



Dyad's consonance and dissonance: combining the compactness and roughness approaches

Isabella Masina^{1,2,a} , Giuseppe Lo Presti³, Domenico Stanzial^{1,4}

¹ Department of Physics and Earth Science, Ferrara University, Ferrara, Italy

² Istituto Nazionale di Fisica Nucleare, Sez. Ferrara, Ferrara, Italy

³ Information Technology Department, CERN, Geneva, Switzerland

⁴ CNR Institute for Microelectronics and Microsystems, Bologna, Italy

Received: 20 May 2022 / Accepted: 3 November 2022
© The Author(s) 2022

Abstract At present, there are two approaches that aim at explaining on physical grounds the psychoacoustic perception of consonance and dissonance for dyads, whose pioneers have been, respectively, Galilei and Helmholtz: One is based on the “compactness” of the waveform of the combined signal, while the other on the absence of “roughness” due to possible beats. We perform a detailed study of each approach and find that none of the associated model versions, not even the more refined ones, is fully satisfactory when faced to perceptual data on dyads. We show that combining the two approaches results instead in a surprisingly successful agreement with perceptual data: This demonstrates that compactness and roughness are both necessary ingredients for a phenomenological description of consonance and dissonance.

Graphical abstract



1 Introduction

Providing an explanation to account for the auditory perceptions of consonance and dissonance (C &D) in music, in particular for intervals (dyads) and triadic chords (triads), is an extremely fascinating issue, which attracted the attention of many generations of scientists, and which is still an open arena for scientific debate, the actual functioning of our hearing system being not fully understood [1]. The present paper aims at giving a contribution to the scientific discussion about C &D, from the point of view of physics and its methodologies. C &D are defined here according to their literal meaning (in Latin, the prefixes *con* and *dis* stand for unity and division, respectively), that is whether two sounds are perceived to mix well or not.

The problem of justifying on a physical basis the perceptions of C &D started to be formulated in a quantitative way since the beginning of modern science, with the insights of G. Galilei about dyad's periodicity, under the form of a commensurability (or coincidence) theory [2], which triggered contributions and debates from many other scientists [3–5], also in relation to the later discoveries about higher harmonics. A breakthrough in the field was represented by the proposal of H. Helmholtz [6], who rather suggested an approach based on the absence of the roughness sensation associated with beats (but see also Ref. [5] about the insights by F. Foderà). As we will argue, these two pioneering approaches gave rise to the two categories of explanations for C &D that we will denote for short “compactness” and “roughness.” The former might also be denoted “periodicity/harmonicity,” according to the definitions given in the extensive models review of Ref. [7], where the additional category of “cultural familiarity” is introduced.

^a e-mail: masina@fe.infn.it (corresponding author)

In general, these two categories have been considered to be alternative and somewhat competing. While for the models based on roughness, the C & D indicator is a continuous function of the dyad's frequency ratio, for models based on compactness, the C & D indicator is rather naturally discontinuous, so that some procedure might be introduced to render it more physical. Actually, none of the representative models of these two categories emerged as a fully satisfactory explanation of the perceptual data: There are indeed many models among these two categories displaying a comparable degree of correlation with perceptual observations [7, 8]. Recently, the hypothesis that some model resulting from a mixing of other models might be more successful than the constituent ones was considered [7, 8]. For instance, the "composite model" of Ref. [7] results from the mixing of three models, two from the categories above, and one cultural familiarity model: The numerical studies based on correlation coefficients revealed however that, for dyads, its performance is marginally better than that for the pure roughness model.

In this paper, we carry out a complete analysis to assess whether a single refined model among the two categories of compactness and roughness,¹ or rather a combination of them, might provide a satisfactory explanation of the perceptual data on C & D for dyads. We directly collect these data by performing a test on a group of 20 people. Our data for the 12 dyads of the just intonation chromatic scale turn out to be in good agreement with previous similar datasets. In addition, we collect datasets for the corresponding 12 compound intervals and for 14 microtonal intervals: This allows to have an extended dataset of 38 dyads, to be compared with model predictions by means of the chi-square test.

We first study the two categories separately, discuss their strengths and weaknesses in accounting for the perceptual data and isolate a few best models for each; by means of the chi-square test, we show that even the best models have intrinsic difficulties in explaining some relevant features of the perceptual observations. As for the category of compactness, we provide a unified description, based on physical grounds (and which at the best of our knowledge is original), of the zoo of proposed models, from the oldest one by G. Galilei [2] to some of the most recent ones [7, 9, 10]; we also explicitly demonstrate why periodicity and harmonicity models are practically equivalent. As for the roughness category [6], after a reanalysis of some of the former models [11–13], we discuss the importance of including, in addition to the first-order beatings of the mistuned unison, also the most relevant second-order ones, namely the mistuned octave and fifth [14].

We then build explicit models combining the two approaches of compactness and roughness. We find that all such combined models (and not just a particular selection of them) are highly successful when confronted with the perceptual observations and perform significantly better than the constituent models. This is a non-trivial result and shows that when one approach underestimates the C & D at some datapoint, the other approach overestimates it by precisely the same amount, so that their cooperation results in an agreement with perceptual observations. This demonstrates that compactness and roughness are both fundamental ingredients for an effective explanation of C & D. In the C & D combined model function, the consonance peaks turn out to be mainly related to the compactness of the perceived signal, while the dissonance minima turn out to be mainly related to the roughness due to beatings.

The paper is organized as follows. In Sect. 2, after a review of some previous datasets on dyads C & D, we describe the methodology used to perform our 38 intervals test and comment on the associated results. In Sect. 3, we focus on the compactness approach, giving a unified description of the many periodicity and harmonicity models proposed; we then evaluate their performance with respect to the results of our test. Section 4 is devoted to the roughness approach: After a review of the previous models, we discuss the relevance of the weight functions for the harmonics and estimate the impact of including the second-order beats. In Sect. 5, we explicitly build and test many combined models of compactness and roughness. Finally, we draw our conclusions in Sect. 6. In Appendix 1, we display the results of our test, while Appendix 2 contains an analysis of some mathematical models related to the periodicity approach.

2 Testing consonance and dissonance

According to the classical tradition, the first systematic attempts to test dyad's C & D were carried out on the monochord by Pythagoras and his school. The findings were that dyads's consonance is related to the simplest ratios of integer numbers, or equivalently to the simplest ratios of the lengths of the two monochord's string segments.² The recognized consonances were only three³: calling the dyad's frequencies f_1 and f_2 , with $f_1 < f_2$, they correspond, in order of decreasing consonance, to the ratios $f_2/f_1 = 2/1$ (*diapason*), $3/2$ (*diapente*), $4/3$ (*diatessaron*). In the present musical notation, they belong to the just intonation (or natural) scale and are denoted as perfect octave (P8), fifth (P5) and fourth (P4), respectively. These dyads were used for the sake of tuning and as building blocks of the ancient Greek modes (*harmonies*).

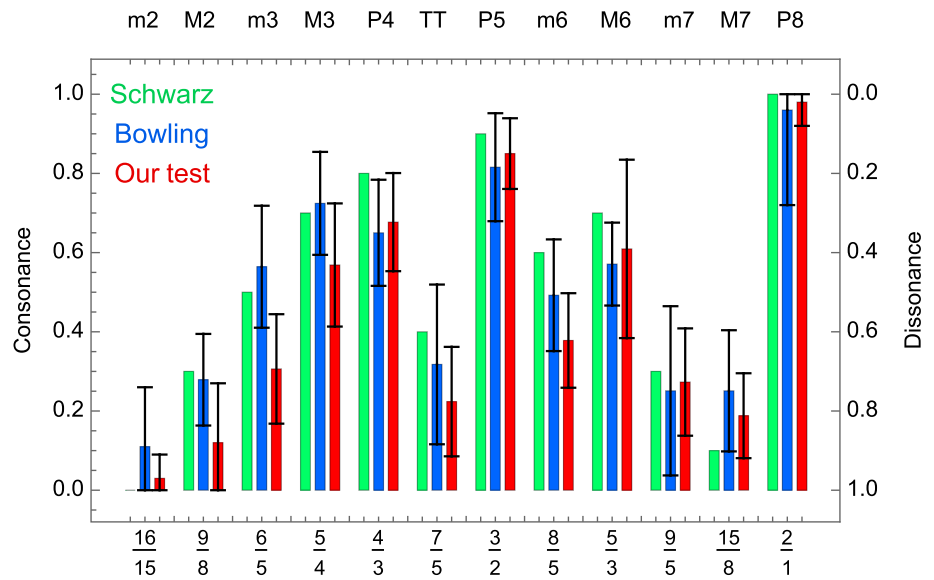
After a few centuries, the major sixth (M6) and major third (M3) were upgraded to the status of imperfect consonances, later followed by the minor sixth (m6) and minor third (m3). In the just scale, these dyads correspond, respectively, to the ratios

¹ We do not consider the category of cultural familiarity since, as will be discussed in more details later, this aspect is not independent on the methodology followed to perform the test. Its impact can be minimized by using a neutral timbre for the sounds and by asking participants not to interpret C & D according to the subjective concept of pleasant/unpleasant or familiar/unfamiliar, but rather according to the more literal and objective concept of unity/division.

² The formal relation between frequency, string length, tension and string mass density was to be derived much later by the scientists of the 17th century, including M. Mersenne and G. Galilei.

³ The Pythagorean school admitted only integers up to 4. The unison (P1) is omitted as it is not strictly speaking a dyad.

Fig. 1 Comparison of the results of three 12 intervals tests: Schwartz et al. [15], Bowling et al. [17] and our test. The error bars represent the standard deviation



$f_2/f_1 = 5/3, 5/4, 8/5, 6/5$. Dissonant intervals were considered to be the major and minor seconds and sevenths (M2, m2, M7, m7) and the tritone (TT), which in the just scale are characterized by the ratios displayed in Fig. 1.

Apart from the dyad’s classifications encoded in Western musical theory and composition manuals (which might slightly differ according to the authors and their epochs), the first attempts to provide an objective ranking to the degree of C & D of a dyad date back to the beginning of the 19th century, as reviewed in Ref. [5]. In the following, we review the results of a few former tests, discuss the motivations which led us to perform our own test, and comment on the associated results.

2.1 Former tests

In their work, Schwartz et al. [15] review the results of a few tests, performed in the years from 1897 to 1918; the most relevant among these old tests is probably the one carried out in 1918 by Malmberg [16] (it was for instance adopted as reference test in the work by Hutchinson and Knopoff [13] about roughness models). Schwartz et al. [15] provide median values for the results of these old tests: We report them in Fig. 1, after normalization by using a scale from 0 to 1. The rankings, in order of decreasing consonance (or, equivalently in order of increasing dissonance), are P8, P5, P4, M3 and M6, m6, m3, TT, m7 and M2, M7, m2. The steps in decreasing consonance being all equal, it is unclear if the data just represent some ordering in ranking or if they can be interpreted as absolute values. In addition, there are no estimates of error bars, nor information about the type of sounds used. In any case, the results turn out to agree quite well with the classical ordering mentioned previously.

A detailed recent test is the one carried out in 2018 by Bowling et al. [17], which deserves to be mentioned and used as reference for comparison (as done for instance in the many models analysis of Ref. [7]). They tested all 12 dyads, 66 triads, and 220 tetrads that can be formed using the intervals specified by the chromatic scale over one octave. In order to generate the sounds, they used a synthesized piano, with just intonation ratios; the fundamental frequency of the tones in each chord was adjusted so that the mean value of the fundamental frequencies of the tones was middle C. They asked their 30 participants, 15 musicians and 15 non-musicians to give a rank from 1 to 4, in increasing order of consonance, defined as the musical pleasantness or attractiveness of a sound. Their results for the mean values and the associated standard deviations can be found in the supplemental material of Ref. [17]; we reproduce their results for the dyads in Fig. 1, adapting them to a scale from 0 to 1. One can notice that the rankings are quite standard, except for the low scores of the P4, M6 and m6 intervals. So, they find the ordering: P8, P5, M3, P4, M6 and m3, m6, TT, M2, M7 and m7, m2. This results in relatively high scores attributed to the M3, considered to be more consonant than the P4, and to the m3, considered to be as consonant as the M6. We think that the exceptional performances of the M3 and m3 might be explained as follows: Since the high rankings for the M3 and m3 were attributed especially by the group of musicians, the pleasantness of the thirds might be due to a bias associated with cultural familiarity for this group (especially for pianists, as we directly experienced in our test).

It is worth to emphasize that, in the perceptual tests, it is assumed that C & D are not independent quantities: They are related in such a way that their numerical values are complementary (and not, for instance, inversely proportional). In particular, using values in the range [0, 1], C & D turn out to be complementary to one, $C + D = 1$.

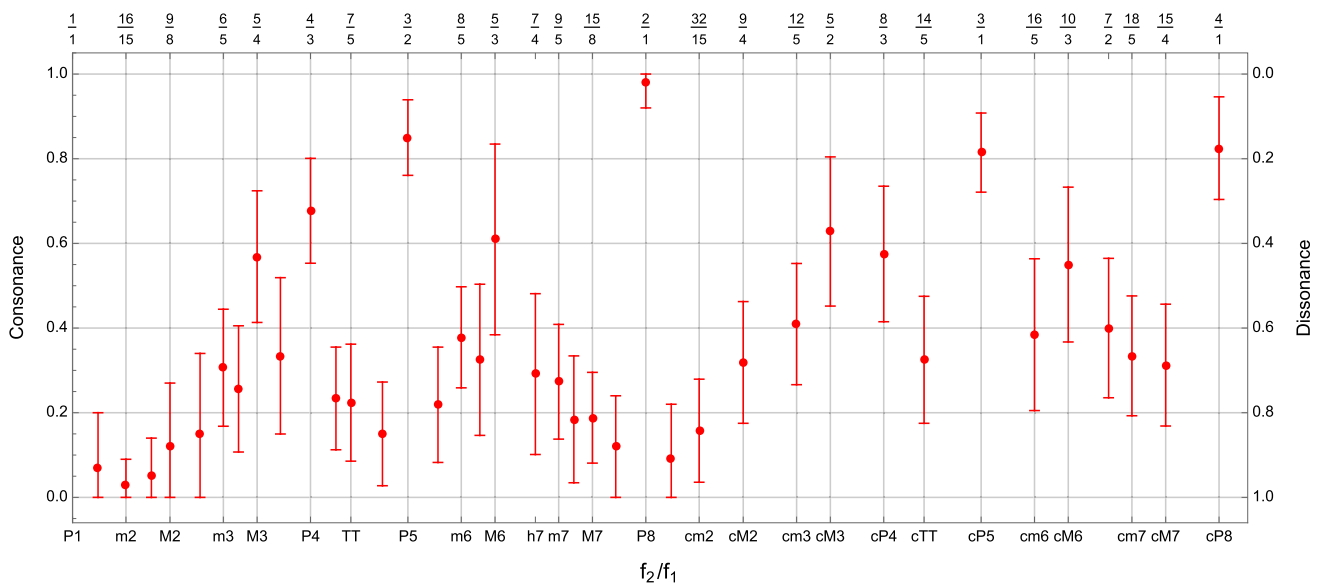


Fig. 2 Results of our 38 intervals test. The error bars correspond to the standard deviation

2.2 Our test

In order to avoid possible biases due to a peculiar timbre of an instrument, we carried out our test on dyads by generating complex tones with 10 harmonics, having weights (Fourier coefficients) decreasing as $1/n$. This choice allows to have a tone with a richer harmonic content than for instance, in the case of a plucked string, for which the weights decrease as $1/n^2$. A different decay time for each harmonic was modeled by an exponential, $e^{-nt/\tau}$, with $\tau = 2$ s. The latter exponential factor is found also in the case of a plucked string and is associated to energetic reasons (the shortest wavelength modes dissipate energy more rapidly than the lowest ones). Still, the choice of $\tau = 2$ s allows to have a tone with a sufficiently persistent harmonic content in its attack. The sounds lasted 2.7 s, with the last 100 ms of the sound smoothed by a Gaussian envelope (to avoid “click” effects in the end). These tones were then used as reference stimuli for our test. Since their timbre does not correspond to the timbre of any real instrument, we denote such timbre as a “neutral” one, for short.

We generated 38 dyads, taking the fundamental frequency of f_1 as middle C (namely $f_1 = 261.63$ Hz), and adjusting f_2 according to the frequency ratios shown in Fig. 2. As we will show in the following, for the sake of discriminating among the models of C & D, it is useful to extend the test to intervals beyond the octave and to microtonal intervals. For the first octave, we selected the 12 microtonal intervals with frequency ratios⁴: 33/32, 11/10, 7/6, 11/9, 9/7, 11/8, 16/11, 14/9, 18/11, 7/4, 11/6, 27/14. We thus collected 24 data points for the first octave. For the second octave, we selected the 12 compound intervals (for which we adopt the notation cm2, cM2, etc.), plus a couple of particularly significant microtonal intervals⁵ with frequency ratios 33/16 and 7/2.

We considered 20 participants, of varying ages and musical education, among our students and colleagues. Each was asked to evaluate the C & D according to a scale from 1 to 5, corresponding to very dissonant, mildly dissonant, not dissonant nor consonant, mildly consonant, very consonant. To minimize the impact of cultural familiarity, we emphasized not to use the subjective categories of pleasant/unpleasant and familiar/unfamiliar, but rather to judge about the (in principle more objective) degree of unity/division of the sound stimulus, namely whether the two sounds are perceived to mix well together or not. Upon rescaling the results using a scale from 0 to 1, we reproduce in Fig. 2 the mean value of the score given to each interval, with the error bar corresponding to the standard deviation. The error bar is asymmetric when it would exceed the range from 0 to 1. The numerical values are collected in Table 5 of Appendix 1. Due to the psychoacoustic nature of the test, we expect that the error bars would not be reduced significantly by adding more participants.

Before analyzing the results for the microtonal and compound intervals, it is worth to compare our results for the standard 12 intervals with the results of the previously mentioned tests, as done in Fig. 1. In our test, the intervals that obtained a score higher than 0.3 are, in decreasing order of consonance, P8, P5, P4, M6, M3, m6: We thus obtained the standard ordering, similarly to Schwartz et al. [15]. The dissonant intervals are, in increasing order of dissonance, m7, m3, TT and M7, M2, m2. Notice the modest performance of the m3 and M3 in our test, with respect to the other tests.

We now focus on the results for the microtonal and compound intervals, shown in Fig. 2. The pattern of scores for the second octave is not equal to that of the first octave: The cP8 is less consonant than the P8; the cM3 is more consonant than the cP4; the

⁴ We selected these ratios as they are relatively simple fractions and correspond to tones which are quite in the middle of two adjacent tones of the just intonation chromatic scale.

⁵ For the second octave, we did not add other microtonal intervals in order to make the test not too heavy.

dissonances of the compound intervals are softer than the corresponding simple ones. Compactness and roughness models have in general different predictions for simple and corresponding compound intervals; however, the periodicity/harmonicity models based on the concept of “pitch class” [18–20] do not, by construction: For this reason, we will not consider them in the following analysis.

The most suitable statistical test to check the agreement between our datasets and the predictions of a specific (compactness or roughness) theoretical model is the chi-square test. Previous analysis, like for instance Ref. [7], considered the partial correlation between model predictions and average consonance ratings: This methodology is less robust than the chi-square, as it loses the important information on the error bars. Considering the linear correlation between the ordering in the ranks of the model and the one of a dataset, as done in Ref. [21], is an even weaker statistical test.

3 The compactness approach and related models

As already mentioned, the compactness approach includes two categories of models that can be denoted as periodicity and harmonicity, for short. The periodicity category corresponds to the models that relate the perception of C & D to the compactness of the waveform of the dyadic signal, according to the criterion that the shorter (longer) is the “global period” of the dyadic signal with respect to the periods of the composing sounds, the more consonant (dissonant) is the interval. This idea was first quantitatively formulated by G. Galilei under the form of a commensurability (or coincidence) theory [2], building on previous arguments by G. B. Benedetti [4]. As a result of the induced scientific debate [4, 5], this was later declined in a variety of models: Some of them going in the direction of arithmetical guesses [9, 21, 22] some others being fully rooted on physical grounds, like for instance the autocorrelation recurrence model of Ref. [23]. The hypothesis that consonance increases when the number of coincident harmonics of the dyad is the largest gave rise to the harmonicity category, to which the model of Ref. [10] belongs. As we will argue, these two approaches give practically the same indicators and can be considered equivalent.

The basic results for the periodicity/harmonicity models are quite general: The most consonant intervals are predicted to be those whose frequency ratios can be expressed as a fraction involving the smallest integer numbers. All other intervals share more or less the same degree of dissonance.

An important feature of the C & D indicators for compactness models is that they are discontinuous functions of the ratio f_2/f_1 . Historically, this gave rise to criticism against these indicators [5], but also led to developing procedures to modify them in such a way to obtain a smooth, thus more physical, curve. Another criticism that was raised is related to the (apparently negligible) role of the phase differences in the tones composing the dyad [4], which is still an open issue [1, 24].

In the following, we first review in a somewhat original way the periodicity models, showing that the many indicators that might be suggested are different declinations of the same basic idea. Then, we turn to the harmonicity models, showing that the associated indicators are essentially the same as for periodicity.

3.1 Periodicity models

We now develop in some detail the approach based on periodicity by considering two sounds with frequencies f_1 and f_2 , such that $f_1 \leq f_2$. The sounds might be complex or simple, that is they might have or not higher harmonics; in case of complex sounds, we consider a harmonic spectrum. In particular, we take the lower frequency f_1 at some fixed value (for instance, the middle C of an equally tempered piano, like in our test), and let the higher frequency f_2 be equal to M/Nf_1 , where M and N are integer numbers. We can express the ratio $f_2/f_1 = M/N \geq 1$ in an equivalent way, but using the smallest possible integer numbers. Defining indeed

$$n = \frac{N}{\text{GCD}[N, M]}, \quad m = \frac{M}{\text{GCD}[N, M]}, \tag{1}$$

where GCD stands for the greatest common divisor, the ratio of the frequencies of the dyad becomes

$$\frac{f_2}{f_1} = \frac{m}{n} \geq 1, \tag{2}$$

with m and n not having prime factors in common.

With this notation, the frequency of the dyad’s waveform, f_0 , is simply

$$f_0 = \frac{f_1}{n} = \frac{f_2}{m}, \tag{3}$$

while its global period is $T_0 = 1/f_0$. Notice that f_0 corresponds to the missing fundamental of the dyad: Obviously, it is not present in its spectrum, but its relevance as *fundamental bass* in music theory is well known [9]. The basic idea of the periodicity approach is that the highest is the degree of consonance, the shorter is the period T_0 or, equivalently, the higher is the frequency f_0 . Of course, f_0 must be compared with the composing frequencies f_1 and f_2 . Since there are various ways to establish this comparison, there will be many possible periodicity indicators. We now introduce and study quantitatively a few of them.

Comparing f_0 with the dyad's lowest sound frequency, f_1 , we have as periodicity indicator the ratio

$$I_1^P = \frac{f_0}{f_1} = \frac{1}{n}. \quad (4)$$

This extremely simple indicator, taking values from 0 to 1, suggests that the P1, P8, cP8 and all higher octaves have the highest mark, 1; the P5 follows with mark 1/2; then the P4 and the M6 get the same mark, 1/3; the M3 has 1/4; the m3, m6 and m7 have mark 1/5; the M2 and M7 have mark 1/8. The results for the 24 chromatic intervals of the first two octaves are shown in Fig. 3. This indicator turns out to be not fully satisfactory: As for the first octave, the equal ranking of P4 and M6 is for instance problematic; furthermore, the high values that the indicator assumes for the cP5 and cP8 are not supported by the results of our test. The comparison of the model results with the perceptual data will be better discussed in the following.

Comparing f_0 with the dyad's highest sound frequency, f_2 , one has as periodicity indicator the ratio

$$I_2^P = \frac{f_0}{f_2} = \frac{1}{m}. \quad (5)$$

This number corresponds to the fraction of “concordant pulses,” proposed *in nuce* as an indicator by G. Galilei [2] himself⁶: Such fraction is indeed the inverse of the “number of strikes” made by f_2 inside a period of f_0 . The highest mark, 1, is obtained only by the P1; the P8 gets 1/2; the P5 gets 1/3; the P4 gets 1/4; the M6 and M3 get 1/5; the m3 gets 1/6; the m6 gets 1/8; the m7 and M2 get 1/9; the M7 gets 1/15. The results for the first two octaves are shown in Fig. 3. Notice that the cP8 gets mark 1/4, as the P4. These rankings are quite acceptable and the Galilei indicator, in its simplicity, turns out to be a surprisingly good one.

One might however guess that a better choice would be to consider the ratio of f_0 and some mean value of f_1 and f_2 . Various indicators can be obtained, according to the choice of the mean: For the arithmetic, geometric and harmonic means, $f_A = (f_1 + f_2)/2$, $f_G = \sqrt{f_1 f_2}$ and $f_H = f_2^2/f_G$, we have

$$I_A^P = \frac{f_0}{f_A} = \frac{2}{n+m}, \quad I_G^P = \frac{f_0}{f_G} = \frac{1}{\sqrt{nm}}, \quad I_H^P = \frac{f_0}{f_H} = \frac{1}{2} \left(\frac{f_0}{f_1} + \frac{f_0}{f_2} \right) = \frac{n+m}{2nm}, \quad (6)$$

as shown in Fig. 3. It can be seen that these three indicators give quite similar predictions, which are (not surprisingly) intermediate with respect to the predictions of the two indicators I_1^P and I_2^P . The harmonic mean indicator turns out to correspond, up to a factor 1/2, to the indicator suggested by A. Frova in Ref. [9], where it is introduced as an effective quantity related to energetic arguments about higher harmonics, thus with a different rationale with respect to ours.

In order to get more insights, we have to directly compare each model predictions with the test results. So, we have to i) normalize the indicators to a scale from 0 to 1, in such a way that the P8 and the most dissonant interval have, respectively, mark 1 and 0; ii) extend the indicator to the “continuum,” namely to non-integer numbers N and M , possibly implementing the effect of the frequency discrimination power of the hearing system.

The first task is straightforward, as it just requires to introduce the normalized consonance indicator as

$$\tilde{I}_X^P(f_2/f_1) = I_X^P(f_2/f_1)/I_X^P(2), \quad (7)$$

where $X = 1, 2, A, G, H$.

The second task is more complex and was addressed in many different ways, from adopting arbitrary mathematical simplifications [9, 21] to exploiting signal processing techniques [19]. We now discuss in some detail our analytical procedure for the extension to the continuum, based on the concept of discrimination limen (DL) [25]. Suppose that m and n take all integer values from 1 up to 50, for instance. We select the k ratios of the type m/n falling in the interval [1, 4], and we denote them by x_i , with $i = 1, \dots, k$. The associated normalized consonance indicator, $\tilde{I}_X^P(x_i)$, is thus defined only for the k discrete values x_i . Our aim is to extend the indicator to any value of the x domain in the interval [0, 4], thus rendering it more “physical.” We recall that the ear has a DL of about 3 Hz at the frequency of middle C (or C₄), which increases up to 6 Hz two octaves above [25]. We propose to simulate the effect of the DL by smoothing the peaks with a Gaussian⁷ characterized by a standard deviation equal to the DL at frequency f_2 , $\sigma = f_{DL}(f_2)/f_1$ (the latter choice will be supported in the following, in connection with equal temperament). Now, the DL turns out to be about 1/30 of the critical bandwidth (CB) [26], as derived by Zwicker et al. [25]. As the CB is frequency dependent, we

⁶ From the Day One of Ref. [2] (our translation): “Consonants, and with delight received, will be those pairs of tones which will strike with some order the eardrum; which order looks for, first, that the pulses made within the same interval of time, shall be commensurable in number, so as not to keep the eardrum in perpetual torment, [...]. The first and most pleasing consonance will be, therefore, the octave since, for every pulse given to the tympanum by the lower string, the sharper string delivers two, so that they strike once simultaneously and once not, for each vibration of the sharper string, so that among the entire number of pulses **one-half** agree to strike simultaneously. [...]. The fifth is also a pleasing, since for every two vibrations of the lower string, the sharper one gives three, so that numbering the vibrations of the sharper string, **one-third** of them agree to strike simultaneously, i.e., two single pulses intervene between each pair of concordant pulses; and in the diatesseron (fourth), three single pulses intervene. In the second, that is in the sesquioctave (with ratio 9/8), it is only every **ninth** pulses of the sharper string which reaches the ear in concordance with one of the lower; all the others are discordant and produce a harsh effect upon the recipient ear which interprets them as dissonants”.

⁷ Our choice of a Gaussian is reasonable but, of course, it is not the unique possibility.

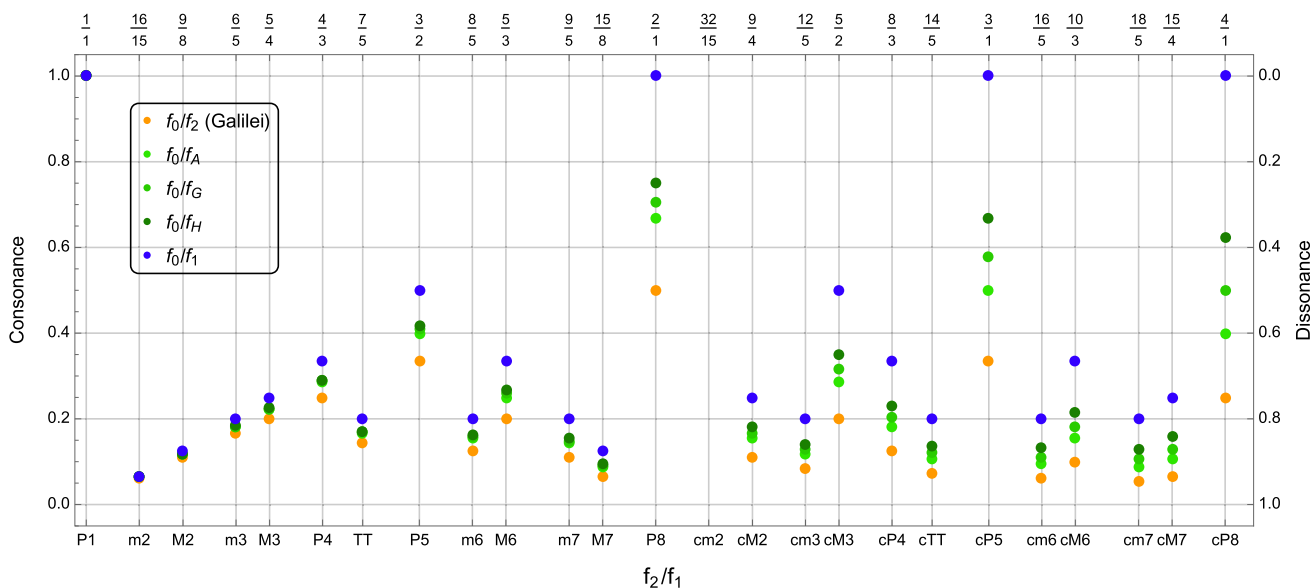


Fig. 3 Periodicity indicators of C & D, as defined in Eqs. (4), (5) and (6): from top to bottom, I_X^P is displayed, with $X = 1, H, G, A, 2$

fix for definiteness f_1 to C_4 . The extension to the continuum can be made by calculating the distance $|x - x_i|$, for all x_i such that $|x - x_i| < 2\sigma(x)$, and evaluating the corresponding value for the periodicity consonance indicator $C_X^P(x)$, defined as

$$C_X^P(x) = \text{Max}_i \tilde{I}_X^P(x_i) e^{-\frac{(x-x_i)^2}{2\sigma(x)^2}}. \tag{8}$$

The result is a continuous function, with smoothed peaks such that, within (beyond) about 3 (6) Hz from a peak, the consonance function does not (might) change significantly. We performed a psychoacoustic test to check that this is indeed the case: For example, focusing on the peak of the P5, within 3 Hz from the peak we found, according to our judgment, that the quality of the perceived consonance is not much affected despite the appearance of some roughness, while beyond about 6 Hz, we assessed that the perception turns into dissonance.

The continuous function C_X^P for the harmonic mean and Galilei indicators, having $X = H, 2$, is shown in Fig. 4. We can see that the general trend of the data is well reproduced but, to be precise, there are various weak points. The model’s predictions are in general lower than the data points. As for the first octave, the consonance of the P5 and the P4 is underestimated by all models; the Galilei indicator turns out to be the best one here. In addition, none of the models explains the trend of the dissonances close to P1. As for the second octave, the harmonic mean accounts well for the cP5 and cP8, but the other data points are missed.

These considerations are supported by the reduced chi-square analysis. By using the results of our test with 38 intervals, we calculate the reduced chi-square values for all the periodicity models previously introduced and display them in Table 1. Values of reduced chi-square smaller (larger) than 1 signal a good (poor) agreement between model and observations. The periodicity models turn out to be quite accurate: The values of the chi-square are slightly bigger than 1. In particular, the best models turn out to be the harmonic and geometric mean ones. For comparison, we also show the values of the chi-square obtained using the smaller subset of the 12 standard intervals of the first octave: In this case, the Galileo model turns out to be the best one.

So far, no difference was introduced for simple and complex tones. The harmonic spectrum of two complex harmonic sounds is given by the two sets of frequencies $\{n_i f_i\}$, with $i = 1, 2$ and $n_i = 1, 2, 3, \dots$. The missing fundamental frequency f_0 of all such frequencies is the same as in the case of pure sounds: This justifies the use of the indicators previously introduced for both simple and complex sounds.

However, one might want to slightly modify the indicators in order to account for the increased consonance effect induced by harmonics. For instance, the periodicity indicator with $X = 1$ might be generalized to a sum of all the ratios of the kind $f_0/(n_1 f_1)$, with weights w_n (in some relation with the Fourier coefficients of the complex sound),

$$\frac{f_0}{f_1} \rightarrow \sum_{n_1=1,2,\dots} w_{n_1} \frac{f_0}{n_1 f_1} = \frac{f_0}{f_1} \left(1 + \sum_{n_1=2,\dots} \frac{w_{n_1}}{n_1} \right) \equiv \frac{f_0}{f_1} (1 + N), \tag{9}$$

where $w_1 = 1$ and N is a constant. The normalized generalized indicator is obtained by dividing the above expression by the value that it takes at the octave, namely $1 + N$: One thus ends up precisely with the original indicator, f_0/f_1 . Similar considerations apply for all other periodicity models. This shows that the inclusion of the increased consonance effect induced by harmonics requires a dedicated approach, to be discussed in the following.

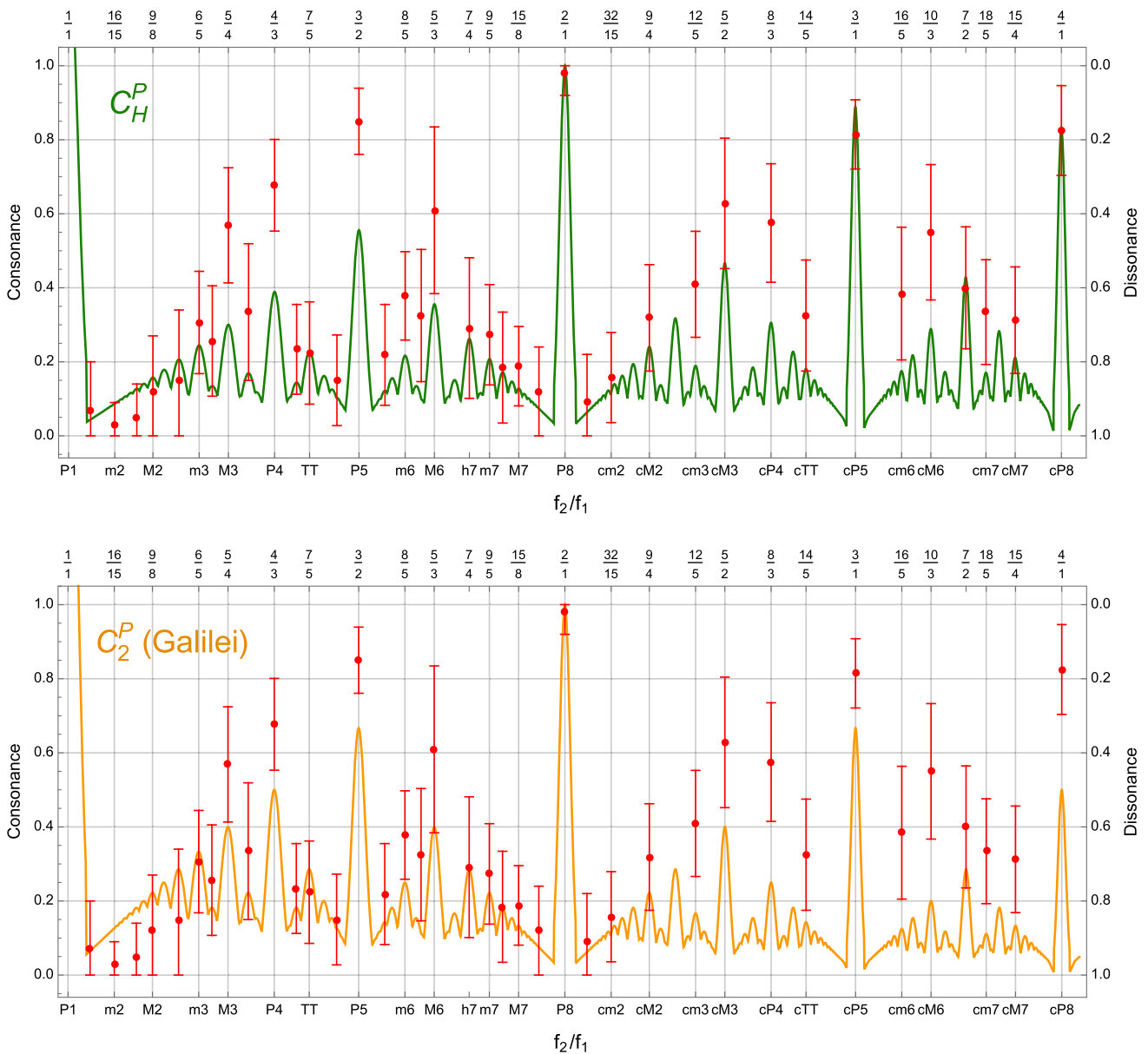


Fig. 4 Extension to continuum for two periodicity models: We show the function C_X^P , with $X = H, 2$, from top to bottom, and compare it with the data from our test

In the appendix, Sect. B, we comment on some other proposed periodicity indicators that cannot be simply rooted on physical grounds (in the sense that they cannot be expressed as simple functions of f_0/f_1 and f_0/f_2) and should rather be seen as mathematical guesses: This is the case for the Euler’s dissonance indicator called “gradus suavitatis” [22], and for the “relative periodicity” dissonance indicator suggested in Ref. [21].

3.2 Harmonicity models

The similarity of the harmonic spectrum, that is to what extent the harmonics of the two complex sounds of the dyad are in common, is another criterion to express the compactness of the dyadic signal. It thus provides an alternative way to define a compactness consonance indicator: This is the harmonicity approach, that we now discuss in detail.

A coincidence for the harmonics happens if the relation $n_1 f_1 = n_2 f_2$ can be fulfilled for some n_1 and n_2 . In the case it is, the lowest coincidence of the harmonics happens for the smallest possible values of n_1 and n_2 , to be denoted by n_1^{c1} and n_2^{c1} . Since the relation above implies

$$\frac{n_1}{n_2} = \frac{f_2}{f_1} = \frac{m}{n}, \tag{10}$$

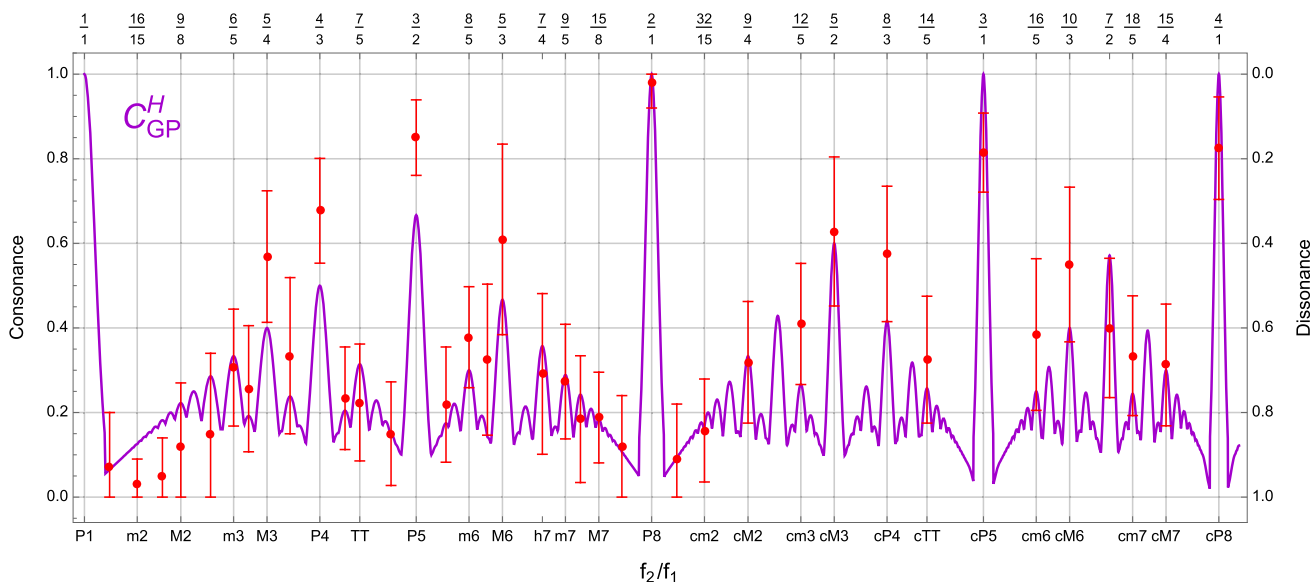


Fig. 5 Extension to continuum for the harmonicity model of GP [10]: We show the function C_{GP}^H and compare it with the data from our test

where m and n are already as small as possible, we learn that the first coincidence happens for $n_1^{c1} = m$ and $n_2^{c1} = n$. The second coincidence happens for $n_1^{c2} = 2m$ and $n_2^{c2} = 2n$, the third for $n_1^{c3} = 3m$ and $n_2^{c3} = 3n$, and so on.

A consonance indicator based on harmonicity is thus some function that increases when $n_1^{c1} = m$ and $n_2^{c1} = n$ are as small as possible or, equivalently, when $1/n_1^{c1} = 1/m$ and $1/n_2^{c1} = 1/n$ are as large as possible. This is precisely the same request as in the case of the periodicity indicator: This shows that the two approaches are basically equivalent.

A couple of explicit examples of harmonicity indicators are for instance the functions

$$\frac{1}{2} \left(\frac{1}{n_1^{c1}} + \frac{1}{n_2^{c1}} \right) = \frac{m+n}{2mn} \quad , \quad \frac{1}{n_1^{c1} n_2^{c1}} = \frac{1}{mn} \quad , \tag{11}$$

where the first is indeed equal to the harmonic mean periodicity indicator, $I_H^P = f_0/f_H$, while the second is equal to the square of the geometric mean periodicity indicator, $I_G^P = f_0/f_G$. These two examples explicitly show that the criterion of having n_1^{c1} and n_2^{c1} as small as possible is equivalent to asking f_0 to be as large as possible when compared to f_1 and f_2 .

Anyway, one might guess that the similarity of the harmonic content of the two sounds, hence the associated indicator, should also depend on the amplitudes of the harmonics, namely on timbre. Denoting the weight function for the contribution of the n_i -th harmonics by w_{n_i} with $i = 1, 2$, one can build an indicator along these lines as

$$I_w^H = \sum_{n_1, n_2=1,2,\dots} w_{n_1} w_{n_2} \delta(n_1 f_1 - n_2 f_2) = w_{n_1^{c1}} w_{n_2^{c1}} + w_{n_1^{c2}} w_{n_2^{c2}} + \dots \quad , \tag{12}$$

where δ is the Dirac delta function. In particular, using $w_{n_i} = 1/n_i^\alpha$, where α is a positive real number, one has

$$I_\alpha^H = \frac{1}{(n_1^{c1} n_2^{c1})^\alpha} + \frac{1}{(n_1^{c2} n_2^{c2})^\alpha} + \dots = \frac{1}{(mn)^\alpha} \left(1 + \frac{1}{4^\alpha} + \frac{1}{9^\alpha} + \dots \right) \equiv \frac{1}{(mn)^\alpha} N_\alpha. \tag{13}$$

The largest contribution to the sum comes from the first coincidences, n_1^{c1} and n_2^{c1} . This harmonicity indicator is thus related to the geometric mean periodicity indicator, $I_\alpha^H = (I_G^P)^{2\alpha} N_\alpha$. The associated normalized and extended to continuum consonance indicators are related by

$$C_\alpha^H = (C_G^P)^{2\alpha}. \tag{14}$$

As an explicit example, we take $\alpha = 1/3$ and estimate the related chi-square for our 38 intervals test: The agreement between this model and the perceptual observation is even better than for periodicity models, as can be seen from Table 1.

The general result is thus that harmonicity models are equivalent, even numerically, to some combination of periodicity ones. This is the case also for the ‘‘percentage similarity,’’ an harmonicity consonance indicator proposed by Gill and Purves [10]

$$C_{GP}^H = \frac{m+n-1}{mn}, \tag{15}$$

which is already normalized to range $[0, 1]$. Notice that $C_{GP}^H = 1$ for all the frequency ratios such that $n = 1$, namely those of the type $m/1$, where $m = 1, 2, 3, \dots$. This indicator turns out to be related to the harmonic and geometric mean periodicity indicators as follows:

$$C_{GP}^H = 2I_H^P - (I_G^P)^2. \quad (16)$$

It was however proposed as harmonicity indicator, with the physical meaning of expressing, in relation to the harmonic expansion of the missing fundamental f_0 , the fraction of harmonics that are in common with the dyad, over the total number of them, up to the first coincidence between f_1 and f_2 . This can be seen as follows. The first coincidence happens at $n_1^{c1} = m$ and, since $f_1 = nf_0$, it corresponds to the (mn) -th harmonic of the series of f_0 . The harmonic series of f_0 and f_1 have m overlapping harmonics below mf_1 , while the harmonic series of f_0 and f_2 have n overlapping harmonics below nf_2 ; in both cases, the highest overlap is the first coincidence of the harmonic series of f_1 and f_2 : The number of common harmonics among f_0 and the dyad is thus $m + n - 1$. The ratio among the common harmonics and all harmonics, up to the first coincidence,⁸ is thus given by Eq. (15): It represents the strength in the perception of the missing fundamental f_0 (that for the sounds of our test corresponds also to the “virtual pitch” [18]).

The extension to the continuum for the Gill and Purves harmonicity consonance indicator, C_{GP}^H , is shown in Fig. 5. One can see that there is an overall agreement with the data points, but that the model underestimates some consonances in the first octave (like the P5 and P4), while it overestimate others in the second octave (like cP5 and cP8). The related chi-square values are displayed in Table 1. It turns out that, for the full set with 38 intervals, this model is even slightly better than the best periodicity models.

Other models related to harmonicity are those that, building on the concept of virtual pitch [18], define some distance between “pitch class sets” [19, 20]. As already mentioned, these models consider intervals and compound intervals to be equivalent by construction, ending up with the same predictions for the first and second octave. As this is against observational evidence from our test and common experience, here, we do not further consider these models.

4 The roughness approach and related models

The roughness approach to explain C & D was proposed by Helmholtz [6] and later refined by many subsequent studies, among which Refs. [11–13, 24, 27–31]. The related consonance function is naturally continuous but, as will be discussed in the following, the roughness models reproduce the perceptual data more poorly than the compactness models. Here, we review the main results of the previous literature models and study two directions for possible improvements.

Plomp and Levelt (PL) [11] incorporated into the Helmholtz roughness approach the effect of the CB [26], relying on the data collected by Zwicker, Flottorp and Stevens (ZFS) [25]. According to these data, the CB is well fitted by the curve

$$b(\bar{f}) = (0.003) \bar{f}^{1.47} + 90) \text{ Hz}, \quad (17)$$

where $\bar{f} = (f_1 + f_2)/2$, and f_1, f_2 are the beating frequencies; in the range 100–500 Hz, the CB is nearly constant and equal to 100 Hz, and then increases proportionally with frequency. PL found that the maximal roughness between two pure tones is not fixed at 33 Hz as assumed by Helmholtz, but it rather corresponds to intervals of about 25% of the CB. They modeled the dissonance of a dyad of pure tones by introducing a function (see Fig. 10 of Ref. [11]), to be called $g(z)$, such that

$$d(f_1, f_2) = g(z), \quad (18)$$

where $z = |f_1 - f_2|/b(\bar{f})$. The curve $g(z)$ vanishes in $z = 0$, has a maximum at $z = 0.25$ and reaches smoothly zero at $z = 1$. Multiple parameterizations of such curve have been given by various authors [28, 31, 32], but none of them (including the PL original curve) takes into account the DL, as they display a non-vanishing value for $dg(z)/dz$ in $z = 0$. The DL is approximately 1/30 of the CB and corresponds to $z_{DL} = 1/30$. In order to account for the DL, here, we adopt a slight modification of the polynomial fit proposed by Dillon [31], such that

$$g(z) = 0.761(1 - e^{-\frac{z^2}{0.06^2}}) \quad \text{for } 0 \leq z \leq 2z_{DL}, \quad g(z) = 4.906z(1.2 - z)^4 \quad \text{for } 2z_{DL} < z \leq 1.2, \quad (19)$$

with $g(z) = 0$ for $z > 1.2$.

Following the approach of Helmholtz [6], the dissonance of two complex tones with harmonic series $\{n_1 f_1\}$ and $\{n_2 f_2\}$, where n_1 and n_2 are integer numbers taking values from 1 up to n_{\max} , is obtained by adding the dissonance values due to the beatings between

⁸ One might wonder what happens to the indicator of Eq. (15) by considering harmonics up to the k -th coincidence. Actually, there is no change. Indeed, the k -th coincidence happens at $n^{ck} = km$ and, since $f_1 = nf_0$, it corresponds to the (kmn) -th harmonic of the harmonic series of f_0 . The harmonic series of f_0 and f_1 have km overlapping harmonics below kmf_1 , while f_0 and f_2 have kn overlapping harmonics below knf_2 ; the highest overlap is the k -th coincidence of the harmonic series of f_1 and f_2 : The number of common harmonics among f_0 and the dyad is thus $km + kn - k = k(m + n - 1)$. Both the numerator and the denominator in Eq. (15) get a factor k , so that finally C_{GP}^H remains precisely the same.

Table 1 Reduced chi-square values for representative periodicity and harmonicity models

	C_1^P	C_H^P	C_G^P	C_A^P	C_2^P	$C_{1/3}^H$	C_{GP}^H
$\tilde{\chi}_{38}^2$	1.4	1.1	1.1	1.2	1.3	0.8	0.7
$\tilde{\chi}_{12}^2$	2.7	2.0	1.8	1.6	1.1	1.4	1.0

The first line refers to our 38 intervals test, the second to its subset with the 12 standard intervals of the first octave

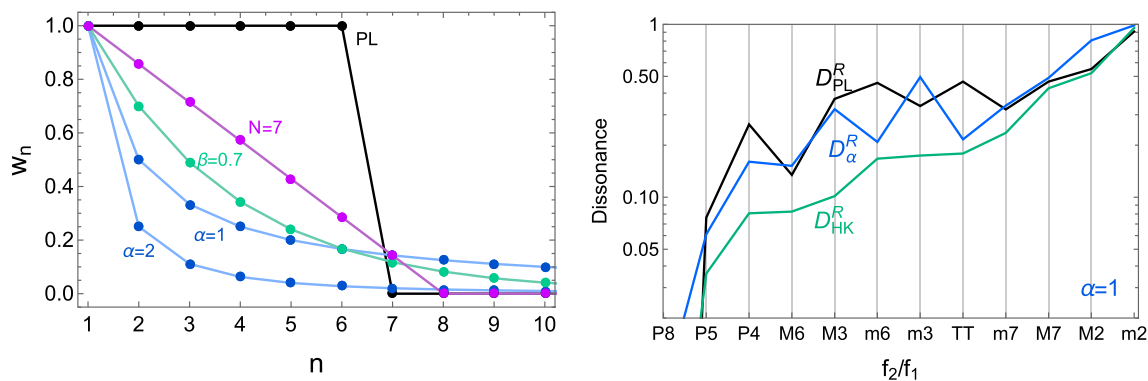


Fig. 6 Left: Values of the weight functions (described in the text), for $\alpha = 2, \alpha = 1, \beta = 0.7, N = 7$. Right: Dissonance values, in logarithmic scale, for the models of Plompl and Levelt (PL) [11], Hutchinson and Knopoff (HK) [13], and for the model with $\alpha = 1$ described in the text

all pairs of harmonics. For the sake of generality, we allow for different weights for the contribution of the various harmonics: The roughness dissonance indicator for the Y model is thus

$$D_Y^R = \frac{1}{N_Y} \sum_{n_1, n_2=1, 2, \dots}^{n_{\max}} w_{n_1}^Y w_{n_2}^Y d(n_1 f_1, n_2 f_2), \tag{20}$$

where Y is some parameter characterizing the weight function w_n^Y , and the denominator N_Y is a proper normalization introduced to obtain values for D_Y^R in the range $[0, 1]$. Without losing generality, one can always choose the weight of the dyad’s fundamental frequencies f_1 and f_2 to be equal to one. According to the roughness approach, the consonance encodes the absence of roughness due to first-order beatings between all pairs⁹ of harmonics:

$$C_Y^R = 1 - D_Y^R. \tag{21}$$

In their work [11], PL considered two complex tones with six harmonics of equal weights, as shown in Fig. 6, and used the CB of ZFS [25]. The consonance function for the PL model, C_{PL}^R , is shown in Fig. 7, fixing f_1 to the frequency of the middle C. Consonant intervals turn out to be related to simple frequency ratios and the hierarchy of decreasing consonance is quite reasonable; however, the model predictions do not agree with the results of the test, being often too high. In addition, the rankings display three problems: The M6 is too consonant, even more than the P4, while it is common to expect the opposite; the M3 is more dissonant than the m3, while it is common to have the inverse; the consonance increases too much for widely separated intervals spanning more than one octave. On top of this, as for the consistency of the model, the choice of using only six harmonics implies that some simple frequency ratios (those having an integer greater than 6 in the numerator or denominator) are missed, because there is no local enhancement corresponding to them [28]; by increasing the number of harmonics while keeping equal weights for all of them, more peaks appear, but at the price of having an even less realistic hierarchy in order of decreasing consonance. For all these reasons, many attempts were made to improve the PL model.

Among the many studies appeared after PL [11], a relevant contribution was done by Hutchinson and Knopoff (HK) [13], who addressed some of the problems above by including weights for the harmonics. In particular, they considered $n_{\max} = 10$ and the weight function $w_n^\alpha = 1/n^\alpha$, with $\alpha = 1$, as shown in Fig. 6. They did not use the ZFS [25] CB, as done by PL, but a much smaller one, based on older data sets.¹⁰ We denote the consonance function for this model by C_{HK}^R and display it in Fig. 7, for a direct comparison with the roughness model of PL. One can see that the HK curve is even more squeezed toward the highest values than the PL one.¹¹ In order to better visualize the results of the HK model, we show in the right panel of Fig. 6 a plot (inspired by a similar one of Ref. [13]) showing the dissonance in a logarithmic scale, while the abscissa displays the intervals according to a classical ranking. The HK model predicts a better ordering for the rankings with respect to the PL model: The rankings for the P4 and M6 are equivalent, and the M3 is more consonant than the m3. From the work of HK, one cannot infer whether this is due to the introduction of the weights or rather to the different choice of CB.

⁹ We do not include the auto-dissonance of a complex tone as this would provide a small constant term. For instance, even assuming $w_n = 1$, with $n_{\max} = 7$, the auto-dissonances of C_4 and C_5 are $\mathcal{O}(10^{-3})$ and $\mathcal{O}(10^{-2})$, respectively.

¹⁰ For pure tones for instance, taking a m3 interval with f_1 fixed at C_4 , the PL and HK models predict, respectively, a significant and a negligible amount of roughness: $d(f_1, f_2) = 0.56$ for PL, and $d(f_1, f_2) = 0.13$ for HK. Our direct experience with beatings at the frequencies of such example supports the PL choice of the CB.

¹¹ Despite this, the model of HK turned out to be the most successful one among all roughness and periodicity models studied in Ref. [7], when confronted to the test results of Bowling et al. [17]. The reason might be the following: If one calculates the logarithm of the HK dissonance model prediction and, after a suitable normalization and rescaling, compares it with the linear Bowling data without the data point of the P8, then one finds a reasonable agreement for the other data points.

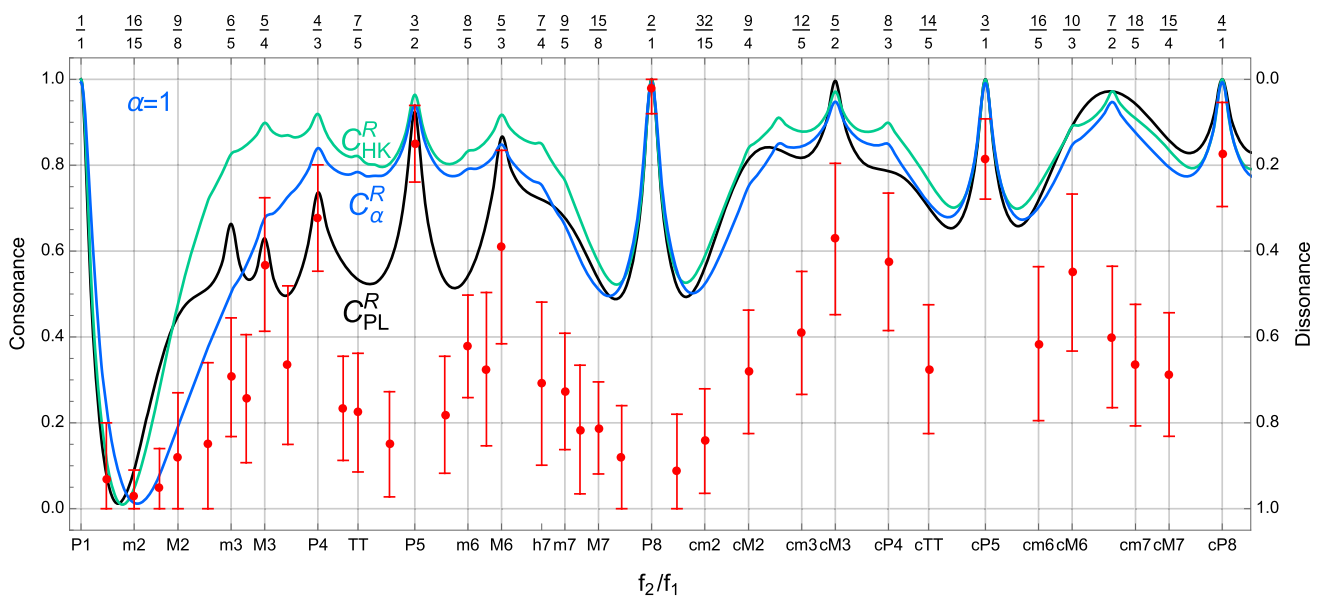


Fig. 7 A few roughness models: Plomp Level (PL) [11], Hutchinson and Knopoff (HK) [13], and the model with $\alpha = 1$ described in the text. The frequency f_1 is fixed to C_4

In order to clarify this point, we calculate the consonance of Eq. (21), taking $n_{max} = 10$, weights of the type $w_n^\alpha = 1/n^\alpha$ with $\alpha = 1$ and use the CB of ZFS [25]: The consonance function C_α^R of this model is shown in Fig. 7. The problems with the ordering of P4 and M6 are alleviated with respect to the PL model, while those of the ordering between M3 and m3 are solved: These improvements are thus due to the introduction of a more suitable weight function, rather than to the choice of CB made by HK. Yet, some of the predictions of the α model with $\alpha = 1$ are very problematic: For instance, the TT and the m6 are more consonant than the M3, as can be seen also by inspecting the right panel of Fig. 6. So, it is clear that this model is not fully satisfactory either. Notice also from Fig. 7 that all three models predict a too large consonance for many dyads, in particular those of the second octave.¹²

Table 2 displays the results of the calculation of the reduced chi-square for the three roughness models introduced so far. The values obtained are definitely larger than those obtained for the compactness models.

Notice that the α model with $\alpha = 1$ can be considered to be the reference roughness model in relation to our data points, which were obtained by using sounds with 10 harmonics and Fourier coefficients decreasing as $1/n$. The relation between the “perceived timbre” given by the weights in Eq. (20) and the true timbre of the sounds (given by its Fourier coefficients) being unknown, the simplest working hypothesis that can be formulated is that they are equal. One should anyway be aware that the weight functions in Eq. (20) are of psychoacoustic type, as they encode all the processing done by the hearing system, also at neuronal level [1].

In the following, we study whether a particular model or some generalization of the roughness approach can provide better results. We first explore various weight functions, with the property that the consonance results are stable with respect to the addition of more harmonics. As a second direction, we consider the possibility of accounting not only for the primary beats of the mistuned unison, as done so far, but also for the well-known secondary beats of mistuned consonances [14], like the mistuned octave and fifth.

We start by considering various forms for the weight functions, in particular those of the types $w_n^\alpha = 1/n^\alpha$, $w_n^\beta = 1/\beta^{(n-1)}$ and $w_n^N = 1 - (n - 1)/N$, which are shown in Fig. 6 for the particular values $\alpha = 2$, $\alpha = 1$, $\beta = 0.7$ and $N = 7$. With respect to the reference weight function with $\alpha = 1$, the weight function with $\alpha = 2$ simulates a perceived timbre poorer in harmonics, while the weight functions with $\beta = 0.7$ and $N = 7$ simulate a perceived timbre richer in harmonics. Figure 8 shows the predicted consonance for these models. One can see that the richer is the perceived timbre, the more the peaks are defined; in addition, the results of the models with a richer perceived timbre and those of the reference model are not very different, apart from the regions close to the P5 and P8. As a general result, the P4 turns out to be at most as consonant as the M6, while the M3 results more consonant than the m3. Even with a richer perceived timbre, these models predict too much consonance, in particular for the second octave, so that many of the data points of our test are missed and the results of the chi-square do not improve significantly, as can be seen in Table 2.

Secondly, we consider the effect of including the mistuned consonances associated with the P8 and P5 [14]: Indeed, one can guess that the impact of adding these secondary beats might be significant. Such inclusion can be made by adding to the dissonance associated to first-order beatings, $d(f_1, f_2) = g(z)$, the dissonance associated to the second-order beatings, which we model by

$$d_{B2}(f_1, f_2) = c_8 [g(z|_{f_1 \rightarrow f_1/2}) + g(z|_{f_1 \rightarrow 2f_1})] + c_5 [g(2z|_{f_1 \rightarrow 2/3f_1}) + g(2z|_{f_1 \rightarrow 3/2f_1})], \tag{22}$$

¹² An ingenious (but ad hoc) solution was proposed in Ref. [27], where it was suggested to introduce a global weight factor with the function of suppressing the predictions for the largest intervals.

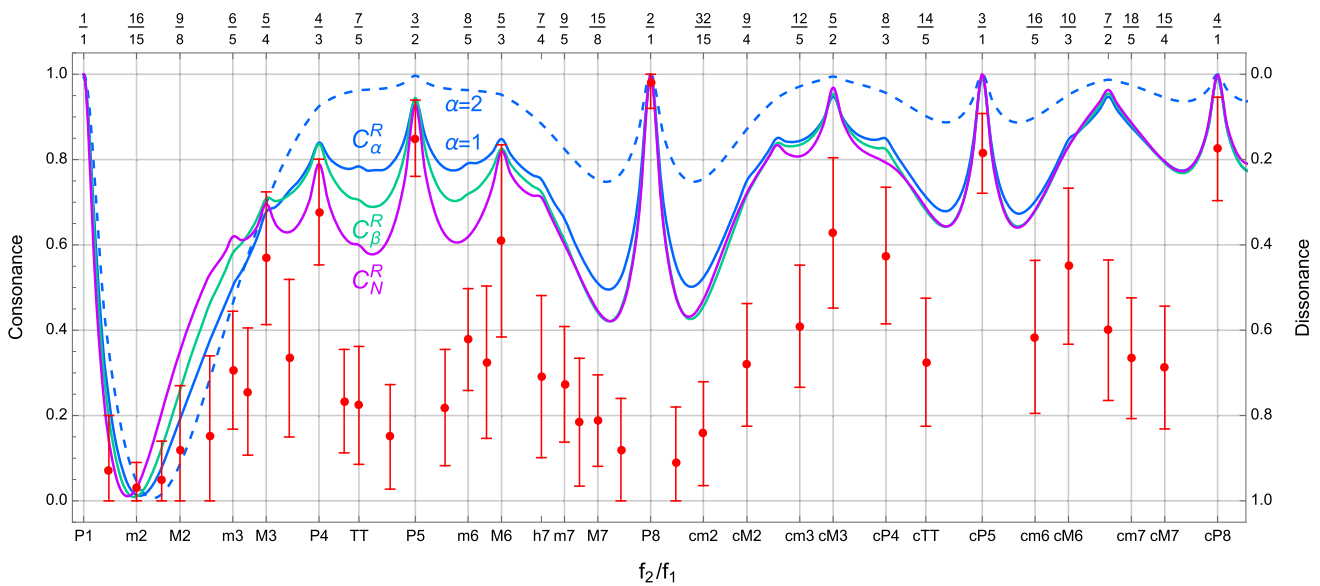


Fig. 8 Roughness models and the effect of various weight functions: From top to bottom $\alpha = 2, \alpha = 1, \beta = 0.7, N = 7$. We take $n_{\max} = 10$ and use the CB of ZFS

where c_8 and c_5 are real coefficients, to be determined experimentally. We performed a related perceptual test and found that reasonable values can be considered to be $c_8 \sim 1/2$ and $c_5 \sim 1/4$. Notice that, while the CB for the mistuned unison and the mistuned octave are experimentally comparable, the CB for the mistuned fifth is smaller. According to our test, the latter is about $1/2$ of the CB for the mistuned unison: This is the origin of the factor $2z$ in the last two terms of the equation above. We do not introduce other well-known mistuned consonances, like for instance the mistuned fourth as, according to our test, its effect is too weak. The dissonance of complex tones can be evaluated by replacing d with $d + d_{B2}$ in Eq. (20) and properly normalizing the resulting function to the range $[0, 1]$:

$$D_{YB2}^R = \frac{1}{N_Y} \sum_{n_1, n_2=1, 2, \dots}^{n_{\max}} w_{n_1}^Y w_{n_2}^Y (d(n_1 f_1, n_2 f_2) + d_{B2}(n_1 f_1, n_2 f_2)). \tag{23}$$

To assess the impact of this extension, we consider the α and N models, with $\alpha = 1$ and $N = 7$. In both cases, we take $n_{\max} = 10$ and use the CB of ZFS. The bottom panel of Fig. 9 displays three models denoted by C_α^R , for which $c_8 = c_5 = 0$; $C_{\alpha 8}^R$, for which $c_8 = 1/2$ and $c_5 = 0$; $C_{\alpha 85}^R$, for which $c_8 = 1/2$ and $c_5 = 1/4$. We can see that the introduction of the mistuned octave is effective in explaining the data points around the P8, where there is a sharp peak emerging over many dissonant intervals; has a good but insufficient impact on the peak of the P5 and cP5; helps in suppressing the M6 with respect to the P4. The further introduction of the mistuned fifth has a negligible impact in general, apart from the intervals around the P5 itself, that emerges better over the intervals between the TT and the m6; the suppression is however not enough to explain the data (further increasing the coefficient c_5 would not be justified as the beats of the mistuned fifth are subdominant). In the bottom panel of Fig. 9, we focus on the three models denoted by C_N^R, C_{N8}^R and C_{N85}^R , corresponding to the same choices of c_8 and c_5 done in the upper panel. In this case, the peak around the P5 is better defined, but there is no significant improvement in reproducing the data points.

Table 2 shows the results of the calculation of the reduced chi-square for these improved models. Considering the full data sets with the 38 intervals, one can see that the impact of the mistuned octave and fifth induces a significant decrease in the value of the chi-square which, however, does not go below 2.7, even for C_{N85}^R , the best performing model. We conclude that, globally, the results of the roughness models, when improved by using proper weight functions and especially by including the effects of the mistuned octave and fifth, are encouraging, but cannot explain the data points in a fully satisfactory way.

5 The combined effect of compactness and roughness

We just concluded that, despite the many good features, roughness models do not reproduce the data points in a satisfactory way. It seems that some other ingredient has to be introduced: Indeed, the focus of roughness models is to assign penalties, rather than prizes. On the contrary, the compactness models provide essentially prizes for the simplicity of the waveform of the signal, but do not assign increasing penalties to increasingly non simple ratios of dyad's frequencies. A complete model of C & D should both give prizes for the compactness of the signal and penalties for the presence of beats.

Table 2 Reduced chi-square values for representative roughness models, for our 38 intervals test and its subset with the 12 standard intervals

	C_{PL}^R	C_{HK}^R	C_{α}^R	$C_{\alpha=2}^R$	C_{β}^R	C_N^R	$C_{\alpha 8}^R$	$C_{\alpha 85}^R$	$C_{N 8}^R$	$C_{N 85}^R$
$\tilde{\chi}_{38}^2$	5.9	9.5	6.9	12.7	5.9	5.2	4.1	2.9	3.3	2.7
$\tilde{\chi}_{12}^2$	3.4	7.6	4.4	9.0	3.5	2.8	2.1	1.3	1.4	1.2

Where unspecified, we took $\alpha = 1, \beta = 0.7, N = 7$

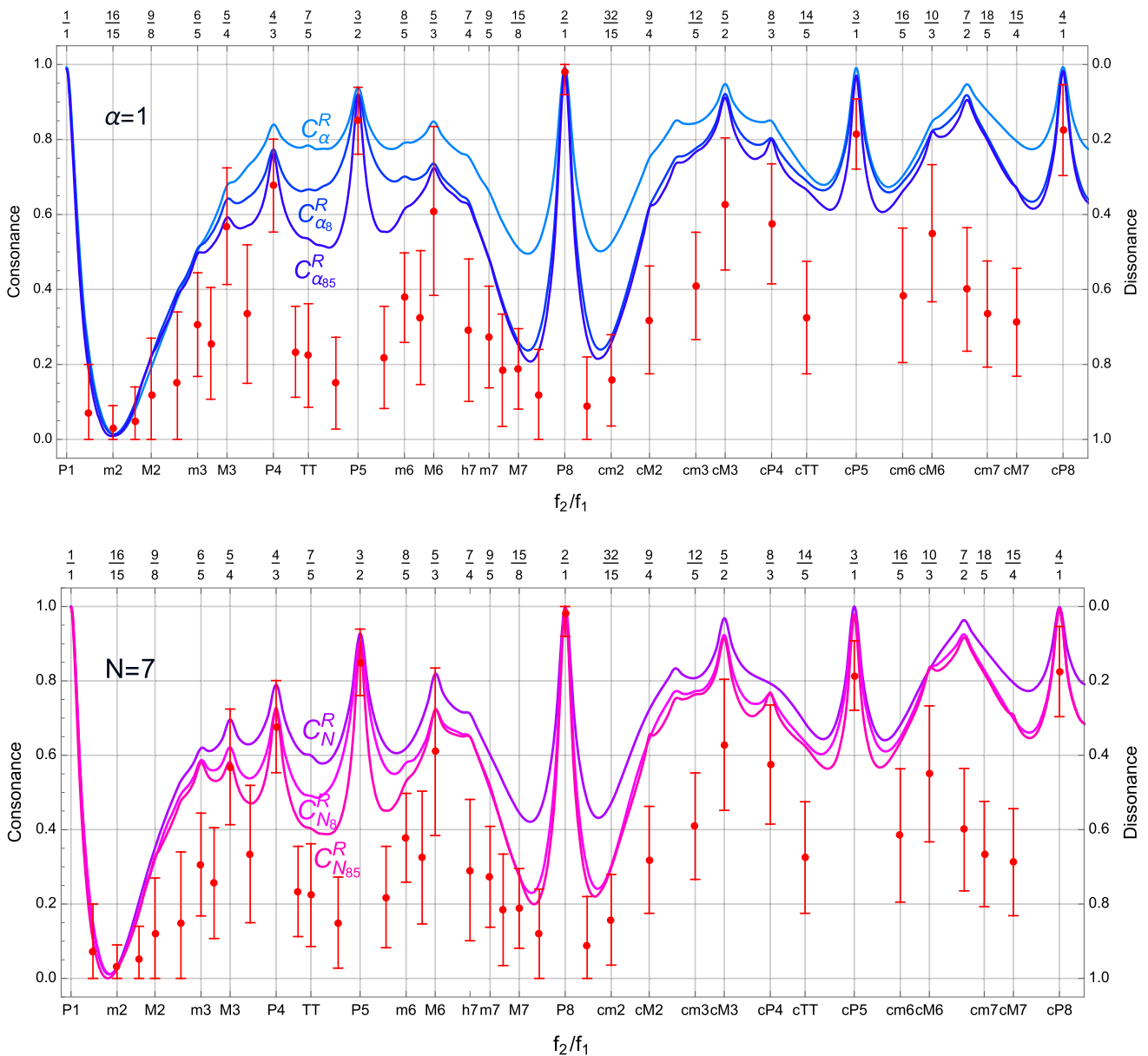


Fig. 9 Roughness models and the effect of the second-order beats. Top: The C_α^R model with $\alpha = 1$ and from top to bottom, $c_8 = c_5 = 0$, $c_8 = 1/2$ and $c_5 = 0$, $c_8 = 1/2$ and $c_5 = 1/4$. Bottom: The same as before but considering the model C_N^R with $N = 7$. We take $n_{\max} = 10$ and use the CB of ZFS

We thus propose to directly combine the two approaches, simply summing two representative indicators in each category. The combined consonance indicator is given by the weighted sum of a compactness model, of the periodicity or harmonicity type, and a roughness one:

$$C_{X,Y}^{\text{tot}} = \frac{F C_X^{P/H} + (1 - F) C_Y^R}{N_{X,Y}}, \tag{24}$$

where X and Y specify the particular compactness and roughness models, respectively, F is the fractional contribution of periodicity/harmonicity with respect to roughness, and $N_{X,Y}$ is a normalization factor. Of course it is not obvious that this automatically corresponds to a better model than its constituents. Only the comparison with the perceptual data can reveal whether this is the case.

We consider all possible combinations of the periodicity/harmonicity models $C_{1,H,G,A,2}^P$ and C_{GP}^H , with the roughness models C_α^R and $C_{\alpha_{85}}^R$, taking $\alpha = 1$: We show the results of the reduced chi-square in Fig. 10, as a function of the parameter F . The solid and dashed lines refer to our 38 intervals test and its subset with the 12 standard intervals, respectively. For large (small) values of F , the combined model is dominated by the compactness (roughness) model constituent.

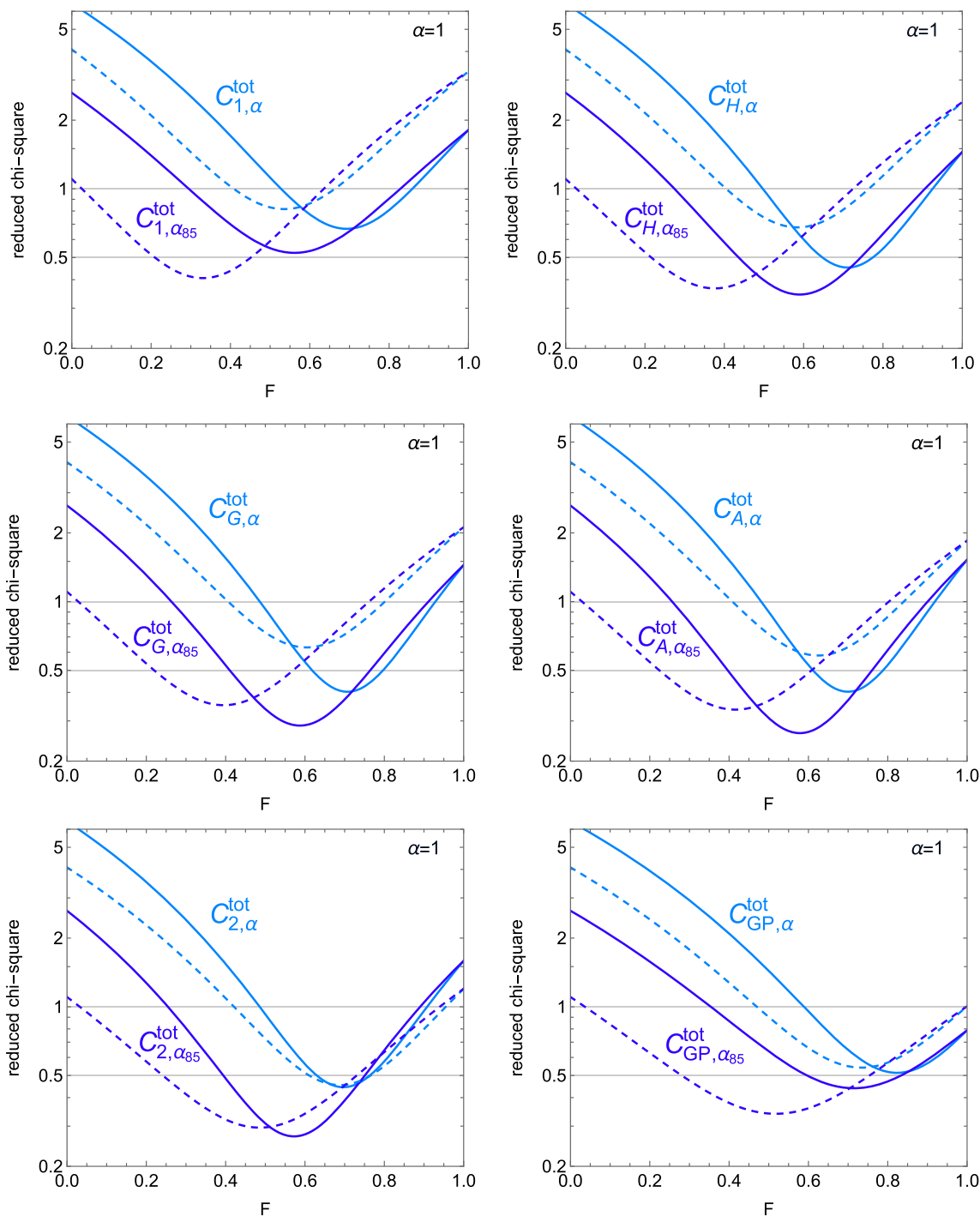


Fig. 10 Values of the reduced chi-square for the indicated combined models, as a function of the parameter F , representing the percentage of the contribution of compactness with respect to roughness. We select all possible combinations of the periodicity/harmonicity models $C_{1,H,G,A,2}^P$ and C_{GP}^H , with the roughness models C_{α}^R and $C_{\alpha_{85}}^R$, taking $\alpha = 1$. Solid and dashed lines refer to our 38 intervals test and its subset with the 12 standard intervals, respectively

We first focus on the results associated to the 38 intervals test (solid lines in Fig. 10). In all the considered cases, it turns out that the reduced chi-square of the combined model has a minimum which is significantly smaller than the chi-square of the constituent models; the inclusion of the mistuned octave and fifth reproduces the data slightly better and is characterized by a lower value of F . The more pronounced minimum is found for the combined models having C_G^P , C_A^P and C_2^P as constituent models, together with $C_{\alpha_{85}}^R$. For these three best performing models, including the effect of the mistuned octave and fifth, the minimum of the chi-square is found at $F \approx 58\%$ and is slightly smaller than 0.3, signaling an impressive agreement between such theoretical models and perceptual

Table 3 Reduced chi-square values for the indicated combined models with $\alpha = 1$ and the indicated range of F , for our 38 intervals test and its 12 intervals subset

	$C_{X,\alpha_{85}}^{\text{tot}}, X = G, A, 2$	$C_{X,\alpha}^{\text{tot}}, X = G, A, 2$	$C_{1,\alpha}^{\text{tot}}$
$\tilde{\chi}_{38}^2$	$\lesssim 0.3, 50\% \lesssim F \lesssim 65\%$	$\lesssim 0.5, 65\% \lesssim F \lesssim 80\%$	$\lesssim 0.7, 65\% \lesssim F \lesssim 75\%$
$\tilde{\chi}_{12}^2$	$\lesssim 0.4, 30\% \lesssim F \lesssim 50\%$	$\lesssim 0.7, 50\% \lesssim F \lesssim 80\%$	$\lesssim 0.8, 50\% \lesssim F \lesssim 60\%$

Fig. 11 Values of the reduced chi-square for our 38 intervals test, as a function of the parameter F , for the three models corresponding to the combination of the arithmetic mean model with three different roughness models including the second-order beats: the models with $\alpha = 1$ (solid blue), $\alpha = 2$ (dashed blue) and $N = 7$ (solid pink)

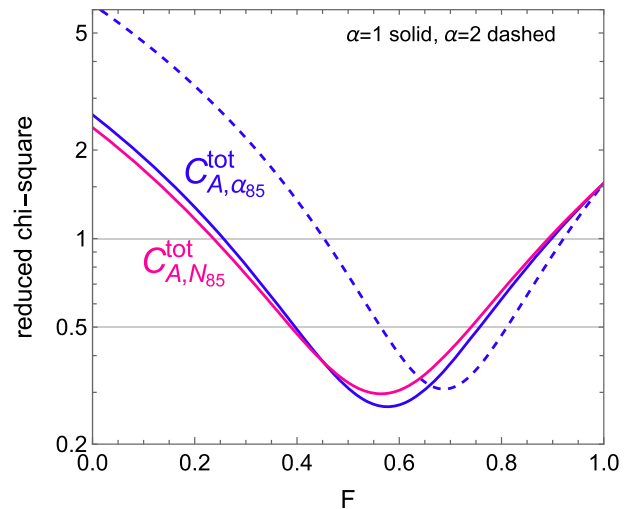


Table 4 Difference between the frequency f_2 of a tone in the equal temperament and the corresponding frequency in the just intonation scale, keeping fixed $f_1 = 261.63$ Hz

	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B
Δf_2 [Hz]	- 1.88	- 0.66	- 2.82	2.60	0.39	3.72	- 0.44	- 3.30	3.96	- 4.76	3.33

observations. The less pronounced minimum is found for the combined models having C_1^P and C_{GP}^H as constituent models, together with C_α^R ; in any case, the minimum is smaller than 0.7. Table 3 summarizes these results. Notice also that, for all models, the minimum of the reduced chi-square is reached for $50\% \lesssim F \lesssim 80\%$.

It is interesting to compare the latter values for the reduced chi-square, with those derived by using only the subset corresponding to the 12 standard intervals (dashed lines in Fig. 10). Also, for this case, it is found that the reduced chi-square of all combined models has a minimum which is significantly smaller than the chi-square of the corresponding constituent models; in addition, the inclusion of the mistuned octave and fifth reproduces the data slightly better and is characterized by a lower value of F . Notice that, in general, $\tilde{\chi}_{12}^2$ is worse than $\tilde{\chi}_{38}^2$ for the compactness-dominated models with large F (except $C_{2,\alpha_{85}}^{\text{tot}}$), while it is better than $\tilde{\chi}_{38}^2$ for the roughness-dominated models with small F . For intermediate values of F , it turns out that the minima of the reduced chi-square are obtained for slightly smaller values of F with respect to those of the 38 intervals case. In any case, the values of the reduced chi-square minima do not vary significantly using the standard 12 rather than the full set with 38 intervals, especially for the best performing models, like $C_{2,\alpha_{85}}^{\text{tot}}$. This is precisely what one expects from a good theoretical model: The reduced chi-square should not vary much by varying the number of experimental points used in the analysis.

These findings show in a robust way that, independently on the model details, *the perception of C & D relies in a cooperation of compactness and roughness, with similar weights*. As discussed, this conclusion holds for both our data set with 38 intervals and its subset with 12 intervals, with the former requiring a slightly larger contribution of the compactness model of choice.

To simulate the uncertainty in modeling the perceived timbre of the sounds of the test, we study the reduced chi-square for the combination of the arithmetic mean model with different roughness models including the second-order beats, like the models with $\alpha = 2$ and $N = 7$, as shown in Fig. 11, where we consider the full 38 intervals test. The model with $\beta = 0.7$ is not shown as it gives results overlapping with the model with $\alpha = 1$. One can conclude that making different choices for the weight functions has a quite small impact on the chi-square. In particular, the choice of the α model with $\alpha = 1$ is a good representative one.

Explicit examples of the C & D combined model functions are shown in Figs. 12 and 13. In particular, in the upper panel of Fig. 12, we show the combination with $F = 58\%$ of the arithmetic mean periodicity model with the $\alpha = 1$ roughness model, with or without mistuned octave and fifth. In general, the agreement with the data points is quite impressive, apart from the slight underestimate of the P5. One can also appreciate the small but relevant effect of including the mistuned octave and fifth. The results

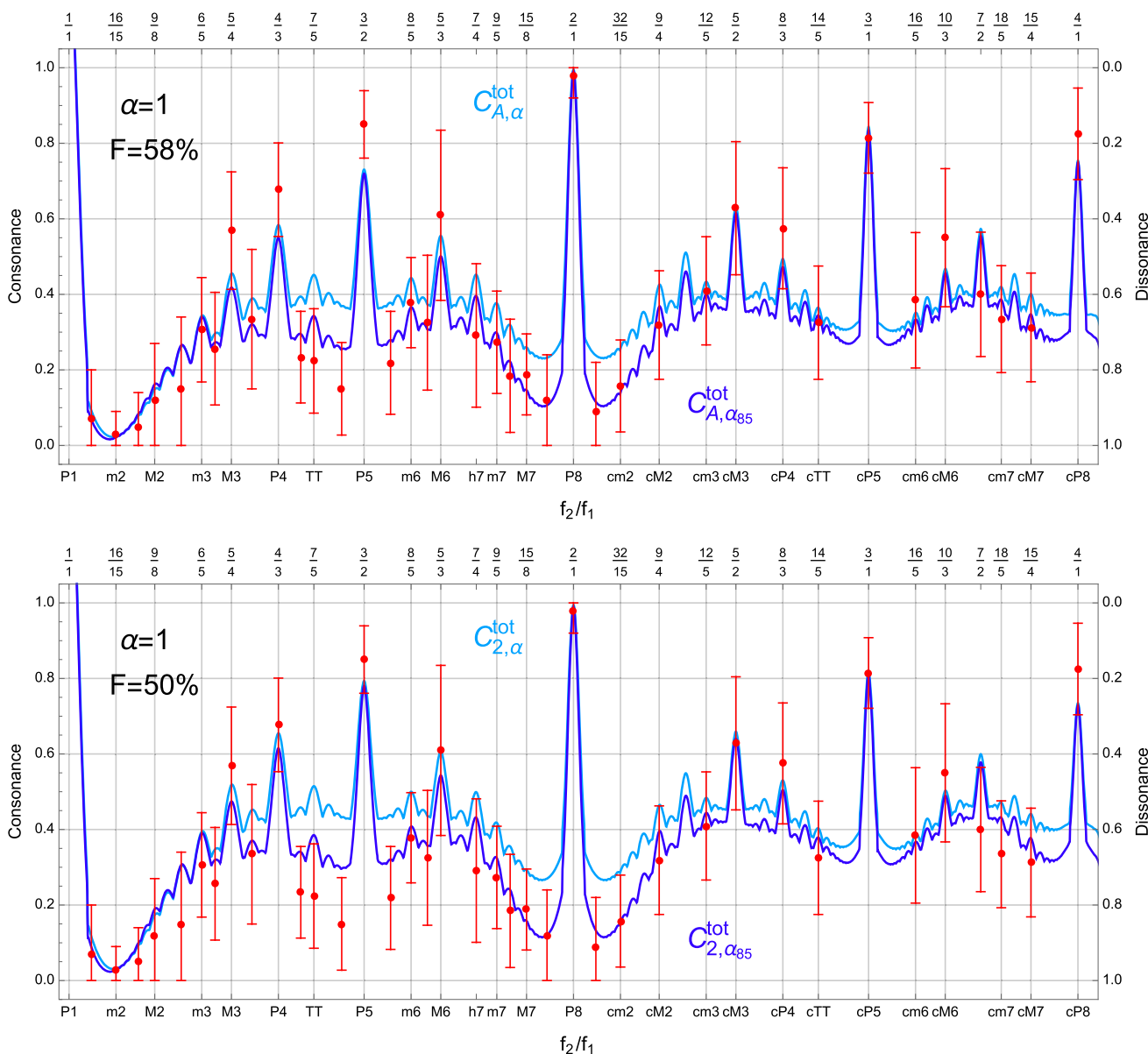


Fig. 12 Combined models of compactness and roughness: The top (bottom) panel shows the consonance function obtained by summing, with the indicated weight F , the arithmetic mean (Galileo) periodicity model with the roughness model with $\alpha = 1$, with and without the inclusion of the mistuned beats of the octave and fifth

of the reduced chi-square (for the 38 intervals test) are 0.3 and 0.6, respectively, with and without the inclusion of second-order beats. Another interesting combined model is shown in the bottom panel of Fig. 12: Here, the compactness model is taken to be the Galileo periodicity model. With the choice $F = 50\%$, the consonance peaks of the first octave are better reproduced. The results of the reduced chi-square (for the 38 intervals test) are again extremely good, 0.3 and 0.9, respectively, with and without second-order beats. This simple combined model turns out to be, in our opinion, very satisfactory.

As further examples, in the upper (lower) panel of Fig. 13 we show the combination with $F = 60\%$ ($F = 50\%$) of the harmonic mean (Galileo) periodicity model with the $\beta = 0.7$ ($N = 7$) roughness model, with or without mistuned octave and fifth. Also in these cases, the agreement with perceptual data is very satisfactory.

We did not try to optimize the values of c_8 and c_5 for the sake of reducing as much as possible the chi-square. Our goal here is to show the importance of combining the compactness and roughness approaches to explain C & D, so that the optimization of these parameters is beyond the goals of this work. However, we checked that c_8 should be quite close to $1/2$ in order to fit the perceptual data; c_5 can be taken in the range $1/6 - 1/3$, as the model predictions depend mildly on such parameter.

A few comments are in order about how to interpret our results in the case of a 12-tone equally tempered scale. We focus for definiteness on the combined model $C_{2, \alpha_{85}}^{tot}$ with $F = 50\%$. In Fig. 14, the values of the consonance function corresponding to the

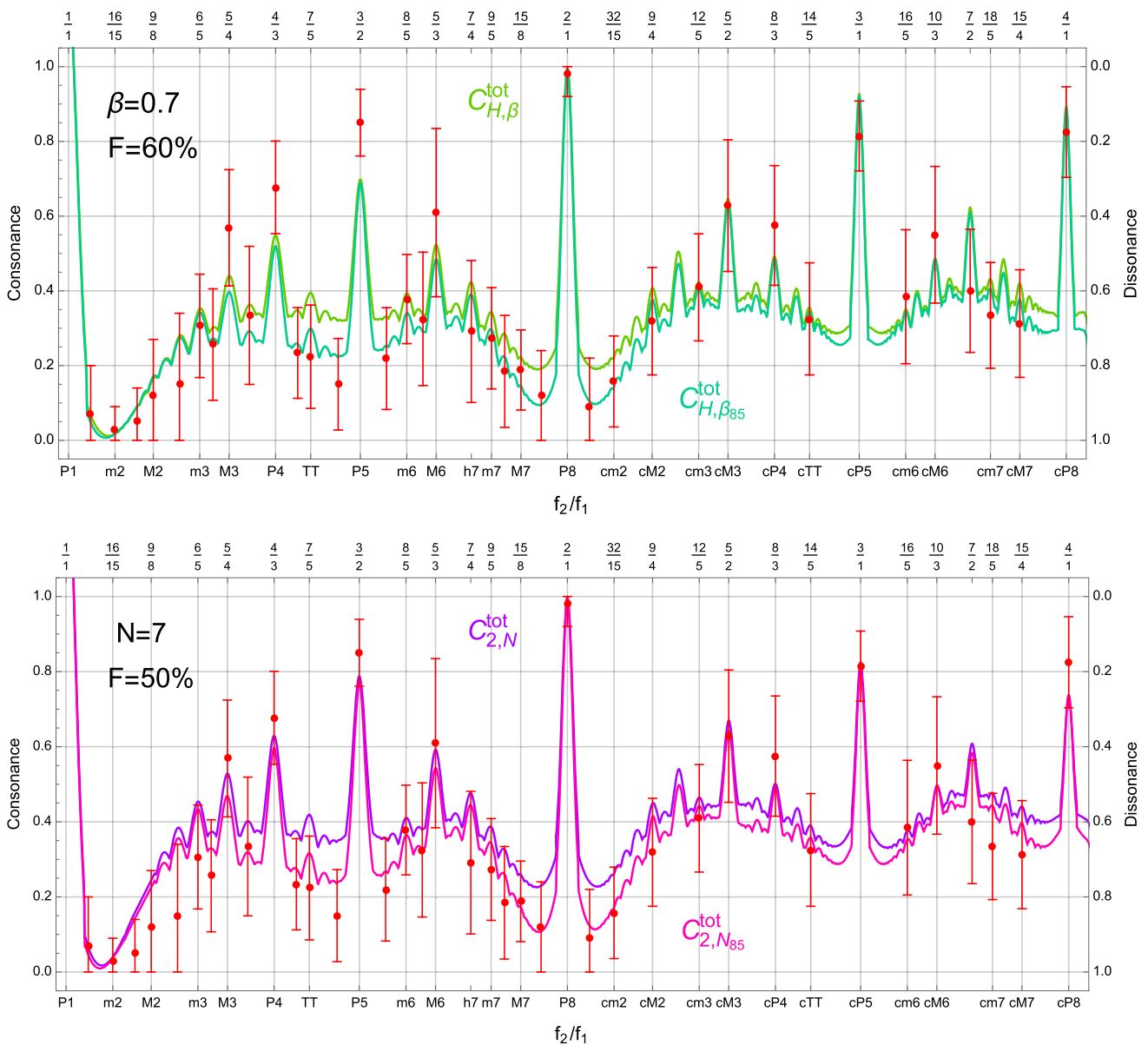


Fig. 13 Combined models of compactness and roughness: The top (bottom) panel shows the consonance function obtained by summing, with the indicated weight F , the harmonic mean (Galileo) periodicity model with the roughness model with $\beta = 0.7$ ($N = 7$), with and without the inclusion of the mistuned beats of the octave and fifth

notes of the 12-tone equal temperament are emphasized with circles. Keeping fixed f_1 at the frequency of middle C, in Table 4, we show, for each tone of the equal temperament in the first octave, the quantity Δf_2 , defined as the difference between the frequency f_2 of the tone in the equal temperament and the corresponding one in the just intonation scale. One can see from Fig. 14 that in general the equal temperament circles are quite close to the just intonation peaks, the larger differences occurring, among the perfect and imperfect consonances, for the M6, M3 and m3: In any case, even these circles remain within the error bars of the test. This result is meaningful because, by direct experience, it is well known that the quality of a consonance is slightly—but not dramatically—worsened in the case of the equal temperament with respect to the just intonation scale, and that, among the perfect and imperfect consonances, the largest effect indeed occurs for the M6, M3 and m3.

Notice that these findings also support our choice of smoothing the peaks with a Gaussian whose standard deviation is identified with the DL, which is about 3 Hz for the first octave above middle C. In the case of a standard deviation significantly smaller than the DL, the peaks in Fig. 14 would be so sharp that the consonance values of the equally tempered intervals would decrease too much, down to the values outside the peaks. In the case of a standard deviation significantly larger than the DL, the peaks would be broader, and the consonance values of the equally tempered intervals would get too close to the peak values of just intonation.

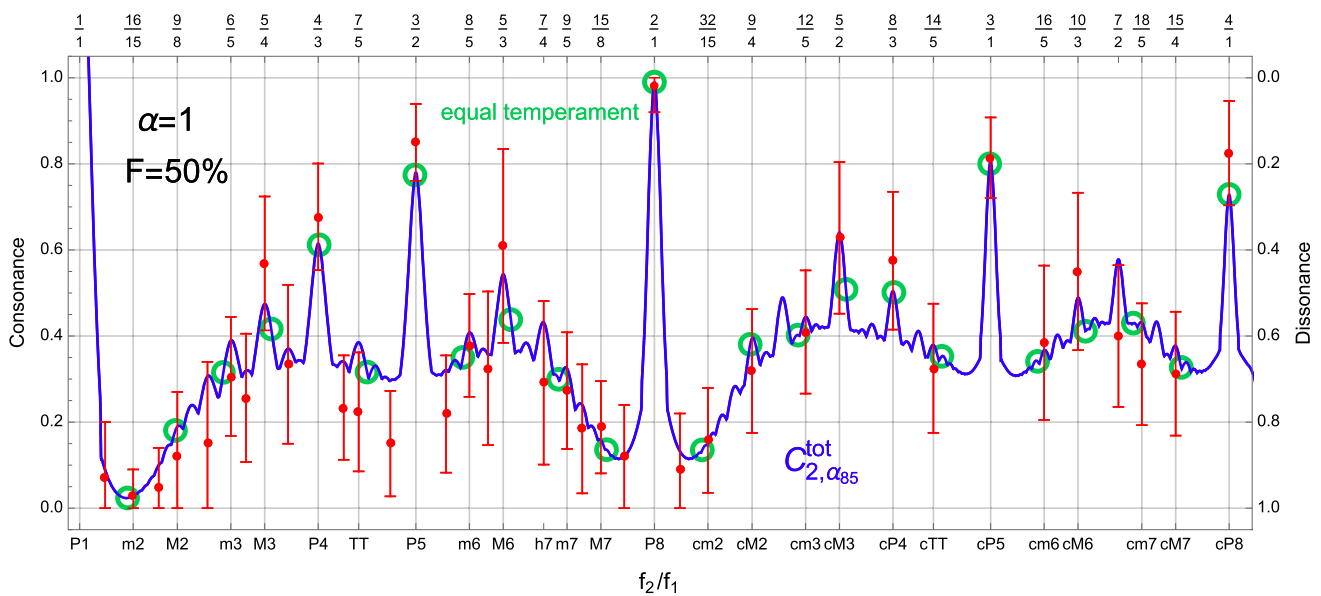


Fig. 14 Combined models of compactness and roughness and 12-tones equally tempered scale: The consonance function $C_{2, \alpha_{85}}^{tot}$ with $F = 50\%$ is shown, emphasizing with circles the values corresponding to the 12 tones of the equal temperament

Finally, one might wonder about the dependence of our results on the tone spectrum. Clearly, a slight change in the harmonic content of the tones is expected to have a slight effect on the test results, so that the chi-square values of the consonance indicators studied here should be slightly affected. On the other hand, expanding the test to include, for example, timbres from several well-known instruments would require to generalize the models, especially the roughness ones, with weights describing arbitrary timbres. This is beyond the goal of the present work and would deserve a dedicated study. In any case, the present results about the necessity of including both compactness and roughness in building a consonance indicator would not be affected.

6 Overview and conclusions

The actual functioning of our hearing system is still mysterious [1]. In Western music, the definition of C & D for a dyad is related to the psychoacoustic perception of its degree of unity and division. This perceived degree should correspond to some objective characteristic of the dyad’s sound signal. Many generations of scientists have been fascinated by this correspondence and took over the challenge of explaining on physical grounds the rankings in the C & D of musical intervals [3–5]. The present work has the goal of giving a contribution to the long standing and still open debate about the effectiveness of physics arguments in explaining the sensations of C & D.

We showed that the proposed approaches are in general based either on some sort of compactness of the dyadic signal waveform or on the absence of roughness due to beatings. The compactness of the waveform can be seen either as the shortness of the waveform period or as the highest number of overlapping harmonics. Notice that the waveform period is called, in psychoacoustics, the period of the missing fundamental, while in musical theory, it corresponds to the period of the *fundamental bass*.

Deeming insufficient the former tests based on the 12 dyads of the natural chromatic scale only, we conducted a test on 38 dyads, including compound and microtonal intervals. The sounds were generated with a “neutral” timbre, composed of 10 harmonics with amplitudes falling as $1/n$. We used the results of this detailed test in order to compare theoretical models with perceptual observations, by means of the chi-square test.

We reviewed in a critical way the proposed approaches based on compactness and roughness, discussing many model realizations for each category. The models were in part taken from the former literature, while others were originally developed here.

As for the periodicity models, we provided a unified description of many possible indicators, including the one based on Galilei’s argument [2], which turns out to be quite accurate. We gave mathematical proofs of the fact that the periodicity and harmonicity models are essentially equivalent. We introduced a simple analytical way to extend the compactness consonance indicators to the continuum, in a way to include the effect of the discrimination limen. As can be seen from Table 1, the values of the reduced chi-square for our 38 intervals test are encouraging, as they are close to unity. However, the compactness consonance functions, as those represented in Figs. 4 and 5, turn out to underestimate many data points of our test. This approach assigns prizes for the simplest frequency ratios, but does not give increasing penalties for increasingly complex ratios.

As for the roughness models [6], we first reviewed in a critical way a few historical models from the literature [11, 13] and then discussed directions for further improvements. In particular, we investigated the role of the weight functions in association with the perceived timbre and the impact of including second-order beatings related to the mistuned octave and fifth. The latter turns out to be particularly relevant. As one can see from Table 2, the results of the reduced chi-square test for our set of 38 intervals, even including second-order beats, are worse than those of compactness models. Figures 8 and 9 show that this is due to the fact that the roughness consonance functions largely overestimate many data points of our test.

We then explored the effect of combining the compactness and roughness approaches, studying many combinations of representative models from the two categories. For this sake, we introduced the parameter F , which encodes the relative weight of the selected compactness model over the roughness one. Figure 10 shows that, for all possible combinations, the value of the reduced chi-square has a minimum smaller than one and significantly lower than the reduced chi-square of the constituent models. Including the beats of the mistuned unison, the minimum is found for $F \sim 50\%$. As it appears from Figs. 12 and 13, the consonance functions of the combined models accurately account for the data points of our test. In addition, as can be seen from Fig. 14, the consonance results for the equal tempered tones are meaningful and realistic.

We thus conclude that the physical properties represented by compactness and roughness are both necessary ingredients for an effective phenomenological explanation of the psychoacoustic sensations of C & D.

Acknowledgements We warmly thank our students and colleagues that kindly accepted to participate in our psychoacoustic test.

Funding Open access funding provided by Università degli Studi di Ferrara within the CRUI-CARE Agreement.

Data Availability Statement The authors declare that data supporting the findings of this study are available within the article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1: Results of our 38 intervals test

The numerical results of our 38 intervals test for a group of 20 participants are collected in Table 5. For each frequency ratio, f_2/f_1 , the mean value and the associated standard deviation are given.

Table 5 Results of our 38 intervals test

f_2/f_1	Mean	SD	Name
33/32	0.067	0.126	
16/15	0.025	0.064	m2
11/10	0.047	0.084	
9/8	0.123	0.152	M2
7/6	0.153	0.186	
6/5	0.306	0.138	m3
11/9	0.256	0.149	
5/4	0.569	0.155	M3
9/7	0.334	0.185	
4/3	0.677	0.124	P4
11/8	0.234	0.178	
7/5	0.224	0.138	TT
16/11	0.15	0.122	
3/2	0.85	0.089	P5
14/9	0.219	0.136	
8/5	0.378	0.119	m6
18/11	0.325	0.178	
5/3	0.609	0.225	M6
7/4	0.291	0.190	h7
9/5	0.273	0.135	m7
11/6	0.184	0.150	
15/8	0.188	0.107	M7
27/14	0.117	0.121	
2/1	0.979	0.058	P8
33/16	0.086	0.130	
32/15	0.157	0.122	cm2
9/4	0.319	0.144	cM2
12/5	0.409	0.143	cm3
5/2	0.628	0.176	cM3
8/3	0.575	0.160	cP4
14/5	0.325	0.150	cTT
3/1	0.814	0.093	cP5
16/5	0.384	0.179	cm6
10/3	0.550	0.183	cM6
7/2	0.400	0.165	ch7
18/5	0.334	0.141	cm7
15/4	0.312	0.144	cM7
4/1	0.825	0.121	cP8

Appendix 2: Other periodicity models

We discuss here a couple of interesting indicators whose expressions are not clearly rooted on physical grounds and, in our opinion, rather represent mathematical guesses.

Despite its name, the Euler’s *gradus suavitatis* [22], d_E , is a dissonance indicator, as it increases with the dissonance. For dyads whose frequency ratio is $f_2/f_1 = m/n$, it is defined as

$$d_E = 1 + \sum_{i=1, \dots, k} m_i (p_i - 1), \tag{25}$$

where p_1, p_2, \dots are the prime factors in the factorization of the product $m n = p_1^{m_1} p_2^{m_2} \dots p_k^{m_k}$, and m_i are integers and nonzero. For the frequency ratios considered in our 38 intervals test, d_E ranges from 2 to 18; one can define a normalized consonance indicator C_E , with values in the range [0, 1], as $C_E = 1 - (d_E - 2)/16$. We calculated the associated reduced chi-square and found that it is equal to 3.0, hence worse with respect to the values displayed in Table 1, found in relation to other compactness models.

Reference [21] introduces two “relative periodicity” dissonance indicators, $d_{S_1} = n$ and $d_{S_2} = \log_2 n$. Considering our 38 intervals, the highest n is 32, while the lowest is 1: One can thus normalize the first indicator by introducing the consonance indicator $C_{S_1} = 1 - (d_{S_1} - 1)/31$. Also the second dissonance indicator can be turned into a normalized consonance indicator, given by $C_{S_2} = 1 - d_{S_2}/5$. These consonance indicators have a high value of the reduced chi-square, respectively, 13.9 and 2.3. Despite this, according to Ref. [21], these indicators should be slightly better than many others (including the percentage similarity of Gill and Purves [10]), as for the linear correlation with some ranks extracted by Schwarz et al. [15]: The reason is that Ref. [21] correlates the ordering of the ranks, not the ranks themselves.

References

1. M. Schroeder, Models of hearing. *Proc. IEEE* **63**(9), 1332–1350 (1975)
2. G. Galilei, *Discorsi e dimostrazioni matematiche intorno a due nuove scienze*, Leida
3. P. Barbieri, *Acustica accordatura e temperamento nell’illuminismo veneto*, Ed. Torre d’Orfeo, Roma
4. P. Barbieri, Galileo’s coincidence theory of consonances, from Nicomachus to Sauveur. *Recercare* **13**, 201–232 (2001)
5. P. Barbieri, La nascita delle teorie ‘continue’ della consonanza. La ignorata curva di Draghetti e Foderà, poi di Helmholtz (1771-1837). *Acta Musicol.* **74**(1), 55–75 (2002)
6. H. von Helmholtz, *On the Sensation of Tone*
7. P. Harrison, M. Pearce, Simultaneous consonance in music perception and composition. *Psychol. Rev.* **127**(2), 216–244 (2020)
8. T. Eerola, I. Lahdelma, The anatomy of consonance/dissonance: Evaluating acoustic and cultural predictors across multiple datasets with chords, *Music Sci.* (2021)
9. A. Frova, *Fisica nella musica*, Zanichelli Ed
10. K. Gill, D. Purves, A biological rationale for musical scales. *PLoS One* **4**(12):e8144
11. R. Plomp, W.J. Levelt, Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* **38**(4), 548–560 (1965)
12. R. Plomp, H. Steeneken, Interference between two simple tones. *J. Acoust. Soc. Am.* **43**(4), 883–884 (1968)
13. W. Hutchinson, L. Knopoff, The acoustic component of western consonance. *Interface* **7**(1), 1–29 (1978)
14. R. Plomp, Beats of mistuned consonances. *J. Acoust. Soc. Am.* **42**, 462 (1967)
15. D. Schwartz, C. Howe, D. Purves, The statistical structure of human speech sounds predicts music universals. *J. Neurosci.* **23**(18), 7160–7168 (2003)
16. C. Malmberg, The perception of consonance and dissonance. *Psychol. Monogr.* **25**(2), 93–133 (1918)
17. D. Bowling, D. Purves, K. Gill, Vocal similarity predicts the relative attraction of musical chords. *PNAS* **115**(1), 216–221 (2018)
18. E. Terhardt, Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* **55**, 1061 (1974)
19. A. Milne, S. Holland, Empirically testing tonnetz, voice-leading, and spectral models of perceived triadic distance. *J. Math. Music* **10**(1), 59–85 (2016)
20. P. Harrison, M. Pearce, 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*, Paris, France
21. F. Stolzenburg, Harmony perception by periodicity detection. *J. Math. Music* **9**(3), 215–238 (2015)
22. L. Euler, *Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae*, Petropoli
23. L.L. Trulla, N. Di Stefano, A. Giuliani, Computational approach to musical consonance and dissonance. *Front. Psychol.* **9**, Article 381
24. P. Vassilakis, *Perceptual and physical properties of amplitude fluctuation and their musical significance*, PhD thesis
25. E. Zwicker, G. Flottorp, S.S. Stevens, Critical band width in loudness summation. *J. Acoust. Soc. Am.* **29**, 548 (1957)
26. H. Fletcher, Auditory patterns. *Rev. Mod. Phys.* **12**, 47–65 (1940)
27. J. Vos, Ratings of tempered fifths and major thirds. *Music Percept. Interdiscip. J.* **3**(3), 221–57 (1986)
28. W. Sethares, Local consonance and the relationship between timbre and scales. *J. Acoust. Soc. Am.* **94**, 1218–1228 (1993)
29. K. Mashinter, Calculating sensory dissonance: some discrepancies arising from the models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff. *Empir. Musicol. Rev.* **1**, 65–84 (2006)
30. R. Krantz, J. Douthett, The role of higher harmonics in musical interval perception. *Bull. Am. Phys. Soc.* **56**
31. G. Dillon, Calculating the dissonance of a chord according to Helmholtz theory. *Eur. Phys. J. Plus* **128**, 90 (2013)
32. E. Bigand, R. Parncutt, F. Lerdahl, Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* **58**, 125–141 (1996)