Universiteit Leiden

Master Creative Intelligence and Technology

Calculating Individual Notes' Contribution to the Consonance/Dissonance Perception of a Vertical Note Set

Name:	Emile Velthuis
Student ID:	3037894
Date:	29/06/2026
1st supervisor:	E. F. van der Heide
2nd supervisor:	F. Stolzenburg

Master's Thesis in Creative Intelligence and Technology

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

Calculating Individual Notes' Contribution to the Consonance/Dissonance Perception of a Vertical Note Set

Emile Velthuis

Graduation Thesis, June 2026

Creative Intelligence & Technology MSc, Leiden University

Primary supervisor: E. F. van der Heide
Second supervisor: F. Stolzenburg

Abstract. With our study, we research approaches for creating note-specific consonance/dissonance models for analysing vertical sets of notes. We do this by proposing adaptations to the best-performing existing consonance/dissonance models for simultaneous sounding notes. We show that the roughness models from Hutchinson & Knopoff, Sethares, and Vassilakis can be transformed using the general rule that the roughness caused by interacting partials can be equally divided among the notes to which the partials belong. We have fully adapted the Hutchinson & Knopoff model, including accounting for coincident partials. For harmonicity/periodicity, the root-based models from Parncutt and Milne are seen as unsuitable since their calculations rely on a strongest (virtual) root, which causes the proposed note-specific method to be unstable and exception-prone. The Harrison & Pearce model seemed unsuitable since it has a strong negative correlation with the cardinality of the chord. For the Stolzenburg model, a note-specific variant has been proposed that has a comparable performance to the original model and has a strong correlation with the note-specific variant of the Hutchinson & Knopoff model, although the correlation is less strong than the correlation of those original models. For the familiarity models, it is argued that the calculation is not decomposable other than that every pitch class has an equal contribution in defining a unique chord class. Furthermore, two other methods have been proposed that use the output of the models instead of transforming their calculation. However, since they compare subsets of the note set, the context of the measures changes, which can lead to distorted note-specific values.

Key words: consonance–dissonance; note-specific analysis; roughness; harmonicity; periodicity; music perception;

1. Introduction

Consonance and dissonance (from now on C/D) is a topic in music that has been explored since the Pythagoreans. It has been described as pleasant vs unpleasant, positively valenced vs negatively valenced, agreeable vs disagreeable, stable vs unstable, notes that fit well together vs notes that do not fit well together, euphonious and in need of resolution (Harrison & MacConnachie, 2024; Harrison & Pearce, 2020; Lahdelma & Eerola, 2020; Parncutt, 1989; Plomp & Levelt, 1965).

In "A History of 'Consonance' and 'Dissonance'", Tenney (1988) describes how different notions of C/D have been used throughout the years. In present studies, the distinction is often made between horizontal and vertical C/D. The former defines C/D as a result of a progression of musical sounds such as a melody or harmonic progression. In such a case, the C/D of a note of a chord is dependent on the notes and chords that came before, making it context-dependent. Vertical C/D defines the consonance and dissonance based on sound in isolation, independent of the musical context. This paper will specifically focus on the last type of C/D. Therefore, the term C/D will be used in describing vertical C/D.

Past studies have investigated the determining factors for C/D. To the current knowledge, there are three main category contributors: harmonicity/periodicity, roughness, and familiarity. For each category, various models have been developed aiming to quantify the determinants. Harrison and Pearce (2020) developed an open-source R package (the incon package) featuring 16 models that have been developed in the past.

Each model has been developed to measure the C/D of a given note set. Thereby, it only gives a score for the total C/D of a given set of notes. What has less been explored is how the individual notes within a note set contribute to the total perceived C/D. By knowing the contribution of each note to the total C/D, the C/D analysis of a musical work can be performed on a higher resolution.

Earlier studies performed by Schwär et al. (2025) show the potential for such a note-specific analysis. Their study provided insight into the C/D distribution among the different tracks from four voiced Baroque chorales. Focusing on roughness-based sensory dissonance, they developed a measure for relative sensory dissonance, which can be seen as a note-specific calculation. The methods for calculating sensory dissonance are based on methods from Von Helmholtz (1912) and the study from Plomp and Levelt (1965). As an input source, they used audio recordings of the tracks.

This study aims to further explore how the available models from the incon package can be used for a note-specific C/D (abbreviated as NsCD) calculation. The main approach is to transform the models into note-specific variants. Additionally, two other methods are explored that rely on the output of the models without modifying the original calculation.

First, the main three determinants will be discussed. Then, the aim for NsCD models will be explained. The candidate models from the incon package will be briefly explained and evaluated on their potential for being transformed into note-specific variants. For the best suitable models, a note-specific variant will be proposed. Lastly, the two alternative methods will be explored and discussed.

2. Determinants

Various theories have been proposed that aim to explain the underlying mechanisms of what causes the perception of C/D. Roughness, harmonicity/periodicity, and familiarity are the most prominent to date. These mechanisms have been modelled in various ways and, when combined in a composite model, they have shown to explain up to 62% of the variance in pooled empirical C/D ratings (Eerola & Lahdelma, 2021). For these three determinants, this section will explain what they are and how they contribute to the perception of C/D.

Roughness is caused by neighbouring partials in a sound. As described by Sethares (2005), roughness can be explained when listening to a pair of sine waves where they start at the same frequency and one gradually increases in frequency. When they are very close in frequency, they are perceived as a single frequency. When there is a slight difference between the sine waves, they are still perceived as the same frequency and the resulting wave will have a slow beat caused by alternating constructive and destructive interference, which is experienced as pleasant. When the frequencies start separating more, the rate of the beat increases. At this point, the ear may be confused by which frequency it is hearing, and the sound is perceived as rough. When the frequencies of the sine waves separate even more, the sine waves are perceived as two distinct tones, and the perceived roughness decreases until it fades away.

This phenomenon, commonly referred to as sensory dissonance, is investigated by a study from Plomp and Levelt (1965). They presented sound stimuli of simultaneously sounding sine waves, where the participants had to judge on the perceived C/D on a 7-point scale.

They found that the peak of perceived roughness is closely related to the critical bandwidth. Maximal roughness is reached when the frequency separation is at approximately 25% of the critical bandwidth.

Figure 1 shows the roughness curve for an absolute lower frequency of 400 Hz based on the findings from Plomp and Levelt (1965).

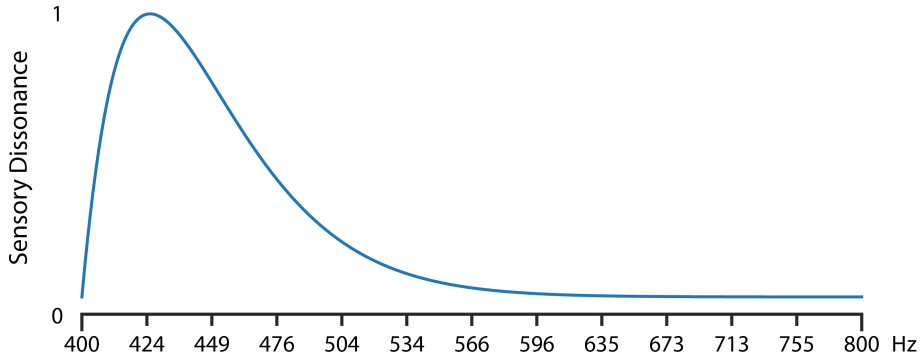

Figure 1: Roughness curve for the interval of two sine waves where one sine wave is 400 Hz and the horizontal axis represents the frequency of the other sine wave. The vertical axis represents the normalised sensory dissonance. *Reproduced with permission from Sethares (2005).*

Harmonicity and periodicity are both concepts that describe the same underlying mechanism. Periodicity, as observed in the time domain, is the rate at which the resulting wave of the sounding tones repeats. Harmonicity, which can be observed in the frequency domain, is how well the frequency components of a sonority's spectrum match the harmonic series of a fundamental frequency, which is possibly a virtual frequency. The (virtual) fundamental frequency of the frequency components of a sound is the same as the repetition rate of a sound. Therefore, harmonicity and periodicity can be seen as describing the same phenomenon.

Since humans have a certain tolerance for pitch discrimination, the ratios for the frequencies do not have to be exact integer ratios. Fastl and Zwicker (1990) shows that the just-noticeable difference in frequency is about the order of a few cents, depending on the frequency, sound level, and stimulus duration. Some models, such as Harrison and Pearce (2018) and Stolzenburg (2015), account for this limited resolution of pitch perception.

There are two extant theories for how our brains analyse harmonicity and periodicity, which are autocorrelation theories and pattern-matching theories (de Cheveigné, 2005).

Familiarity is the phenomenon where a certain sound will be perceived as more consonant, the more the listener is exposed to the sound. (McDermott et al., 2016) found that the Tsimane society, which has minimal exposure to Western music in which harmony plays an important role, rated consonant and dissonant chords and vocal harmonies as equally pleasant. The difference in preference is hypothesised by exposure to musical harmony through culture. Moreover, in an earlier study, (McDermott et al., 2010) found that people with more experience in playing an instrument rate sounds with higher harmonicity as more consonant than people with less musical experience. This suggests that people who have become more familiar with sounds of high harmonicity have an increased consonance perception for these sounds.

Although from the mentioned determinants it is the least explored in research, familiarity has been shown to be a substantial component for explaining the variance of C/D ratings. In their analysis including the models from the incon package, Eerola and Lahdelma (2021) found that familiarity emerged as the dominant component in a principal component analysis. The familiarity component accounted for 46.2% of the explained variance in the C/D ratings against 19.3% for the harmonicity/roughness component.

Other theories such as vocal similarity (Bowling et al., 2018), fusion (Stumpf, 1890), combination tones (Hindemith, 1945), sharpness (Fastl & Zwicker, 1990), and evenness (Cook, 2009) have either a lack of empirical evidence, contribute to one of the earlier mentioned phenomena, or only have a small contribution to the perception of C/D. Therefore, they will not be further explored for the note-specific models.

3. NsCD

The aim of the study is to develop and explore approaches for calculating the note-specific contributions to the overall perceived C/D of a note set. Three methods are proposed.

The first method is to transform the original calculations from the models in the incon package. The best-performing models for each category determinant are evaluated (as evaluated by Harrison and Pearce (2020) and Eerola and Lahdelma (2021)). Initially, only models using symbolic input are explored. In some cases, the logic applied to the symbolic version of a model can also be applied to the version of the model using audio as input.

Every model has its own way of capturing the underlying mechanism of the determinant that the model is based on. The method used by the researcher(s) to capture the underlying mechanism is used as the starting point for creating the note-specific variant.

The models which seem most suitable for being transformed into a note-specific variant are developed into a fully specified note-specific model.

The second and third methods use the output of the original model instead of transforming its calculation. The former is based on leaving a note out and taking the difference with the original measure as the measure for the note-specific contribution; the latter uses Shapley values to determine the note-specific contributions to the overall C/D.

4. Candidate models

This section explores the models available in the incon package (Harrison & Pearce, 2020). For each model, an initial screening is performed to determine whether the model would be suitable for being transformed into a note-specific variant. Based on the findings, a selection is made based on the models that will be further explored.

4.1. Roughness models

Building on the dissonance curve from Plomp and Levelt (1965), **Hutchinson and Knopoff (1978)** developed a model that sums the roughness caused by all partial pair combinations of the sound spectrum. In the final step, the model normalises to the total power of the sound.

Sethares (1993) and **Vassilakis (2001)** each developed a model that sums the raw dissonance measures from the pairwise interactions of the partials without normalising. The Sethares model in the incon package sums the pairwise roughness by either the lowest amplitude of the two frequencies or by the product of the two amplitudes. The model from Vassilakis is a variant that is more sensitive to the imbalance of the amplitude of the frequencies by down-weighting partial pairs with imbalanced amplitudes.

The model from **Wang et al. (2013)** is not considered due to its low score in the Harrison and Pearce (2020) evaluation.

Bowling et al. (2018) developed a model that calculates the minimum distance between notes in a note set. Chords with higher overall minimum distances are assumed to be more consonant.

The model from **Huron (1994)** calculates a C/D measure by aggregating the dyadic C/D measures of the note pairs in a note set. A downside of the model is that it does not model intervals wider than an octave since it collapses the notes to pitch classes.

For the Hutchinson and Knopoff (1978), Sethares (1993), Vassilakis (2001), and Huron (1994) models, a note-specific measure can be obtained by dividing the measured roughness of a partial pair or note pair equally among the notes contributing to the pair measure. In both the evaluation from Harrison and Pearce (2020) and Eerola and Lahdelma (2021), the Hutchinson & Knopoff model is the best predictor for the perceptual data sets on which it was tested. In the former evaluation, it was tested in isolation, and in the latter, it was tested within a composite model. Since the aim of NsCD is to match empirical data as closely as possible, the Hutchinson & Knopoff model will be the model of choice to explore. Section 5.1 will elaborate on how the model can be transformed to a note-specific variant.

4.2. Harmonicity/Periodicity models

The model from **(Parncutt, 1988)** is based on Terhardt (1982), framing consonance as the chord-root perception. According to the theory, consonance arises from a clear root and dissonance from root ambiguity. Instead of using the absolute pitches, it uses pitch classes and assumes octave equivalence. The model uses harmonic templates quantised to a 12-tet scale. To determine the ambiguousness, it uses pattern recognition using the harmonic series. Every pitch class receives a weight based on how well a sound matches the harmonic template of that root. Root ambiguity is then calculated by dividing the total root support by the maximum root support. Parncutt developed another model **Parncutt and Strasburger (1994)** which is an updated version of **Parncutt (1989)**. Similar to the 1988 model, pattern matching is used in the model, but the later models preprocess the sound based on the limitations of the auditory system. Instead of calculating root ambiguity, one of its measures for harmonicity is complex sonorousness, which is based on the strongest virtual root.

For calculating each note's contribution to the strongest (virtual) root, a note's contribution can be calculated to this fundamental. However, following this logic, there are two special cases. First, when there is a tie between two roots, the note's highest contribution is taken. Which weight will then be considered to be contributing to the strongest root? Or when the note has zero contribution to the strongest root, it would calculate infinite ambiguity due to a zero division. Moreover, this approach is very unstable. When there are two candidate strongest fundamentals with one having just more weight than the other, the notes contributing to the strongest fundamentals with just more weight are considered to contribute to more harmonicity. However, when a note is added that results in the other candidate root having more weight, a major shift occurs where the notes of that candidate root are considered to contribute to more harmonicity and the other notes will be considered to contribute to less harmonicity. Due to this unstable nature of the note-specific interpretation of the model and the exceptions that it raises, these models are seen as unsuited for NsCD.

(Milne, 2013) developed a model that is similar to the Parncutt and Strasburger (1994) in the sense that it also relies on the strongest (virtual) root. In this model, the sonority spectrum is displayed on a template in a pitch-class domain, representing the strength of each pitch class. Then, the template of a root note is swept over this spectrum using cosine similarity to see how well a root note matches the spectrum. The highest peak of the cosine similarity is taken as the measure for harmonicity. Since this method also relies on a strongest root, it faces the same problems as the Parncutt models and is seen as unsuited for NsCD.

The model from **(Harrison & Pearce, 2018)** uses Milne's approach as a starting base. Instead of taking the peak of the distribution of the cosine similarity, the KL-divergence is

taken with a uniform distribution. This is a more stable approach since this measure captures the whole contribution of a note to the ambiguity of a sound. A measure of ΔH can be taken as the note-specific harmonicity value of a note. Here the note-specific harmonicity is measured as the difference between the harmonicity value with and without the note. Important to take into account is that the model does have a large negative correlation with the amount of notes in a note set. This means that to be able to interpret the values, the amount of notes must be considered in the calculation. This can be done by taking the zero mean, unit-variance for each cardinal. Since every value will be relative only to the other measures with the same amount of notes, it does not allow for cross-comparability of note sets with different amounts of notes.

(Gill & Purves, 2009) developed a model that is not considered since it requires notes that are just-intonated. The present research will focus on a well-tempered scale which contains non-perfect ratios.

As compared to spectral pattern matching from the previous models, the model from **(Stolzenburg, 2015)** relies on ratio simplicity as the basis of the calculation. In a given note set, every note makes a certain frequency ratio with another note in the set. By looking at the least common multiple of the denominators of these ratios, the period of the resulting wave can be calculated. The main idea for the periodicity measure is to divide the resulting period by the period of the lowest note and then take the base-2 logarithm of the resulting ratio.

Because the Stolzenburg model provides detailed information about every specific note, given information on the frequency ratios that a note has with the other notes, and since this is the best-performing model in both the Harrison and Pearce (2020) and Eerola and Lahdelma (2021) evaluation studies, this model will be further explored for its potential for a note-specific variant.

4.3. Familiarity models

The familiarity models are based on the prevalence of a chord class in a corpus. The two best-performing familiarity models are from **Harrison and Pearce (2020)** and **Eerola and Lahdelma (2021)**. Both rely on the probability of a chord representation in a corpus.

Since every pitch class has an equal contribution in making the chord class distinct, using the logic of the model as a base, the C/D measure is not decomposable into note-specific measures other than that every pitch class has an equal contribution.

5. Note-specific variants

This section explores in detail how the selected candidate models can be transformed into note-specific variants.

5.1. Hutchinson and Knopoff

The selected model for roughness is from Hutchinson and Knopoff (1978). It calculates the roughness measure for a set of partials by the formula in Equation (1).

$$D = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} A_i A_j g_{ij}}{\sum_{r=1}^{N} A_r^2} \tag{1}$$

The formula sums the interaction of every partial pair weighted by the product of the partials' amplitudes and by the dissonance curve g_{ij}, which is dependent on the absolute frequency distance between partials i and j and the critical bandwidth at the mean frequency

of the two partials. Since the roughness of every partial interaction is measured twice (both x-y and y-x), the roughness calculated in the numerator is halved. Moreover, the formula also includes the measure of same partial pairs x-x which physically does not happen. Since g_{ij} gives 0 for g_{xx} (an interval of 0 gives zero dissonance), that is not a problem.

The note-specific roughness can be calculated by adding the internal roughness (roughness created by interactions of a note's own partials) to half of the roughness that the partials of a note create with the partials of other notes (2). In the paper from Schwär et al. (2025), a similar approach is taken. Here, a split of four different partial interactions is considered. [1] A note's internal partial interactions, [2] the partial interactions with the partials of the note with the partials of the other notes, [3] the partial interactions of the partials of the other notes with the partials of the note in question, and [4] all the other partial interactions excluding partials of the note in question. A note's contribution to roughness can then be seen as the roughness caused by the interacting partials from [1] and [2]. Besides that this method considers every partial interaction twice (x-y and y-x), the general approach is the same. Adding a weighting factor for assigning roughness to a note gives:

$$D_k = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^{(k)} A_i A_j g_{ij}}{\sum_{r=1}^{N} A_r^2} \tag{2}$$

$$w_{ij}^{(k)} = \begin{cases} 1, & \text{if both partials belong to note } k \\ \frac{1}{2}, & \text{if exactly one partial belongs to note } k \\ 0, & \text{if neither of the partials belongs to note } k \end{cases} \tag{3}$$

Where $w_{ij}^{(k)}$ is the weighting factor that is either 1, 1/2, or 0, depending on whether both, one, or none of the partials belong to the note k.

The choice for distributing the roughness of cross-note partials equally among the notes to which the partials belong is a modelling choice. Another approach could have been to divide the roughness proportionally to the amplitudes of the partials.

In the original model, first the full sound spectrum is calculated. Within this process, coincident partials are combined as the root-sum-square of their amplitudes (4), which assumes arbitrary phase relations between the partials.

$$A_i = \sqrt{A_1^2 + A_2^2 + \cdots + A_{n_i}^2} \tag{4}$$

The contribution of a note to a merged partial is calculated as the proportion of the squared amplitude of the note's partial contributing to the merged partial to the total sum-square of the partials contributing to the merged partial. The contribution factor of note k to partial i becomes:

$$m_i^{(k)} = \frac{\left(A_i^{(k)}\right)^2}{\sum_{r=1}^{n_i}\left(A_i^{(r)}\right)^2} \tag{5}$$

Where $A_i^{(k)}$ is the amplitude of the partial of note k contributing to the merged partial i and the denominator sums the squared amplitudes of all partials belonging to merged partial i. To integrate the contribution factor in the formula for D_k, $w_{ij}^{(k)}$ is redefined as the following formula:

$$w_{ij}^{(k)} = \frac{1}{2}\left(m_i^{(k)} + m_j^{(k)}\right) \tag{6}$$

It satisfies the old logic since $w_{ij}^{(k)}$ becomes 1 when both partials fully belong to note k, 1/2 when one of the partials fully belongs to note k, and 0 when none of the partials belong to note k. In the case when one of or both partials are only partly contributed by note k, $m_i^{(k)}$ and/or $m_j^{(k)}$ becomes a fraction of 1, only accounting for the contribution as a proportion of the total sum-square of the amplitudes as calculated by Equation 5. The final formula with the new definition for $w_{ij}^{(k)}$ becomes:

$$D_k = \frac{\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}^{(k)} A_i A_j g_{ij}}{\sum_{r=1}^{N} A_r^2} \tag{7}$$

A property of this note-specific variant is that the sum of the note-specific roughness values is equal to the original total roughness value of the Hutchinson and Knopoff (1978) model. The only model choice that has been made is to distribute roughness caused by two interacting partials equally among the notes to which the partials belong. The other choices for the note-specific variant can be seen as a logical consequence of the original model.

The models from Sethares (1993) and Vassilakis (2001) could be adapted in a similar way. These models differ from the Hutchinson and Knopoff (1978) model in their definition of the function g_{ij}, how the amplitudes are weighted, and not having the final normalisation step. Since the Hutchinson & Knopoff model matches the perceptual data more closely, it is expected that the note-specific variant will also more closely match the perceptual C/D contribution of the individual notes to a note set.

5.2. Stolzenburg

The selected model for harmonicity/periodicity is from Stolzenburg (2015). It calculates the logarithmic periodicity of a set of notes by taking the base-2 logarithm of the ratio between the resulting period and the period of the lowest note. The resulting period is calculated by considering the frequency ratios that the reference note makes with the other notes. These frequency ratios can be found in Figure 2. Since humans have a certain tolerance for the just noticeable difference of frequencies, the frequency ratios up to a denominator of 16 are considered that have a maximal deviation of 1.1% (the figure shows the frequency ratios with a maximal deviation of 2.0%). If there are multiple candidates, the simplest ratio is chosen.

The period of the resulting wave is calculated by taking the least common multiple of the denominators. Consider the chord C4, F#4, A4. The frequency ratios are $\{1/1, 7/5, 5/3\}$. The lcm of the denominators is $\mathrm{lcm}(1, 5, 3) = 15$. So the period of the resulting wave is 15.

In this example, C4 was taken as a reference note as it was considered the 0th semitone. This set can be seen as $S_0 = \{0, 6, 9\}$. When taking different notes from the note set as the 0th note, the set shifts and the results may be different. For example, when considering the shifted sets $S_{-6} = \{-6, 0, 3\}$ and $S_{-9} = \{-9, -3, 0\}$. The frequency ratios become $\{7/10, 1/1, 6/5\}$ and $\{3/5, 5/6, 1/1\}$. Notice that the denominators for semitones below 0 are doubled since they are an octave lower.

The least common multiples of these sets are $\mathrm{lcm}(10, 1, 5) = 10$ and $\mathrm{lcm}(5, 6, 1) = 30$. Since the periodicity is relative to the lowest note, the outcome of the lcm value for a shifted set is multiplied by the frequency ratio of the lowest note from that shifted set. This converts the period from being relative to the reference note to being relative to the lowest note. The resulting periodicity measures are $\frac{1}{1} \cdot 15 = 15$ for S_0, $\frac{7}{10} \cdot 10 = 7$ for S_{-6}, and $\frac{3}{5} \cdot 30 = 18$ for S_{-9}.

The last step of the Stolzenburg (2015) model is to transform the periodicity values using $\log_2(\cdot)$, which reflects the logarithmic organisation of the neuronal periodicity map in the brain

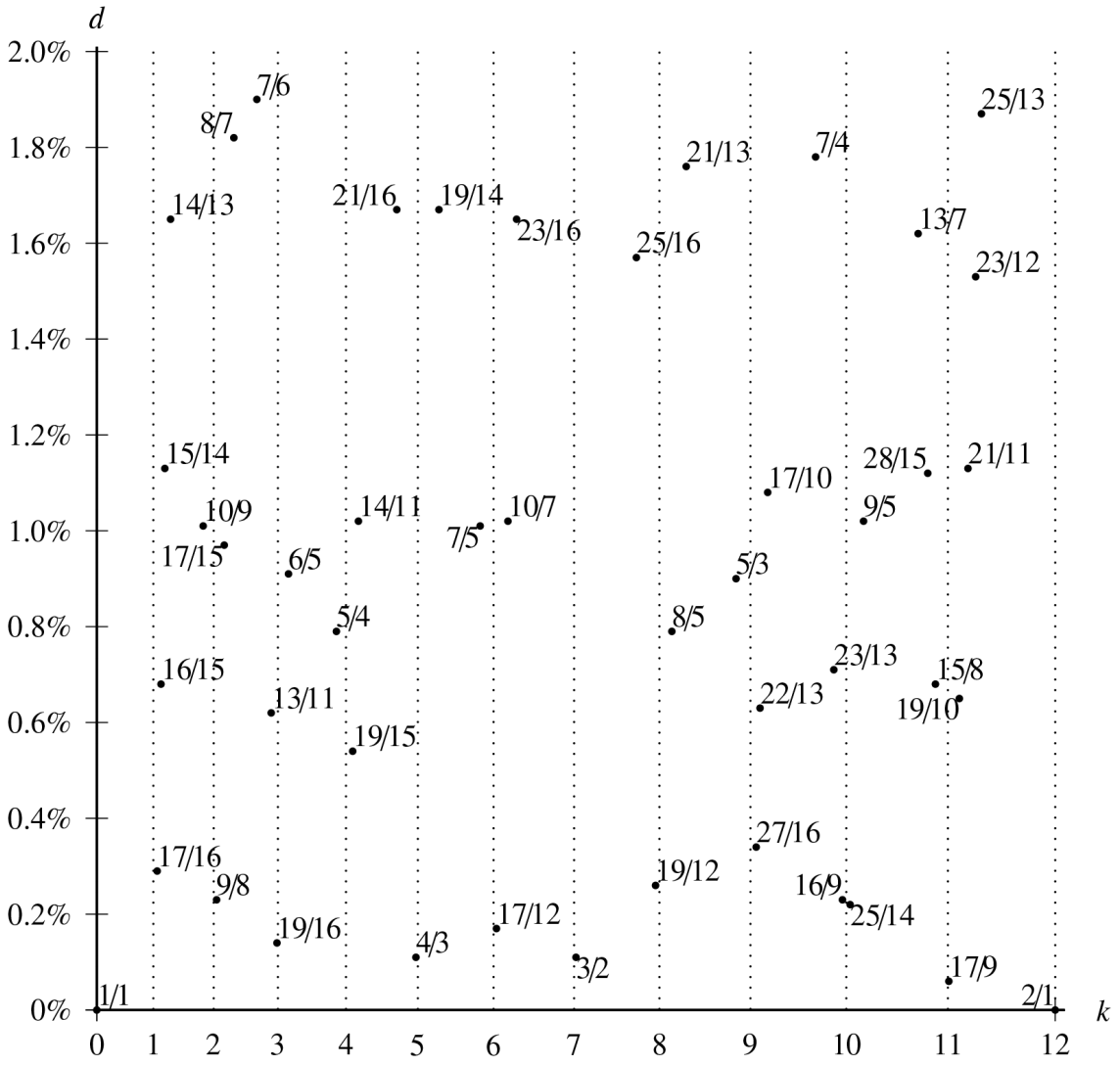

Figure 2: Integer frequency ratios a/b and their deviation d from the kth semitone in equal temperament for $b \leq 16$ and $d \leq 2\%$. *Reprinted with permission from Stolzenburg (2015).*

(Langner et al., 1997). Then, the values are averaged to obtain the smoothed logarithmic periodicity, which is $\big(\log_2(15) + \log_2(7) + \log_2(18)\big)/3 = 3.628057$ for C4, F#4, and A4.

A way of representing the inner structure of the periods of the note pairs within a note set is by Figure 3. Here, the notes C4, C#4, E4, and G4 are displayed in the 0th set: S_0. The frequency ratios are 1/1, 16/15, 5/4, 3/2. The periods that a note forms with another note relative to the lowest note with frequency ratios A = a/b and B = c/d are equal to A/B = (a/b) / (c/d). This ratio is then written in the lowest terms p/q. To calculate the period relative to the lowest note, first the period is calculated relative to the period of the reference note (the note with a frequency ratio of 1/1): (p/a) * b (or (q/c) * d). Then, this period is multiplied by the ratio of the lowest note to make the period relative to the lowest note. For example, in the note set from Figure 3, E4 and G4 have frequency ratios 5/4 and 3/2 with reference note C4. Their pair ratio is (5/4) / (3/2) = 10/12, which is simplified to 5/6. The period relative to the reference note is 5 / 5 * 4 = 4. The period length relative to the lowest note is 4 * 1/1 = 4. So the period of E4 and G4 relative to the period of the lowest note is 4. Figure 3 shows all period pairs relative to the lowest note.

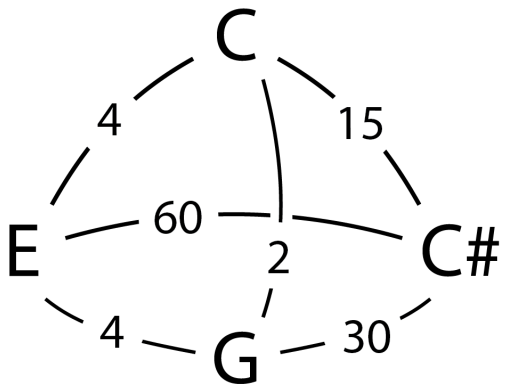

Figure 3: Periodicity structure for note set C4, C#4, E4, and G4 in S_0.

This overview of the pair periodicities allows for deeper insight into the inner periodic structure of periods that the notes make with each other. The lcm of these periods, in this case 60, is always the same value as the original periodicity measure of that shifted set. From the structure, a list of pair periods can be derived for each individual note.

C: 4, 2, 15
C#: 30, 60, 15
E: 4, 60, 4
G: 4, 2, 30

It can be observed that the longest period is between E4 and C#4. Another observation is that C#4 has the longest period pairs in this 0th shifted set.

The way that a note-specific value is derived from this structure by taking the mean of the periodicities that a note makes with the other notes. As in the original model, first the $\log_2(\cdot)$ of the periodicity value is taken.

C: $mean(log2(15), log2(4), log2(2)) = 2.302297$
C#: $mean(log2(15), log2(60), log2(30)) = 4.906891$
E: $mean(log2(4), log2(60), log2(4)) = 3.302297$
G: $mean(log2(2), log2(30), log2(4)) = 2.635630$

This value reflects the average logarithmic period pair that the note makes with the other notes. Figure 4 shows the inner structure of periods of the note pairs for the other shifted sets.

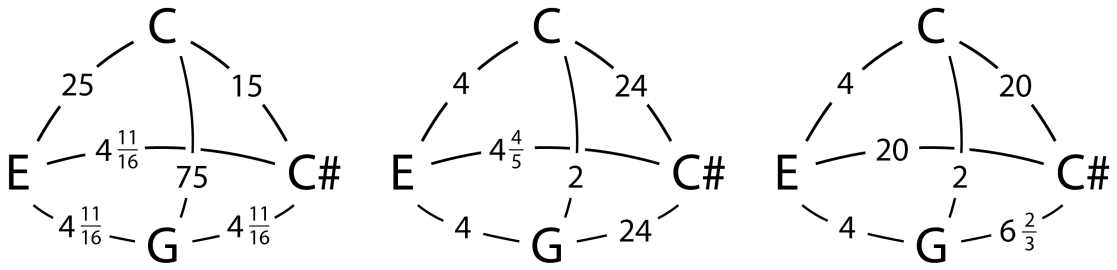

Figure 4: Periodicity structure for note set C4, C#4, E4, and G4 in S_{-1}, S_{-4}, and S_{-7}.

Applying the same steps as before, the smoothed note-specific periodicity measure can be calculated.

C: $mean(2.302297, 4.926522, 2.528321, 2.440643) = 3.049446$
C#: $mean(4.906891, 2.788176, 3.810986, 3.793607) = 3.824915$
E: $mean(3.302297, 3.033831, 2.087678, 2.773976) = 2.799446$
G: $mean(2.635630, 3.562152, 2.528321, 1.912322) = 2.659606$

In this note set, C# has the highest note-specific periodicity value, followed by C, E, and G.

6. Evaluating the Stolzenburg variant

The Hutchinson & Knopoff variant from Section 5.1 (hereafter Hutch per-note) can be seen as a direct descendant of the original model with the only modelling choice that it distributes the roughness by two partials equally among the notes to which the partials belong. Moreover, the sum of the note-specific measures equals the C/D measure for the complete note set.

The proposed note-specific variant for the Stolzenburg model (hereafter Stolz pair-average) is not a direct consequence of the model. It still considers logarithmic periodicity and the shifted sets, but it uses a different method for calculating the note-specific values than how the original smoothed logarithmic periodicity is measured.

The proposed variant therefore requires evaluation before it can be considered a suitable note-specific measure. Ideally, the model would have to be tested on empirical measurements of how individual notes contribute to the perceived C/D of a given note set. As such data is currently unavailable, a different approach is taken.

The model will be evaluated using three different analyses. (1) a comparison of its predictive performance on C/D rating of complete note sets with that of the original Stolzenburg model; (2) an assessment of the correlation between the note-specific variant and the original Stolzenburg model; and (3) an assessment of the correlation between the note-specific variant and the note-specific variant of the Hutchinson & Knopoff model.

6.1. Stolz original vs Stolz pair-average

We first evaluate the original Stolzenburg (2015) (hereafter Stolz original) model and the summed values of the Stolz pair-average model. It is assumed that the values from the note-

specific model can be summed to obtain the overall C/D score. Three tests are performed using the Bowling et al. (2018) dataset. The first evaluates the Stolz original model against the ratings of the data set to test its performance. Then, the summed Stolz pair-average is tested against the same data. Finally, both measures are calculated for all complete chords in the dataset, and the correlation between their outputs is evaluated.

The Bowling et al. (2018) dataset contains C/D ratings for 12 dyads, 66 triads, and 220 tetrads, giving a total of 298 stimuli. The sounds were presented using a piano timbre. Thirty participants judged the stimuli on a four-point C/D scale from 1 = "quite consonant" to 4 = "quite dissonant".

As a performance metric, the Pearson partial correlation between the model outputs and the C/D ratings is calculated. The correlation is controlled for the number of notes, which is treated as a categorical variable. The number of notes is controlled for because the stimuli were presented in order of chord size, potentially allowing participants to adapt their ratings to chord size.

The Stolz original model has a Pearson partial correlation of 0.72 (95% CI [.66, .77]). The summed Stolz pair-average achieves a Pearson partial correlation of 0.74 (95% CI [.68, .78]). Although the value is slightly higher, the overlap between the confidence intervals does not provide clear evidence that the summed Stolz pair-average model outperforms the original model. Comparing the Stolzenburg model directly against the summed values of the note-specific variant using the chords from the Bowling dataset as input gives a Spearman correlation of 0.94. The corresponding plots are shown in Figure 5.

6.2. Stolz pair-average vs Hutch per-note

Since past research has shown that there is a high collinearity between the roughness and the harmonicity/periodicity models, it is expected that there should also be a correlation between Stolz pair-average and the Hutch per-note model.

First, the correlation between the Hutch original and Stolz original models is calculated. For the chords of the Bowling dataset, the models have a Spearman correlation of 0.77. Since the Bowling dataset consists of just-intonated stimuli, the correlation was also measured for the corresponding chords in 12-TET. It is expected that the roughness values would be slightly higher in the 12-TET version for some chords, as the partials no longer align perfectly and therefore cause more roughness with neighbouring partials. Evaluating the models on equal temperament indeed gives an increased Spearman correlation of 0.81, see Figure 6 for the plots.

These results confirm a strong correlation between the models. Testing for the note-specific variants, now having 1102 observations since every chord has multiple measures, gives a Spearman correlation of 0.68 for the just-intonated chords and 0.74 when the chords are in equal temperament (see Figure 7).

Because multiple note-specific observations originate from the same chord, these observations are not statistically independent. To verify that this dependency does not substantially affect the results, an additional analysis using one note per chord was performed (see Appendix A).

When analysing the top 50 notes with the highest differences using z-difference and the top 50 notes with the highest rank disagreement, there are a few factors that play a role. The first is the distance to the nearest note. Compared to the Stolz pair-average model, the Hutch per-note model scores higher for neighbouring notes, especially the minor second. This can be explained by the fact that many partial pairs fall close enough in frequency to cause strong roughness. The opposite is also the case when the note has a higher distance to the nearest

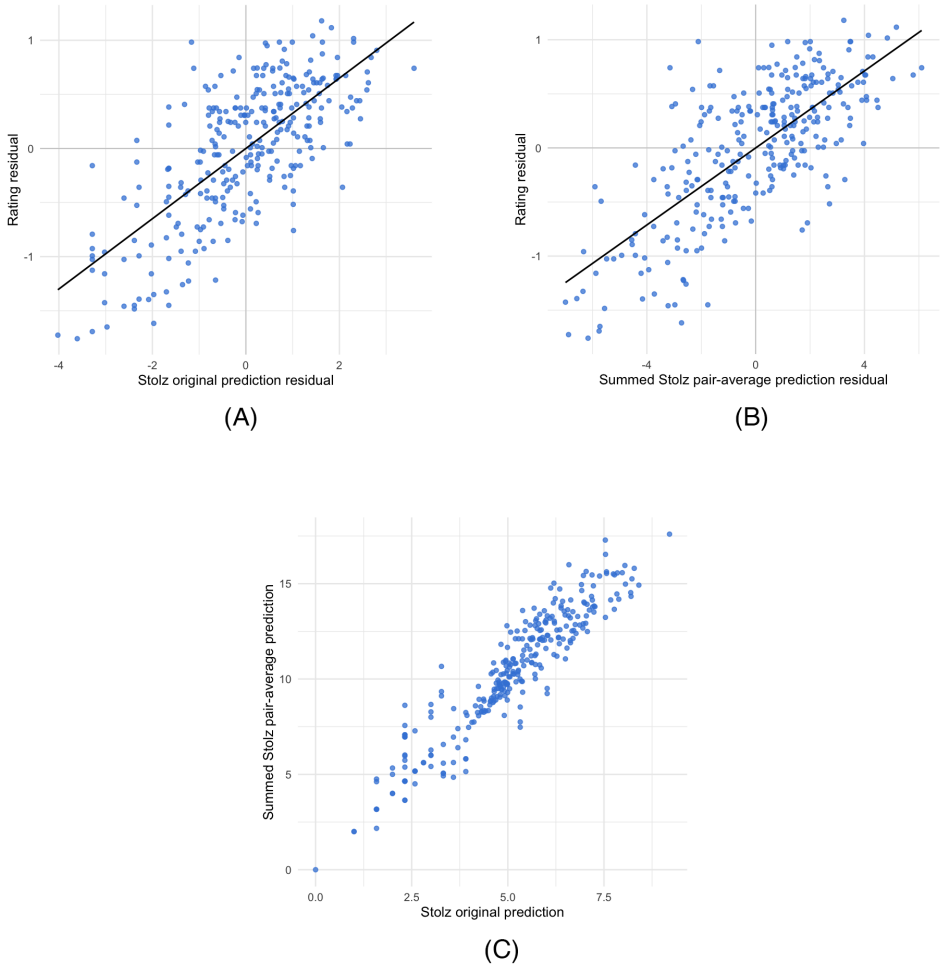

Figure 5: Evaluation of the Stolz original model and the summed Stolz pair-average model on the Bowling et al. (2018) dataset. (A) Relationship between the residualised C/D ratings and the residualised Stolz original model outputs after controlling for chord size (categorical); Pearson partial correlation $r = .72$. (B) Relationship between the residualised C/D ratings and the residualised summed Stolz pair-average model outputs after controlling for chord size (categorical); Pearson partial correlation $r = .74$. (C) Agreement between the raw outputs of the Stolz original and summed Stolz pair-average models across chords; Spearman $\rho = .94$. Each point represents one chord.

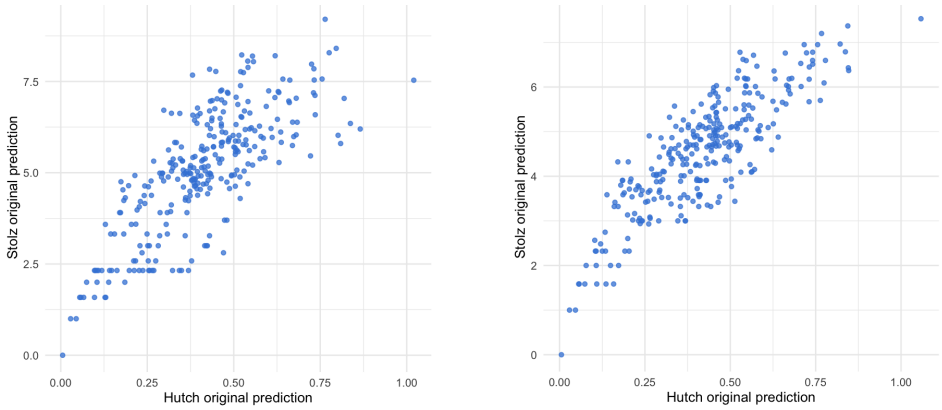

Figure 6: Agreement between the Hutch original model and the Stolz original model on the Bowling et al. (2018) chords. (A) Model outputs for the original just-intonated chord representations; Spearman $\rho = .77$. (B) Model outputs for the corresponding 12-TET chord representations; Spearman $\rho = .81$. Each point represents one chord.

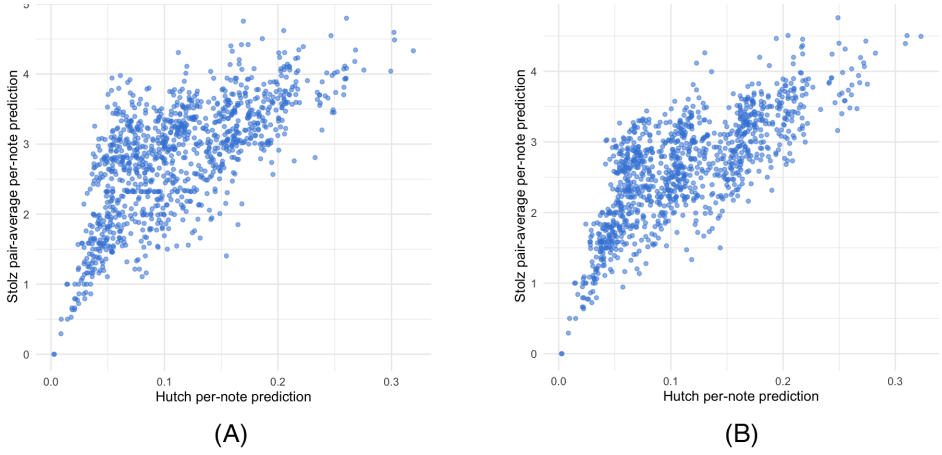

Figure 7: Agreement between the Hutch per-note model and the Stolz pair-average model on the Bowling et al. (2018) dataset. (A) Note-specific outputs for the original just-intonated chord representations; Spearman $\rho = .68$. (B) Note-specific outputs for the corresponding 12-TET chord representations; Spearman $\rho = .74$. Each point represents one note occurrence.

note. In such a case, the Hutch per-note model scores lower since the partials cause less roughness, but the notes can still receive a high periodicity value. This pattern is especially visible for the highest notes of a chord.

Moreover, the roughness model scores lower than the harmonicity model for intervals within wide or smooth chords, for example when a note has an interval of a major third, perfect fourth, or major sixth. In such cases there are fewer strong close partial interactions. The Stolz pair-average model however may still have a relatively larger value since some shifted sets contain large periods for these intervals.

7. LOO and Shapley

In addition to transforming existing models into note-specific variants, two other methods have been explored that instead rely on the output of the models.

The first method is to observe the difference between the output of the model with and without the note for which the note-specific C/D contribution is being calculated. It thus compares the complete set minus the subset where the note is left out. This method, we refer to as Leaving One Out (LOO).

The problem with this method is that it compares C/D measures obtained from different contexts, which can give a distorted measure. For example, when considering the Hutchinson & Knopoff model, here the measure for C/D is calculated by dividing the roughness of the partial pairs by the total power of the sound. When leaving a note out, not only the amount of roughness changes for the numerator, but also the total power of the sound. In other words, the amount of roughness for the subset without the note is calculated within a different context. That is why in the note-specific calculation, the total amount of power in the sound remains constant for all note-specific measures.

Another problem which can be observed for the Hutchinson & Knopoff model is that when leaving a note out, all roughness arising from the interactions between the partials from that note and the partials from the other notes is fully attributed to the removed note. Whereas in the note-specific calculation this measure is split among the notes to which the partials belong.

Given these observations, it is clear that also the sum of the note-specific values using this method does not equal the total C/D value of the original measure.

Similar issues arise for the Stolzenburg model. Here, the periodicity value is calculated by dividing by the period of the lowest note in the note set. When the lowest note is removed, a new reference note is assigned, thereby changing the context in which the periodicity measure is calculated. Hence, in the note-specific calculation for the Stolzenburg model, all period pairs are calculated relative to the same lowest note.

Therefore, LOO cannot be interpreted as a note-specific measure but only as a deletion effect: how much the C/D measure would change when removing a note.

Another method that was considered was to use Shapley values, which originate from cooperative game theory and provide a way to distribute a value created by a group among its individual components.

The procedure is to calculate the marginal contribution of a component to every possible subset of the coalition. So, in the case of a note set, it would calculate the marginal C/D contribution of a note to every possible subset of the note set. For example, for note A in note set ABC, the contribution of A to an empty set would be calculated (an empty set can be set to the baseline of most consonant); the contribution of A to the subset B and subset C; and finally, A to the subset (B-C). The total number of ways that the notes can enter the note set is 3! (A-B-C, A-C-B, B-A-C, B-C-A, C-A-B, C-B-A). Since there are two ways in which A

can enter the coalition as a first note (A-B-C and A-C-B), the marginal contribution as a first note is weighted as 1/3. Since there is only one way in which A can enter after B (B-A-C), this marginal contribution is weighted by 1/6. The same applies to A after C. A entering as a last note is weighted by 1/3 since there are two possible orderings (B-C-A, C-B-A). By adding the weighted marginal contributions, a final value for A is calculated.

The benefit of this method is that, unlike the LOO method, this method calculates values from which the sum is equal to the original measure. Moreover, it gives a more balanced output since it considers all possible marginal contributions that a note has to every possible subset of the note set.

However, similar to the LOO method, it relies on comparing different contexts with each other that, despite having a more balanced outcome, give a distorted measure for the note-specific values. But when there is no note-specific measure for a model, Shapley values offer the best alternative.

8. Conclusion and Discussion

To calculate the contribution of individual notes to the consonance/dissonance perception of a vertical note set, existing C/D models were transformed into note-specific variants.

The note-specific values for roughness models from Sethares, Vassilakis, and Hutchinson & Knopoff can be measured by allocating the pairwise roughness measure to the notes to which the partials belong, for which a similar approach has been explored by Schwär et al. (2025). This can be seen as a logical consequence of the construct of the models. The only modelling choice that has been made is to divide the roughness measure equally between the notes to which the partials belong. The Hutchinson & Knopoff model has been fully worked out, also accounting for coincident partials and leaving the normalisation step intact.

For the harmonicity/periodicity models, the Parncutt models and the model from Milne were seen as unsuited. Their reliance on a winning root model resulted in an unstable model containing exceptions. The Harrison and Pearce model seemed unsuited since it had a large negative correlation with the note set size. Consequently, the resulting note-specific values would not permit direct comparisons between notes belonging to chords of different cardinalities.

For the Stolzenburg model, a variant has been proposed that captures the logarithmic average period pair of a note averaged over the shifted sets. Since it is not a direct consequence of the original model, it still needs to be evaluated on empirical note-specific data, which is currently unavailable. Alternatively, it has first been compared to the original Stolzenburg model by summing the values, showing similar correlations for the ratings of the Bowling et al. (2018) dataset (Pearson correlations of r = .72 and r = .74 for the Stolz original and Stolz pair-average model respectively). Moreover, the models show a high degree of agreement when their output on the chords from the Bowling dataset is compared (Spearman correlation of $\rho = .94$). These findings show that the model's output is comparable to the original Stolzenburg model and has similar performance on empirical data.

The model is then compared to the note-specific variant of the Hutchinson & Knopoff model. Here, a correlation is expected since the original Hutchinson & Knopoff model and the Stolzenburg model show a high degree of collinearity (Spearman correlation of $\rho = .77$ and $\rho = .81$ for the chords from the Bowling et al. (2018) dataset represented in just intonation and 12-TET respectively). The note-specific variants show Spearman correlations of $\rho = .68$ and $\rho = .74$ for the just-intonated chords and the 12-TET representations respectively for the note-specific outputs. Thus, there is still agreement between the models, although it is weaker than the measures for the complete note sets. When analysing the outliers, the

differences are mainly caused by isolated notes where the roughness is lower since there are fewer strong partial interactions while the Stolz pair-average model still has a higher value for the z-difference and ranking for the notes since the notes still have a higher average pair period. The same holds for wide or smooth chords where the partials coincide and give low roughness measures but still have a higher average pair period. Conversely, the Hutch per-note model scores higher on closely neighbouring notes compared to the Stolz pair-average model, owing to the strong roughness caused by the partial interactions of the notes.

Based on this preliminary evaluation, the Stolz pair-average model can be considered a potential note-specific variant of the original Stolzenburg model. To test its real performance, it still has to be evaluated on note-specific empirical data.

For the familiarity models, it is argued that the models cannot be decomposed to a note-specific variant other than that every pitch class contributes equally to the C/D measure of a chord class in the corpus. This is because each pitch class within a note set contributes equally to making the chord class unique.

Where the previous approach aims to transform the original calculations, two other methods have been proposed that directly use the output of the original models.

The methods are leave-one-out (LOO) and the Shapley method. The values from the LOO method give a measure for how much the C/D value would change when the note is removed. The Shapley values provide a more balanced measure for the marginal contribution to every possible subset of the note set. Both methods can be applied to every C/D model, also to composite models. However, since these methods compare different subsets of the note set, they can produce distorted values that may not reflect the actual note-specific contribution to the overall note set within the original context of the complete note set.

The note-specific variants of the Hutchinson & Knopoff model and the proposed Stolzenburg model have been created using the original calculations as their basis. Less attention has been given to the underlying biological mechanisms of interacting partials and periodicity detection in simultaneously sounding notes, and to how individual notes contribute to these processes. For example, for the Stolzenburg model, a deeper insight would be gained when the actual neuronal model for the analysis of periodic signals (as outlined by Langner et al. (1997)) is used as a starting point for calculating a note-specific contribution to the period detection in the brain.

9. Further Research

The present research took the existing C/D models as a starting point for calculating note-specific C/D. Another approach would be to develop a model that incorporates the individual note's contribution from the start based on the underlying mechanism.

Currently, there is only data available for the overall C/D perception of a given note set. To be able to test the note-specific models, a method would have to be constructed to measure the note-specific contributions to the overall perceived C/D. Note-specific models could then be evaluated on the empirical data.

Once such models have been developed and validated on empirical data, they may be used for musical analysis. Such analysis could reveal how composers embed consonance and dissonance within their music at the level of individual notes.

10. Appendix A

For the comparison between the note-specific variants of the Hutchinson & Knopoff and Stolzenburg models an additional evaluation was performed. The main analysis used all note-

specific observations from the chords in the Bowling et al. (2018) data set. However the notes within a chord are not independent, so using all notes from a chord to measure the correlation between the models would violate the assumption of independent observations. Therefore, an additional analysis was performed that only uses one note per chord as input for measuring the correlation between the models. Three sampling strategies were considered: only the lowest note of the chords, only the highest note of the chords, and one randomly selected note per chord. The random sample analysis is the most informative for overall model agreement; the other two are to show how the position of the note in the chord influences the correlation. Each analysis contains 298 observations, corresponding to the 298 chords in the dataset. The results are shown in Figure 8.

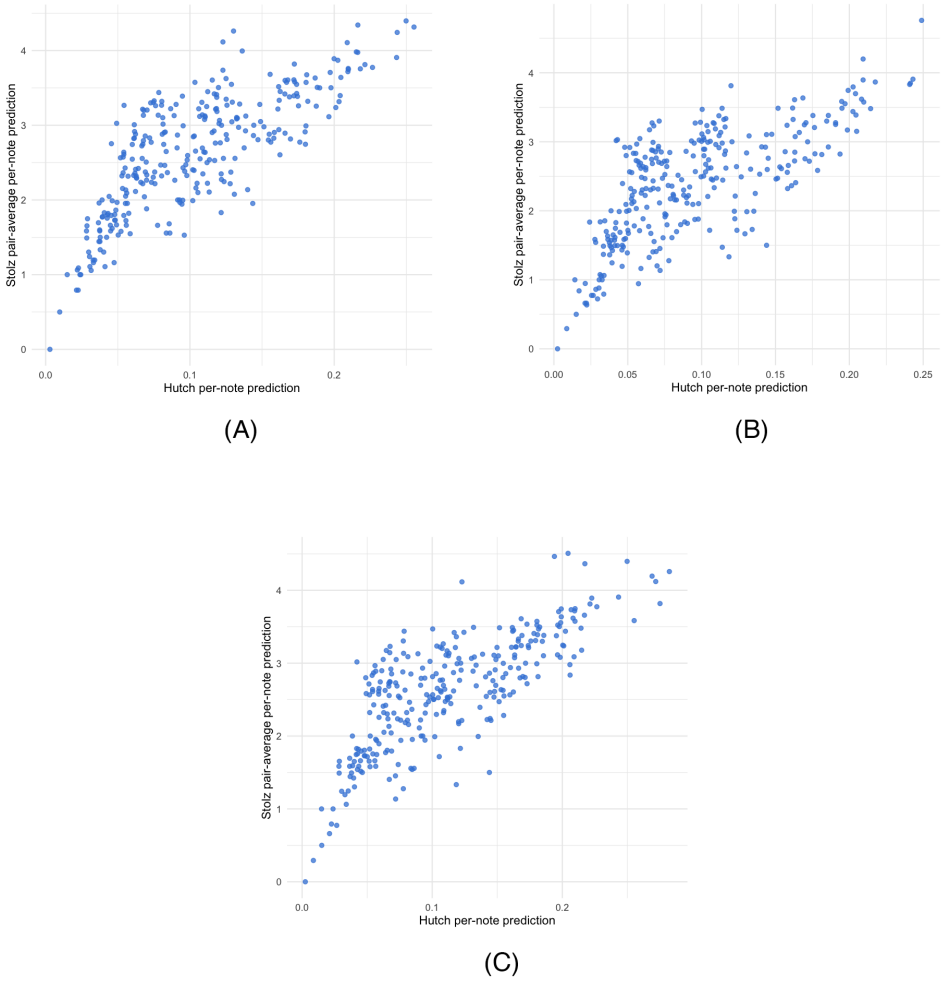

Figure 8: Agreement between the Hutch per-note model and the Stolz pair-average model on the Bowling et al. (2018) dataset with one note per chord. (A) Note-specific outputs for the corresponding 12-TET chord representation of the lowest notes from the chords; Spearman $\rho =$ .78 (B) Note-specific outputs for the corresponding 12-TET chord representation of the highest notes from the chords; Spearman $\rho = .70$ (C) Note-specific outputs for the corresponding 12-TET chord representation of one randomly selected note per chord; Spearman $\rho = .76$. Each point represents one note occurrence.

The additional evaluation showed that using one randomly selected note per chord results in a slightly higher correlation ($\rho = .76$). Furthermore, the lowest notes show an even stronger correlation ($\rho = .78$), whereas the highest notes have a weaker correlation ($\rho = .70$). The latter is consistent with the findings in Section 6.2, where the highest notes tend to receive

lower scores from the roughness model than from the periodicity model, resulting in greater disagreement between the models for the highest notes.

Regarding the random sample analysis, now using observations which are independent of one another, it can be concluded that the correlation between the note-specific variant of the Stolzenburg model and the Hutchinson & Knopoff model is slightly stronger than the main evaluation ($\rho = .76$ versus $\rho = .74$), further supporting the proposed Stolzenburg note-specific variant as a suitable note-specific model for periodicity.

References

Bowling, D. L., Purves, D., & Gill, K. Z. (2018). Vocal similarity predicts the relative attraction of musical chords. *Proceedings of the National Academy of Sciences*, *115*(1), 216–221. https://doi.org/10.1073/pnas.1713206115

Cook, N. D. (2009). Harmony perception: Harmoniousness is more than the sum of interval consonance. *Music Perception*, *27*(1), 25–42. https://doi.org/10.1525/mp.2009.27.1.25

de Cheveigné, A. (2005). Pitch perception models. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 169–233). Springer New York. https://doi.org/10.1007/0-387-28958-5_6

Eerola, T., & Lahdelma, I. (2021). The anatomy of consonance/dissonance: Evaluating acoustic and cultural predictors across multiple datasets with chords. *Music & Science*, *4*, 20592043211030471. https://doi.org/10.1177/20592043211030471

Fastl, H., & Zwicker, E. (1990). Psychoacoustics: Facts and models. https://api.semanticscholar.org/CorpusID:144689350

Gill, K., & Purves, D. (2009). A biological rationale for musical scales. *PloS one*, *4*, e8144. https://doi.org/10.1371/journal.pone.0008144

Harrison, P. M. C., & MacConnachie, J. M. C. (2024). Consonance in the carillon. *The Journal of the Acoustical Society of America*, *156*(2), 1111–1122. https://doi.org/10.1121/10.0028167

Harrison, P. M. C., & Pearce, M. T. (2018). An energy-based generative sequence model for testing sensory theories of western harmony. *CoRR*, *abs/1807.00790*. http://arxiv.org/abs/1807.00790

Harrison, P. M. C., & Pearce, M. T. (2020). Simultaneous consonance in music perception and composition. *Psychological review*, *127*(2), 216.

Hindemith, P. (1945). *The craft of musical composition*. Associated Music Publishers.

Huron, D. (1994). Interval-class content in equally tempered pitch-class sets: Common scales exhibit optimum tonal consonance. *Music Perception*, *11*(3), 289–305.

Hutchinson, W., & Knopoff, L. (1978). The acoustic component of western consonance. *Interface*, *7*(1), 1–29. https://doi.org/10.1080/09298217808570246

Lahdelma, I., & Eerola, T. (2020). Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Scientific reports*, *10*(1), 8693.

Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of comparative Physiology A*, *181*(6), 665–676.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2010). Individual differences reveal the basis of consonance. *Current Biology*, *20*(11), 1035–1041.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550.

Milne, A. J. (2013). *A computational model of the cognition of tonality* [Doctoral dissertation, The Open University]. https://oro.open.ac.uk/38787/

Parncutt, R. (1988). Revision of terhardt's psychoacoustical model of the root(s) of a musical chord. *Music Perception: An Interdisciplinary Journal*, *6*(1), 65–93. Retrieved April 8, 2026, from http://www.jstor.org/stable/40285416

Parncutt, R. (1989). Psychoacoustics. In *Harmony: A psychoacoustical approach* (pp. 19–47). Springer.

Parncutt, R., & Strasburger, H. (1994). Applying psychoacoustics in composition: "harmonic" progressions of "nonharmonic" sonorities. *Perspectives of New Music*, *32*(2), 88–129. Retrieved April 8, 2026, from http://www.jstor.org/stable/833600

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *The journal of the Acoustical Society of America*, *38*(4), 548–560.

Schwär, S., Balke, S., & Müller, M. (2025). Measuring sensory dissonance in multi-track music recordings: A case study with wind quartets. *Ismir 2025 Hybrid Conference*.

Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale. *The Journal of the Acoustical Society of America*, *94*(3), 1218–1228.

Sethares, W. A. (2005). *Tuning, timbre, spectrum, scale*. Springer.

Stolzenburg, F. (2015). Harmony perception by periodicity detection. *Journal of Mathematics and Music*, *9*. https://doi.org/10.1080/17459737.2015.1033024

Stumpf, C. (1890). *Tonpsychologie* (Vol. 2). S. Hirzel.

Tenney, J. (1988). A history of consonance and dissonance. https://api.semanticscholar.org/CorpusID:117901311

Terhardt, E. (1982). Die psychoakustischen grundlagen der musikalischen akkordgrundtöne und deren algorithmische bestimmung. In C. Dahlhaus & M. Krause (Eds.), *Tiefenstruktur der musik* (pp. 23–50). Technische Universität Berlin.

Vassilakis, P. N. (2001). *Perceptual and physical properties of amplitude fluctuation and their musical significance*. University of California, Los Angeles.

Von Helmholtz, H. (1912). *On the sensations of tone as a physiological basis for the theory of music*. Longmans, Green.

Wang, Y., Shen, G., Guo, H., Tang, X., & Hamade, T. (2013). Roughness modelling based on human auditory perception for sound quality evaluation of vehicle interior noise. *Journal of Sound and Vibration*, *332*(16), 3893–3904.