonic numbers 1–10, 1–12, or 1–15. For the low-frequency harmonic dyads the HNRs follow the rankings of the pleasantness ratings in the experiment, with larger HNRs for the perfect fifth and minor third intervals. For the high-frequency harmonic dyads the HNRs also follow the rankings of the pleasantness ratings in the experiment, except for the fact that the tritone has a higher HNR than the minor third. As expected the HNRs for the inharmonic dyads are generally lower than for the harmonic ones. The HNR profiles across the various intervals are also flatter, and except for a small peak for the tritone in the low-frequency conditions with harmonics sieves consisting of 12 or 15 harmonics, generally follow the pattern of the pleasantness ratings. In particular, the minor second dyad consistently shows the lowest HNRs in the inharmonic conditions.

The fact that in the inharmonic conditions the minor second dyad had the lowest HNR in our harmonic sieve modeling could explain why this dyad was rated lower than the other dyads in the inharmonic conditions of the pleasantness rating test. However, given that there is no standard way to measure HNRs these results should be interpreted cautiously. We tried to choose reasonable parameters for the harmonic sieves on the basis of known constraints. However, without more definitive knowledge of the psychophysiological mechanisms used by the auditory system to assess harmonicity, results from harmonic sieve models remain necessarily tentative. In any case, it should be remarked that in the inharmonic conditions the minor second dyad was given lower ratings than the "consonant" dyads both in the low, and in the high frequency region. Therefore this result is unlikely to be due to some idiosyncrasy of the high-frequency diads. Instead, this result supports the view that the pleasantness ratings were determined by the same mechanisms in the low, and in the high frequency regions.

Interestingly in the high-frequency harmonic condition the HNR rankings of the tritone and minor third dyads are reversed compared to the pleasantness ratings. This could be taken as evidence against the idea that pleasantness ratings are determined by harmonicity. However, it is possible that a learned association between pleasantness and a given dyad with all its lower harmonics as they occur naturally is transferred to a dyad with only a subset of those harmonics, as is the case for the dyads filtered in the high-frequency region of our experiment. It is also possible that given that the dyads were presented in noise, the lower harmonics, even if absent in the stimulus, are nonetheless perceived through spectral completion effects (McDermott and Oxenham, 2008). Pleasantness ratings for inharmonic stimuli may be more directly related to HNRs given that both learned associations and spectral completion effects are unlikely for this kind of stimuli.

V. DISCUSSION

We found that two consonant intervals were rated higher than two dissonant intervals even when they were presented in a high frequency region where neural phase locking to individual harmonics is thought to be severely degraded or absent. Given that the envelope repetition rates for our stimuli were higher than the highest rates at which the ability to perceive pitch on the basis of purely envelope rate cues has been observed (Burns and

Viemeister, 1976; Macherey and Carlyon, 2014), it was a priori unlikely that the perception of consonance for our stimuli could be mediated by such cues. The finding that consonance preference in the high frequency region was higher for harmonic stimuli than for stimuli that had the same envelope repetition rate, but were made inharmonic by shifting their component frequencies by a fixed offset, further dispels this possibility. This finding also rules out the possibility that preference ratings could have been dictated by the detection of binaural envelope beats, rather than by the detection of harmonic relations between the components of the stimuli. If ratings reflected the detection of binaural envelope beats, they should have been similar for the harmonic and inharmonic stimuli in the high frequency region, given that these stimuli had the same envelope repetition rates. The possibility that preference ratings were mediated by binaural envelope beats in our study seems, in any case, a priori unlikely given that such beats are difficult to detect for interaural envelope rate differences above about 3–5 Hz, well below the interaural envelope beat rates of our stimuli in the high-frequency region, and the ability to detect such beats declines with increasing monaural envelope rate, and is already very poor at 640 Hz (Bernstein and Trahiotis, 1996; McFadden and Pasanen, 1975). An additional reason why binaural envelope beats are unlikely to explain the results of the current study is that the perception of roughness for monaural envelope beats disappears for envelope rates exceeding ~ 300 Hz (Plomp and Steeneken, 1968; Terhardt, 1978). Neurophysiological studies suggest that this upper limit may be related to the upper limit of phase locking of auditory cortex neurons to envelope beats (Fishman *et al.*, 2000, 2001). The binaural envelope beats for some of our stimuli were completely outside the $\sim 20 - 300$ Hz range over which roughness can be perceived

(Terhardt, 1974a,b); for example both the minor second and the perfect fifth dyads in the high frequency harmonic condition, which respectively received the lowest and the highest, pleasantness ratings, did not contain any difference frequencies in this range. Therefore, the differences in pleasantness ratings given to these dyads cannot be attributed to perceived roughness caused by envelope beats.

Overall, our results indicate that pleasantness ratings in our experiment were determined by pitch relations between the tones forming the dyads rather than by beats. Our results do not shed light on the debate between the "harmonicity", and the "cultural" theories of consonance, because both theories predict preferences for certain dyads on the basis of the pitch combinations of their component tones. What our results clearly show, is that these pitch combinations can be readily perceived for dyads presented in a high frequency region, where neural phase locking to individual harmonics is either severely degraded or absent.

On the basis of the poor performance observed in the detection of mistuning of a single harmonic of a complex tone presented at high frequencies, Gockel and Carlyon (2018) hypothesized that harmonic templates at high frequencies may either have wider tolerances than at low frequencies, or even though they may have similar tolerances, the mechanism that leads to the perceptual segregation of the mistuned harmonic is absent at high frequencies. Our results suggest that harmonic templates at high frequencies have sufficiently narrow tolerances to support consonance judgments for the dyads used in the study. Although determining how narrow these tolerances are from pleasantness ratings data is not straightforward, as it is dependent on several modeling assumptions of harmonic sieves (see sec. IV) it is quite clear that they should be narrower than 100 cents, which corresponds to a mistuning of $\sim 6\%$ that was very difficult to detect in Gockel and Carlyon (2018)'s study. The reasoning behind this is that given that the distance between the root and interval notes of a minor second dyad is 100 cents, a harmonic template at the F0 of the root note with a tolerance ≥ 100 cents would pass through all components of a minor second dyad, just as it would pass through all components of a unison dyad. Given that the unison, together with the octave typically receive the highest pleasantness ratings amongst all musical intervals, the fact that the minor second received the lowest pleasantness ratings in our study clearly shows that it was treated differently than a unison. Therefore our results, combined with those of Gockel and Carlyon (2018) suggest that harmonic templates at high frequencies may not be larger than at low frequencies, but the mechanism that leads to the perceptual segregation of the mistuned harmonic may be absent at high frequencies.

A. Is neural phase locking necessary for the perception of consonance?

Although phase locking is thought to be severely degraded or absent above ~ 5 kHz, some computational models suggest that, theoretically, some residual temporal information usable for pitch coding may be available up to frequencies as high as 10 kHz (Heinz *et al.*, 2001; Recio-Spinoso *et al.*, 2005). Additionally, Moore and Ernst (2012) have shown that pure tone FDLs increase as a function of frequency up to 8 kHz, and then show a plateau, suggesting that a transition from a temporal to a place code may occur ~ 8 kHz rather than ~ 5 kHz as once commonly thought. On the basis of this evidence it has been argued that, although phase locking may be too weak to support musical pitch perception for individual pure tones above 5 kHz, the combined temporal information across several > 5

kHz harmonics of a complex tone may be sufficient to support musical pitch perception. Lau *et al.* (2017), however, measuring FDLs for pure tones > 8 kHz and F0DLs for complex tones with harmonics > 8 kHz, found that the F0DLs were better than predicted from an optimal combination of peripheral information from each of their component frequencies. This finding poses two additional difficulties to the theory that pitch perception at high frequencies is supported by temporal coding: 1) it pushes the upper limit at which phase locking information would be viable for pitch perception above 8 kHz, the point at which a putative transition between a temporal to a place code would occur according to the data of Moore and Ernst (2012), 2) if pitch were nonetheless coded temporally at such high frequencies, pure tone FDLs would have to be limited by additional central noise sources rather than by peripheral limitations due to degraded phase locking.

Another factor to consider when evaluating the possible role of a temporal code for human pitch perception at high frequencies is how the limits of neural phase locking in humans compare to those of other mammalian species for which direct single neuron recordings are available (Johnson, 1980; Palmer and Russell, 1986; Winter, 2005). Recordings of the compound action potential using a technique that separates the auditory nerve neurophonic from the cochlear microphonic, indicate that this limit is at best similar, and probably lower than the 5 kHz limit recorded in the cat (Verschooten *et al.*, 2018).

Given the results of our study, the question of whether neural phase locking is necessary for the perception of consonance hinges on the issue of whether a temporal code may be used for frequency coding in the high frequency region where the stimuli in our study were presented. The lowest component of the dyads in the high frequency region was the 5^{th}

harmonic of the root note, just above 5.8 kHz. However, in order to differentially rate the pleasantness of the consonant and dissonant dyads listeners needed to also perceive at least the pitch of the first audible component of each interval note. This pushes the minimum frequency needed to differentially rate the consonant and dissonant dyads to at least 7 kHz. The evidence reviewed above strongly points to the use of a place code rather than a temporal code for frequency coding at such high frequencies.

Assuming that the frequency components of our stimuli could not be coded via phase locking, the results of this study indicate that temporal coding is not necessary for the perception of consonance. Hence, models of consonance perception based on neural periodicity detection would be either incorrect, or at best incomplete, because they could not explain the perception of consonance at high frequencies observed in the current study. However, our results are not inconsistent with the notion that temporal coding may play an role in the perception of consonance in low frequency regions, and that inter-individual differences in temporal coding (Bones and Plack, 2015b), which can be partly due to factors such as musical experience (Bones *et al.*, 2014) and aging (Bones and Plack, 2015a) may lead to changes in the perception of consonance. For example, it is possible to envisage a model in which the perception of consonance is based on a central harmonic template matching unit similar to the models proposed by Goldstein (1973) and Srulovicz and Goldstein (1983) for the perception of pitch. This template matching unit could receive input from both temporal and place frequency representations. In the low frequency region, where phase locking is good, temporal frequency representations may be dominant. If these frequency representations are degraded, the input to the central harmonic template matching unit will be degraded as well, and the perception of consonance may be affected even though temporal processing plays no direct role in the neural computations determining consonance.

The issue of whether sound frequencies are represented via a temporal code based on neural phase locking, or via a "rate" code based on cochlear tonotopy represents a fundamental aspect of auditory neurophysiology that remains still partly unsolved (Oxenham, 2018). The studies of Oxenham *et al.* (2011) and Lau *et al.* (2017) indicate that musical pitch perception is possible at frequencies that are highly unlikely to be coded via neural phase locking. Overall, the results of these studies, together with those of the current study, strongly suggest that a tonotopic rate code is sufficient to convey pitch and consonance information that is crucial for the perception of melody and harmony in music. However, a recent collection of viewpoints on the topic indicates a lack of consensus on the upper limit of phase locking in humans (Verschooten *et al.*, 2019). This consensus may not be reached until further experimental data is available, including direct recordings from the human auditory nerve. The results of our experiment provide further data that is relevant to this debate. Comprehensive neurophysiological models of consonance should be able to explain consonance perception at high frequencies, whether they are based on rate-place or on temporal frequency coding.

ACKNOWLEDGMENTS

The results of this study were presented at the 3^{rd} Joint Meeting of the Acoustical Society of America and European Acoustics Association (Boston, Massachusetts, June 2017). The research was supported by a grant (BB/M007243/1) from the Biotechnology and Biolog-

ical Sciences Research Council, UK. Author CP was supported by the NIHR Manchester Biomedical Research Centre. We would like to thank Laurent Demany for valuable comments on a previous version of this manuscript. The underlying data in this paper are available from
https://osf.io/vxzbc/?view_only=967c43c5a1084d3bb2874c2073d7b312.

[1] See supplementary material at [please insert URL] for additional methods details, additional figures, and additional tables.

[2] Due to an error during the setup of the experiment the stimuli were presented in sine phase rather than in random phase as in the pleasantness rating test. Presenting the stimuli in random phase reduces the crest factor, or "peakiness" of their envelope, thus reducing (but not eliminating) the usefulness of potential envelope cues (Bernstein and Oxenham, 2005). Performance in the melody discrimination experiment of Oxenham *et al.* (2011) was high with random-phase harmonic complex tones filtered in a high frequency region, but approached chance level when only envelope cues were available. Even though potential envelope cues could have been more salient in the current experiment, envelope-based pitch perception declines dramatically at high frequencies (Burns and Viemeister, 1976; Macherey and Carlyon, 2014). Therefore it is unlikely that envelope cues could account for the high performance levels observed in the current study.

Attneave, F., and Olson, R. K. (**1971**). "Pitch as a medium: a new approach to psychophysical scaling," Am. J. Psychol. **84**(2), 147–166, doi: 10.2307/1421351.

Bernstein, J. G., and Oxenham, A. J. (**2005**). "An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination," J. Acoust. Soc. Am. **117**(6), 3816–3831, doi: 10.1121/1.1904268.

Bernstein, L. R., and Trahiotis, C. (**1996**). "Binaural beats at high frequencies: Listeners' use of envelope-based interaural temporal and intensitive disparities," J. Acoust. Soc. Am. **99**(3), 1670–1679, doi: 10.1121/1.414689.

Bidelman, G. M., and Heinz, M. G. (**2011**). "Auditory-nerve responses predict pitch attributes related to musical consonance-dissonance for normal and impaired hearing," J. Acoust. Soc. Am. **130**(3), 1488–1502, doi: 10.1121/1.3605559.

Bidelman, G. M., and Krishnan, A. (**2009**). "Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem," J. Neurosci. **29**(42), 13165–13171, doi: 10.1523/JNEUROSCI.3900-09.2009.

Bones, O., Hopkins, K., Krishnan, A., and Plack, C. J. (**2014**). "Phase locked neural activity in the human brainstem predicts preference for musical consonance," Neuropsychologia **58**, 23–32, doi: 10.1016/j.neuropsychologia.2014.03.011.

Bones, O., and Plack, C. J. (**2015**a). "Losing the music: aging affects the perception and subcortical neural representation of musical harmony," J. Neurosci. **35**(9), 4071–4080, doi: 10.1523/JNEUROSCI.3214-14.2015.

Bones, O., and Plack, C. J. (**2015**b). "Subcortical representation of musical dyads: individual differences and neural generators," Hear. Res. **323**, 9–21, doi: 10.1016/j.heares.2015.01.009.

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, doi: 10.1525/mp.2017.35.1.118.

Bowling, D. L., and Purves, D. (**2015**). "A biological rationale for musical consonance," Proc. Natl. Acad. Sci. U.S.A. **112**(36), 11155–11160, doi: 10.1073/pnas.1505768112.

Burns, E., and Viemeister, N. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**(4), 863, doi: 10.1121/1.381166.

Cousineau, M., McDermott, J. H., and Peretz, I. (**2012**). "The basis of musical consonance as revealed by congenital amusia," Proc. Natl. Acad. Sci. U.S.A. **109**(48), 19858–19863, doi: 10.1073/pnas.1207989109.

Darwin, C. J. (**2005**). "Pitch and auditory grouping," in *Pitch: Neural Coding and Perception*, edited by R. Fay, A. N. Popper, (Series Eds.), C. J. Plack, and A. J. Oxenham (Springer, New York), pp. 278–305.

Duifhuis, H., Willems, L. F., and Sluyter, R. J. (**1982**). "Measurement of pitch in speech: an implementation of Goldstein's theory of pitch perception," J. Acoust. Soc. Am. **71**(6), 1568–1580, doi: 10.1121/1.387811.

Ebeling, M. (**2008**). "Neuronal periodicity detection as a basis for the perception of consonance: a mathematical model of tonal fusion," J. Acoust. Soc. Am. **124**(4), 2320–2329, doi: 10.1121/1.2968688.

Feeney, M. P. (**1997**). "Dichotic beats of mistuned consonances," J. Acoust. Soc. Am. **102**(4), 2333–2342, doi: 10.1121/1.419602.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2000**). "Complex tone processing in primary auditory cortex of the awake monkey. I. Neural ensemble correlates of roughness," J. Acoust. Soc. Am. **108**(1), 235–246, doi: 10.1121/1.429460.

Fishman, Y. I., Volkov, I. O., Noh, M. D., Garell, P. C., Bakken, H., Arezzo, J. C., Howard, M. A., and Steinschneider, M. (**2001**). "Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans," J. Neurophysiol. **86**(6), 2761–2788, doi: 10.1152/jn.2001.86.6.2761.

Gelman, A., Hill, J., and Yajima, M. (**2012**). “Why we (usually) don’t have to worry about multiple comparisons,” J. Res. Educ. Eff. **5**, 189–211, doi: 10.1080/19345747.2011.618213.

Gockel, H. E., and Carlyon, R. P. (**2018**). “Detection of mistuning in harmonic complex tones at high frequencies,” Acta Acust. united Ac. **104**, 766–769, doi: 10.3813/AAA.919219.

Goldstein, J. L. (**1973**). “An optimum processor theory for the central formation of the pitch of complex tones,” J. Acoust. Soc. Am. **54**(6), 1496–1516, doi: 10.1121/1.1914448.

Hautus, M. J., and Lee, A. J. (**1998**). “The dispersions of estimates of sensitivity obtained from four psychophysical procedures: implications for experimental design,” Perception & Psychophysics **60**(4), 638–649, doi: 10.3758/BF03206051.

Heinz, M. G., Colburn, H. S., and Carney, L. H. (**2001**). “Evaluating auditory performance limits: I. one-parameter discrimination using a computational model for the auditory nerve,” Neural Comput. **13**(10), 2273–2316, doi: 10.1162/089976601750541804.

Helmholtz, H. (**1954**). *On the Sensations of Tone* (Dover, New York).

Johnson, D. H. (**1980**). “The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones,” J. Acoust. Soc. Am. **68**(4), 1115–1122, doi: 10.1121/1.384982.

Kameoka, A., and Kuriyagawa, M. (**1969**a). “Consonance theory part I: consonance of dyads,” J. Acoust. Soc. Am. **45**(6), 1451–1459, doi: 10.1121/1.1911623.

Kameoka, A., and Kuriyagawa, M. (**1969**b). "Consonance theory part II: consonance of complex tones and its calculation method," J. Acoust. Soc. Am. **45**(6), 1460–1469, doi: 10.1121/1.1911624.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**(3 Pt 1), 1170–1180, doi: 10.1121/1.1287843.

Kruschke, J. K. (**2010**). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*, 1st ed. (Academic Press, Inc., Orlando, FL, USA).

Kruschke, J. K. (**2014**). *Doing Bayesian data analysis, a tutorial with R, JAGS, and Stan*, 2^{nd} ed. (Elsevier, London).

Lau, B. K., Mehta, A. H., and Oxenham, A. J. (**2017**). "Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies," J. Neurosci. **37**(37), 9013–9021, doi: 10.1523/JNEUROSCI.1507-17.2017.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), 467–477, doi: 10.1121/1.1912375.

Macherey, O., and Carlyon, R. P. (**2014**). "Re-examining the upper limit of temporal pitch," J. Acoust. Soc. Am. **136**(6), 3186, doi: 10.1121/1.4900917.

Macmillan, N. A., and Creelman, C. D. (**2004**). *Detection theory: A user's guide*, 2^{nd} ed. (Lawrence Erlbraum Associates, London).

Malmberg, C. F. (**1918**). "The perception of consonance and dissonance," Psychol. Monogr. **25**, 93–133.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2010**). "Individual differences reveal the basis of consonance," Curr. Biol. **20**(11), 1035–1041, doi: 10.1016/j.cub.2010.04.019.

McDermott, J. H., and Oxenham, A. J. (**2008**). "Spectral completion of partially masked sounds," Proc. Natl. Acad. Sci. U.S.A. **105**(15), 5939–5944, doi: 10.1073/pnas.0711291105.

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, doi: 10.1038/nature18635.

McFadden, D., and Pasanen, E. G. (**1975**). "Binaural beats at high frequencies," Science **190**(4212), 394–396, doi: 10.1126/science.1179219.

McLachlan, N., Marco, D., Light, M., and Wilson, S. (**2013**). "Consonance and pitch," J. Exp. Psychol. Gen. **142**(4), 1142–1158, doi: 10.1037/a0030830.

Meddis, R., and O'Mard, L. P. (**2006**). "Virtual pitch in a computational physiological model," J. Acoust. Soc. Am. **120**(6), 3861–3869, doi: 10.1121/1.2372595.

Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**(3), 610–619, doi: 10.1121/1.1913640.

Moore, B. C. J., and Ernst, S. M. (**2012**). "Frequency difference limens at high frequencies: evidence for a transition from a temporal to a place code," J. Acoust. Soc. Am. **132**(3), 1542–1547, doi: 10.1121/1.4739444.

Moore, B. C. J., Huss, M., Vickers, D. A., Glasberg, B. R., and Alcantara, J. I. (**2000**). "A test for the diagnosis of dead regions in the cochlea," Br. J. Audiol. **34**(4), 205–224, doi: 10.3109/03005364000000131.

Oxenham, A. J. (**2018**). “How we hear: The perception and neural coding of sound,” Annu. Rev. Psychol. **69**, 27–50, doi: 10.1146/annurev-psych-122216-011635.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (**2011**). “Pitch perception beyond the traditional existence region of pitch,” Proc. Natl. Acad. Sci. U.S.A. **108**(18), 7629–7634, doi: 10.1073/pnas.1015291108.

Palmer, A. R., and Russell, I. J. (**1986**). “Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells,” Hear. Res. **24**(1), 1–15, doi: 10.1016/0378-5955(86)90002-X.

Patterson, R. D. (**1986**). “Spiral detection of periodicity and the spiral form of musical scales,” Psychol. Music **14**(1), 44–61, doi: 10.1177/0305735686141004.

Plack, C. J., and Oxenham, A. J. (**2005**). “The psychophysics of pitch,” in *Pitch: Neural Coding and Perception*, edited by R. Fay, A. N. Popper, (Series Eds.), C. J. Plack, and A. J. Oxenham (Springer, New York), pp. 7–55.

Plomp, R., and Levelt, W. J. (**1965**). “Tonal consonance and critical bandwidth,” J. Acoust. Soc. Am. **38**(4), 548–560, doi: 10.1121/1.1909741.

Plomp, R., and Steeneken, H. J. (**1968**). “Interference between two simple tones,” J. Acoust. Soc. Am. **43**(4), 883–884, doi: 10.1121/1.1910916.

Plummer, M. (**2003**). “JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling,” in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, edited by K. Hornik, F. Leisch, and A. Zeileis, Vienna, Austria.

R Core Team (**2019**). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/.

Recio-Spinoso, A., Temchin, A. N., van Dijk, P., Fan, Y. H., and Ruggero, M. A. (**2005**). "Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers," J. Neurophysiol. **93**(6), 3615–3634, doi: 10.1152/jn.00882.2004.

Schwartz, D. A., Howe, C. Q., and Purves, D. (**2003**). "The statistical structure of human speech sounds predicts musical universals," J. Neurosci. **23**(18), 7160–7168, doi: 10.1523/JNEUROSCI.23-18-07160.2003.

Semal, C., and Demany, L. (**1990**). "The upper limit of "musical" pitch," Music Percept. **8**(2), 165–176, doi: 10.2307/40285494.

Srulovicz, P., and Goldstein, J. L. (**1983**). "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum," J. Acoust. Soc. Am. **73**(4), 1266–1276, doi: 10.1121/1.389275.

Tenney, J. (**1988**). *A history of 'consonance and dissonance'* (Excelsior Music Publishing Company, New York).

Terhardt, E. (**1974**a). "On the Perception of Periodic Sound fluctuations (Roughness) ," Acta Acust. united Ac. **30**(4), 201–2013.

Terhardt, E. (**1974**b). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**(5), 1061–1069, doi: 10.1121/1.1914648.

Terhardt, E. (**1978**). "Psychoacoustic evaluation of musical sounds," Perception & Psychophysics **23**(6), 483–492, doi: 10.3758/BF03199523.

Terhardt, E. (**1984**). "The Concept of Musical Consonance: A Link between Music and Psychoacoustics," Music Perception: An Interdisciplinary Journal **1**(3), 276–295, doi: 10.2307/40285261.

Terhardt, E., Stoll, G., and Seewann, M. (**1982**). "Algorithm for extraction of pitch and pitch salience from complex tonal signals," J. Acoust. Soc. Am. **71**(3), 679–688, doi: 10.1121/1.387544.

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. N. Y. Acad. Sci. **930**, 92–116, doi: 10.1111/j.1749-6632.2001.tb05727.x.

Verschooten, E., Desloovere, C., and Joris, P. X. (**2018**). "High-resolution frequency tuning but not temporal coding in the human cochlea," PLOS Biol. **16**(10), e2005164, doi: 10.1371/journal.pbio.2005164.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C. J., Joris, P. X., Heinz, M. G., and Plack, C. J. (**2019**). "The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints," Hear. Res. **377**, 109–121, doi: 10.1016/j.heares.2019.03.011.

Winter, I. M. (**2005**). "The neurophysiology of pitch," in *Pitch: Neural Coding and Perception*, edited by R. Fay, A. N. Popper, (Series Eds.), C. J. Plack, and A. J. Oxenham (Springer, New York), pp. 99–146.

TABLES

Name	Cents	F0 (Hz)
Minor Second (m2)	100	1244.508
Minor Third (m3)	300	1396.913
Tritone (TT)	600	1661.219
Perfect Fifth (P5)	700	1760.000

TABLE I. Harmonic intervals used in the pleasantness rating experiment. The F0 of the root note was always 1174.659 Hz. The first column shows the name of the interval and its abbreviation, in parentheses. The second column shows the size of the interval in cents. The third column shows the F0 of the interval note.

FIGURE LEGENDS

FIG. 1. (Color online) Spectra for the harmonic dyads. The solid blue line plots the spectrum of the root note. The dotted red line plots the spectrum of the interval note. The root and interval notes were always presented each to a different ear.

FIG. 2. (Color online) Mean standardized pleasantness ratings ±1 s.e.m.

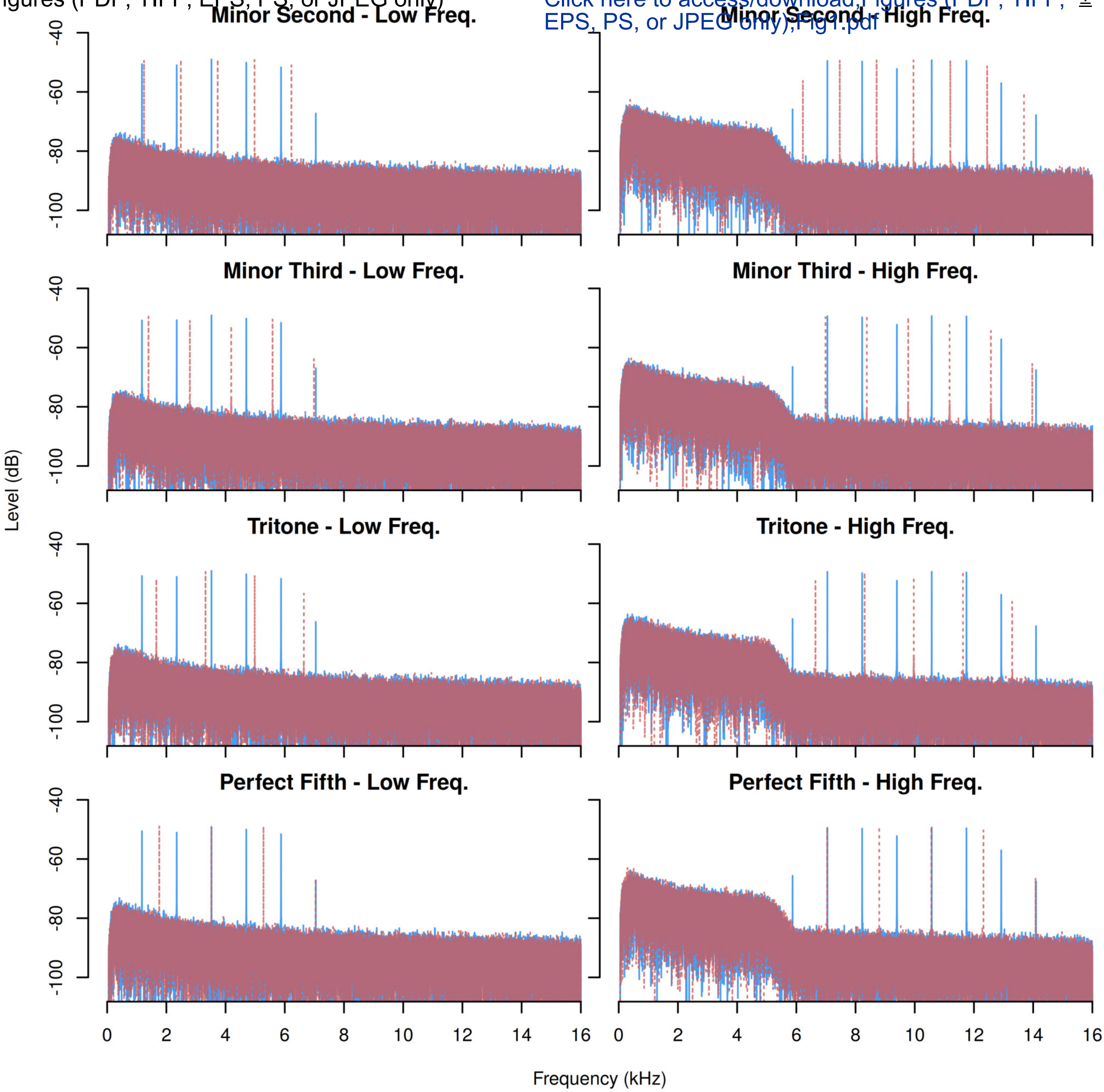
FIG. 3. Mean consonance preference scores ±1 s.e.m.

FIG. 4. (Color online) Posterior distributions estimated by the Bayesian model for consonance preference. The four distributions at the bottom show effects at the low and high frequency regions, for harmonic and inharmonic stimuli, separately. The two top distributions show the effect difference between the harmonic and inharmonic stimuli for each frequency region. Circles denote the mode of the distribution. Horizontal segments mark the 95% CIs.

FIG. 5. (Color online) Posterior modes and 95% CIs for contrasts between each consonant and dissonant interval, by frequency region and harmonicity.

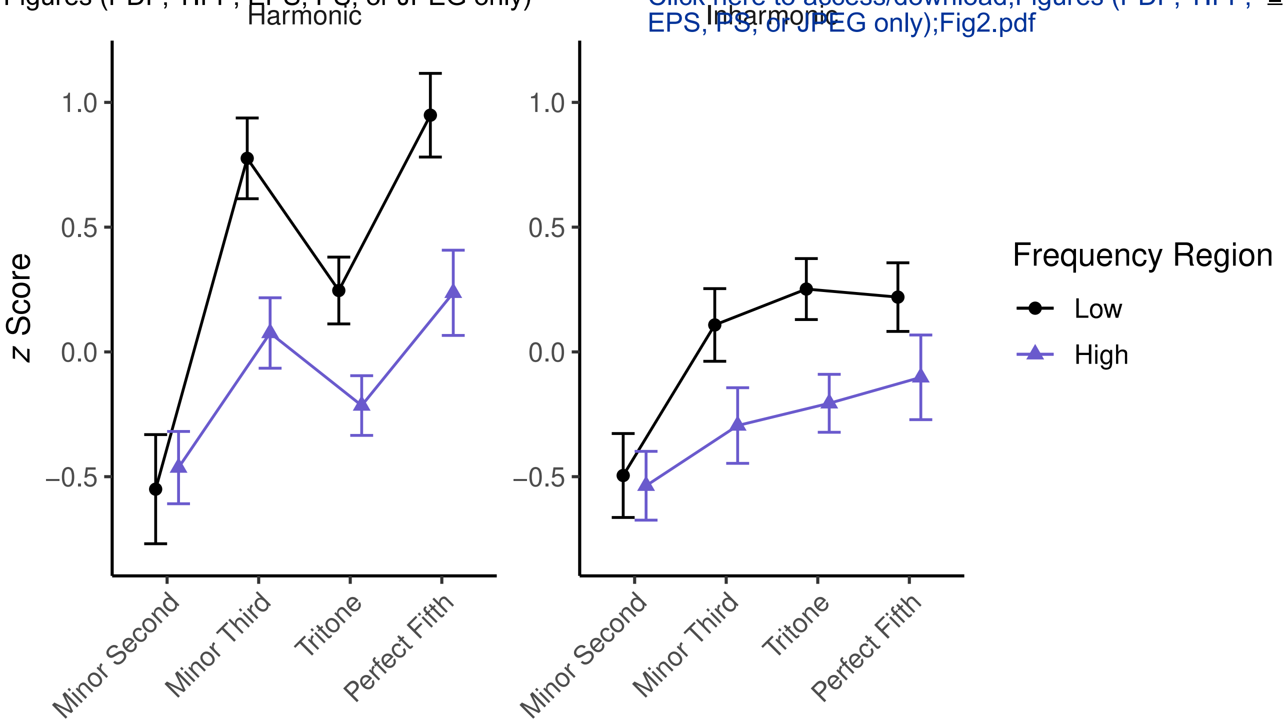
FIG. 6. (Color online) Results of the melody discrimination experiment. Points indicate the d' values estimated by the Bayesian model for each individual listener, and are jittered for clarity. The vertical segments around these points enclose their 95% CIs. The wide horizontal bar indicates the group-level d' estimated by the Bayesian model, and the narrow horizontal bars enclose its 95% CI.

FIG. 7 (color online) Harmonic to noise ratio for the stimuli used in the experiment. The different line colors denote the results for harmonic sieves with harmonic numbers 1–10, 1–12, or 1–15.



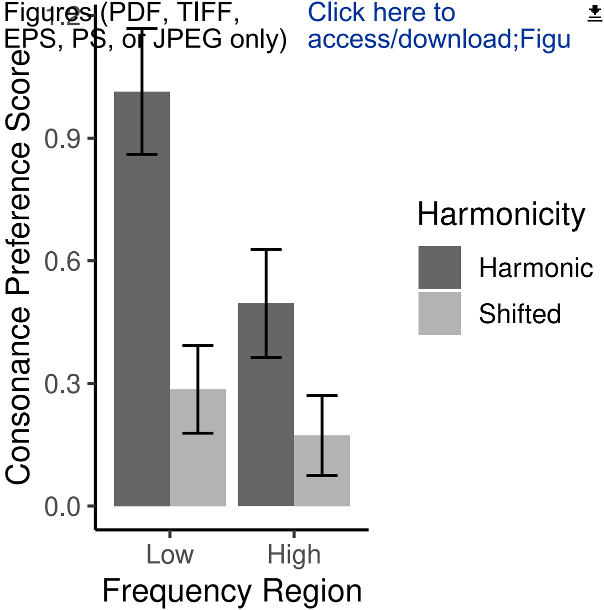
Minor Second - Low Freq.
Minor Second - High Freq.
Minor Third - Low Freq.
Minor Third - High Freq.
Tritone - Low Freq.
Tritone - High Freq.
Perfect Fifth - Low Freq.
Perfect Fifth - High Freq.
Level (dB)
-40
-60
-80
-100
0
2
4
6
8
10
12
14
16
Frequency (kHz)

Figures (PDF, TIFF, EPS, PS, or JPEG only)



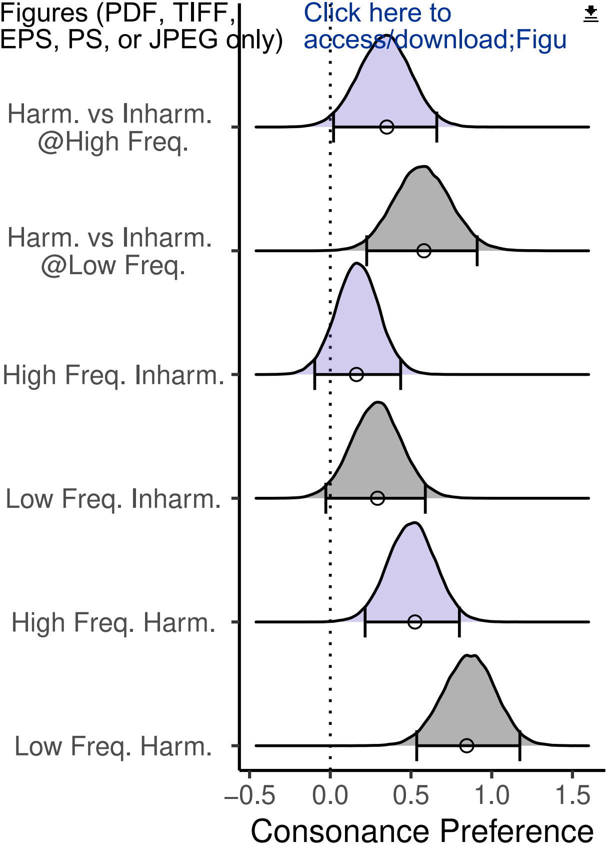
Figures (PDF, TIFF, EPS, PS, or JPEG only)

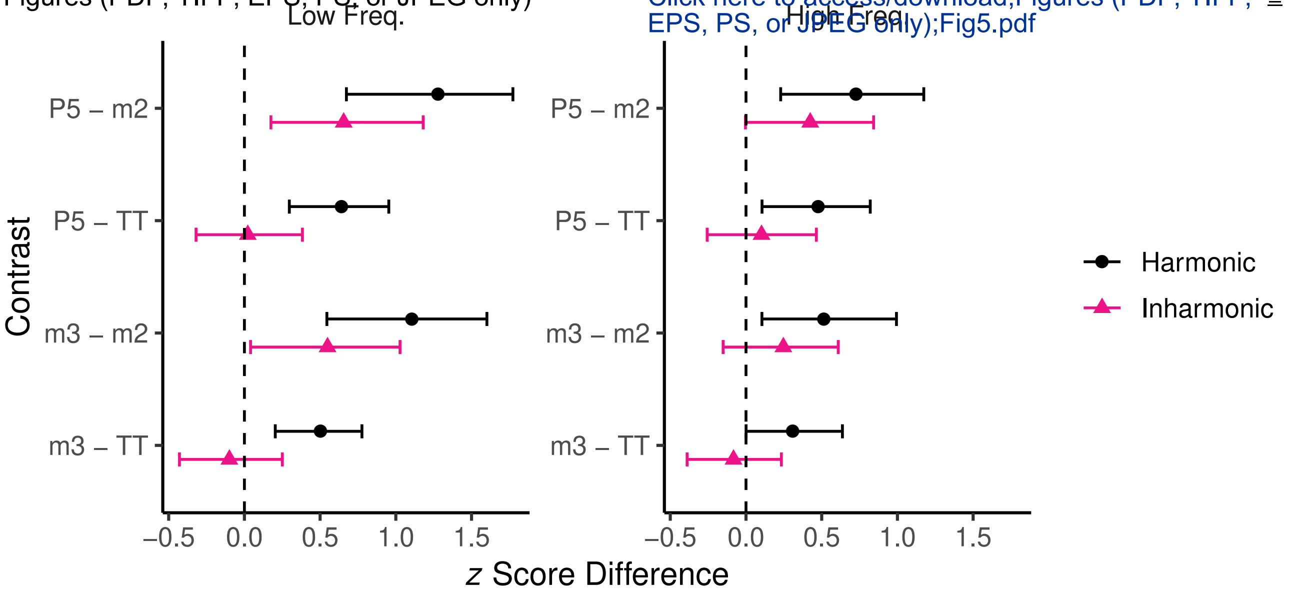
Click here to access/download;Figu



Figures (PDF, TIFF, EPS, PS, or JPEG only)

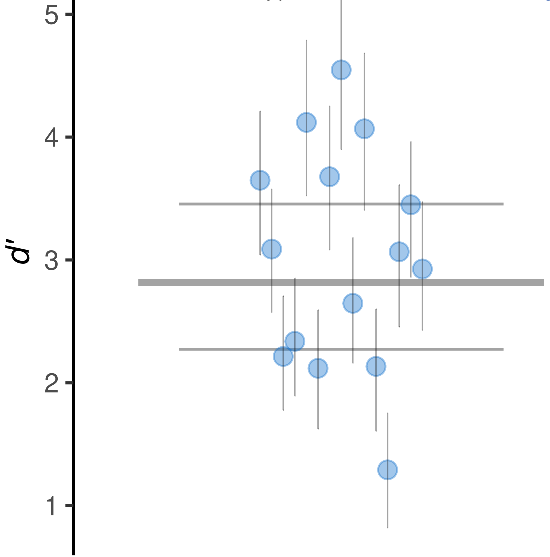
Click here to access/download;Figu




Low Freq.
High Freq.
Contrast
P5 – m2
P5 – TT
m3 – m2
m3 – TT
–0.5
0.0
0.5
1.0
1.5
z Score Difference
Harmonic
Inharmonic

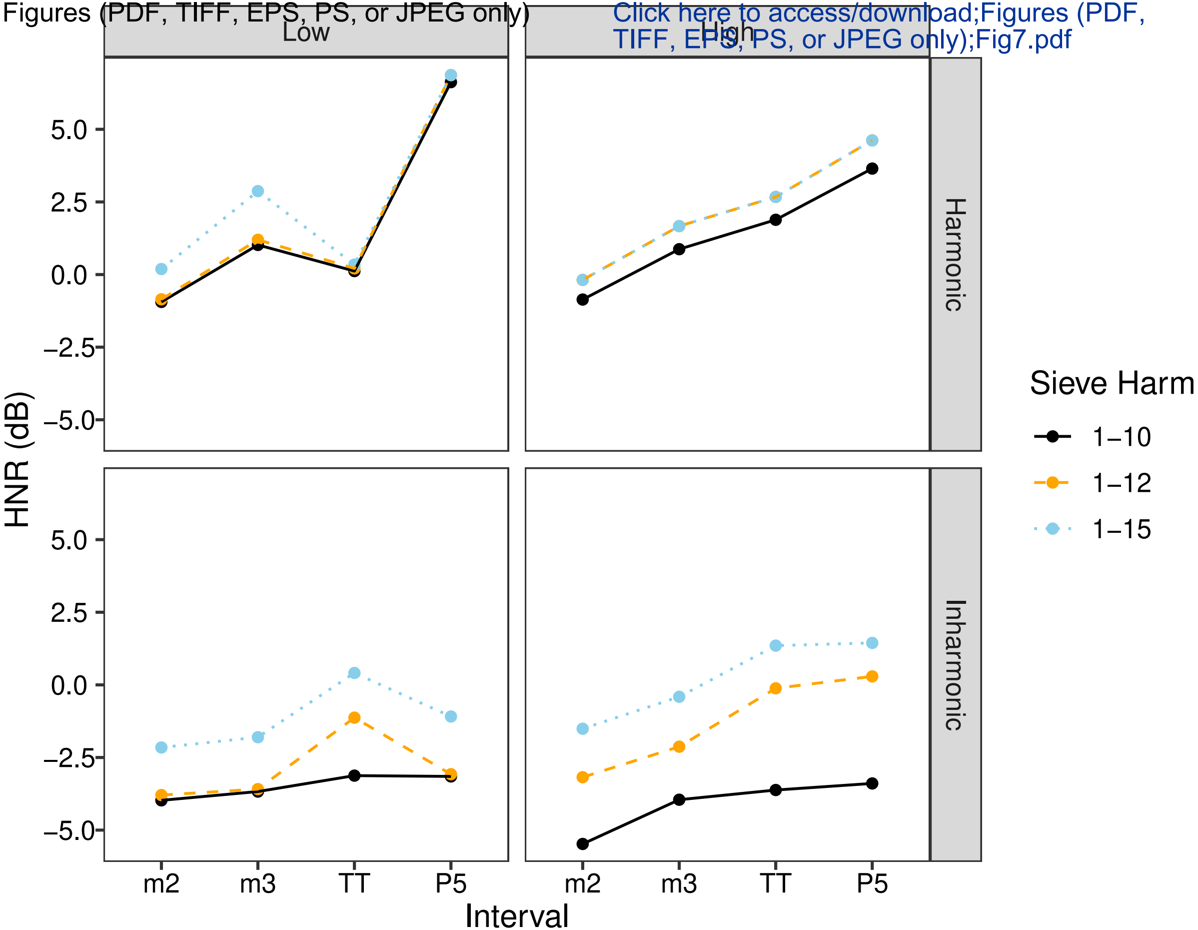
Figures (PDF, TIFF, EPS, PS, or JPEG only)

Click here to access/download;Figu



Figures (PDF, TIFF, EPS, PS, or JPEG only)

Click here to access/download;Figures (PDF, TIFF, EPS, PS, or JPEG only);Fig7.pdf



Rebuttal letter

Manuscript with tracked changes

Click here to access/download

Helpful/Supporting Material for Reviewer

Manuscript_tracked_changes.pdf