

Article

# Geometry of Music Perception

Benjamin Himpel 

Department of Computer Science, Reutlingen University, 72762 Reutlingen, Germany;  
benjamin.himpel@reutlingen-university.de

**Abstract:** Prevalent neuroscientific theories are combined with acoustic observations from various studies to create a consistent geometric model for music perception in order to rationalize, explain and predict psycho-acoustic phenomena. The space of all chords is shown to be a Whitney stratified space. Each stratum is a Riemannian manifold which naturally yields a geodesic distance across strata. The resulting metric is compatible with voice-leading satisfying the triangle inequality. The geometric model allows for rigorous studies of psychoacoustic quantities such as roughness and harmonicity as height functions. In order to show how to use the geometric framework in psychoacoustic studies, concepts for the perception of chord resolutions are introduced and analyzed.

**Keywords:** music psychology; music theory; Riemannian geometry; Whitney stratification; music perception; voice leading; dissonance; transitional harmony; chord progression

**MSC:** 53Z05; 58A35; 57Z10; 35Q92; 92C20; 92B05



**Citation:** Himpel, B. Geometry of Music Perception. *Mathematics* **2022**, *10*, 4793. <https://doi.org/10.3390/math10244793>

Academic Editor: Luca Andrea Ludovico

Received: 10 October 2022

Accepted: 8 December 2022

Published: 16 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Jacob Collier's fascinating a cappella arrangement of "In The Bleak Midwinter" [1] modulates from the key of E to the key of G half-sharp between the third and fourth verses. This is by design, and he explains this choice in his own metaphorical language [2]. In response to the question "Why does music theory sound good to our ears?" on Wired.com Tech Support (on 26 May 2021), Jacob Collier answers "Music theory doesn't really sound like anything. It sounds like parchment. Music sounds like stuff though, and the truth is no one knows. It's a bit of a mystery." [3]. This work addresses precisely the question of how to geometrically model what music sounds like. We approach this question like a theoretical physicist would: the world consists of physical objects governed by differential equations.

### 1.1. Background

Music is based on a temporal sequence of pitched sounds. Over time, theorists have analyzed patterns in musical works and described some classes of tones, sounds and sequences thereof as pitches, chords (harmonies) and melodies/chord progressions, respectively. The resulting theory is used in turn by composers to describe their musical inventions and allow musicians to reproduce them. The theory of harmonies is also used by jazz musicians as a common basis for spontaneous musical creations.

There is a lot of research related to our differential-geometric approach to music perception. However, music psychology and music theory remain practically distinct as it was already noted by Carol Krumhansl in 1995 [4]. She empirically develops in [5] a tonal hierarchy in specific musical contexts such as scales and tonal music. Frieder Stolzenberg [6] presents a formal model for harmony perception based on periodicity detection which is compatible with prior empirical results. Harrison and Pearce [7] reanalyse and formalize consonance perception data from four previous major behavioral studies by way of a computer model written in R. Their conclusion is that simultaneous consonance derives in a large part from three phenomena: interference, periodicity/harmonicity, and cultural familiarity. This suggests that chord pleasantness is a multi-dimensional phenomenon, and

experiment design in the study of pleasantness in chord perception is highly problematic. They extend their ideas to introduce a new model for the analysis and generation of voice leadings [8]. Marjeh et al. [9] provide a detailed analysis of the relationship between consonance and timbre. A speculative account on the evolutionary aspect of consonance has been discussed in [10] with the conclusion that understanding evolutionary aspects require elaborate cross-cultural and cross-species studies. Chan et al. [11] combine the ideas of periodicity and roughness in the language of wave interferences in order to define stationary subharmonic tension (essentially the ratio of a generalization of roughness to different frequencies and periodicity) and use it to develop a new theory of transitional harmony, also known as tension and release. Tonal expectations have been analyzed from a sensoric and cognitive perspective in [12]. Dmitri Tymoczko [13,14] provides a geometric model of musical chords. He also analyzed three different concepts of musical distance and observed that they are in practise related [15]. Since our pitch perception is rather forgiving and imprecise, pitch perception corresponds to a probability distribution and therefore a smoothing should be applied to frequencies as in [16], which gives a rigorous way of evaluating similarity of chords or more generally pitch collections using expectation tensors. Differential geometry has also been used in mathematical musicology by way of gauge theory with the aim of explaining tonal attraction [17,18]. Music was viewed as a dynamical system in order to study tonal relationships [19] or musical performances [20]. On the level of audio signals, [21,22] use Hopf bifurcation control to study sound changes in music. In [23,24], music theory for classical and jazz music is formalized by providing a mathematical model for tonality, voice leading and chord progressions, which is very different from the geometric and psychoacoustic approach presented in this paper but could help in further developing it. Recent work by Wall et al. [25] analyzes voice leading and harmony in the context of musical expectancy which is precisely the motivation for our geometric model. Some very interesting vertical ideas on a scientific approach to music can be found in [26], even though there are—strictly speaking—no new results in that specific article: the brain’s exceptional ability for soft computing and pattern recognition on incomplete or over-determined data is relevant for our model. Microtonal intervals have been discussed in the context of harmony by [27,28]. Several results from cognitive neuroscience studies in the context of music perception also need to be considered for a geometric model [29–33]. A more conventional and more elaborate account on a scientific approach to music can be found in [34]. William Sethares wrote a comprehensive analysis on musical sounds based on roughness [35].

### 1.2. Aims

We hypothesize that there exists a simple underlying mathematical model and mechanism which is responsible for the harmonic and melodic development in music, in particular Western music. In order to study changes in sound and time, and since sound and time are best modelled as continuous spaces, we need differential geometry in order to study or construct musical trajectories on these spaces. Since the brain has not been understood well enough, there is currently no way of rigorously proving the correctness of a geometric model by deducing it from the way our brain processes music, even though there is a bit of work in this direction [36,37]. Instead, the goal of this research project is to validate the model by verifying its music theoretic implications. Our aim is to provide a framework from a differential geometer’s point of view in the spirit of [14,38] which is flexible enough to allow for various existing and forthcoming approaches to studying perceptive aspects of the space of notes and chords. In particular, this will remedy all the limitations of geometric models mentioned in [5] (119ff.) by making the relations between notes and chords depend on the context and the order. A focal point for this study is the cadence, “a melodic or harmonic configuration that creates a sense of resolution” [39] (pp. 105–106), which is an important basis for a lot of modern western music and has a long history in human evolution. It reduces tensions in chords, is related to falling fifths and minimizes voice leading distances. For us, it will serve as a guiding principle for the development of a

differential-geometric model. While we hope that this model generalizes to other kinds of music because of its generalist approach, our focus will be on Western music. Note that there are many approaches to analyzing and developing music based on machine learning. For the time being, we will stay away from machine learning, even though we may later use techniques from parameter optimization or artificial neural networks to narrow down the model.

Despite numerous studies on music perception, there is a need for a holistic approach by way of a common computational framework in order to study and compare various psychoacoustic quantities such as tension, consonance and roughness in a given context. In the spirit of theoretical physics, we make use of mathematical models, abstractions and generalizations in order to create a geometric framework consistent with prevalent neuroscientific theories and results. We show how to rationalize, explain and predict psychoacoustic phenomena as well as disprove psychoacoustic theories using tools from differential geometry. We will not be able to reach as far as explaining Jacob Collier's specific modulation to a half-key, but we will show why half-keys appear naturally from a psychoacoustic point of view. We will describe how this simple yet powerful differential-geometric model opens up new research directions.

### *1.3. Main Contribution*

The problem with the abundance of competing approaches to dissonance and tension, apart from the great number of different terminologies, is that they are related but not the same, the neuronal processing behind the perception of music has not been understood and the music theory does not yet have a satisfactory explanation based on existing approaches to dissonance and tension. Our geometric model has been constructed in order for these approaches to be studied, compared, and combined. Despite numerous statistical evaluations of models for dissonance and tension, none of these models can be used directly to compose music or develop music further. The main contribution is therefore to present a new approach to music perception by combining the above approaches to music cognition and geometric modelling in a simple differential-geometric model which can be used together with suitable concepts of consonance and tension to deduce the laws of music theory, lends itself to further research and musical developments, as well as provide a flexible framework to relate the perception of music and music theory. This allows for systematically and quantitatively studying the perception of music and music theory with or without just intonation or various equally or not equally tempered systems and describing new approaches to composition and improvisation in the universal language of mathematics and with the tools provided by geometric analysis. It is general enough and modular so that some or all of the concrete sensoric functions presented here can be replaced with alternative ones. A possible outcome is, that we are able to use certain gradient vectors of psychoacoustic quality functions on the space of chords that explain which chord progressions sound good (at which speed and why) and thereby provide an effective tool for composers.

### *1.4. Implications*

Therefore, the aim is not to provide yet another bottom-up approach, but to follow a top-down construction of a convenient model, which integrates roughness, consonance, tension with voice leading in order to be useful for analysing music, composing music and ultimately developing music further. In order to study time-dependent aspects of music, we need to be able to consider derivatives of psychoacoustic functions on a space of musical chords with a Riemannian structure. In particular, we want to associate musical expectation to tension on the space of chords. Even though many of the underlying ideas can be generalized, we restrict ourselves to Western music for reasons of accessibility and convenience with an octave spanning 12 semitones.

## 2. Foundations of Music Perception

In order to be able to construct a geometric framework which is consistent with the prevalent neuroscientific theories, let us first review and briefly discuss the most relevant results. Human evolution has optimized the ability of our sensory nervous system and our brain to process signals efficiently in order to quickly and easily produce the most useful interpretations and implications. Logarithmic perception of signals [40] and pattern recognition [41] are at the heart of this optimized mechanism and also provide the basis for music cognition. Signal detection theory provides a mathematical foundation for constructing psychometric functions as models for music perception [42–45]. Those readers who are not interested in the underlying mechanisms of music perception are welcome to continue with the mathematical part in Section 3.

### 2.1. Neural Coding

Sensory organs such as eye, ears, skin, nose, and mouth collect various stimuli for transduction, i.e., the conversion into an action potential, which is then transmitted to the central nervous system [46] and processed by the neuronal network in the brain as combination of spike trains [47]. Both sensation and perception are based on a physiological process of recognizing patterns in the spike trains [48].

### 2.2. Logarithmic Perception of Signals

By the Weber–Fechner law [49] the perceived intensity  $p$  of a signal is logarithmic to the stimulus intensity  $S$  above a minimal threshold  $S_0$ :

$$p = k \log\left(\frac{S}{S_0}\right).$$

Varshney and Sun [40] gave a compelling argument, why this is due to an optimization process in biological evolution where the relative error in information-processing is minimized. Quantization in the brain due to limited resources forces a continuous input signal to be perceived logarithmically. The Weber–Fechner law applies to the perception of pressure, temperature, light, time, distance and—most importantly for us—to the frequency and amplitude of sound waves.

### 2.3. Phase Locking

Synchronization and phase locking is a mechanism in the brain for organizing data, recognizing patterns and soft computing. It has also been proposed and confirmed by Langner in the case of pitches [50,51]. Phase locking for multiple frequencies has been studied in [52]. These pattern recognition capabilities can be explained by human evolution [41]. In [53] (pp. 193–213), it is argued how pattern recognition has improved over millions of years in order to allow for better predictions. It is even suggested that the current age of digitalization adds another layer of neurons to recognize new patterns. Pattern recognition is essential for living beings and humans in particular.

We immediately recognize shapes of objects and rhythmic repetitions of signals. Even if we do not see something clearly, because it is too far away, we can predict the shape within a context and thereby recognize the object. Pattern recognition in signals is based on phase-phase synchronizations. This applies for simultaneously emitted signals such as pictures and chords, but also for temporally adjacent patterns such as moving pictures and chord progressions. Signal predictions and expectations are based on a continuation of patterns. The more patterns diverge from the predicted patterns the more unexpected a signal is. Arguably, our brain prefers signals where patterns can be detected. Again, possible reasons for this can be found in evolution:

- Patterns allow us to predict events, and correctly predicting events allows us to evade dangers or kill prey.

- More abstractly, changes of patterns cause a rise of information, and we want to minimize the information we need to process,

According to [54] processes of our working memory are accomplished by neural operations involving phase-phase synchronization. We can think of working memory as an echo of firing neurons in our brain. Temporally adjacent sounds yield synchronized firings, which not only allow us to detect a rhythm but enable us to detect pitches and relate pitches to each other in chord progressions and melodies.

Quantifying phase synchronizations had been addressed by [55] which showed that phase-locking values provide better estimation of oscillatory synchronization than spectral coherence. There are other possible explanations for the relevance of simple ratios and periodicity as described in Section 4.4 based on neural coding such as cross entropy and minimizing sensoric quantities in the context of estimating distances and other measures. At this point, phase-locking seems to be as good an explanation as any for all kinds of sensoric phenomena and pattern recognition, even though it will eventually be necessary to confirm this or find better explanation for the signal expectation on the neuronal level. As the mechanism for expectation will be similar for different signals, a geometric model will help to reject explanations and find suitable ones based on psychoacoustic observations.

An example of a popular loss function is cross-entropy which is minimized for the training of artificial neural networks. Since we have a metric on each stratum we can study any height function from a differential geometric point of view. For example we can compute the differential or gradient of the dissonance function by way of which we can find the optimal direction in the space of chords to reduce dissonance as fast as possible.

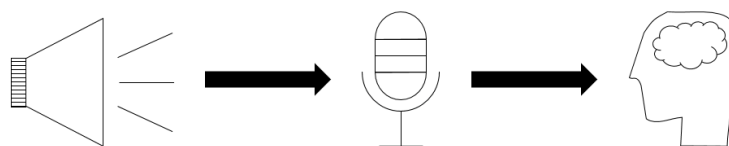
Cross-entropy might be a good mathematical concept for the purpose of pattern recognition, where we match information received with the information already stored in the brain.

#### 2.4. Audio Signals

A vibrating object causes surrounding air molecules to vibrate. As long as the kinetic energy is sustained it spreads as a wave by way of a chain reaction. This sound wave travels through the ear canal into the cochlea. Hair cells inside the cochlea convert the wave into an electrical signal, which then travels along the auditory nerve into the brain.

The audio signal goes through various stages of existence from the moment of creation to the perception in the brain. Due to a limited resolution of human perception frequency and amplitude is quantized, and the brain logarithmically perceives patterns thereof as certain sound features. These characteristics enable us to quickly recognize and describe instruments, voices and other sounds. We want to distinguish three major stages of an audio signal's existence as shown in Figure 1:

1. The produced sound, e.g., the vibrating molecules in the air as they are stimulated by a musical instrument or a loudspeaker.
2. The received sound, e.g., the vibrating microphone diaphragm or the hair cells in the cochlea, at which point the sound wave is converted into an electric signal, before it reaches the brain or different analog or digital recording devices.
3. The perceived sound, e.g., the interpretation by a person's brain.



**Figure 1.** Three major stages of the audio signal's existence.

#### 2.5. Spectrum

The shape of an object is an important factor in the way it can vibrate [56]. It can be modeled by differential equations involving the geometry of the object. There are several

such possibilities known as eigenmodes, each of which moves at a fixed frequency and amplitude as long as the energy is sustained. These eigenmodes are called partials, and the collection of all partials is known as the overtone spectrum of the audio signal. For example, the partials of an ideal vibrating string of length  $L$  fixed at both its ends are  $n/L$  for  $n \in \mathbf{N}$ . In this case, the overtone spectrum is called the harmonic spectrum.

Pattern recognition and logarithmic signal perception seem instrumental for the qualitative analysis of sound and music: A musical instrument can play different notes, but our brain detects the same spectral pattern which enables us to identify the sound as coming from the same instrument. This sound quality is also known as timbre and the process of merging several frequencies tonal fusion. Analogous mechanisms apply to voice recognition. Depending on certain deviation patterns in the spectral pattern we can classify and compare different members of the same instrument family (saxophone, clarinet, flute, string, trombone, etc.). It is also exactly this spectral pattern which allows us to recognize the different tones that are played by various sources simultaneously and to determine which instruments are playing which notes, depending on how much training we have.

### 2.6. Pitch Detection

Upper partials cannot be easily singled out, only a fundamental frequency can usually be detected by humans. Sounds, where a fundamental frequency can be detected, are called pitched sounds. The process in our brain that detects the pitch is phase locking. The same mechanism is responsible for detecting a pitch in several octaves played together and for detecting a pitch in a tone with a missing fundamental, which seems compatible with autocorrelation [57]. Several pitched tones can be played together to produce a chord, where each pitch can be detected.

Notice that different people might detect different fundamental frequencies depending on the context. This can be seen by considering the ascending Shepard's scale [58] constructed by a series of complex tones which is circular even though the pitch is perceived as only moving upward.

### 2.7. Interference

Simultaneously emitted Soundwaves interfere with each other. The interference between sine waves with slightly differing frequencies result in beatings which can be computed explicitly. Arbitrary sound waves such as those from pitched tones can be approximated by sums of sine waves. The various beatings between slightly different sine wave summands combine to a quality called roughness. Sethares [35,59] uses the Plomp–Levelt curves to provide a formula for measuring roughness and argues that this sound quality is behind tuning and scales. In particular, he suggests that some aspects of music theory can be transferred to compressed and stretched spectra, when played in compressed and stretched scales. This has been confirmed by recent results [9].

It has been shown by Hinrichsen [60] that the tuning of musical instruments such as pianos based on minimizing Shannon entropy of tone spectra is compatible with aural tuning and the Rainsback curve. While the tuning of harmonic instruments approximating twelve-tone equal temperament using coinciding partials will work, tuning inharmonic instruments in the context of Western music is more challenging [61].

Overtone singing is also an interesting aspects of interference. Possibly, overtones are sometimes not what you want to hear, maybe you want to stay away from them, because they are an unwanted artefact.

### 2.8. Just-Noticeable Difference and Critical Bandwidth

The probability for detecting a pitch change between two succeeding tones can be described rigorously using signal detection theory [42,45]. It is a collection of psychophysical methods based on statistics for analyzing and determining how signals and noise are perceived.

The just-noticeable difference (JND) also known as difference limen is often described as the minimal difference between two stimuli that can be noticed half of the time. Let us adapt the concise definition and method of computation from psychometric function analysis provided by [62] to pitch changes. Suppose a subject is presented two succeeding tones as part of a pitch discrimination task. One of the tones is called the reference pitch  $p$ , the other the comparison pitch  $c$ . Responses  $R_1$  and  $R_2$  correspond to the choices  $c < p$  and  $c > p$ , respectively. There is no option  $c = p$ . A small set of tone pairs are repeated a number of times (15 to 20), and the subject has to choose one of the two responses. A psychometric function models the proportion of either  $R_1$  or  $R_2$ . For a fixed reference pitch  $p$  the psychometric function for  $R_2$  should be a monotonically increasing function in the comparison pitch  $c$  with values between 0 and 1, because for  $c$  much bigger than  $p$  the correct response  $R_2$  should be obvious. We will assume for simplicity that the shape of the curve fitted to the data follows a cumulative Gaussian as in Figure 2, even though other functions such as sigmoid, Weibull, logistic or Gumbel are also a possibility [63]. The point of subjective equality (PSE) is the comparison pitch at which the two responses in this discrimination task are equally likely, i.e., the median. Then, the JND is defined to be half its interquartile range, i.e.,

$$\text{JND} = \frac{c_{0.75} - c_{0.25}}{2},$$

where  $c_{0.25}$  and  $c_{0.75}$  represent the comparison pitches, at which a change is detected with probability 0.25 and 0.75, respectively.

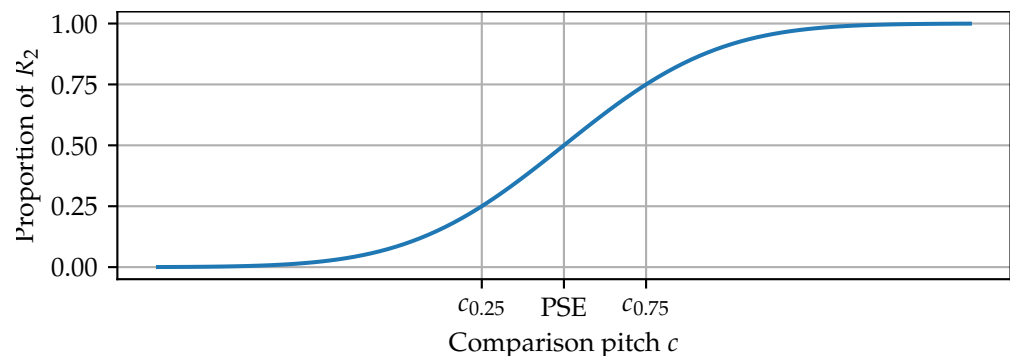


Figure 2. Psychometric function with quartiles  $c_{0.25}$ , PSE and  $c_{0.75}$ .

Notice that when two tones are played in succession the JND is bigger than when the two notes are played simultaneously. This is due to the interference discussed in Section 2.7. Astonishingly, Section 7.2.2 of [64] states that the JND for two succeeding tones with a pause (difference) is three times higher than the without a pause (modulation). Figure 7.2 in [65] shows that the just-noticeable frequency modulation is approximately 3 Hz below 500 Hz and 0.7% of the frequency above 500 Hz. Clearly, the JND depends on the observer as well as other circumstances (noise) that might interfere with the perception of the signal.

The critical band is the frequency bandwidth within which the interference between two tones is perceived as beats or roughness, not as two separate tones. The JND is a lot smaller than the critical bandwidth. According to [66] “a critical band is 100 Hz wide for center frequencies below 500 Hz, and 20% of the center frequency above 500 Hz”. A comparison between the critical band and the JND can be seen in Figure 7.2 of [65] which in turn is based on Figure 12 in [67].

In the context of periodicity, Stolzenburg [6] uses the JND of 1% and 1.1% or, equivalently,  $\log_2(1.01) \cdot 12 = 0.014355 \cdot 12 \approx 17.23$  cent and  $\log_2(1.011) \cdot 12 = 0.015783 \cdot 12 \approx 18.94$  cent. In [16], a standard deviation of 3 cent has been used due to experimentally obtained frequency difference limens of supposedly 3 cent [68], even though the value of 1% in [68] corresponds to about 18 cent as we have just seen. Still, the fact that they used the standard deviation of 3 cent for the Gaussian smoothing is an interesting aspect that we will revisit in Section 4.4. It will be necessary to design experiments and perform further studies along the lines of [69] to

collect data for periodicity discrimination in the light of pitch and roughness correlations for tones within chords and between different chords, determine the best model and describe the dependency on noise [43,44,70], which is beyond the scope of this work. Due to a lack of such a study, we will assume that cultural familiarity lets us associate slightly mistuned pitches with an ideal pitch and thereby detect and use the implied pitch for the perception of music.

### 2.9. Music Perception

Let us define music to be a temporal sequence of pitched sounds created by a formal system. Formal systems obey a set of rules for sound and rhythm, which are ultimately based on physics and mathematics respectively. Different cultures developed and are continuing to develop a variety of systems and scales besides the ones used in Western music [71], for example Gamelan music [35,72], Arabic music [73,74], Turkish music [75] and classical Indian music [76]. In Western music, there are major subsystems such as classical and jazz music. Enculturation is an important factor in the listener's musical expectation and perception [77,78], but we want to focus on a specific prevalent and in some way universal aspect of music, namely, pitch [79,80]. While the space of received sounds lends itself to a mathematical model, e.g., by using the frequencies and amplitudes computed by Fourier analysis, the spaces of produced and perceived sounds can be compared to it. Given a good microphone connected to some recording device and a good understanding of particle physics the space of produced sounds should be more or less the same as the space of received sounds. Our brain transforms sound waves of music by applying additional filters and perceiving pitch, timbre and loudness. There is also a short term memory effect in the brain, which we hypothesize to be responsible for the sense of resolution in certain chord progressions.

The perception of every person is different and can change via training or degradation. Sound and music are therefore very subjective and can be compared to food, in the sense that the chemical content of food corresponds to the Fourier decomposition of a sound, food can be analysed using chemistry just like we can analyse sound using Fourier analysis or harmony theory, different tastes can be analyzed using signal detection theory and can be described using various characteristics such as spiciness, sweetness, sourness, temperature, etc., just like sounds can be characterized as warm, loud, sweet, rough, etc., via a psychoacoustic analysis. In addition there is an after-taste to food, which might influence the characteristics of food-to-come, just like chord progressions need to be viewed within a musical context.

Chords are also called harmonies and play a key role in Western music. These can sound consonant or dissonant, and the change in this characteristic is an important aspect of musical pieces. Composers build up tension and resolve it subsequently by way of cadences. Notice that it clearly is not only a question of how consonant or dissonant chords sound in a chord progression: the precise way or direction of chord movement is important. It is this kind of aspect in music, that we want to illuminate by geometrically modeling the perception of chords. To this end, we revisit the geometric model of chords [13] with a focus on music perception.

### 2.10. Mathematics and Music

While sound seems to be well-understood by physics and mathematical structures can be found at every point in music, neither one gives a deep understanding by providing a general principle of how music is perceived by humans. On the other hand, music itself is in reality a mathematical concept based on the brain's perception of sound, put into action in a creative and aesthetically pleasing way: Any kind of scale has been developed mathematically to be compatible with some acoustic observations, rhythm is a time-dependent structure governed by elementary mathematics. Western music theory is a formal system consisting of an assortment of rules that have been deduced from various psychoacoustic preferences. An account of the major aspects surrounding mathematics



and music can be found in [81]. We want to emphasize the difference between two types of mathematical structures:

The first kind consists of superimposed formal systems in order to give music more structure and to make it more interesting. It starts with simple structures such as note lengths and bars to organize rhythm. Other examples include composition procedures such as the fugue characterized by imitation and counterpoint as well as various special techniques such as Kanon, Krebs, Umkehrung. Then, there is the twelve-tone technique invented by Arnold Schoenberg [82]. For some of these structures we assume a twelve-tone equal temperament, which is itself a mathematical structure superimposed on pitched sounds, not accidentally but deliberately based on a second type of mathematical structure.

This second kind is more subtle, originally due to an evolutionary process and a preference for patterns but ultimately caused by psycho-physical mechanisms such as phase locking. It captures the structure inherent in music. It covers temporal structures such as rhythmic repetitions. Most Western instruments have approximately a harmonic overtone spectrum. Guided by the simultaneous or sequential perception of intervals and chords humans developed scales, instruments and music theory. Already Pythagoras discovered that simple rational relationships between fundamental frequencies correlate with pleasant sounding intervals. The twelve tones in an octave are also the result of simple rational relationships between frequencies, even though the two physical psychoacoustic qualities harmonicity/periodicity and roughness/interference have been shown to be fundamentally different [9,35]. Music theory is a formal system which captures more subtle perceptual aspects in Western music. It developed over centuries by the efforts of countless musicians and theorists, mainly however due to observed perceptive qualities of chord progressions.

Concise models of physical observations can be formulated in the universal language of mathematics, whose powerful tools allow us to deduce complex facts from simple ones. Therefore, the goal is to find a simple way of modeling sounds in the context of music perception, from which we can for example deduce good sounding chord progressions independent of functional harmony, create a music theory in other less common music systems as well as ultimately explain the established Western music theory of harmonies.

### 3. Riemannian Geometry of Chords

Tymoczko [13] viewed the space of chords with  $n$  notes as an orbifold [83]. In [84], the orbifold of chords had been generalized from a topological point of view, while we focus on the geometry. We argue that it is a Riemannian orbifold [85] and show that the space of chords  $\mathcal{C}$  with an arbitrary numbers of notes is a Whitney stratified space [86] endowed with a metric given by the geodesic distance. The metric provides voice leading distance across different strata. Chord progressions can formally be viewed as sections of the (trivial)  $\mathcal{C}$ -bundle over the real line. While our motivation is its use for Western music with its twelve-tone equal temperament, it can readily be adopted to other music. For simplicity, the geometric model represents the chords that can be played using a single instrument which can produce musical tones at any frequency (like a violin) but cannot duplicate notes (like a piano).

Pitches and frequencies can formally be identified with integers via  $B3 = -1$ ,  $C4 = 0$ ,  $C\sharp4 = 1$ , etc. Therefore unit distance corresponds to a pitch distance of 100 Cent, which is compatible with the musician's perception of distance between musical tones. The identification between frequency and pitch numbers is given by the function

$$\begin{aligned} \text{pitch} : \mathbf{R} &\rightarrow \mathbf{R}, \\ f &\mapsto 12 \cdot \log_2(f/f_0), \end{aligned}$$

where  $f_0 = 261.626\text{Hz}$  corresponds to  $\text{pitch}(f_0) = 0 = C4$ . Chords can then be identified with integer tuples  $(p_1, \dots, p_n) \in \mathbf{Z}^n$ . Instead, we will identify chords with tuples  $(p_1, \dots, p_n) \in \mathbf{R}^n$  for the following reasons:

- There are usually minor pitch adjustments to make chords sound “better”.
- The fundamental frequency  $f_0$  can assume different values.
- Quarter tones are entirely legitimate.
- There are other tuning systems.
- In particular, not even the piano is tuned using twelve-tone-equal temperament but their stretched tuning follows the Rainsback curve [60,87].
- We assume that instruments play pitches and that the perceived pitch is most relevant for our purpose. We do not include the overtone spectrum with all its amplitudes. When it becomes necessary it can easily be introduced.

Since chord notes are played simultaneously, the order of pitches  $p_i$  in a chord is irrelevant. For example, the dominant seventh chord (0, 4, 7, 11) needs to be identified with (4, 0, 7, 11).

**Lemma 1.** Let  $S_n$  be the finite symmetric group of all bijective functions  $\{1, \dots, n\} \rightarrow \{1, \dots, n\}$ .

1. The permutation

$$s : \mathbf{R}^n \rightarrow \mathbf{R}^n$$

$$(p_1, \dots, p_n) \mapsto (p_{s(1)}, \dots, p_{s(n)}).$$

is a left action on  $\mathbf{R}^n$ .

2. The relation

$$\tilde{c}_1 \simeq \tilde{c}_2 :\Leftrightarrow \exists s \in S_n : s(\tilde{c}_1) = \tilde{c}_2 \quad \text{for } \tilde{c}_j \in \mathbf{R}^n$$

is an equivalence relation on  $\mathbf{R}^n$ .

**Proof.** Bijective functions of finite sets form a group.

1. We compute that

$$s_2(s_1(p_1, \dots, p_n)) = s_2(p_{s_1(1)}, \dots, p_{s_1(n)}) = (p_{s_2(s_1(1))}, \dots, p_{s_2(s_1(n))})$$

$$(p_{(s_2 \circ s_1)(1)}, \dots, p_{(s_2 \circ s_1)(n)}) = (s_2 \circ s_1)(p_1, \dots, p_n).$$

Therefore  $S_n$  acts on  $\mathbf{R}^n$  from the left.

2. Clearly, the relation is reflexive since  $\tilde{c}_1 = \tilde{c}_1$ . If  $\tilde{c}_1 \simeq \tilde{c}_2$ , we have  $s(\tilde{c}_1) = \tilde{c}_2$  for some  $s \in S_n$ . Since  $S_n$  is a group we have  $\tilde{c}_2 = s^{-1}(\tilde{c}_1)$ . Therefore  $\tilde{c}_2 \simeq \tilde{c}_1$ , and symmetry is satisfied. If  $\tilde{c}_1 \simeq \tilde{c}_2$  and  $\tilde{c}_2 \simeq \tilde{c}_3$ , then  $s_1(\tilde{c}_1) = \tilde{c}_2$  and  $s_2(\tilde{c}_2) = \tilde{c}_3$  for some  $s_1, s_2 \in S_n$ . Therefore  $(s_2 s_1)(\tilde{c}_1) = \tilde{c}_3$  and  $\tilde{c}_1 \simeq \tilde{c}_3$  so that the relation is transitive.  $\square$

Then the quotient by the symmetric group action is given by  $\mathbf{R}^n / S_n := \mathbf{R}^n / \simeq$ , and its elements are written as  $[p_1, \dots, p_n]$ . This space  $\mathbf{R}^n / S_n$  is known as the  $n$ -the symmetric power of  $\mathbf{R}$  and is an example of an orbifold [83], a generalization of a manifold which is locally a quotient of a differentiable manifold by a finite group action. We can also identify notes with the same name but in different octaves before we consider the quotient by  $S_n$ . Then, we get the toroidal orbifold  $(\mathbf{R}/12\mathbf{Z})^n / S_n$  considered by Tymoczko [13,14] in order to study efficient voice leading. From a mathematical point of view, this orbifold does not behave differently from  $\mathbf{R}^n / S_n$ , but this model is not suitable for music perception. More importantly for us, Theorem 1 shows that it is a Riemannian orbifold [85,88–90] and a Riemannian orbit space [91–94].

**Definition 1.** A Riemannian orbifold is a metric space which is locally isometric to orbit spaces of isometric actions of finite groups on Riemannian manifolds. A Riemannian orbit space is the quotient of a Riemannian manifold by a proper and isometric Lie group action.

**Proposition 1.** Consider the  $L^p$  metric on Euclidean space  $\mathbf{R}^n$ . Then, the symmetric group  $S_n$  acts on  $\mathbf{R}^n$  by isometries.

**Proof.** Let  $(p_1, \dots, p_n), (q_1, \dots, q_n) \in \mathbf{R}^n$ . Then, we obtain for any  $s \in S_n$  by commutativity of the sum

$$\begin{aligned} d((p_1, \dots, p_n), (q_1, \dots, q_n)) &= \sum_{k=1}^n (q_k^p - p_k^p)^{1/p} = \sum_{k=1}^n (q_{s(k)}^p - p_{s(k)}^p)^{1/p} \\ &= d((p_{s(1)}, \dots, p_{s(n)}), (q_{s(1)}, \dots, q_{s(n)})). \quad \square \end{aligned}$$

This yields the following.

**Theorem 1.** *The quotient space  $\mathcal{S}_n := \mathbf{R}^n / S_n$  is a Riemannian orbifold and a Riemannian orbit space.*

In order to study chord progressions it is necessary to consider chords of varying size. We need to construct a metric space of chords with an arbitrary number of tones that is useful for describing music. The metric should provide a sensible voice leading distance, in particular for chord progressions of the form  $[0, 3] \rightarrow [0, 3, 4]$  or  $[0, 3] \rightarrow [0, 3, 3]$ . Multiple same pitches as well as transitions between chords with a different number of tones can be dealt with by considering multiple same pitches in a chord only once, just like a piano plays chords. For example,  $[0, 0, 4, 7, 11]$  is identified with  $[0, 4, 7, 11]$ .

**Proposition 2.** *Consider the set of chords*

$$\mathcal{S} := \left( \bigcup_{k=1}^{\infty} \mathcal{S}_k \right).$$

*The relation*

$$[p_1, p_2, \dots, p_k] \sim [p_1, p_2, \dots, p_{k-1}] \Leftrightarrow [p_1, p_2, \dots, p_k] \simeq [p_1, p_1, p_2, \dots, p_{k-1}]$$

*for all  $k = 2, \dots, n$  is an equivalence relation on  $\mathcal{S}$ .*

**Proof.** This is an immediate consequence of  $\simeq$  being an equivalence relation.  $\square$

This allows us to define the space of all chords.

**Definition 2.** *Let  $\mathcal{C} := \mathcal{S} / \sim$  the space of all chords and  $\mathcal{C}_n := (\bigcup_{k=1}^n \mathcal{S}_k) / \sim$  the space of chords with at most  $n$  pitches. Let  $\mathcal{U}_n := \{(p_1, \dots, p_n) \subset \mathbf{R}^n \mid p_i \neq p_j \text{ for } i \neq j\}$ . Let  $\pi : \mathbf{R}^n \rightarrow \mathcal{C}_n$  be the quotient map  $(p_1, \dots, p_n) \mapsto [p_1, \dots, p_n]$ .*

**Remark 1.** *Notice that  $\mathcal{C}_n \setminus \mathcal{C}_{n-1}$  is the set of chords with exactly  $n$  different pitches.*

**Example 1.** *The space  $\mathcal{C}_2$  is the Euclidean plane as shown in Figure 3, where the points are identified with their mirror image when reflected across the diagonal, essentially equivalent to the lower (or the upper) triangle of the plane.  $\mathcal{C}_1$  consists of the singular points with respect to this reflection and is the boundary of  $\mathcal{C}_2$ .*

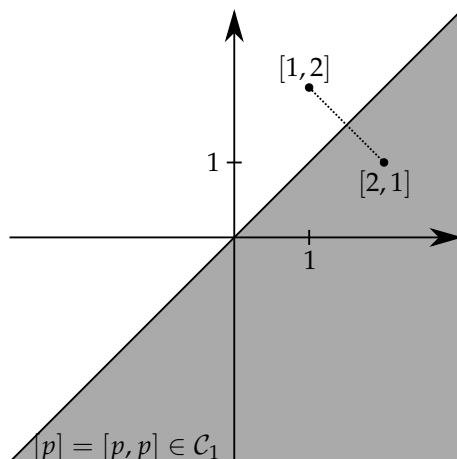


Figure 3. The space  $\mathcal{C}_2$ .

**Lemma 2.** The Stabilizer  $S_p$  of the action of  $S_n$  on  $\mathbf{R}^n$  is trivial for each  $p \in U_n$ .

**Proof.** The Stabilizer  $S_p$  of action of  $S_n$  on  $\mathbf{R}^n$  is given by  $\{s \in S_n \mid s(p) = p\}$ . If  $s \neq 1$  then  $s(i) = j$  for some  $i \neq j$ . Then,  $p_i \neq p_{s(j)}$  and therefore  $s(p) \neq p$ . Therefore  $S_p$  is trivial for each  $p \in U_n$ .  $\square$

**Proposition 3.**  $\mathcal{C}_n \setminus \mathcal{C}_{n-1}$  is a Riemannian manifold of dimension  $n$ . The space of chords  $\mathcal{C}$  is the disjoint union of  $\mathcal{C}_n \setminus \mathcal{C}_{n=1}$

$$\mathcal{C} = \bigsqcup_{n=1}^{\infty} (\mathcal{C}_n \setminus \mathcal{C}_{n=1}).$$

**Proof.** Due to Lemma 2 we have  $[\tilde{c}_1] \sim [\tilde{c}_2] \Leftrightarrow \tilde{c}_1 \simeq \tilde{c}_2$  for  $\tilde{c}_1, \tilde{c}_2 \in U_n$ . Therefore,  $\pi : U_n \rightarrow \mathcal{C}_n \setminus \mathcal{C}_{n-1}$  is a canonical bijection, and  $\mathcal{C}_n \setminus \mathcal{C}_{n-1}$  inherits the Riemannian metric from  $\mathbf{R}^n$ .  $\square$

**Remark 2.** The family of chords  $\{\mathcal{C}_n\}_{n \in \mathbf{N}}$  is an example of a filtration

$$\mathcal{C}_1 \subset \mathcal{C}_2 \subset \dots \subset \mathcal{C}_n \subset \dots$$

By Proposition 3 the filtration  $\{\mathcal{C}_k\}_{k=1, \dots, n}$  of  $\mathcal{C}$  is an infinite-dimensional stratification, and  $\mathcal{C}_k \setminus \mathcal{C}_{k-1}$  are the strata of dimensions  $k$ .

**Remark 3.** The Riemannian metric  $g_n$  provides a norm  $\|v\|_n$  for every  $v \in T_p S_n$ . Furthermore, the Riemannian metric  $g_n$  makes the orbifold  $S_n$  into a metric space using the geodesic distance defined by

$$d(p, q) := \inf \left\{ \int_a^b \|\rho'(t)^p\|_n^{1/p} dt \mid \begin{array}{l} \rho : [a, b] \rightarrow S_n \text{ piecewise smooth,} \\ \rho(a) = p, \rho(b) = q \end{array} \right\} \text{ for } p, q \in S_n.$$

**Proposition 4.** The distance on  $S_n$  can be computed via

$$d_n(p, q) = \min_{s \in S_n} \tilde{d}_n(p, s(q))$$

where  $\tilde{d}$  is an  $L^p$ -metric on  $\mathbf{R}^n$ .

**Proof.** In Euclidean space, the geodesic distance is given by the  $L^p$ -metric. Let  $p, q \in S_n$ . Consider two representatives  $(p_1, \dots, p_n), (q_1, \dots, q_n) \in U_n$  of  $p$  and  $q$  with  $p_i < p_{i+1}$  and  $q_i < q_{i+1}$ . Then,  $tq_i + (1-t)p_i < tq_{i+1} + (1-t)p_{i+1}$  for  $t \in [0, 1]$  which implies

$t(q_1, \dots, q_n) + (1 - t)(p_1, \dots, p_n) \in U_n$ . Therefore  $U_n$  is convex. Since  $U_n$  is a fundamental domain for  $U_n/S_n$ , the distance on  $U_n/S_n$  is equal to the Euclidean distance in  $U_n$ . Since the canonical projection  $U_n \rightarrow \mathcal{S}_n \setminus \mathcal{S}_{n-1}$  is an isometric bijection, it follows that for  $p, q \in \mathcal{S}_n \setminus \mathcal{S}_{n-1}$  we have  $d_n(p, q) = \min_{s \in \mathcal{S}_n} \tilde{d}_n(p, s(q))$ . Since the closure of  $\mathcal{S}_n \setminus \mathcal{S}_{n-1}$  is also convex the same formula holds for all  $p, q \in \mathcal{S}_n$ .  $\square$

Chord progressions in  $\mathcal{S}_n$  with a small distance  $d_n$  correspond to efficient voice leading. The metric  $d_n$  on  $\mathcal{S}_n$  clearly yields a metric on each stratum  $\mathcal{C}_n \setminus \mathcal{C}_{n-1}$ . Finding a suitable distance on all of  $\mathcal{C}$  is problematic. We can define the following functions  $d_n$  and  $d$  and  $\mathcal{C}_n$  and  $\mathcal{C}$ , respectively:

$$d_n(c_1, c_2) := \min_{\tilde{c}_j \in \mathcal{C}_j} d_n(\tilde{c}_1, \tilde{c}_2) \quad \text{and} \quad d(c_1, c_2) := \min_{n \in \mathbf{N}} d_n(c_1, c_2).$$

For example, we compute

$$\begin{aligned} d_3([0, 1, 7], [0, 6, 7]) &= 5, \\ d_4([0, 1, 7], [0, 6, 7]) &= d_4([0, 1, 7, 7], [0, 0, 6, 7]) = 2. \end{aligned}$$

Even if this is considered to be suitable for determining efficient voice leading, the following shows that this is not a metric on  $\mathcal{C}_n$ .

**Proposition 5.** *The functions  $d_n$  and  $d$  on  $\mathcal{C}_n$  and  $\mathcal{C}$  do not satisfy the triangle inequality.*

**Proof.** Since we have

$$d([0], [0, 1]) + d([0, 1], [0, 1, 2]) = 1 + 1 < 3 = d([0], [0, 1, 2]), \tag{1}$$

(see Figure 4) this generalization does not satisfy the triangle inequality. The same holds for  $d_n$ .  $\square$

Since the aim is to do differential geometry on  $\mathcal{C}$  the following result is important. See [86] for a detailed treatment of stratified spaces from a geometric analysis point of view.

**Theorem 2.** *For each  $n \in \mathbf{N}$ , the filtration  $\{\mathcal{C}_k\}_{k \in \mathbf{N}}$  is a Whitney stratification of  $\mathcal{C}$ .*

**Proof.** We show that Whitney’s condition B is satisfied. Consider the strata  $X := \mathcal{C}_k \setminus \mathcal{C}_{k-1}$  and  $Y := \mathcal{C}_l \setminus \mathcal{C}_{l-1}$  for  $k > l$  and embed them in some  $\mathbf{R}^N$  via a map  $\iota : \mathcal{C}_k \rightarrow \mathbf{R}^N$ . Let  $x_1, \dots$  and  $y_1, \dots$  be sequences of points in  $X$  and  $Y$ , respectively, both converging to the same point  $y \in Y$ , such that the sequence of secant lines  $L_i$  between  $x_i$  and  $y_i$  converges to a line  $L \subset \mathbf{R}^N$  in real projective space  $\mathbf{RP}^N$  and the sequence of tangent planes  $T_i$  to  $X$  at the points  $x_i$  converges to a  $k$ -dimensional plane  $T$  of  $\mathbf{R}^N$  in the Grassmannian  $\mathbf{Gr}(k, \mathbf{R}^N)$  as  $i$  tends to infinity. The points  $x_1, \dots$  uniquely lift to a sequence  $\tilde{x}_1, \dots$  in  $\mathbf{R}^k$ . Let  $\tilde{y}$  be the lift of  $Y$  to  $\mathbf{R}^k$  so that  $\tilde{x}_1, \dots$  converges to  $\tilde{y}$ . Choose the lift  $\tilde{Y} \subset \mathbf{R}^k$  of  $Y$  such that  $\tilde{y} \in \tilde{Y}$ . Then,  $y_1, \dots$  uniquely lifts to a sequence  $\tilde{y}_1, \dots$  of points in  $\tilde{Y}$  that converge to  $\tilde{y}$ . Each tangent plane  $T_i$  pulls back to the only plane in  $\mathbf{R}^k \in \mathbf{Gr}(k, \mathbf{R}^k)$ . The secant lines between  $(\iota \circ q)^{-1}(x_i)$  and  $(\iota \circ q)^{-1}(y_i)$  converge to a line  $\tilde{L}$  in  $\mathbf{RP}^k$  which is contained in  $\mathbf{R}^k$ . This implies that its push-forward  $L = d(\iota \circ q)_{\tilde{y}} \tilde{L}$  is contained in  $d(\iota \circ q)_{\tilde{y}} \mathbf{R}^k = T$ .  $\square$

Since every stratum of  $\mathcal{C}$  is a metric space and a Riemannian manifold, and the notion of piecewise smooth paths makes sense in  $\mathcal{C}$ , we can define the geodesic distance on  $\mathcal{C}$  as follows.

**Definition 3.** *We call a continuous path  $\rho : [a, b] \rightarrow \mathcal{C}$  piecewise smooth, if there exists a partition  $a = x_1 < \dots < x_N = b$  of  $[a, b]$  such that  $\rho$  restricted to  $(x_i, x_{i+1})$  is a smooth path in  $\mathcal{C}_{n_i} \setminus \mathcal{C}_{n_i-1}$  for*

some  $n_i \in \mathbf{N}$ . Let  $\rho : [a, b] \rightarrow \mathcal{C}$  be a piecewise smooth path, then we define  $\|\rho'(t)\| := \|\rho'(t)\|_{n_i}$  if  $\rho(t) \in \mathcal{C}_{n_i} \setminus \mathcal{C}_{n_i-1}$ . The geodesic distance on  $\mathcal{C}$  is

$$d(p, q) := \inf \left\{ \int_a^b \|\rho'(t)\|^p dt \mid \begin{array}{l} \rho : [a, b] \rightarrow \mathcal{C} \text{ piecewise smooth,} \\ \rho(a) = p, \rho(b) = q \end{array} \right\} \text{ for } p, q \in \mathcal{C}.$$

**Theorem 3.** The function  $d$  is a metric on  $\mathcal{C}$ . It can be computed via

$$d(p, q) = \inf \left\{ \sum_{i=1}^{n-1} d_i(x_i, x_{i+1}) \mid x_i \in \mathcal{C}_i \setminus \mathcal{C}_{i-1} \right\}$$

**Proof.** Clearly,  $d(p, p) = 0$ . Since every stratum is a metric space and we have only a finite number of strata, we obtain  $d(p, q) > 0$  for  $p \neq q$  and  $d(p, q) = d(q, p)$ . The concatenation of any piecewise smooth path from  $p$  to  $q$  and from  $q$  to  $r$  in  $\mathcal{C}$  is a piecewise smooth path from  $p$  to  $r$ , so that the triangle inequality holds. Therefore, the function  $d$  is a metric.

Let  $\rho_i : [a, b] \rightarrow \mathcal{C}$  be a sequence of piecewise smooth paths with  $\rho_i(a) = p$  and  $\rho_i(b) = q$  with a partition  $a = x_1 < \dots < x_N = b$  of  $[a, b]$  such that  $\rho$  restricted to  $(x_i, x_{i+1})$  is a smooth path in  $\mathcal{C}_{n_i} \setminus \mathcal{C}_{n_i-1}$  for some  $n_i \in \mathbf{N}$  whose length converges to  $d(p, q)$ . Since  $\mathcal{C}_{n_i} \setminus \mathcal{C}_{n_i-1}$  is convex, this implies

$$d(p, q) = \sum_{i=1}^{N-1} d_{n_i}(x_i, x_{i+1}).$$

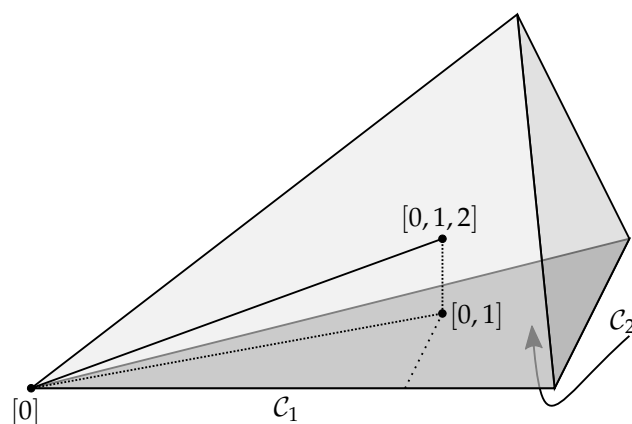
Furthermore, we can assume that  $n_i > n_{i-1}$  because of this convexity.  $\square$

The metric on  $\mathcal{C}$  can be considered as a voice leading distance for music theory.

**Example 2.** Let us compute the distance between  $[0]$  and  $[0, 1, 2]$ . It can be computed by minimizing the concatenation of geodesic paths within  $\mathcal{C}_3 \setminus \mathcal{C}_2$  and  $\mathcal{C}_2 \setminus \mathcal{C}_1$ , and we obtain

$$\begin{aligned} \delta([0], [0, 1, 2]) &= \min_{p \geq 0} (d_2([0], [0, p]) + d_3([0, p], [0, 1, 2])) = \min_{p \geq 0} (|p| + |p - 1| + |p - 2|) \\ &= 1 + 0 + 1 = 2 \end{aligned}$$

In particular, we confirm together with  $\delta([0], [0, 1]) = 1$  and  $\delta([0, 1], [0, 1, 2]) = 1$  that the triangle inequality has not been violated as it was in Equation (1). See Figure 4.



**Figure 4.** The stratum  $\mathcal{C}_3$ .

In summary, Theorems 1–3 show that  $\mathcal{C}$  is a well-behaved differential-geometric space:

1.  $\mathcal{C}$  is a metric space,
2.  $\mathcal{C}$  is a Whitney stratified space, and

3. each stratum of  $\mathcal{C}$  is a Riemannian manifold.

This provides a rich structure for quantitative studies of psychoacoustic models with a voice leading distance on all of  $\mathcal{C}$ . The Riemannian metric allows us to study the shape of melodies and chord progressions by differentiating psychoacoustic functions and computing directional derivatives of paths in  $\mathcal{C}$ . This model is universal in the sense that it allows note and chord progressions in any musical system.

**4. Sound Perception of Chords**

We relate psychometric functions to psychoacoustic height functions on  $\mathcal{C}$ . Contour plots of psychoacoustic functions on  $\mathcal{C}$  provide us with insightful visualizations of different models for consonance such as roughness and periodicity. Timbre and loudness are also important perceptive quantities, which can be addressed later. The Riemannian structure on  $\mathcal{C}$  allows us to study the shape of melodies and chord progressions as paths in  $\mathcal{C}$  and the perception thereof by differentiating psychoacoustic lifts of the paths in  $\mathcal{C}$  in Section 5.

*4.1. Psychoacoustic Functions on the Space of Chords*

While the space of musical chords can be modelled geometrically, independently of the listener, and a music score can be viewed as a sequence of points or a path in this space, sound perception varies and corresponds to different psychoacoustic functions on this space: dissonance, musical expectation, sense of resolution, root of chord, interference/roughness, all of which depend on both player and listener. Usually, these functions are real-valued on the space of chords (with a given spectrum/timbre) and quantify the individual sensation. This kind of a function turns out to be an example for an important mathematical tool in geometry, analysis and optimization known as a height function on a surface, manifold or more generally a Whitney stratified space. Since the psychoacoustic function varies with the listener and noise it is natural to analyze them using psychometric functions introduced in Section 2.8.

Assuming that the JND for pitch discrimination is the same for every reference pitch, let us give a different perspective on the psychometric function from Figure 2. Consider the Gaussian distribution  $\phi_{\mu,\sigma}$  given by given by

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

with mean  $\mu = \text{PSE}$  and standard deviation  $\sigma = \text{JND}/0.674490$ , as well as the Heaviside step function

$$f(x) := \begin{cases} 0 & \text{if } < 0 \\ 1 & \text{if } \geq 0. \end{cases}$$

Then the function in Figure 2 is equal to the convolution  $f * \phi_{\mu,\sigma}$  given by

$$(f * \phi_{\mu,\sigma})(p) := \int f(x)\phi_{\mu,\sigma}(p - x)dx.$$

Consider now a task which is slightly different from the one presented in Section 2.8: For a given reference pitch  $p$  a subject has to say, whether a tone with pitch  $c$  has the same pitch as  $p$  or not. Let us reformulate the task using random variables. Let  $X_{p,c}$  be the random variable which is 1 (yes) when a comparison pitch  $c$  is perceived as the reference pitch  $p$  and 0 (no) otherwise. We can go one step further and consider the continuous random variable  $X_p$  which equals  $c$  when  $p$  is perceived as  $c$ . Then, the probability distribution of  $X_p$  is given by the normal distribution  $\phi_{\mu,\sigma}$  with  $\sigma$  and  $\mu$  as above. For our purpose let us assume that PSE is equal to  $p$ .

Again, we can view the probability distribution  $\phi_{\mu,\sigma}$  as a convolution of  $\phi_{\mu,\sigma}$  with point mass at 0 or, equivalently, as a convolution of  $\phi_{0,\sigma}$  with point mass at  $\mu = p$ . We observe that  $\phi_{\mu,\sigma}(c) = 0.5$  for  $c = p \pm \text{JND}$ . Now that we have set up the notation, we can

ask which pitch we expect to hear when a tone with pitch  $p$  is played. Clearly, it should be  $p$ , and we can confirm this by computing the expectation value of  $X_p$ :

$$E(X_p) = \int_{-\infty}^{\infty} P(X_p = c) \cdot c \, dc = \int_p^{\infty} \phi_{\mu,\sigma}(c) \cdot (c + (2p - c)) \, dc = \frac{1}{2}2p = p.$$

We will use this as a basis for modeling psychoacoustic functions as an expectation value of certain random variables associated to psychometric functions. In [16], this viewpoint has been used in order to model perceived distance between pairs of pitch collections, where the perceived dissimilarity was reformulated as a metric between expectation tensors.

As we will see in Section 4.4, consonance of dyads and chords is likely to be determined by certain nearby pitches with low periodicity which in turn is due to the phase locking and pattern recognition principle described in Section 2.3. Let us therefore discuss the following multi-variate scenario. Given a fixed set of  $N$  pitches  $\mathcal{P} = \{p_1, \dots, p_N\}$  with  $p_i < p_{i+1}$  and  $JND < |p_{i+1} - p_i| < 2 \cdot JND$ , a subject has to choose one pitch from  $\mathcal{P}$  which is equal or closest to a given pitch  $c$ . Let  $X_{\mathcal{P},c}$  be the random variable which equals  $p_i$  if a perceived pitch  $c$  is closest to  $p_i$ . Clearly, we expect a smoothed version of a step function for the expectation value  $E(X_{\mathcal{P},c})$  as a function of  $p$  where the steps are located at  $(p_i + p_{i+1})/2$ . One might be tempted to use the convolution of the step function with  $\phi_{\mu,\sigma}$  as above, but by doing so we have neglected the subtle interplay of the random variables and possible dependencies. If we interpret  $E(X_{\mathcal{P},c})$  as

$$E(X_{\mathcal{P},c}) = E \left( \bigvee_{i=1}^N \left( X_{p_i,c} \wedge \bigwedge_{j \neq i} \overline{X_{p_j,c}} \right) \right),$$

we take into account the knowledge that  $c$  is not perceived as  $p_j$  for  $j \neq i$ , but we neglect terms of the form  $X_{p_i,c} \wedge X_{p_j,c}$  or  $\overline{X_{p_1,c}} \wedge \dots \wedge \overline{X_{p_N,c}}$ . If we assume that  $X_{p_i,c}$  and  $X_{p_j,c}$  are independent random variables for  $i \neq j$  we can apply the product formula for independent random variables.

Under the premise that common chord progressions in music theory and their psychoacoustic properties find their justification in certain sound qualities, the chord model  $\mathcal{C}$  together with its sound qualities given by certain height functions on  $\mathcal{C}$  is not only interesting for the music theorist and psychoacoustic analyst, but can become a powerful tool in the hands of composers and computer programs emulating composers because of its conceptual simplicity and quantitative control. Even though  $\mathcal{C}$  could theoretically extend to include the whole overtone spectrum, we hypothesize that different spectra will simply change the psychoacoustic height functions, as long as the spectra consistently have almost the same pattern.

Since there are instruments that do not produce a harmonic series in overtones, it will be interesting to analyse how music and music theory changes for these instruments. A change in the interference scheme due to a different overtone spectrum will promote different note systems. This can be observed in history and other cultures because of the construction of different scales for instruments, which do not produce a harmonic series. Possibly, the relationship of periodicity/harmonicity and consonance needs to be re-evaluated: Is it due to the almost harmonic spectrum of the notes produced by most musical instruments, is it connected to the way human beings interpret periodicity of chords, or are there other more basic concepts at work such as logarithmic perception and pattern recognition? However, if it depends on our interpretation of chords, is this due to enculturation or our physical and chemical processing of sound?

In summary, height functions based on mathematical quantitative models for psychoacoustic quantities on the space of chords allow for rigorous studies on music perception. Once the correctness of mathematical models has been confirmed they will yield new music theories. In our work we focus on the psychoacoustic concepts of consonance and tension/release in music. From a psychometric point of view it will be necessary to conduct



further studies regarding these psychoacoustic quantities. We will see that experiments must be carefully designed as in [9] due to the fortunate (from a Western musical point of view) and at the same time the undesirable (from a scientific point of view) correlations between roughness and periodicity.

#### 4.2. Consonance

Consonance is a psychoacoustic quality of perceived chords considered to be an important factor in Western music with the usual twelve-tone equal temperament system. Two or more musical tones are considered consonant/dissonant, if they sound pleasant/unpleasant together, and there are a variety of explanations for this phenomenon [95]. The most important ones go back to roughness (interference) by Helmholtz [96] and tonal fusion (neural periodicity) by Stumpf [97,98]. The discussion in [7] carefully analyses various different psychoacoustic interpretations, evaluates data from previous studies, provides a code for several computational models and shows their correlation with consonance ratings. They conclude that consonance depends on interference/roughness, periodicity/harmonicity, and cultural familiarity. While the first two are based on physically justifiable phenomena independent of the individual, cultural familiarity is different for every person by way of musical expertise and cultural conditioning in the following ways:

1. Musical training actively and systematically changes your perception. In particular, it allows to better differentiate how consonant chords sound.
2. The cultural context passively changes your perception by repetition. In particular, it determines how consonant chords sound. E.g. certain jazz chords sound dissonant to people who are unfamiliar with the jazz idiom, while they sound pleasant to jazz musicians.

Tension, a concept of horizontal harmony between consecutive chords, had also been linked to dissonance [99], but [100] suggests that tension is less subjective to cultural familiarity and musical expertise than consonance, pleasantness and harmoniousness of chords. A recent study [101] determined that roughness influences automatic responses in a simple cognitive task while harmonicity did not. Furthermore, [11] argues that tension is independent of harmonicity because it has been shown in [102] that it is possible for a more consonant chord to resolve into a more dissonant chord. Even though we expect tension to be related to harmonicity, it is apparently fundamentally different from the vertical quality of consonance and should be reflected in the model accordingly. The difficulty in this discussion surrounding consonance and tension is that in reality they are a conglomeration of different psycho-acoustic phenomena. Furthermore, the terminology might be misleading: Horizontal harmony needs to be viewed in musical context, therefore we will call it the *resolve* instead of tension.

Dichotic presentation (different ears for different tones) of chords preserves harmonicity and reduces roughness [103], therefore roughness cannot be responsible for the psychological effect of consonance for chord resolutions, even though roughness and consonance are highly correlated during diotic presentations (same ear for all tones) and will increase the respective effects. The difference of harmonicity and roughness has also been studied in [36]. It is legitimate to say that interference plays a role for the construction of scales, tuning and the quantification of sensory dissonance [35], but we hypothesize that there is a fundamental mechanism in the brain that is responsible for the effect of consonance and tension (for a given scale) in the context of chord resolutions and for the way Western music has developed. In particular, such a mechanism should in principal not depend on how badly in tune the notes of a chord sound as long as the chord is approximately correct, and it should not depend on whether the chord tones are presented diotically or dichotically. Therefore, we can ignore roughness and beatings for the purpose of studying the mechanism behind chord resolutions. Nevertheless, roughness will strengthen the effect harmonicity has on the listener and will play a role for more subtle variations and fine-tuning of ideal chord progressions.

From a neurophysiological point of view, we hypothesize that roughness, harmonicity and the resolve all find their neural coding origin in the same phase locking principle:

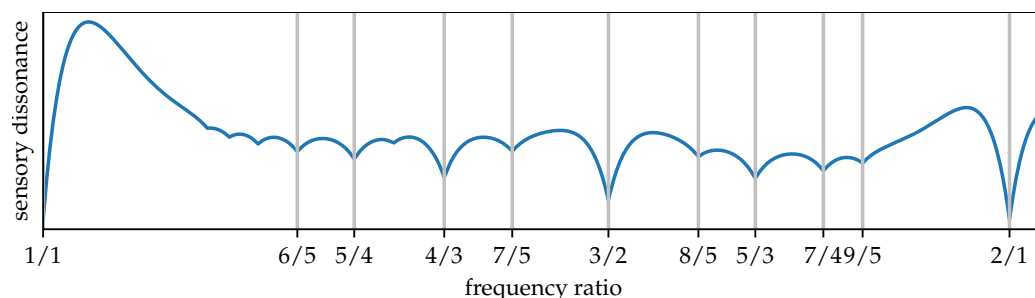
- Roughness is based on the interference of sine waves and can be perceived even during dichotic presentation of dyads. It is usually determined using a spectral analysis which will be reviewed briefly in Section 4.3, but it can also be modeled by the synchronization index model using the degree of phase locking to a particular frequency within the neural pattern [104] and [35] (Appendix G).
- Harmonicity can be modeled via periodicity [6], which is based on phase locking of perceived pitches and will be discussed in Sections 4.4 and 4.5.
- The resolve has not been studied much with respect to the phase locking principle, but we hypothesize that it depends on the interplay between the working memory and harmonicity. Not only has harmonicity been successfully computed via neural periodicity, but working memory has also been linked to phase-phase synchronization [54]. Some ideas are developed in Section 5.1.

In summary, the three physically justifiable phenomena roughness, harmonicity and resolve are correlated, and their respective psycho-acoustic effects on the listener are amplified by this correlation and by cultural familiarity. These mechanisms are often presented as explanations of consonance, even though they address different issues within the perception of music. The aim of the following sections is therefore to define, distinguish and elaborate upon the individual psycho-acoustic phenomena related to consonance in the context of  $\mathcal{C}$ .

#### 4.3. Roughness

Nineteenth-century physicist Herman von Helmholtz [96] was the first to notice a relation between the harmonic series and the pleasantness of chords, based on which he proposed a theory of consonance and dissonance. In short, he argued, that each tone played by a musical instrument consists of a series of partials determined by the harmonic series: The fewer partials the spectra share, the more dissonant they should be. The interaction between sound waves is called interference, and the interference between two different but similar sine waves create beatings within and roughness outside a critical bandwidth of frequency.

While Western music is usually based on twelve-tone equal temperament, this specific tuning is really a compromise for musical instruments whose pitches are fixed. The pitches of notes for more flexible instruments such as the violin or the saxophone are usually adjusted slightly in order to produce chords with minimal or the right amount of roughness. Even pianos are not tuned using twelve-tone-equal temperament but their stretched tuning follows the Rainsback curve [87]. Sethares [35] describes how roughness between complex notes can be computed based on the interference between their partials. He argues that this is one of the main reasons for having a twelve-tone equal temperament system, and that it is a useful tool for tuning and intonating instruments. However, while roughness might be behind tuning, and you want to mostly reduce roughness, it is simply an acoustic artifact that you need to take into account in order to have exactly the correct amount of roughness, just like some coinciding partials whose audibility you want to control. A graph of roughness for dyads can be seen in Figure 5 (adapted from [105]). The roughness function fits very well into our geometric framework. A contour graph of roughness for triads can be seen in Figure 6.21 of [35]. One small issue is the fact that the model is not differentiable at its local minima. A possible remedy is the modeling approach by [104]. Its roughness graph of a harmonic tone complex can be seen in [104] (Figure 4). It remains to be seen how deep we have to dive into other aspects such as cochlear hydrodynamics [106] in order to improve the roughness model for further studies on music perception.



**Figure 5.** Sensory dissonance for dyads in terms of their frequency ratio.

It will be interesting to study roughness together with the geometric model  $\mathcal{C}$  to compute and visualize entropy and determine ideal tunings. For details, formulas and graphs of roughness we refer to [35]. More importantly for us we need to study roughness in combination with harmonicity, because both types of consonance are relevant for music in their own way, but their specific psycho-acoustic effects independently of each other are not clear yet.

#### 4.4. Harmonicity of Dyads

In order to create a suitable harmonicity height function  $p : \mathcal{C} \rightarrow \mathbf{R}$  on musical chords, we will focus on the harmonicity model determined by relative (logarithmic) periodicity as presented by Stolzenburg [6]. His explanations based on the neuronal model by Langner [50,51] using phase locking are convincing, even if probabilistic implications of the psychometric functions apart from the JND have not been included and other aspects such as roughness and cultural familiarity clearly alter the perception of chords. In short, if the ratio of two pitch frequencies  $f_1$  and  $f_2$  with  $f_2 \geq f_1$  is given by  $f_2/f_1 = p/q$  with  $\gcd(p, q) = 1$ , then the periodicity for this dyad is  $q$ . In other words, the period of the sound wave for this dyad equals  $q$  periods of the first (lower) note.

Since the above periodicity  $q$  will change a lot for small changes of the ratio  $f_2/f_1 = p/q$ , our brain will pick the smallest  $q$  within a JND for harmonicity through phase locking as discussed in Section 2.8. It is chosen to be 1% and 1.1% in [6] based on related results by [64,67,68,107–113]. As we have seen in Section 2.8 this corresponds to approximately 18 cent.

A naive model for periodicity is therefore given by a step function with a JND of 18 cent for periodicity as shown in Figure 6, but we need to keep in mind that depending on the listener, the loudness and distracting noise the JND might vary. Furthermore, in order to compute JND for harmonicity of simultaneously played tones we need to design a new experiment, where we can analyze the effects of roughness and harmonicity separately.

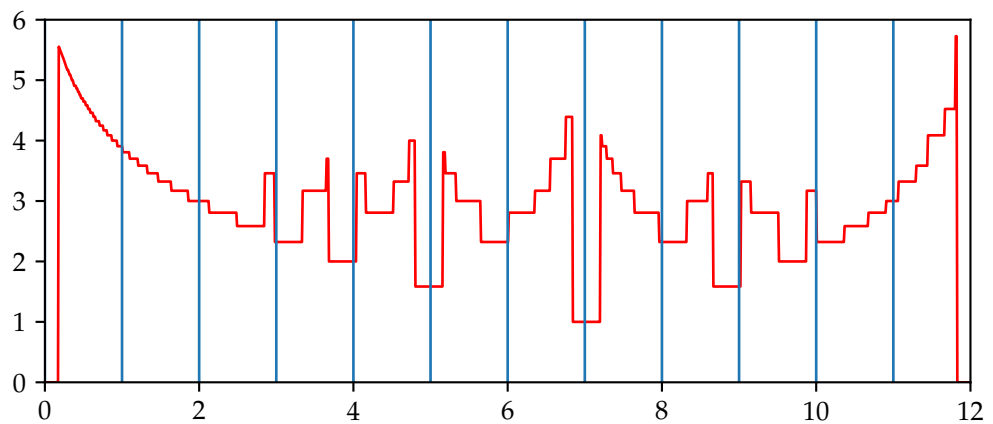
Notice however, that by incorporating probabilistic aspects via Gaussian smoothing after first constructing a step function resembling the periodicity based on [6] we commit a conceptual error, which we are not able to correct in this work but which is hopefully small enough to still provide useful results. The perceived periodicity of a chord is determined by the period that is the best fit for the given spike train induced by the audio signal. The brain either chooses the smallest period it can detect or it detects a mixture of periodicities as an average. It is also possible that different periods are detected at different times within a small time interval due to small variations in the spike sequence or in the pitch. Spike trains with low periodicities are more likely to be detected than spike trains with high periodicities. In order to create a better model we need to take into account these probabilistic issues already within the phase locking stage and make use of probabilistic tools such as cross entropy and coherence in the time domain along the lines of [55].

Let us consider a dyad in 12-tone equal temperament as discussed in Section 3. If the lower note is fixed, a dyad spanning at most one octave is determined by the number of separating semitones  $i \in \{0, \dots, 12\}$ . Its frequencies  $f_i$  within a JND of 1.1%, its relative periodicities  $L_i$  and its logarithm are given in Table 1.

**Table 1.** Frequencies and periodicities relative to pitch 0.

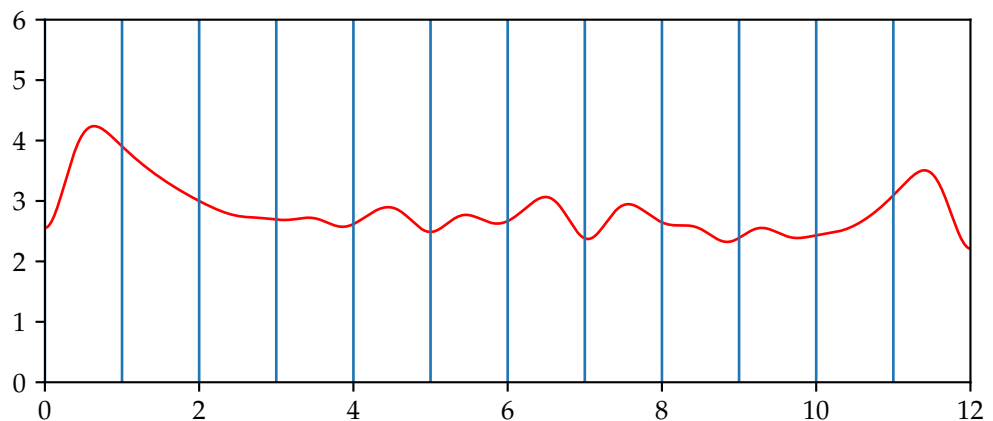
$i$	0	1	2	3	4	5	6	7	8	9	10	11	12
$f_i$	1	$\frac{16}{15}$	$\frac{9}{8}$	$\frac{6}{5}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{7}{5}$	$\frac{3}{2}$	$\frac{8}{5}$	$\frac{5}{3}$	$\frac{9}{5}$	$\frac{15}{8}$	$\frac{2}{1}$
$L_i$	1	15	8	5	4	3	5	2	5	3	5	8	1
$\log_2(L_i)$	0	3.91	3	2.32	2	1.58	2.32	1	2.32	1.58	2.32	3	0

Observe that the concept of voice leading is also related to the periodicity of an octave being 1. It allows the player to change the voicing of a chord without changing the psychological effect of its sound by much. Certainly, periodicities can be computed for all intervals as a function  $f_p$  for all dyads  $[0, p]$ , where  $p \in [0, 12]$ . Its graph is shown in Figure 6, where the JND is 18 cent. Notice that the step functions has jumps very close to some of the integers.



**Figure 6.** Logarithmic periodicities of dyads spanning at most one octave.

As we have discussed above, in order to obtain a smooth height function on the space of chords in the spirit of psychometric functions we can consider the convolution with a Gaussian. A standard deviation of  $\sigma = \text{JND}/0.674490$  which we discussed in Section 4 to be the correct value in the context of psychometric functions seems much too big. When applied to the step function the resulting graph can be seen in Figure 7.



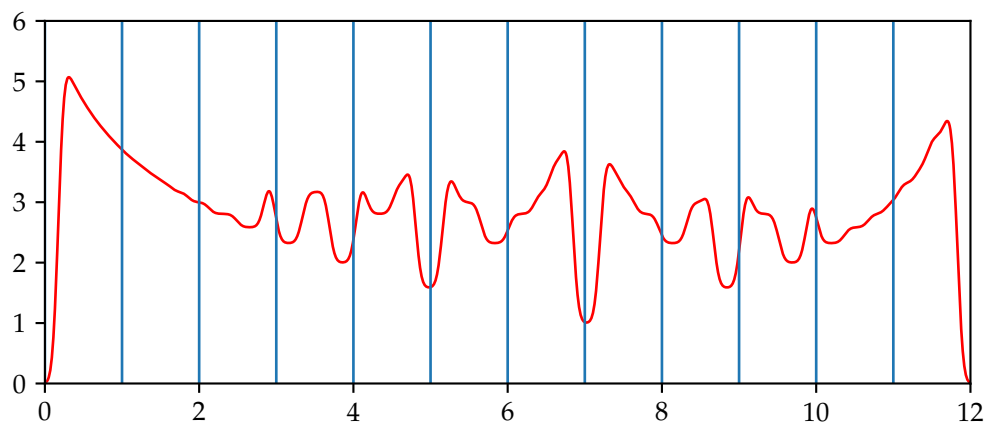
**Figure 7.** Logarithmic periodicities smoothed by a Gaussian with standard deviation of  $\sigma = \text{JND}/0.674490 \approx 26.69$  cent.

In order to keep the appropriate maxima and minima of the step function a standard deviation of  $\sigma = \text{JND}/3 = 6$  cent seems better. The result is shown in Figure 8. There are a few reasons why this smaller  $\sigma$  is more appropriate. First of all, [16] suggests a

minimum standard deviation of 3 cent. Even though this alone is not a good enough reason, especially because [16] refers to [68], where the difference limen has been computed to be approximately 1%, it suggests that a careful psycho-metric analysis of harmonicity, roughness and pitch needs to be conducted that sheds some light on their interdependence. We hypothesize that a side effect of roughness is the increase of phase locking precision for the detection of harmonicity. In combination with pitch detection the conditional probability for detecting the correct harmonicity will also increase, because the product of two Gaussians with standard deviations  $\sigma_1$  and  $\sigma_2$  is again a Gaussian with (smaller) standard deviation

$$\sigma = \sqrt{\frac{\sigma_1^2 \cdot \sigma_2^2}{\sigma_1^2 + \sigma_2^2}}$$

We realize that these reasons need to be elaborated on, treated more rigorously and their effects quantified, but this needs to be conducted elsewhere. Instead we will lift the periodicity function with its visually and subjectively satisfactory parameters to higher dimensions.



**Figure 8.** Logarithmic periodicities smoothed by a Gaussian with standard deviation of  $\sigma = \text{JND}/3 = 6$  cent.

#### 4.5. Harmonicity of Arbitrary Chords

The definition of periodicity generalizes to chords with more than two notes by letting the periodicity be the smallest positive integer  $q$  satisfying  $q/f_1 = p_2/f_2 = p_n/f_n$  for some  $p_2, \dots, p_n \in \mathbf{N}$ , where  $f_1$  is the frequency of the lowest note. Equivalently, periodicity is the smallest positive integer satisfying  $f_2/f_1 = p_2/q, f_3/f_1 = p_3/q, \dots, f_n/f_1 = p_n/q$ , in other words,  $q$  is the least common multiple of the denominators in the irreducible fractions representing the frequencies relative to  $f_1$ . Let  $\mathcal{C}_n^{[0,12]} \subset \mathcal{C}_n$  be the subspace of all chords where each tone is contained in the octave  $[0, 12]$  and the base note is equal to 0. (This can easily be generalized to chords spanning more than an octave.) Define the chords  $\mathcal{C}_n^p \subset \mathcal{C}_n^{[0,12]}$  with periodicity  $p$  via

$$\mathcal{C}_n^p := \left\{ \left[ 0, 12 \cdot \log_2\left(\frac{p_2}{q_2}\right), \dots, 12 \cdot \log_2\left(\frac{p_n}{q_n}\right) \right] \mid \forall_i \left( 1 \leq \frac{p_i}{q_i} \leq 2 \wedge \text{gcd}(p_i, q_i) = 1 \right) \wedge \text{lcm}(q_2, \dots, q_n) = p \right\}$$

Again, we assume a JND of 18 cent between every two notes of a chord. Even though relative periodicity resembles harmonicity well qualitatively, Stolzenburg [6] considers logarithmic periodicity as a computational model for harmonicity because of the Weber-Fechner law as discussed in Section 2.2. We generalize JND to chords by determining a polyhedral neighborhood  $N_c \subset \mathcal{C}_n$  for each chord  $c = [c_1, \dots, c_n] \in \mathcal{C}_n \setminus \mathcal{C}_{n-1}$  in which there is no noticeable difference compared to  $c$ . Formally, we have for JND = 18 cent

$$N_c := \{[c_1 + d_1, \dots, c_n + d_n] \mid \forall_{i,j=1,\dots,n} (d_i \in [-\text{JND}, \text{JND}] \wedge |d_i - d_j| \leq \text{JND})\}. \quad (2)$$

We could try generalizing periodicity to arbitrary chords using Table 1. Following Example 10 in [6], the first inversion of the diminished triad can be written as  $[0, 3, 9]$ ,  $[-3, 0, 6]$  and  $[-9, -6, 0]$ . These representations have relative periodicities 15, 25 and 6 depending on which note in this triad is considered to be pitch 0 in Table 1. To illustrate the computation we compute frequency ratios  $[5/6, 1, 7/5]$  from Table 1 for  $[-3, 0, 6]$  which translates to  $[1, 6/5, (6/5) \cdot (7/5)]$  for  $[0, 3, 9]$  and results in an overall periodicity of  $\text{lcm}(1, 5, 5 \cdot 5) = 5 \cdot 5 = 25$ . In [6], this problem of potentially having different periodicities for the same chord has been solved by computing the average of the three periodicities (both raw and logarithmic), i.e., raw  $(15 + 25 + 6)/3 \approx 15.3$  and logarithmic  $(\log_2(15) + \log_2(25) + \log_2(6))/3 \approx 3.7$ . Even though this gives good empirical results, it seems to contradict the “rational tuning” principle, which uses the fractions with the smallest denominator approximating equal temperament within a certain error margin. For example,  $19/16$ ,  $13/11$  and  $6/5$  all approximate the frequency ratios for the minor third, and 5 is the relative periodicity. For the same reason, the relative periodicity for the the first inversion of the diminished triad should be 6 and not an average. In summary, the algorithm for finding the rational tuning provided by [6] should be generalized to arbitrary chords rather than using the frequencies in Table 1.

Let us therefore modify the computation of periodicity slightly and not use the proposed smoothing from [6]. We define for  $c \in \mathcal{C}_n^{[0,12]}$

$$p(c) := \min\{p \mid N_c \cap \mathcal{C}_n^p \neq \emptyset\}. \quad (3)$$

Informally, we choose the best fit of periodicity for each chord within a JND for every two notes, rather than averaging over periodicities. Algorithm 1 yields the periodicity of a chord with  $n$  notes as a step function on  $\mathcal{C}_n^{[0,12]}$ . We have used a resolution of 100 cent per semitone. For  $n = 4$  the array size of  $\mathcal{C}_n^{[0,12]}$  is therefore  $1200^4 \approx 2 \cdot 10^{12}$  which was our computational limit. This can be implemented more efficiently, e.g., by only considering all  $c \in \mathcal{C}_n^p$  in a neighborhood of *RemainingChords* and by reducing the resolution.

---

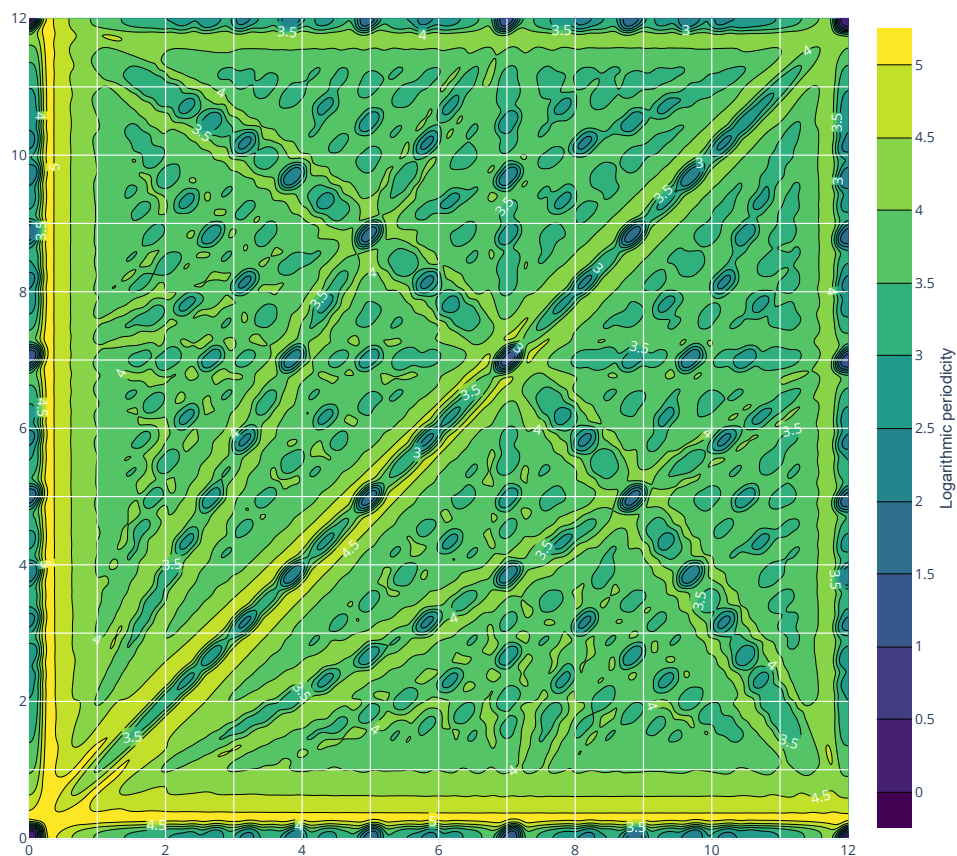
**Algorithm 1** Determine periodicity step function  $p : \mathcal{C}_n^{[0,12]} \rightarrow \mathbf{R}$

---

**Require:**  $n \geq 1$  ▷  $n$ =number of chord tones  
 $q \leftarrow 1$  ▷  $q$ =periodicity index  
*RemainingChords*  $\leftarrow \mathcal{C}_n^{[0,12]}$  ▷ Consider all chords of the form  $[0, c_2, \dots, c_n]$   
**while** *RemainingChords*  $\neq \emptyset$  **do** ▷ While there are chords without periodicity  
    **for all**  $c \in \mathcal{C}_n^q$  **do** ▷ For all chords with periodicity  $q$   
        **for all**  $d \in N_c \cap \textit{RemainingChords}$  **do** ▷ For all new chords within JND  
             $p(d) \leftarrow q$  ▷ Set periodicity to  $q$   
            *RemainingChords*  $\leftarrow (\textit{RemainingChords} \setminus N_c)$  ▷ Update new chords  
        **end for**  
    **end for**  
     $q \leftarrow q + 1$  ▷ Increase periodicity index by 1  
**end while**

---

This can be smoothed as discussed above using a Gaussian with standard deviation 6 cent. Figure 9 visualizes the resulting logarithmic periodicity function  $\log_2 p(c)$  for triads spanning at most one octave; we normalize a triad in continuous pitch space to be of the form  $[0, x, y]$  with  $x, y \in [0, 12]$  and draw the graph as a contour plot in the  $xy$ -plane with the height  $z$  given by the logarithmic periodicity. The intersection points of the grid lines correspond to chords in twelve-tone equal temperament,  $\mathcal{C}_2$  diagonally embeds into  $\mathcal{C}_3$ , and the second inversion  $[0, 5, 9]$  of the major triad appears to be the most consonant chord consisting of 3 different tones.



**Figure 9.** Contour plot for logarithmic periodicities of  $[0, x, y] \in C_3^{[0,12]}$  using JND = 18 cent smoothed by a Gaussian with standard deviation  $\sigma = 6$  cent cent

## 5. Music Perception

Time adds a layer of complexity to sound perception by way of considering paths and sequences in  $\mathcal{C}$ . We will focus on musical expectation as well as on tension and release. Our ability to anticipate future events is another vital aspect of human evolution. Perceptual expectation has been studied in cognitive neuroscience [69], and it is also a fundamental part of music perception [25,114–118]. Clearly, musical expectation depends on the listener, or, more precisely, on his brain and its musical training [119]. It involves recognizing and predicting patterns both in sound and time set within a context. Among musicians this is also known as the concept of tension and release. Some aspects of its neuro-acoustic mechanism have been studied in the literature [11,120,121]. While roughness plays a role in tuning, scales and the sound perception of chords, it is not audible in the psychoacoustic interaction between consecutive chords. We will demonstrate in this section why we should and how we can analyze a periodicity approach to tension and release using tools from differential geometry.

### 5.1. Tension and Release

Concepts related to the harmonic transitions are the circle of fifths, the Tonnetz model by Euler [122,123], the tonal pitch space by Lerhdahl [124], the tonal hierarchy by Bharucha and Krumhansl [5,125], the gauge-theoretic approach to tonal attraction [17,18], and similar geometric structures describing harmonic relationships [126–128]. Plus, the bass line clearly plays an important role in the perception of chord progressions [129,130]. There have been several studies about the relationship between these horizontal approaches and roughness, (vertical) harmonicity and as well as cultural familiarity [100,119,121,131,132]. Both the models by Lerhdahl [124] and by Bharucha and Krumhansl [125] capture and describe distances for harmonic motions, and can both be viewed as a metric space in the

mathematical sense [133]. We will not view perceptive distances in harmonic transitions as a metric, because the order of tones or chords matters. Instead, we will show how to make the relations between notes and chords depend on the context and the order, and thereby remedy the limitations of geometric models mentioned in [5] (119ff.). The intricate interplay between the voice leading distance presented in Section 3 and harmonic transition is important, even though harmonies and harmony theory are often discussed without considering the individual note movement. The cognitive mechanisms are related and interfere with each other to create the sensation of transitional harmony. This is hinted at in [127,128].

Let us discuss the potential factors that affect transitional harmonicity. On the one hand we expect the basic principles behind harmonicity to play a role. We have not found much empirical evidence, but we hypothesize that the neuronal network mechanisms behind many sensations and particularly between horizontal and vertical harmony should be the same, and Tramo et al. [36] (p. 96) also suggest that the vertical and the horizontal dimension of harmony is related. Therefore phase locking or an even more fundamental physiological principle will be behind transitional harmony. On the other hand we expect pattern recognition and the universal ability of detecting minima in sensoric input to be important. Minimization is implicitly used in defining the voice leading distance presented in Section 3 given by the geodesic distance.

For our purpose there are two fundamentally different kinds of expectations:

1. If you listen to a piece of music, you can predict how it continues. You might be able to anticipate a few notes and chords depending on your training and background. Your anticipation will be based on tempo, meter, rhythm, melody, dynamics, form, chord progression which are time-dependent aspects of context. In the language of our geometric model: From a path in  $\mathcal{C}$  we want to anticipate its continuation. This will depend on its speed and its shape, including its direction, its curvature and other geometric aspects. You can compare a musical piece to a roller coaster ride, which you should construct or analyze using differential geometry. However, it will also depend on a second, time-independent kind of expectation.
2. Assume you are listening to a single chord, and you have to predict which chord could follow. You might wonder, which context this is in, and this might partially be responsible for your expectations. As before, it is based on your training and background. However, there are physical reasons for your anticipations as well. This certainly has to do with consonance of chords and voice leading, but also with the order in which two different chords are played. We can view this expectation either as a psychoacoustic evaluation of difference vectors on  $\mathcal{C}$  or of ordered pairs of chords. We call this time-independent psychoacoustic quality for an ordered pair of chords the resolve. This time-independent quantity has been studied under different names in [100,134–136], but we would like to emphasize its dependence on its contextual reference by giving it this new name.

A progression of notes and chords with or without additional bass notes can be described as a sequence of points in  $\mathcal{C}$ , which can be viewed as a discretization of a path in  $\mathcal{C}$  parameterized by time. It can be approximated by a differentiable path. Either way, we can study differential geometric properties such as speed, momentum, acceleration, and angular speed in order to analyze and understand chord progressions better. Furthermore, we can consider differential geometric properties of the path after applying suitable height functions. We hypothesize that the time-independent expectation can be deduced from the resolve by way of differential geometry. If we are at a point  $p \in \mathcal{C}$ , we can quantify the resolve as a height function on  $\mathcal{C}$ . In other words, the resolve is a function on  $\mathcal{C} \times \mathcal{C}$ .

Some interesting questions arise, which we do not attempt to answer here: Is the resolve the result of a priming with ordered pairs of chords based on cultural familiarity and training, or is it a multi-dimensional vector intrinsic to the starting chord or a local neighborhood of the starting chord, i.e., without the necessity of having ever heard the second chord? Is the training happening on the level of some basic neuronal mechanism



for any chord progression or do we need all kinds of pairs of chords as training data? Do musicians and composers imagine the succeeding chord or do they sense the direction in which they have to move the notes?

Research from [11] attempts to quantify the resolve. They call it transitional harmony and compute it via

$$\Delta\Delta\hat{f} := \frac{\Delta t_p - \Delta t_s}{T_{sub}}, \text{ where } \Delta t = [k_i t_i]_{max} - [k_i t_i]_{min}, \tag{4}$$

where  $[k_i t_i]_{max}$  and  $[k_i t_i]_{min}$  are the largest and smallest multiples of the chord tone periods that (nearly) coincide with the chord periodicity  $T_{sub}$  which we introduced in Section 4.5 and where the indices  $s$  and  $p$  correspond to the succeeding and nearest preceding chord, respectively. Even though the authors have found some strong correlations shown in Table 3 of [11] supporting the validity of  $\Delta\Delta\hat{f}$ , we question the definition due to its strong dependence on small pitch changes: music perception should not change a lot by small pitch variations, but it does in the definition given by Equation (4). We hypothesize that the correlations found in Table 3 of [11] are due to the correlation between harmonicity and roughness for instruments with harmonic spectra.

### 5.2. Two-Chord Progressions Starting with a Tritone

In order to motivate various approaches to the resolve we consider two-chord progressions of dyads within 12TET starting on a tritone  $[F3, B3]$  where at least one note changes and each note does not move more than a semitone. Let us ignore the choice of octave in this section. There is a total of eight such chord movements.

The two parallel tritone movements considered on their own and out of context do not sound like they resolve anything, but adding the bass lines  $C\sharp3 \rightarrow F\sharp2$  or  $G2 \rightarrow C2$  yields the standard chord progression  $II^7 \rightarrow V^7$  as  $[C\sharp3, E\sharp3, B3] \rightarrow [F\sharp2, E3, A\sharp3]$  and  $[G2, F3, B3] \rightarrow [C2, E3, B\flat3]$ , respectively, where the tonalities are clearly very far away from each other. The first note in this notation always corresponds to the bass note. On the one hand this simple example confirms the well-known assumption that chords should always be viewed in a context, but on the other hand it represents the charm behind the technique of modulation in music. It is therefore nevertheless necessary to consider chord progressions without a given tonality or context. It just leaves chord progressions ambiguous, and probabilistic methods can be employed.

The strongest resolution from the perspective of periodicities or ratios should clearly be the progression to the perfect fifth  $[F3, B3] \rightarrow [E3, B3]$  or  $[F3, B3] \rightarrow [F3, C3]$ . However, it does not sound like a good way of resolving the tritone. If we think of the notes as attracting or repelling magnets then both the notes should move in opposing directions in order to resolve the dissonance, which we will consider in the next paragraph. However, we can again add bass lines to make the first progression sound like the jazz resolution to the major seventh chord  $V^7 \rightarrow I^\Delta$  given by  $[G2, F3, B3] \rightarrow [C2, E3, B3]$  and the second progression like the resolution  $V^{\dim7} \rightarrow I^\Delta$  or  $V^7 \rightarrow I^{\text{sus}4} \rightarrow I$  partially represented by  $[A\flat2, F3, C\flat4] \rightarrow [D\flat2, F3, C4]$  and  $[G2, F3, B3] \rightarrow [C2, F3, C4] \rightarrow [C2, E3, C4]$ , respectively.

The chord progression into a perfect fourth  $[F3, B3] \rightarrow [F3, B\flat3]$  or  $[F3, B3] \rightarrow [F\sharp3, B3]$  also does not sound like a good way of resolving the tritone. Again, we can put them in a suitable context by adding bass lines. The first progression sounds like the jazz resolution to the major seventh chord  $V^7 \rightarrow I^\Delta$  partially given by  $[D\flat, F, B] \rightarrow [G\flat2, F3, B\flat3]$  and the second progression like the resolution  $V^{\dim7} \rightarrow I^\Delta$  or  $V^7 \rightarrow I^{\text{sus}4} \rightarrow I$  partially represented by  $[D3, F3, B3] \rightarrow [G2, F\sharp3, B3]$  and  $[G2, F3, B3] \rightarrow [C2, F3, C4] \rightarrow [C2, E3, C4]$ , respectively.

The best sounding dyad progression is the tritone  $[F3, B3]$  resolving into the major third  $[G\flat3, B\flat3]$  or the minor sixth  $[E3, C4]$ . Even though these progressions already sound like resolutions, it helps to view them in a context and a tonality in order to relate them to music theory. Possibly, our brain has already been primed for possible tonalities, and some tonalities are more probable than others. Clearly, the corresponding chord progressions are  $V^7 \rightarrow I$  partially represented by  $[D\flat3, F3, C\flat4] \rightarrow [G\flat2, G\flat3, B\flat3]$  and  $[G2, F3, B3] \rightarrow$

[C2, E3, C4]. Notice that the tonality is already determined by the progression of dyads, the bass line only emphasizes the tonal center. An insightful work by Tom Sutcliffe [137] picks up on the gap in the literature of failing to explain why voice leading in combination with root progressions is used in tonal pieces.

Let us describe a few possible approaches to transitional harmony between two chords, consider the differential geometry and revisit the above example.

### 5.3. Transitive Periodicity from the First to the Second Chord

Musical structures such as rhythmic patterns and periodicity cause phase locking [138]. We therefore assume that the brain relates two chords  $c_1$  and  $c_2$  of a chord progression  $c_1 \rightarrow c_2$  through the working memory based on phase locking as described in Section 2.3. On the one hand this seems compatible with the strong preference to descending fifths and ascending fourths over descending fourths and ascending fifths. On the other hand, if  $c_1$  and  $c_2$  only consist of one note each, a low periodicity of  $c_1$  with respect to  $c_2$  is desirable, because the neuronal firing is synchronized. This seems to be incompatible with voice leading at first, but as soon as you consider small chord movements with respect to voice leading, it is possible to move a short distance while being close with respect to phase synchronization. Therefore, we introduce a transitive periodicity analogously to the periodicity definition given in Equation (3).

The transitive periodicity from  $c_1$  to  $c_2$  is the number of periods of  $c_2$  necessary to match up with a period multiple of  $c_1$ , where the periods for  $c_1$  and  $c_2$  are each due to phase locking. Formally, transitive periodicity from  $c_1$  to  $c_2$  is the periodicity of  $[c_1, c_2]$  relative to  $c_2$ , where  $[c_1, c_2]$  is the (set-theoretic) union of  $c_1$  and  $c_2$ : Figure 10 shows  $c_1 = [0, 4, 7, 10]$ ,  $c_2 = [0, 5, 9]$  and the combined chord  $[c_1, c_2] = [0, 4, 5, 7, 9, 10]$ .

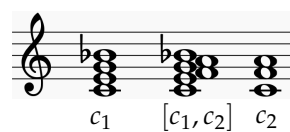


Figure 10. A chord progression  $c_1 \rightarrow c_2$  with the combined chord  $[c_2, c_1]$ .

This corresponds to computing  $p([c_1, c_2]) / p(c_2)$ , but due to the JND from Section 3 the smoothed periodicities are not the correct quantities to be used for computing transitive periodicity. We need to view  $c_1, c_2$  and  $[c_1, c_2]$  in the context of their related periodicities before smoothing. Due to technical difficulties we need to work with  $\mathbf{R}^n$  rather than its quotient  $\mathcal{C}^n$ . In analogy to Section 4.5 we define

$$\mathcal{C}_{m,n}^p := \left\{ \left( 0, 12 \cdot \log_2 \left( \frac{p_2}{q_2} \right), \dots, 12 \cdot \log_2 \left( \frac{p_{m+n}}{q_{m+n}} \right) \right) \left| \begin{array}{l} \forall_i \left( 1 \leq \frac{p_i}{q_i} \leq 2 \wedge \gcd(p_i, q_i) = 1 \right) \\ \wedge \frac{\text{lcm}(q_1, \dots, q_{m+n})}{\text{lcm}(q_2, \dots, q_n)} = p \end{array} \right. \right\} \quad (5)$$

and for pitch tuples  $t \in \mathbf{R}^n$

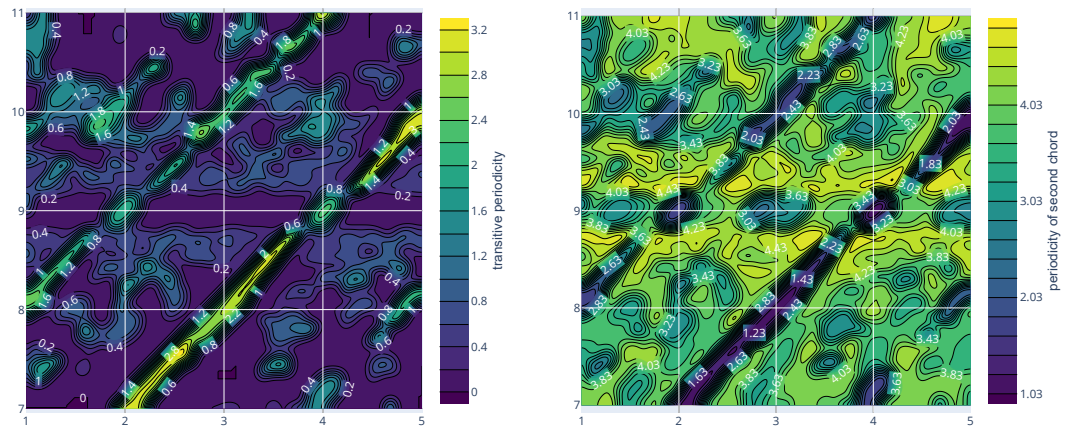
$$N_t := \{ t + (d_1, \dots, d_n) \mid \forall_{i,j=1,\dots,n} (d_i \in [-\text{JND}, \text{JND}] \wedge |d_i - d_j| \leq \text{JND}) \}. \quad (6)$$

Informally,  $\mathcal{C}_{m,n}^p$  contains the  $(m+n)$ -tuples  $(t_1, t_2)$  with representatives  $t_1 \in \mathbf{R}^m$  and  $t_2 \in \mathbf{R}^n$  of  $c_1 \in \mathcal{C}_m$  and  $c_2 \in \mathcal{C}_n^{[0,12]}$  so that the periodicity of  $[c_1, c_2]$  relative to  $c_2$  is  $p$ . In order to improve readability the first  $n$  entries in the elements of  $\mathcal{C}_{m,n}^p$  correspond to  $c_2$ . In order to incorporate approximations within a JND we define transitional periodicity  $p : \mathcal{C}_m \times \mathcal{C}_n^{[0,12]} \rightarrow \mathbf{R}, (c_1, c_2) \mapsto p(c_1 \rightarrow c_2)$  via

$$p(c_1 \rightarrow c_2) := \min \left\{ p \left| \begin{array}{l} N_t \cap \mathcal{C}_{m,n}^p \neq \emptyset, \text{ where } t \in \mathbf{R}^{m+n}, \\ c_1 = [t_1, \dots, t_m], c_2 = [t_{m+1}, \dots, t_{m+n}] \end{array} \right. \right\}.$$

Just like in the case of periodicity in Section 4.5 we can use the logarithmic transitive periodicity. In order to extend  $p$  to  $\mathcal{C}_m \times \mathcal{C}_n$ , it will be necessary to shift a chord  $c_1$  and  $c_2$  so that the lowest note of  $c_2$  is 0. For  $c = [p_1, \dots, p_n]$  let  $s_p(c) := [p_1 - p, p_2 - p, \dots, p_n - p]$  and  $\min(c) := \min\{p_1, \dots, p_n\}$ . Then,  $p(c_1 \rightarrow c_2)$  will be redefined as  $p(s_{\min(c_2)}(c_1), s_{\min(c_2)}(c_2))$ .

Let us revisit the example in Section 5.2. The local neighborhood of the transitional periodicity  $c_1 \rightarrow c_2$  starting with the tritone  $c_1 = [3, 9]$  with the corresponding periodicities of  $c_2$  is shown in Figure 11.



**Figure 11.** Contour plot for logarithmic transitive periodicities  $p(c_1 \rightarrow c_2)$  (left) with the corresponding logarithmic periodicities  $p_{c_1}(c_2)$  of  $c_2 \in \mathcal{C}_2$  (right) starting with the tritone  $c_1 = [3, 9]$  using JND = 18 cent smoothed by a Gaussian with standard deviation  $\sigma = 6$  cent

Notice that while the periodicities of the perfect fourth and fifth are smallest, the transitive periodicities resolving to the perfect fourth and fifth are bigger than the transitive periodicities resolving to other chords. Even if small transitive periodicities play a role in chord resolutions, they do not fully explain them. The algorithm for determining the transitive periodicity step function as in Equation (3) is shown in Algorithm 2, where we define for  $c = [c_1, \dots, c_m] \in \mathcal{C}_m$

$$N_c(\epsilon, n) := \{(t_1 + d_1, \dots, t_n + d_n) \mid t \in \mathbf{R}^n, [t] = c, \forall_{i=1, \dots, n} (d_i \in [-\epsilon, \epsilon])\} \subset \mathcal{C}_n,$$

since we want  $c_1$  and  $c_2$  to be close with respect to voice leading. In the algorithm we need the projection to the last  $n$  coordinates  $l_n(t_1, \dots, t_{m+n}) := (t_{m+1}, \dots, t_{m+n})$ .

We hypothesize that transitive periodicity will play a role in combination with the usual periodicity. A good chord resolution will have a low periodicity for the second chord  $c_2$  as well as a low transitive periodicity  $c_1 \rightarrow c_2$ .

The combined chord offers more possibilities for transitive quantities that can be studied in the context of music perception. For example, we can consider the periodicity  $p_{c_1}([c_1, c_2])$  of  $[c_1, c_2]$  relative to  $c_1$  for the chord progression  $c_1 \rightarrow c_2$ .

---

**Algorithm 2** Determine transitive periodicity step function  $p : \mathcal{C} \times \mathcal{C} \rightarrow \mathbf{R}$

---

**Require:**  $m, n \geq 1$  ▷  $m, n$ =number of tones in  $c_1, c_2 \in \mathcal{C}$   
**Require:**  $c_1 \in \mathcal{C}_m$  ▷  $c_1$ =chord with  $m$  tones  
**Require:**  $\epsilon \geq 0$  ▷  $\epsilon$ =maximal distance between the tones of  $c_1 \in \mathcal{C}_m$  and  $c_2 \in \mathcal{C}_n$   
 $q \leftarrow 1$  ▷  $q$ =periodicity index  
 $RemainingChords \leftarrow N_{c_1}(\epsilon, n)$  ▷ Consider all chords with  $n$  notes close enough to  $c_1$   
**while**  $RemainingChords \neq \emptyset$  ▷ While there are chords without periodicity  
  **for all**  $s \in \mathcal{C}_{m,n}^q$  ▷ For all chords with relative periodicity  $q$   
    **for all**  $t \in N_{I_n(s)} \cap RemainingChords$  ▷ For all new chords within JND  
       $p(c_1, [t]) \leftarrow q$  ▷ Set periodicity to  $q$   
       $RemainingChords \leftarrow (RemainingChords \setminus N_{I_n(s)})$  ▷ Update new chords  
    **end for**  
  **end for**  
   $q \leftarrow q + 1$  ▷ Increase periodicity index by 1  
**end while**

---

#### 5.4. Directional Derivative of Periodicity

Melodies and chord progressions not only have a sense of direction in  $\mathcal{C}$ , but the rate of change in psychoacoustic quantities with respect to these directions should play a role in music perception. Assuming that periodicity is a good measure for consonance, chord resolutions should decrease periodicity while traveling only a small distance with respect to voice leading. If this perceptive quality is truly local then the infinitesimal change in periodicity can be formulated using directional derivatives in chord space.

Due to the soft computing skills of our brain, periodicity on  $\mathcal{C}$  can be considered to be continuously differentiable. Let  $v$  be a tangent vector of  $T_c\mathcal{C}$  at a chord  $c \in \mathcal{C}$ . Let  $\rho : [-\epsilon, \epsilon] \rightarrow \mathcal{C}$  be a continuously differentiable path with  $\rho|_{t=0} = v$ . Then, the directional derivative of the periodicity in the direction  $v \in T_c\mathcal{C}$  at  $c \in \mathcal{C}$  is given by

$$D_v p := \left. \frac{d}{dt} \right|_{t=0} (p \circ \rho).$$

The more negative  $D_v p$  is, the stronger its resolution in the direction  $v$  is perceived. It is not enough that the final chord is more consonant. For example, either resolution from a tritone to a perfect fifth in 12TET does not sound as good as the one to the minor sixth or the major third as we have discussed in Section 5.2.

On the other hand, the parallel movement of chords does not change periodicity. This implies that  $D_v p$  vanishes for  $v = [1, 1]$ . Clearly, maximal infinitesimal change in periodicity (either negative or positive) is provided for  $v = [-1, 1]$  and  $v = [1, -1]$ . If Figure 8 is correct, however, then only a very small repelling movement will reduce periodicity which does not yield harmonic relationships of chords. The progressions  $[3, 9] \rightarrow [2.75, 9.25]$  and  $[3, 9] \rightarrow [3.25, 8.75]$  will increase periodicity. The progressions to the perfect fourth and fifth  $[3, 9] \rightarrow [2.5, 9.5]$  and  $[3, 9] \rightarrow [3.5, 8.5]$  will certainly decrease periodicity. Interestingly enough, the only feasible progressions  $[3, 9] \rightarrow [2, 10]$  and  $[3, 9] \rightarrow [4, 8]$  in 12TET do not reduce periodicity by much or at all. Still, they are the best resolutions available in 12TET.

While the directional derivative presents an interesting approach it needs to be viewed in combination with other aspects of transitional harmony as part of a Pareto optimal solution. Possibly, the periodicity function shown in Figure 8 needs to be corrected as well. However, it shows that, in the case of the tritone, the chord progressions with the maximal effect on periodicity will move each note simultaneously inwards or outwards. Furthermore, it suggests that quarter tone movements will be the best chord resolution when considering periodicity only. This hypothesis is confirmed by the author’s perception.

## 6. Results and Discussion

Music is considered as something real and vital for humankind, but no attempt on a holistic model for music perception has yet been attempted. Clearly, humans do not

perceive musical sounds as the complicated audio waves they are or as the way they are presented in music notation but as something simple and often beautiful. While music theory formalizes the music we perceive, music psychology carries out empirical studies about specific perceptive aspects. We envision that it should also be possible to deduce music theory from music psychology with the correct holistic model for music perception. Using psychoacoustic results and facts from music theory it should be possible to reverse engineer this model. With this in mind, we have introduced mathematical structures that allow for rigorous quantitative studies of music perception based on the mechanics described by physical or neuronal models. We laid an emphasis on a rigorous approach that is not more complicated than absolutely necessary and which can be extended when needed.

We revisit ideas by Tymoczko [13,14] to prove that the space of chords  $\mathcal{C}$  is a metric space and a Whitney stratified space with a Riemannian structure. The geometry of  $\mathcal{C}$  is not much more than Euclidean space itself. However, it allows us to apply calculus across different strata of  $\mathcal{C}$ . Furthermore, the Riemannian metric on  $\mathcal{C}$  allows us to consider the geodesic distance across different strata which yields a voice leading distance satisfying the triangle inequality. The geodesic approach is surprisingly simple and natural considering the common desire that distance functions satisfy certain conditions and in view of more elaborate attempts regarding voice leading distances [127,139,140]. The space  $\mathcal{C}$  only contains the objects for music production, but not any information about music perception.

Psychoacoustic quantities can be viewed, computed and analyzed as height functions on  $\mathcal{C}$ . In particular, we have modified the periodicity approach to consonance by Stolzenburg [6] in order to present a definition of periodicity for arbitrary chords. Roughness is another way of interpreting dissonance. Height functions themselves are static. Music is a dynamic process, so it might be necessary to consider the change in height function as a dynamical system in order to deduce properties of music. All of the psychoacoustic functions can be assumed to be differentiable which enables us to use tools from differential geometry to study them by considering gradient vectors and directional derivatives.

The height function for periodicity led to two possible approaches for transitive harmonicity. In particular, we showed how to use the differential structure of the periodicity graph on  $\mathcal{C}$  to study geometric properties of paths in  $\mathcal{C}$  and their respective lifts to the graph of psychoacoustic functions on  $\mathcal{C}$ . We implicitly assume that the geodesic distance agrees with the psychoacoustic reality. This needs to be verified empirically. Although we do not expect our two approaches for transitive harmonicity to be valid, we expect that other approaches to music perception can be analyzed using the differential geometric framework. Clearly, music works, because we look at a discrete subset of chords with certain properties. It will be interesting to see which tools are the correct ones for discretizing the differential-geometric model.

The differential-geometric structure invites studies that falsify or confirm psychoacoustic models for music. Ultimately, this approach can close the gap between music theory and music psychology. Even though the mechanisms discussed here stem from Western music, they are founded on more general physical and neuronal principles, which are in theory applicable to music from other cultures or sounds with inharmonic spectra. Furthermore, it will be interesting to study, generalize and extend the mathematical structures themselves and to incorporate statistical aspects of music perception in the model.

## 7. Conclusions

For the purpose of analyzing music perception, we have described useful geometric structures for the space of chords  $\mathcal{C}$ . We have rigorously proven properties that are desirable from a mathematical as well as from a music perception point of view. In particular, chords with a different number of notes can be viewed as strata of  $\mathcal{C}$ . The Riemannian metric on each stratum allows to define a geodesic distance on  $\mathcal{C}$ , which makes it into a metric space. The metric is a natural choice for determining efficient voice leading. The Riemannian metric also allows to study shapes in the context of music perception. This enables music

psychologists and music theorists to use tools from differential geometry in order to study music perception.

**Funding:** The article processing charge was funded by the Baden-Württemberg Ministry of Science, Research and Arts and by Reutlingen University in the funding programme Open Access Publishing.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** I am grateful to Dmitri Tymoczko, Matthias Kreck, James R. Hughes and Peter M. C. Harrison for helpful suggestions and discussions.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Collier, J. Youtube, in the Bleak Midwinter. Available online: <https://youtu.be/mPZn4x3uOac> (accessed on 21 July 2022).
2. Nempla Música. Youtube, Jacob Collier Masterclass en NEMPLA—Parte 3 de 5, 2018. Available online: <https://youtu.be/5vrhKl7JHQc?t=946> (accessed on 21 July 2022).
3. Wired.com Support. Jacob Collier Answers Music Theory Questions From Twitter. Available online: <https://www.wired.com/video/watch/tech-support-jacob-collier-answers-music-theory-questions-from-twitter> (accessed on 21 July, 2022).
4. Krumhansl, C.L. Music Psychology and Music Theory: Problems and Prospects. *Music Theory Spectr.* **1995**, *17*, 53–80. [[CrossRef](#)] [[CrossRef](#)]
5. Krumhansl, C.L. *Cognitive Foundations of Musical Pitch*; Oxford Psychology Series; Oxford University Press: New York, NY, USA, 2001.
6. Stolzenburg, F. Harmony perception by periodicity detection. *J. Math. Music* **2015**, *9*, 215–238. [[CrossRef](#)] [[CrossRef](#)]
7. Harrison, P.M.C.; Pearce, M.T. Simultaneous consonance in music perception and composition. *Psychol. Rev.* **2020**, *127*, 216–244. [[CrossRef](#)] [[CrossRef](#)] [[PubMed](#)]
8. Harrison, P.M.C.; Pearce, M.T. A Computational Cognitive Model for the Analysis and Generation of Voice Leadings. *Music Percept. Interdiscip. J.* **2020**, *37*, 208–224. [[CrossRef](#)] [[CrossRef](#)]
9. Marjeh, R.; Harrison, P.M.C.; Lee, H.; Deligiannaki, F.; Jacoby, N. Reshaping musical consonance with timbral manipulations and massive online experiments. *bioRxiv* **2022**. [[CrossRef](#)]
10. Harrison, P.M.C. Three Questions Concerning Consonance Perception. *Music Percept.* **2021**, *38*, 337–339. [[CrossRef](#)] [[CrossRef](#)]
11. Chan, P.Y.; Dong, M.; Li, H. The Science of Harmony: A Psychophysical Basis for Perceptual Tensions and Resolutions in Music. *Research* **2019**, *2019*, 2369041. [[CrossRef](#)] [[CrossRef](#)]
12. Collins, T.; Tillmann, B.; Barrett, F.S.; Delbé, C.; Janata, P. A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychol. Rev.* **2014**, *121*, 33–65. [[CrossRef](#)] [[CrossRef](#)]
13. Tymoczko, D. The Geometry of Musical Chords. *Science* **2006**, *313*, 72–74. [[CrossRef](#)] [[CrossRef](#)]
14. Tymoczko, D. *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*; Oxford Studies in Music Theory; Oxford University Press: New York, NY, USA, 2011.
15. Tymoczko, D. Three Conceptions of Musical Distance. In *Mathematics and Computation in Music*; Chew, E., Childs, A., Chuan, C.-H., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 258–273. [[CrossRef](#)]
16. Milne, A.J.; Sethares, W.A.; Laney, R.; Sharp, D.B. Modelling the similarity of pitch collections with expectation tensors. *J. Math. Music* **2011**, *5*, 1–20. [[CrossRef](#)] [[CrossRef](#)]
17. beim Graben, P.; Blutner, R. Toward a Gauge Theory of Musical Forces. In *Quantum Interaction*; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 99–111. [[CrossRef](#)]
18. Blutner, R.; beim Graben, P. Gauge models of musical forces. *J. Math. Music* **2020**, *15*, 17–36. [[CrossRef](#)]
19. Large, E.W. A Dynamical Systems Approach to Musical Tonality. In *Nonlinear Dynamics in Human Behavior*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 193–211. [[CrossRef](#)]
20. Burrows, D. A Dynamical Systems Perspective on Music. *J. Musicol.* **1997**, *15*, 529–545. [[CrossRef](#)]
21. Gazor, M.; Shoghi, A. Bifurcation control and sound intensities in musical art. *J. Differ. Equ.* **2021**, *293*, 86–110. [[CrossRef](#)]
22. Gazor, M.; Shoghi, A. Tone colour in music and bifurcation control. *J. Differ. Equ.* **2022**, *326*, 129–163. [[CrossRef](#)]
23. del Pozo, I.; Gómez-Martín, F. A Mathematical Model of Tonal Function (I): Voice Leadings. In *Mathematics and Computation in Music*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 218–230. [[CrossRef](#)]
24. del Pozo, I.; Gómez-Martín, F. A Mathematical Model of Tonal Function (II): Modulation. In *Mathematics and Computation in Music*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022; pp. 231–239. [[CrossRef](#)]
25. Wall, L.; Lieck, R.; Neuwirth, M.; Rohrmeier, M. The Impact of Voice Leading and Harmony on Musical Expectancy. *Sci. Rep.* **2020**, *10*, 5933. [[CrossRef](#)]
26. Wilkerson, D.S. Harmony Explained: Progress Towards A Scientific Theory of Music. *arXiv* **2014**, arXiv:1202.4212. [[CrossRef](#)]

27. Bailes, F.; Dean, R.T.; Broughton, M.C. How Different Are Our Perceptions of Equal-Tempered and Microtonal Intervals? A Behavioural and EEG Survey. *PLoS ONE* **2015**, *10*, e0135082. [CrossRef]
28. Bridges, B. Can Harmony be Non-Linear? A response to some of Glenn Branca's '25 Questions'. In Proceedings of the Society for Musicology in Ireland Annual Conference, Waterford, Ireland, 9–11 May 2008.
29. Leino, S.; Brattico, E.; Tervaniemi, M.; Vuust, P. Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Res.* **2007**, *1142*, 169–177. [CrossRef]
30. Zhang, J.; Zhou, X.; Chang, R.; Yang, Y. Effects of global and local contexts on chord processing: An ERP study. *Neuropsychologia* **2018**, *109*, 149–154. [CrossRef]
31. Pagès-Portabella, C.; Toro, J.M. Dissonant endings of chord progressions elicit a larger ERAN than ambiguous endings in musicians. *Psychophysiology* **2019**, *57*, e13476. [CrossRef]
32. Vuust, P.; Heggli, O.A.; Friston, K.J.; Kringelbach, M.L. Music in the brain. *Nat. Rev. Neurosci.* **2022**, *23*, 287–305. [CrossRef] [PubMed]
33. Sauv e, S.A.; Cho, A.; Zendel, B.R. Mapping Tonal Hierarchy in the Brain. *Neuroscience* **2021**, *465*, 187–202. [CrossRef] [PubMed]
34. Feng, J.Q. Music in Terms of Science. *arXiv* **2012**, arXiv:1209.3767. [CrossRef]
35. Sethares, W. *Tuning, Timbre, Spectrum, Scale*; Springer: London, UK, 2005.
36. Tramo, M.J.; Cariani, P.A.; Delgutte, B.; Braid a, L.D. Neurobiological foundations for the theory of harmony in western tonal music. *Ann. N. Y. Acad. Sci.* **2001**, *930*, 92–116. [CrossRef]
37. Dumas, R. Melodies in Space: Neural Processing of Musical Features. Ph.D. Thesis, University of Minnesota, Minneapolis, MN, USA, 2013. Available online: <https://hdl.handle.net/11299/155664> (accessed on 21 July 2022).
38. Rickles, D. Spaces. In *The Oxford Handbook of Philosophy of Science*; Humphreys, P., Ed.; Oxford University Press: Oxford, UK, 2016.
39. Randel, D.M. *The Harvard Concise Dictionary of Music and Musicians*; Harvard University Press: Cambridge, Massachusetts, 1999.
40. Varshney, L.R.; Sun, J.Z. Why do we perceive logarithmically? *Significance* **2013**, *10*, 28–31. [CrossRef]
41. Mattson, M.P. Superior pattern processing is the essence of the evolved human brain. *Front. Neurosci.* **2014**, *8*, 265. [CrossRef]
42. Abdi, H. Signal Detection Theory. In *International Encyclopedia of Education*; Elsevier: Amsterdam, The Netherlands, 2010; pp. 407–410. [CrossRef]
43. Wichmann, F.A.; Hill, N.J. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* **2001**, *63*, 1293–1313. [CrossRef]
44. Wichmann, F.A.; Hill, N.J. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept. Psychophys.* **2001**, *63*, 1314–1329. [CrossRef]
45. Macmillan, N.A.; Creelman, C.D. *Detection Theory: A User's Guide*; Lawrence Erlbaum Associates: Mahwah, NJ, USA, 2004. [CrossRef]
46. Lodish, H. *Molecular Cell Biology*; W.H. Freeman: New York, NY, USA, 2000.
47. Johnson, K.O. Neural Coding. *Neuron* **2000**, *26*, 563–566. [CrossRef]
48. Partridge, L.D.; Partridge, L.D. From Reception to Pattern Recognition and Perception. In *Nervous System Actions and Interactions*; Springer: New York, NY, USA, 2003; pp. 145–174. [CrossRef]
49. Fechner, G. *Elemente der Psychophysik*; 2. Teil; Breitkopf und H artel: Leipzig, Deutschland, 1889.
50. Langner, G. Temporal processing of pitch in the auditory system. *J. New Music Res.* **1997**, *26*, 116–132. [CrossRef]
51. Langner, G.; Benson, C. *The Neural Code of Pitch and Harmony*; Cambridge University Press: Cambridge, UK, 2015. [CrossRef]
52. Sinz, F.H.; Sachgau, C.; Henninger, J.; Benda, J.; Grewe, J. Simultaneous spike-time locking to multiple frequencies. *J. Neurophysiol.* **2020**, *123*, 2355–2372. [CrossRef]
53. Jordan, B. *Advancing Ethnography in Corporate Environments: Challenges and Emerging Opportunities*; Routledge: New York, NY, USA, 2013.
54. Fell, J.; Axmacher, N. The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* **2011**, *12*, 105–118. [CrossRef]
55. Lowet, E.; Roberts, M.J.; Bonizzi, P.; Karel, J.; Weerd, P.D. Quantifying Neural Oscillatory Synchronization: A Comparison between Spectral Coherence and Phase-Locking Value Approaches. *PLoS ONE* **2016**, *11*, e0146443. [CrossRef]
56. Kac, M. Can One Hear the Shape of a Drum? *Am. Math. Mon.* **1966**, *73*, 1. [CrossRef]
57. Cariani, P.A.; Delgutte, B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* **1996**, *76*, 1698–1716. [CrossRef]
58. Shepard, R.N. Circularity in judgments of relative pitch. *J. Acoust. Soc. Am.* **1964**, *36*, 2345–2353. [CrossRef]
59. Sethares, W.A. Local consonance and the relationship between timbre and scale. *J. Acoust. Soc. Am.* **1993**, *94*, 1218–1228. [CrossRef]
60. Hinrichsen, H. Entropy-based tuning of musical instruments. *Rev. Bras. Ensino F is.* **2012**, *34*, 1–8. [CrossRef]
61. Cohen, E.A. Some Effects of Inharmonic Partial on Interval Perception. *Music Percept.* **1984**, *1*, 323–349. [CrossRef]
62. Bausenhardt, K.M.; Luca, M.D.; Ulrich, R. Assessing Duration Discrimination: Psychophysical Methods and Psychometric Function Analysis. In *Timing and Time Perception: Procedures, Measures, & Applications*; BRILL: Leiden, The Netherlands, 2018; pp. 52–78. [CrossRef]
63. Gilchrist, J.M.; Jerwood, D.; Ismaiel, H.S. Comparing and unifying slope estimates across psychometric function models. *Percept. Psychophys.* **2005**, *67*, 1289–1303. [CrossRef]
64. Hugo Fastl, E.Z. *Psychoacoustics*; Springer: Berlin/Heidelberg, Germany, 2006.

65. Rossing, T. *The Science of Sound*; Addison Wesley: San Francisco, CA, USA, 2002.
66. Smith, J.; Abel, J. Bark and ERB bilinear transforms. *IEEE Trans. Speech Audio Process.* **1999**, *7*, 697–708. [[CrossRef](#)]
67. Zwicker, E.; Flottorp, G.; Stevens, S.S. Critical Band Width in Loudness Summation. *J. Acoust. Soc. Am.* **1957**, *29*, 548–557. [[CrossRef](#)]
68. Moore, B.C.J.; Glasberg, B.R.; Shailer, M.J. Frequency and intensity difference limens for harmonics within complex tones. *J. Acoust. Soc. Am.* **1984**, *75*, 550–561. [[CrossRef](#)] [[PubMed](#)]
69. Bubic, A.; von Cramon, D.Y.; Jacobsen, T.; Schröger, E.; Schubotz, R.I. Violation of Expectation: Neural Correlates Reflect Bases of Prediction. *J. Cogn. Neurosci.* **2009**, *21*, 155–168. [[CrossRef](#)] [[PubMed](#)]
70. Hartmann, W.M.; Rakerd, B.; Packard, T.N. On measuring the frequency-difference limen for short tones. *Percept. Psychophys.* **1985**, *38*, 199–207. [[CrossRef](#)] [[PubMed](#)]
71. Gill, K.Z.; Purves, D. A Biological Rationale for Musical Scales. *PLoS ONE* **2009**, *4*, e8144. [[CrossRef](#)]
72. Becker, J. *Traditional Music in Modern Java*; University of Hawaii Press: Honolulu, HI, USA, 2019. [[CrossRef](#)]
73. Boulos, I. Inside Arabic Music: Arabic Maqam Performance and Theory in the 20th Century. By Johnny Farraj and Sami Abu Shumays. *Music Lett.* **2021**, *102*, 171–172. [[CrossRef](#)]
74. Marcus, S. The Interface between Theory and Practice: Intonation in Arab Music. *Asian Music* **1993**, *24*, 39. [[CrossRef](#)]
75. Akkoc, C. Non-Deterministic Scales Used in Traditional Turkish Music. *J. New Music Res.* **2002**, *31*, 285–293. [[CrossRef](#)]
76. Valla, J.M.; Alappatt, J.A.; Mathur, A.; Singh, N.C. Music and Emotion—A Case for North Indian Classical Music. *Front. Psychol.* **2017**, *8*, 2115. [[CrossRef](#)]
77. Balkwill, L.L.; Thompson, W.F. A Cross-Cultural Investigation of the Perception of Emotion in Music: Psychophysical and Cultural Cues. *Music Percept. Interdiscip. J.* **1999**, *17*, 43–64. [[CrossRef](#)]
78. Demorest, S.M.; Morrison, S.J.; Stambaugh, L.A.; Beken, M.; Richards, T.L.; Johnson, C. An fMRI investigation of the cultural specificity of music memory. *Soc. Cogn. Affect. Neurosci.* **2009**, *5*, 282–291. [[CrossRef](#)]
79. Owen, H. *Music Theory Resource Book*; Oxford University Press: Oxford, UK; New York, NY, USA, 2000.
80. Burton, R.L. The Elements of Music: What Are They, and Who Cares? In *ASME XXth National Conference Proceedings*; Rosevear, J.; Harding, S., Eds.; The Australian Society for Music Education Inc: Parkville, VIC, Australia, 2015.
81. Wright, D. *Mathematics and Music*; American Mathematical Society: Providence, RI, USA, 2009.
82. Schönberg, A.; Black, L.; Stein, L. *Style and idea: Selected writings of Arnold Schönberg*; University of California Press: Berkeley, CA, USA, 2000.
83. Thurston, W.P. *Three-Dimensional Geometry and Topology*; Princeton University Press: Princeton, NJ, USA, 1997.
84. Hughes, J.R. Generalizing the Orbifold Model for Voice Leading. *Mathematics* **2022**, *10*, 939. [[CrossRef](#)]
85. Borzellino, J.E. Riemannian Geometry of Orbifolds. Ph.D. Thesis, University of California, Los Angeles, CA, USA, 1992.
86. Pflaum, M. *Analytic and Geometric Study of Stratified Spaces: Contributions to Analytic and Geometric Aspects*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2003.
87. Railsback, O.L. Scale Temperament as Applied to Piano Tuning. *J. Acoust. Soc. Am.* **1938**, *9*, 274–274. [[CrossRef](#)]
88. Ratcliffe, J. *Foundations of Hyperbolic Manifolds*; Springer: New York, NY, USA, 2007. [[CrossRef](#)]
89. Lange, C. Orbifolds from a metric viewpoint. *Geom. Dedicata* **2020**, *209*, 43–57. [[CrossRef](#)]
90. Bettiol, R.G.; Derdzinski, A.; Piccione, P. Teichmüller theory and collapse of flat manifolds. *Ann. Mat. Pura Appl. (1923)* **2018**, *197*, 1247–1268. [[CrossRef](#)]
91. Alekseevsky, D.; Kriegl, A.; Losik, M.; Michor, P.W. The Riemannian geometry of orbit spaces. The metric, geodesics, and integrable systems. *Publ. Math. Debr.* **2003**, *6*, 247–276. [[CrossRef](#)]
92. Michor, P.W. *Topics in Differential Geometry*; American Mathematical Society: Providence, RI, USA, 2008.
93. Huckemann, S.; Hotzand, T.; Munk, A. Intrinsic shape analysis: Geodesic PCA for Riemannian manifolds modulo isometric Lie group actions. *Stat. Sin.* **2010**, *20*, 1–58.
94. Thanwerdas, Y. Riemannian and Stratified Geometries on Covariance and Correlation Matrices. Ph.D. Thesis, Université Côte d’Azur, Nice, France, 2022.
95. Stefano, N.D.; Bertolaso, M. Understanding Musical Consonance and Dissonance: Epistemological Considerations from a Systemic Perspective. *Systems* **2014**, *2*, 566–575. [[CrossRef](#)]
96. Helmholtz, H. *On the Sensations of Tone*; Dover Publications: Mineola, NY, USA, 1954.
97. Stumpf, C. *Tonpsychologie 1*; Cambridge University Press: Cambridge, UK, 2013.
98. Stumpf, C. *Tonpsychologie 2*; Cambridge University Press: Cambridge, UK, 2013.
99. Parncutt, R.; Hair, G. Consonance and dissonance in music theory and psychology: Disentangling dissonant dichotomies. *J. Interdiscip. Music. Stud.* **2012**, *5*, 119–166. [[CrossRef](#)]
100. Lahdelma, I.; Eerola, T. Cultural familiarity and musical expertise impact the pleasantness of consonance/dissonance but not its perceived tension. *Sci. Rep.* **2020**, *10*, 8693. [[CrossRef](#)] [[PubMed](#)]
101. Armitage, J.; Lahdelma, I.; Eerola, T. Automatic responses to musical intervals: Contrasts in acoustic roughness predict affective priming in Western listeners. *J. Acoust. Soc. Am.* **2021**, *150*, 551–560. [[CrossRef](#)] [[PubMed](#)]
102. Cook, N.D.; Fujisawa, T.X. The Psychophysics of Harmony Perception: Harmony is a Three-Tone Phenomenon. *Empir. Musicol. Rev.* **2006**, *1*, 106–126. [[CrossRef](#)]



103. Bidelman, G.M.; Krishnan, A. Neural Correlates of Consonance, Dissonance, and the Hierarchy of Musical Pitch in the Human Brainstem. *J. Neurosci.* **2009**, *29*, 13165–13171. [[CrossRef](#)] [[PubMed](#)]
104. Leman, M. Visualization and calculation of the roughness of acoustic music signals using the Synchronization Index Model. In Proceedings of the COSTG-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, 7–9 December 2000.
105. endolith. Github repository, Sethares dismeasure in Python. Available online: <https://gist.github.com/endolith/3066664> (accessed on 4 October 2021)
106. Vencovský, V. Roughness Prediction Based on a Model of Cochlear Hydrodynamics. *Arch. Acoust.* **2016**, *41*, 189–201. [[CrossRef](#)]
107. Hall, D.E.; Hess, J.T. Perception of Musical Interval Tuning. *Music Percept.* **1984**, *2*, 166–195. [[CrossRef](#)]
108. Roederer, J.G. *The Physics and Psychophysics of Music*; Springer: Berlin/Heidelberg, Germany, 2008.
109. Vos, J. Purity Ratings of Tempered Fifths and Major Thirds. *Music Percept.* **1986**, *3*, 221–257. [[CrossRef](#)]
110. Kopiez, R. Intonation of Harmonic Intervals: Adaptability of Expert Musicians to Equal Temperament and Just Intonation. *Music Percept.* **2003**, *20*, 383–410. [[CrossRef](#)]
111. Moore, B.C.J.; Peters, R.W.; Glasberg, B.R. Thresholds for the detection of inharmonicity in complex tones. *J. Acoust. Soc. Am.* **1985**, *77*, 1861–1867. [[CrossRef](#)]
112. Moore, B.C.J.; Glasberg, B.R.; Peters, R.W. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *J. Acoust. Soc. Am.* **1986**, *80*, 479–483. [[CrossRef](#)]
113. Hartmann, W.M. *Signals, Sound, and Sensation*; American Institute of Physics: Melville, NY, USA, 2004.
114. Huron, D. *Sweet Anticipation*; The MIT Press: Cambridge, MA, USA, 2006. [[CrossRef](#)]
115. Pearce, M.T.; Wiggins, G.A. Auditory Expectation: The Information Dynamics of Music Perception and Cognition. *Top. Cogn. Sci.* **2012**, *4*, 625–652. [[PubMed](#)]
116. Rohrmeier, M. Musical Expectancy. Bridging Music Theory, Cognitive and Computational Approaches. *Z. Ges. Musik. [J.-Ger.-Speak. Soc. Music Theory]* **2013**, *10*, 343–371. [[CrossRef](#)]
117. Schmuckler, M.A. Expectation in Music: Investigation of Melodic and Harmonic Processes. *Music Percept.* **1989**, *7*, 109–149. [[CrossRef](#)]
118. Seger, C.A.; Spiering, B.J.; Sares, A.G.; Quraini, S.I.; Alpeter, C.; David, J.; Thaut, M.H. Corticostriatal Contributions to Musical Expectancy Perception. *J. Cogn. Neurosci.* **2013**, *25*, 1062–1077. [[CrossRef](#)]
119. Bigand, E.; Parncutt, R.; Lerdahl, F. Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Percept. Psychophys.* **1996**, *58*, 125–141. [[CrossRef](#)]
120. TRAMO, M.J. Neurophysiology and Neuroanatomy of Pitch Perception: Auditory Cortex. *Ann. N. Y. Acad. Sci.* **2005**, *1060*, 148–174. [[CrossRef](#)]
121. Lahdelma, I.; Eerola, T. Theoretical Proposals on How Vertical Harmony May Convey Nostalgia and Longing in Music. *Empir. Musicol. Rev.* **2015**, *10*, 245. [[CrossRef](#)]
122. Leonhard, E. *Tentamen Novae Theoriae Musicae*; Typographia Academiae Scientiarum: Saint Petersburg, Russia, 1739. [[CrossRef](#)]
123. Euler, L. De harmoniae veris principiis per speculum musicum repraesentatis. *Novi Comment. Acad. Sci. Petropolitanae* **1774**, *18*, 330–353.
124. Lerdahl, F. *Tonal Pitch Space*; Oxford University Press: New York, NY, USA, 2001.
125. Bharucha, J.; Krumhansl, C.L. The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition* **1983**, *13*, 63–102. [[CrossRef](#)]
126. Cohn, R. Neo-Riemannian Operations, Parsimonious Trichords, and their ‘Tonnetz’ Representations. *J. Music Theory* **1997**, *42*, 1–66. [[CrossRef](#)]
127. Callender, C.; Quinn, I.; Tymoczko, D. Generalized Voice-Leading Spaces. *Science* **2008**, *320*, 346–348. [[CrossRef](#)] [[PubMed](#)]
128. Tymoczko, D. The Generalized Tonnetz. *J. Music Theory* **2012**, *56*, 1–52. [[CrossRef](#)]
129. Schwitzgebel, E.; White, C.W. Effects of Chord Inversion and Bass Patterns on Harmonic Expectancy in Musicians. *Music Percept.* **2021**, *39*, 41–62. [[CrossRef](#)]
130. Hove, M.J.; Marie, C.; Bruce, I.C.; Trainor, L.J. Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10383–10388. [[CrossRef](#)] [[PubMed](#)]
131. Bigand, E.; Parncutt, R. Perceiving musical tension in long chord sequences. *Psychol. Res.* **1999**, *62*, 237–254. [[CrossRef](#)]
132. Lahdelma, I.; Eerola, T. Mild Dissonance Preferred Over Consonance in Single Chord Perception. *i-Perception* **2016**, *7*, 204166951665581. [[CrossRef](#)]
133. Randall, R.R.; Khan, B. Lerdahl’s tonal pitch space model and associated metric spaces. *J. Math. Music* **2010**, *4*, 121–131. [[CrossRef](#)]
134. Geer, J.V.D.; Levelt, W.; Plomp, R. The connotation of musical consonance. *Acta Psychol.* **1962**, *20*, 308–319. [[CrossRef](#)]
135. Maher, T.F. “Need for Resolution” Ratings for Harmonic Musical Intervals. *J. Cross-Cult. Psychol.* **1976**, *7*, 259–276. [[CrossRef](#)]
136. Arthurs, Y.; Beeston, A.V.; Timmers, R. Perception of isolated chords: Examining frequency of occurrence, instrumental timbre, acoustic descriptors and musical training. *Psychol. Music* **2017**, *46*, 662–681. [[CrossRef](#)]
137. Sutcliffe, T. Syntactic Structures in Music. 2011. Available online: <http://www.harmony.org.uk/> (accessed on 21 July 2022).
138. der Nederlanden, C.M.V.B.; Joannis, M.F.; Grahn, J.A. Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *NeuroImage* **2020**, *214*, 116767. [[CrossRef](#)]

- 
139. Milne, A.J.; Holland, S. Empirically testing Tonnetz, voice-leading, and spectral models of perceived triadic distance. *J. Math. Music* **2016**, *10*, 59–85. [[CrossRef](#)]
  140. Genuys, G. Pseudo-distances between chords of different cardinality on generalized voice-leading spaces. *J. Math. Music* **2019**, *13*, 193–206. [[CrossRef](#)]