Modeling the Perception of Similarity in Musical Chords

Konstantin Howard (kdhoward@princeton.edu)
Department of Computer Science, Princeton University

Raja Marjieh (rm5855@princeton.edu)
Department of Psychology, Princeton University

Thomas L. Griffiths (tomg@princeton.edu)
Departments Psychology and Computer Science, Princeton University

Abstract

One domain in cognitive science examines how different sensory stimuli are represented and understood by the mind. Musical chords, composed of simpler stimuli, but more than the sum of its parts, present an interesting stimulus to explore in this domain. The question of how musical chords are represented can be answered by studying studying their comparison. Research dating back centuries has proposed a variety of different models for conceptualize the relationship between subsequent chords. This paper implements standard voice-leading, minimal voice-leading, experimentally-judged consonance difference, and spectral pitch-class distance models of chord similarity. As hypothesized, we found that a combination of minimal voice leading and consonance difference best explains the experimental similarity data collected on dyads and triads of Shepard tones.

Keywords: musical chord; chord similarity; consonance; perception modeling; Shepard tone

Introduction

Ever since humanity first began to create music, it has sought to understand and explain its musical preferences (Milne & Holland, 2016). As a highly subjective, yet consensus-derived phenomenon, it offers a fascinating array of contexts for stimulus perception and decision-making for cognitive research. The narrow facet of psychoacoustics explored in this study is the perceived similarity between musical chords. Chords, collections of two or more individual notes, define *harmony*, which constitutes the backbone of any musical creation. Music theory offers rules for harmonic combinations and movement, but the underlying psychological representations that led to these rules are not settled (Rogers & Callendar, 2006).

Research in the psychological representation of various stimuli has grown tremendously since Shepard's seminal paper (1979). In this context, previous research has proposed various models of chord similarity and tested them on limited harmonic combinations (Milne & Holland, 2016). The key difficulty is that there exist a variety of models and many variables that might influence how a chord is perceived. For example, consider the phenomenon of inversions: as sets, not sequences, and the order (low to high) of pitches in a chord can be changed to significantly change the character of a chord's sound, but not its underlying composition. Another example, is the apparent asymmetry of harmony: if chords are not sequences themselves, their combination to form chord *progressions* most definitely are. Harmony relies on tension and resolution in one direction, such as from a V chord to a I chord, which clearly does not work in the opposite direction (Kamien, 2007).

This paper is an attempt to analyze a unique collection of experimental data on musical chord similarity, already gathered by Raja Marjieh. We implement two existing models of chord comparison, voice leading and spectral pitch-class distance, and synthesize their predictive power with a never before utilized measure of harmony: consonance. The consonance to dissonance range is a subjective tonal quality that only applies to chords. In short, it describes how pleasant a collection of notes sounds when played simultaneously.

The results of our study reveal that a combination of voice-leading and consonance difference is the strongest predictor for chord similarity when using Shepard tones and every possible variation of dyads and triads (with a common root). This paper will first explain previous research in chord and consonance representation. Then, it will describe the models and consonance metric used in the study to compare chords. It will compare the results of these models in predicting the similarity of two chords with the survey data. Lastly, it will conclude by analyzing the implications of the results on our understanding of chord representation and future research in this domain.

Background

We will review the fundamental concepts of chords, harmonic movement, and consonance as they are understood both in music theory and in psychoacoustic literature.

Chord Distance

A chord is any collection of two or more pitches. If a pitch is sound wave at a given frequency, then chords can be described scientifically as the constructive interference of multiple such waves and categorized by their frequency ratios. Representations of chords have taken on many forms over the centuries, beginning with Euler's Tonnetz, a wrapped lattice diagram (Milne & Holland, 2016). Another popular representation is pitches as chromae: natural numbers in which each half-step is a difference of 1. Taken modulo 12, this scale can simplify to represent pitch-classes, the same note across all octaves (Tymoczko, 2016). Chromatic difference is naturally understood as subtraction. More recently, a popular model has become the spectral pitch-class model that repre-

sents pitches as vectors with elements for cents, the smallest interval used to describe pitch where a half-step is 100 cents. The spectral pitch-class model derives its power by representing not only the fundamental frequency of a pitch but also its overtones and comparing these overtones with those of other pitches of chords.

Harmonic Series Crucial to understanding pitches and the spectral model is the concept of the harmonic series. The fundamental frequency of a pitch is really the average of its overtones, a series of notes above the fundamental that can also be heard, with diminishing amplitude, whenever a tone is sounded. The composition of these overtones creates the difference in *timbre* (Marjieh, Harrison, Lee, Deligiannaki, & Jacoby, 2022).

1, 1, 5, 1, 3, 5, b7, 1, 9, 3, ♯ 4, 5, ...

Figure 1: The pitches of the first twelve intervals of the harmonic series with respect to the root pitch or fundamental frequency.

Shepard Tones In an effort to simplify the representation of pitches, stimuli known as Shepard tones are often used in research (Shepard, 1964). Shepard tones consist only of stacked octaves, thereby eliminating the medium-specific variation of overtones and enforcing the circularity of pitch intervals. All octaves of a pitch are the same Shepard tone, a *pitch-class*. The similarity and consonance data used in this study was gathered from experiments that used Shepard tones.

Harmonic Movement

The foundation of musical harmony is the chord, and it is common, though not necessary, that music involves movement between chords, called harmonic progression. What determines an appropriate harmonic progression? The answer to this question depends greatly on a listener's cultural and musical background. Music theory dictates rules about what harmonic motion sounds "good", but such constraints are routinely challenged in the pursuit of artistic creation and the field as a whole is post hoc to the artform of music. In this paper we consider chord similarity, which is but one facet of harmonic motion. By no means does it attempt to explain the entirety of this phenomenon.

Research has found some support for a variety of models that determine chord similarity: voice-leading (Rogers & Callendar, 2006), chord Tonnetz, spectral pitch-class, (Milne & Holland, 2016), topological (Tymoczko, 2016), and others (Milne, 2009). Notably, the experiments used to verify these models often tested on only the 26 possible (accounting for transposition) pairs of minor and major triads (Milne & Holland, 2016; Rogers & Callendar, 2006) or drew from a real piece of music (Smith & Cuddy, 2003). The variety of successful models indicates that there is not a single framework that fully describes the psychoacoustic process that occurs when we hear musical chords. Combinations of these models are likely to perform better by accounting for different features of the representation. Likewise, the limited testing cases leave many more possible chords to be studied. Although the major and minor triads are the foundation of most chords used in music, cluster voicings (notes in close proximity to each other) commonly appear in orchestrated musical pieces in idioms like jazz.

Consonance

Where similarity is the metric for comparing two chords in succession, consonance is the corresponding metric for comparing multiple pitches simultaneously. Also a subjective measure, research has shown that consonance judgements of a given interval vary according to a listener's musical background and the timbre of the instrument sounding the pitches (Marjieh et al., 2022)). Many models have been offered to explain how consonance is perceived by the brain, often using similar models as those used when examining the relationship between different chords (for an analysis of 16 different models of consonance perception, see Harrison & Pearce, 2020).

We consider consonance because it is clearly a quality intrinsic to any chord, and as such must contribute in some way to chord similarity judgement. For example, two "consonant" chords are similar to each other in that they are both considered to be consonant.

Approach

The novelty of this study is the use of consonance as a metric for chord similarity. In addition to this metric, we implement two existing models of distance between musical chords, voice-leading and spectral pitch-class, as they are described by Milne and Holland (2016).

Voice Leading Distance

There are two variants of voice-leading distance. For both, we represent a chord of n pitches as a vector of length n with values $v_i \in \{k \bmod 12, (k+12) \bmod 12\}$. The standard model of voice-leading distance for two chords simply computes the p-norm of the difference between their two vectors.

$$d(u,v;p) = (\sum_{i=1}^{n} |u_i - v_i|^p)^{1/p}$$

In the minimal model, we consider all possible inversions of one chord (i.e. permutations of its vector) and find the minimal p-norm difference among these, as well as minimizing over which octave a note is represented in. For example, for the two dyads (C, G) and (A, B) represented as the vectors (0, 7) and (9, 11), the minimal voice-leading vector would be (1, 2), not the difference (9, 4). This makes sense because B is one half step from C and A is one whole step from G.

$$d(u,v;p) = \min_{\forall s \in S_n} (\sum_{i=1}^{n} \min(|u_{s(i)} - v_i|, 12 - |u_{s(i)} - v_i|)^p)^{1/p}$$

Where $\forall s \in S_n$ gives every permutation of the vector u. The inner minimization ensures that no distance between two notes is greater than 6 (the circularity of the chromatic scale).

Spectral Pitch-class Distance

In this model, we represent each pitch as a 1200-element vector where each index corresponds to the strength of frequency at that cent (the cent is another scale of pitch distance where one half-step is 100 cents) and C is arbitrarily denoted as 0. This representation captures the overtones of the harmonic series discussed in the introduction. Because each successive overtone is quieter, a shrinking factor ρ is used to dampen the higher overtones such that entry for the i^{th} overtone modulo 1200 is $i^{-\rho}$. We do this for the first twelve overtones because any higher overtones are hardly perceptible (Milne & Holland, 2016). For chords, we sum the representative vectors of each component pitch. Next, we smooth the sum vector by convolving it with a random normal distribution with a mean of one and variance σ. This smears the spikes across adjacent cents because frequencies are not so precisely heard in reality (Milne & Holland, 2016). The pitch-class distance is the cosine distance between the two chord vectors, given by

$$d(u,v;\rho,\sigma) = 1 - \frac{uv^T}{\sqrt{uu^T vv^T}}$$

Methods

All analysis was implemented in Python with NumPy and SciPy standard libraries.

Similarity Data

The study played a series of pairs of chords for each participant and asked them to rate their "similarity" on scale from 1 to 7, where 7 is the most similar. For the dyads, the set of chords consisted of every possible combination of two pitches for 347 participants where the pitches played were Shepard tones. For the triads, the set of chords again consisted of every combination of dyads but played with a root of 65 or F for 79 participants. This placed every triad in the same context. Crucially, the data was averaged over the order in which the chords were played, correcting for the asymmetry of harmonic motion discussed in the introduction. Even before the analysis and modeling, it was clear that the data was highly structured. The numbers on the axes are the dyads indexed in the following manner: for each pitch i, for each pitch $j \neq i$ starting from C. See Figure 2.

Consonance Data

The consonance data was collected in a similar manner. Every possible pair of pitches and every possible triad built on top of a root of 65 or F played as Shepard tones were sounded for 99 and 124 participants, respectively, and they rated the chord on a scale of 1 to 7, where 7 is the most consonant. This data was converted into a similarity metric by considering the absolute difference in consonance for each pair of chords, thus generalizing over the order of the two chords.

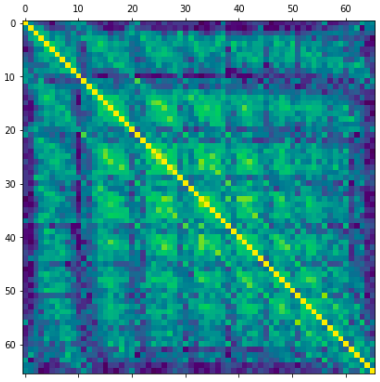

Figure 2: Raw Triad Similarity Data (Symmetric)

The figure below shows that the consonance data is consistent with music theory: half-steps are the most dissonant while a major third, a minor third, and the same tone make up the top three most consonant intervals.

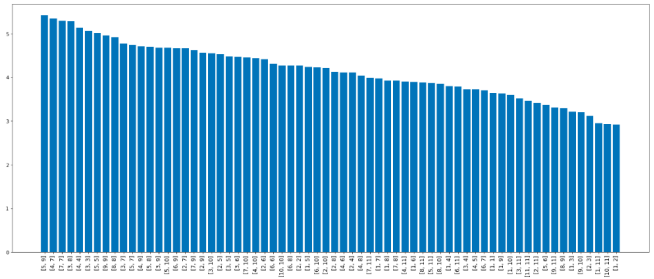

Figure 3: Dyads Ranked by Perceived Consonance

Results

Summary

All models besides the spectral pitch-class proved to be indicators of similarity with varying degrees of success for both the dyad and triad data sets. While the similarities predicted by each model varied little from the similarity data, the combined model performed best when measured by correlation coefficient. These performance metrics are summarized below with the standard voice-leading model included as a baseline.

Table 1: Performance Summary for Dyads

Model	var	r
Standard Voice-leading	0.0099	-0.3984
Minimal Voice-leading	0.0085	-0.5653
Absolute Consonance	0.0103	-0.3719
Gaussian Consonance	0.0102	0.3821
Spectral Pitch-class	N/A	N/A
Combined	0.0076	0.6623

Table 2: Performance Summary for Triads

Model	var	r
Standard Voice-leading	0.0079	-0.4675
Minimal Voice-leading	0.0074	-0.5121
Absolute Consonance	0.0085	-0.3847
Gaussian Consonance	0.0085	0.3671
Spectral Pitch-class	N/A	N/A
Combined	0.0067	0.5878

Voice-leading Distance

Consistent with previous studies, minimal voice-leading distance proved to be an effective predictor of similarity (Milne & Holland, 2016; Rogers & Callendar, 2006). The p-norm used in the distance calculations was optimized by $p = 1$. See Figure 4.

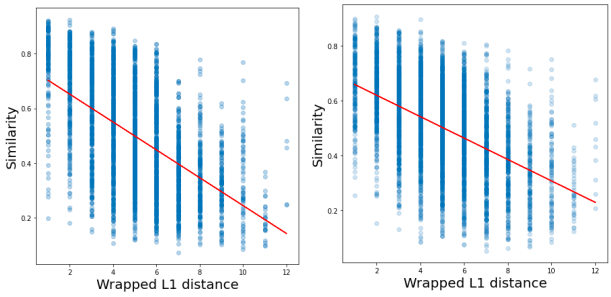

Figure 4: Dyad and Triad Fits for Minimal Voice-leading Model over Similarity Data Plots

Spectral Pitch-class

The spectral pitch-class model provided an inconclusive metric in this study. Not only did the scatter plots visibly lack structure, attempts to find a fit for the data by optimizing the parameters ρ and σ were computationally unsuccessful. This is likely due to the use of Shepard tones as stimuli (more on this in the discussion). See Figure 5.

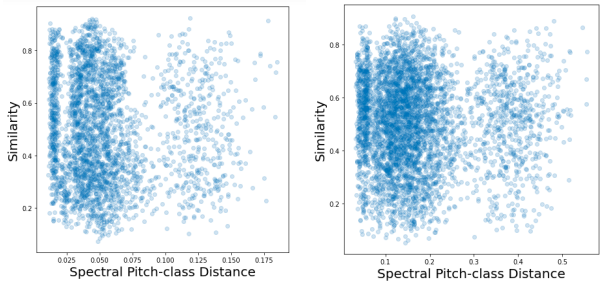

Figure 5: Dyad and Triad Plots for Spectral Pitch-class Model

Consonance Difference

The consonance metric on its own was not a very strong predictor of similarity. In fact, when compared to the standard voice-leading model, it performed slightly worse for the dyads and significantly worse for the triads. Three functions were used to refine the difference data: absolute value, Gaussian, and Sigmoid. The Sigmoid obfuscated the distribution, but the absolute value and Gaussian performed similarly well, with the Gaussian outperforming absolute value for dyads and vice versa for triads. Optimizing for σ in the Gaussian function did not change the fit of the model. Note that the Gaussian function also generalizes over order by squaring the consonance difference. See Figure 6.

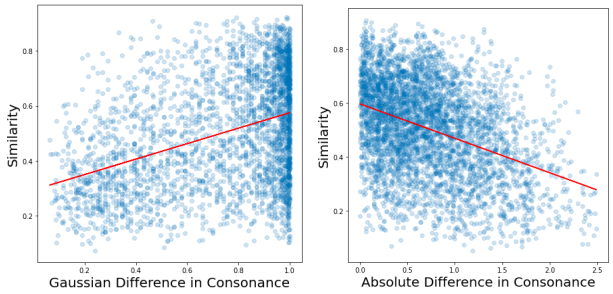

Figure 6: Dyad and Triad Fits for Consonance Difference

Combined Model

As hypothesized, the combined model of minimal voice-leading distance and consonance difference performed the best by combining the predictive powers of both metrics in a linear regression with two variables. The combined models for dyads and triads use the Gaussian and absolute consonant differences, respectively. Combined models including the spectral pitch-class distance were not implemented, given its failure as a predictor. See Figure 7.

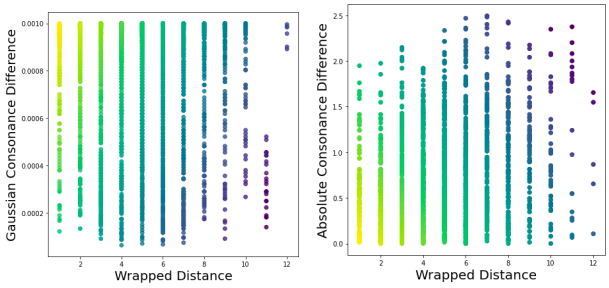

Figure 7: Dyad and Triad Plots for Combined Models where yellow indicates a greater similarity

Discussion

By examining the way humans think about the similarity of musical chords, we are able to learn more about how these complex stimuli are represented in the brain. This study confirmed that the minimal voice-leading model is an effective predictor of similarity judgements, was inconclusive on the spectral pitch-class model, and, crucially, added a new metric to the arsenal of psychoacoustic researchers: consonance

difference. The minimal voice leading model, one of the simplest and most intuitive measures chord distance, is sure to be used in future chord studies as a baseline against which to test new models.

Even without developing a comprehensive model of consonance (a wholly separate research topic that should also be explored), simply using consonance judgement data as this study did provides a powerful quantitative descriptor chords that can be combined with other models. Consonance judgements alone are likely not enough to be a reliable predictor, given that it performed worse in isolation than even the standard voice-leading model in this study. This is because there are many different qualities to a chord than just its consonance. For example, two identical chords a tritone apart (6 half-steps) may sound equally consonant on their own, indeed they should if they consist of the same intervals, but when played one after the other, may sound dissimilar because of the "unpleasantness" of the tritone interval. This explanation is also supported by the results of the combined model. Complex stimuli like chords are likely represented in the mind according to its various qualities, some objective, like a distance measure, and some subjective, like a consonance difference. Future research should to continue develop new models of chord similarity, capturing different qualities of the stimulus, and test them in combination with existing proven models.

The failure of the spectral pitch-class model is on the one hand not surprising and confirms that Shepard tones work as they are intended to: remove all the overtones and timbral impurities of naturally sounded pitches and keep only the stacked octaves. On the other hand, the implications of the failure should be explored in future research. If the spectral pitch-class model is representative of real-world cognition, as Milne and Holland found in their 2016 study that did not use Shepard tones, then studies that do use Shepard tones could not possibly contribute. In fact, if overtone relations are so critical to chord perception, research on this subject cannot be researched with Shepard tones at all. In an effort to simplify the stimulus to make for easier generality, Shepard tones might actually be removing a critical facet of the stimulus. Further research to evaluate this possibility is most urgent, as it may determine how future studies can be conducted.

Acknowledgments

I would like to extend my gratitude to Raja Marjieh for advising throughout this study, my first experience in independent research, and to Tom Griffiths for introducing me to the fascinating field of computational cognitive science.

References

Dean, R., Milne, A., & Bayes, F. (2019). Spectral pitch similarity is a predictor of perceived change in sound- as well as note-based music. *Music an Science*, *2*.

Harrison, P., & Pearce, M. (2020). Simultaneous consonance in music perception and composition. *Psychological Review*, *127*, 216–244.

Kamien, R. (2007). *Music: An appreciation, 6th brief edition*. New York, NY: McGraw-Hill.

Marjieh, R., Harrison, P. M. C., Lee, H., Deligiannaki, F., & Jacoby, N. (2022). Reshaping musical consonance with timbral manipulations and massive online experiments. *bioRxiv*.

Milne, A. (2009). A psychoacoustic model of harmonic cadences. *University of Jyväskylä MA Thesis*.

Milne, A., & Holland, S. (2016). Empirically testing tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, *10*, 59-85.

Rogers, N., & Callendar, C. (2006). Judgements of distance between trichords. In *Proceedings of the ninth international conference on music perception and cognition* (pp. 1686–1691). ICMPC.

Shepard, R. (1964). Circularity in judgments of relative pitch. *The Journal of the Acoustical Society of America*, *36*, 2346-2353.

Shepard, R. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Smith, N., & Cuddy, L. (2003). Perceptions of musical dimensions in beethoven's waldsiein sonata: An application of tonal pitch space theory. *Musicae Scientiae*, *7*, 7-34.

Tymoczko, D. (2016). The geometry of musical chords. *Science*, *313*, 72-74.