TISMIR

Lieck, R., et al. (2020). The Tonal Diffusion Model. *Transactions of the International Society for Music Information Retrieval,* 3(1), pp. 153–164. DOI: https://doi.org/10.5334/tismir.46

RESEARCH

# The Tonal Diffusion Model

Robert Lieck, Fabian C. Moss and Martin Rohrmeier

Pitch-class distributions are of central relevance in music information retrieval, computational musicology and various other fields, such as music perception and cognition. However, despite their structure being closely related to the cognitively and musically relevant properties of a piece, many existing approaches treat pitch-class distributions as fixed templates.

In this paper, we introduce the Tonal Diffusion Model, which provides a more structured and interpretable statistical model of pitch-class distributions by incorporating geometric and algebraic structures known from music theory as well as insights from music cognition. Our model explains the pitch-class distributions of musical pieces by assuming tones to be generated through a latent cognitive process on the Tonnetz, a well-established representation for harmonic relations. Specifically, we assume that all tones in a piece are generated by taking a sequence of interval steps on the Tonnetz starting from a unique tonal origin. We provide a description in terms of a Bayesian generative model and show how the latent variables and parameters can be efficiently inferred.

The model is quantitatively evaluated on a corpus of 248 pieces from the Baroque, Classical, and Romantic era and describes the empirical pitch-class distributions more accurately than conventional template-based models. On three concrete musical examples, we demonstrate that our model captures relevant harmonic characteristics of the pieces in a compact and interpretable way, also reflecting stylistic aspects of the respective epoch.

**Keywords:** Tonnetz; Pitch-Class Distributions; Cognitive Modeling; Tonality; Bayesian Generative Model

## 1. Introduction

*Pitch-class distributions* (PCDs) are fundamental objects of study in many fields of empirical music research, from *music information retrieval* (MIR), mathematical music theory, and computational musicology to music perception, cognition, and corpus studies (Huron and Veltman, 2006). They are closely related to relevant cognitive and musical properties of the pieces, such as their key, degree of consonance and dissonance, and characteristics of the modulation plan. A PCD may thus be understood as a fingerprint of a piece that contains important information about its tonality.

However, despite the meaningful structure of PCDs, many existing approaches do not make the relevant aspects explicit in the way PCDs are modeled and represented. In fact, PCDs are mostly treated as fixed templates, whose structure is only interpreted in a post-hoc manner. This is also reflected in the way PCDs are commonly visualized, namely as categorical distributions of twelve chromatic semitones, where visual proximity does not necessarily reflect musical relations. Ignoring music-theoretical insights about the structure of tonal space for representing and visualizing PCDs conceals deeper structural relations that are reflected in these distributions.

The goal of this paper therefore is to provide a compact, structured and interpretable representation of PCDs. The proposed *Tonal Diffusion Model* (TDM) achieves this in two ways: first, by leveraging relevant music-theoretic and cognitive insights, specifically, the algebraic representation of tonal space called the *Tonnetz* (Hostinský, 1879; Cohn, 1997); and second, by defining an explicit statistical model, which independently generates all tones in a piece by starting at a piece-specific tonal origin and taking steps in different directions on the Tonnetz.

We quantitatively evaluate our model against several baseline models on a corpus of 248 pieces from the Baroque, Classical, and Romantic era, demonstrating that the TDM provides music-theoretically reasonable interpretations. Moreover, it models the empirical PCDs more accurately than the unstructured baseline models as measured by the *Kullback-Leibler divergence* (KLD) to the empirical PCD.

## 2. Related Work

### *2.1 Representations of Pitch-Class Distributions*

Many large corpora are available in the MIDI format, which encodes pitch through an integer representation, commonly interpreted as *twelve-tone equal temperament*

Digital and Cognitive Musicology Lab, Digital Humanities Institute, Collège des Humanités, École Polytechnique Fédérale de Lausanne, CH

Corresponding author: Robert Lieck (robert.lieck@epfl.ch)

(12TET) (Selfridge-Field, 1997). MIDI thus has the major drawback that it does not preserve the musically relevant distinctions between enharmonically equivalent pitch classes that are expressed in pitch spelling. This simplified representation has largely carried over to PCDs, which are commonly visualized in a linear chromatic arrangement.

This implies three shortcomings (see Appendix A for a more detailed discussion): 1) the visual representation in a linear arrangement suggests some implicit ordering and proximity relation that is not explicitly modeled; 2) the commonly used chromatically ascending order does not reflect tonal relations well, especially with respect to harmony and key, where an arrangement along the line or circle of fifths would be better suited; and 3) the space of pitch classes exhibits an inherent cyclic topology, which is not reflected in a linear arrangement. The TDM overcomes these limitations by using the Tonnetz as basis representation for PCDs, which preserves pitch spelling, and by representing proximity relations via steps on the Tonnetz. Moreover, the Tonnetz extends the purely fifth-based representation by additionally allowing for proximity relations based on major and minor third steps.

### 2.2 Empirical Evidence for the Tonnetz

The pervasive categorical representation of PCDs is also used to display the results of psychological probetone experiments where participants rate different degrees of stability of tones in 12TET (Krumhansl and Kessler, 1982). The underlying assumption in these experiments is that enculturated listeners have acquired mental representations of tonality through exposure to music via statistical learning, which approximately corresponds to inferring the note distributions in musical corpora (Krumhansl, 1990; Huron and Veltman, 2006).

Several studies have shown that PCDs suggest an arrangement that is essentially equivalent to the Tonnetz (see Section 3.2). Krumhansl and Kessler (1982) use multidimensional scaling (MDS) to find a toroidal structure of the 24 key profiles, obtained by transposing the major and minor profiles to all twelve keys. The appropriateness to model tonal space as a torus has a long-standing tradition in music theory and is supported by these empirical findings.

The cognitive relevance of this structure has been further investigated by Krumhansl (1998), where the Tonnetz is used to model the perceived distance of triads. Toiviainen and Krumhansl (2003) use a selforganizing map, which essentially reproduces the results of the MDS. Their study extends prior research by modeling listeners' dynamic perception of tonality in a musical piece over time, which is reflected in high activation of certain regions on the Tonnetz.

Milne and Holland (2016) compare three different models for perceived triadic distance, namely the psychoacoustic measure of spectral pitch-class distance, distance on the Tonnetz, and voice-leading distance between the triads. They conclude that the Tonnetz is highly predictive for empirical data of triadic similarity and correlates strongly with the psychoacoustic measure.

Our approach builds upon these prior works on the cognitive relevance of the Tonnetz by using it more generally to represent empirical PCDs.

### 2.3 Topic Models

The basic structure of the TDM (see Section 4) is similar to that of probabilistic topic models (Blei et al., 2003; Steyvers and Griffiths, 2007), which have successfully been employed for the prediction of word frequencies in natural language. Topic models represent a corpus as a bag (an unordered collection) of documents (e.g. texts or pieces) and each document as a bag of items (e.g. words or notes). However, language models do not incorporate the kind of structure present in music, which can be expressed in algebraic (Fiore and Noll, 2011) and geometric (Tymoczko, 2011) models. Most topic models for language assume no relation between words and topics besides their probability of co-occurrence, even though there are some attempts to overcome this limitation (Griffiths et al., 2005).

A topic model for music should therefore go beyond those developed for natural language by explicitly modeling the structural relations of pitch classes in tonal music and incorporating the geometric and algebraic structures known from music theory as well as insights from music cognition.

The only music-specific application of topic models we are aware of is by Hu and Saul (2009a,b). However, while sharing the general topic model structure with our TDM, there are two profound differences (see Appendix B for a more detailed discussion): 1) Hu and Saul address the problem of key finding and learn two static key profiles in twelve enharmonically equivalent pitch classes, whereas our model aims to provide a more fine-grained representation of tonality that goes beyond a binary major/minor classification, also taking into account pitch spelling. 2) Hu and Saul infer separate key labels for different sections of each piece, while our goal is to compactly represent the global harmonic character of the entire piece.

Consequently, while Hu and Saul evaluate their model by comparing its classification accuracy to that of established template-based key finders, we use the KLD to the empirical distribution as a performance measure, which makes a direct comparison of the present approach with that of Hu and Saul difficult. Note, however, that our baseline model with two static key profiles is conceptually very similar to the model by Hu and Saul.

## 3. A Cognitive Interpretation of the Tonnetz

Based on the music-theoretical considerations discussed above, we present a cognitive interpretation of the Tonnetz and describe a model of the latent cognitive process, grounded in the perception of tonal music.

### 3.1 Representation of Tones

We represent tones on the *line of fifths* (Temperley, 2000). That is, the tonal space $\mathscr{T}$ is the set of tones generated by taking an arbitrary number of steps from a given reference tone by either descending or ascending a perfect fifth (see **Figure 1**). Assuming octave equivalence, *tonal pitch classes* (TPCs), used for labeling notes in Western music, have a one-to-one mapping to the line of fifths (second row in **Figure 1**). The set of intervals $\mathscr{I}$, describing the distance between two tones, is the set of directed TPC intervals (third and fourth row in **Figure 1**).

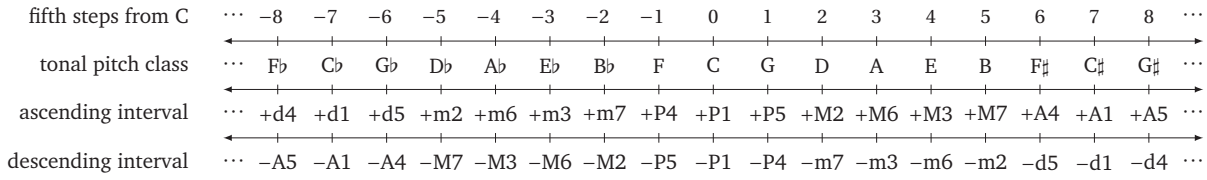



**Figure 1:** Tonal space $\mathscr{T}$ and interval space $\mathscr{I}$ on line of fifths: tonal pitch-classes and tonal pitch-class intervals along with their integer representation in terms of the number of steps along the line of fifths (using C as a reference tone).

We use '+' and '−' to indicate the direction (ascending or descending), describe the *generic intervals* (Clough and Myerson, 1985; Harasim et al., 2016) with Arabic numerals (1: unison, 2: second, 3: third, 4: fourth, 5: fifth etc.), and denote the quality of the intervals (diminished, minor, perfect, major, and augmented) with 'd', 'm', 'P', 'M', 'A', respectively.

Note that in this representation, we are able to distinguish enharmonically equivalent intervals, such as +A4, the ascending augmented fourth (tritone), and +d5, the ascending diminished fifth. This is a musically and cognitively relevant distinction: since the interval +d5 commonly resolves inwards into a third (+M3 or +m3) and +A4 (the tritone) commonly resolves outwards into a sixth (+m6 or +M6), these intervals create different harmonic expectations that imply different cognitive representations. As discussed in Section 2.1, a major drawback of the commonly used MIDI format is that it cannot be used to represent these musically important distinctions.

Our goal in this paper is to provide a statistical model for music that allows to describe tonal pitch class distributions of musical pieces in a compact, structured and interpretable way. In the following, the term *pitch class* always refers to TPCs and the term *interval* refers to directed TPC intervals.

### 3.2 The Tonnetz

While the line of fifths connects all possible pitch classes and is of central importance in tonal music (Gárdonyi and Nordhoff, 2002; Temperley, 2000; Weber, 1851), there are other ways of relating the same two tones via one or more interval steps. Of particular importance for harmonic relations in tonal music are, apart from the perfect fifth (±P5), also the major third (±M3) and the minor third (±m3), both ascending and descending (Aldwell et al., 2010; Haas, 2004). We call these the *primary intervals* and for our evaluation of the TDM we will restrict the set of intervals $\mathscr{I}$ to the primary intervals. **Figure 2** shows these primary interval relations with respect to the central tone C.

The infinite expansion of this graph is called the *Tonnetz* (Bigo and Andreatta, 2016; Cohn, 1997, 2012; Harasim et al., 2019) and has a number of historic precursors (e.g., Euler, 1739; Hostinský, 1879; von Oettingen, 1866; Riemann, 1896). Using the representation as a hexagonal graph (Douthett and Steinbach, 1998), PCDs can be plotted as color-coded heatmaps (Moss, 2019). **Figure 3** shows the distributions of three pieces from three different musical epochs, which we also use for our evaluation below, namely:

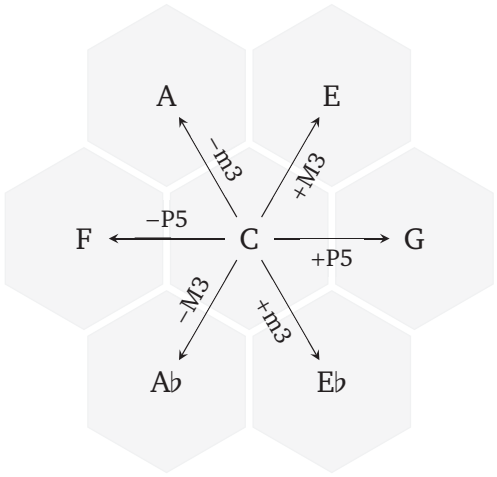
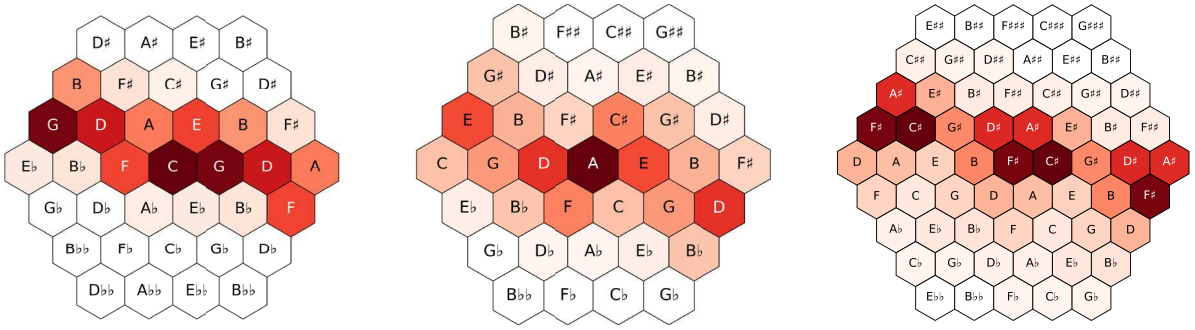



**Figure 2:** Primary intervals with respect to C.

1. Baroque: Johann Sebastian Bach, Prelude in C major, BWV 846 (1722),
2. Classical: Ludwig van Beethoven, Sonata op. 31, no. 2 in D minor 'Tempest', 1st mov. (1802),
3. Romantic: Franz Liszt, *Bénédiction de Dieu dans la Solitude*, S. 173/3 (1853)

In the case of tonal pitch classes, the Tonnetz is topologically equivalent to the *Spiral Array* (Chew, 2000). In the two-dimensional visualizations in **Figure 3**, where this structure is 'unrolled', the same tonal pitch class may appear multiple times and is then colored identically.

When looking at the distributions in **Figure 3** it becomes obvious that the different pieces exhibit characteristic structures along the different axes of the Tonnetz. Bach's piece is largely distributed along the line of fifths (horizontally) with its tonic C occurring most frequently. Beethoven's Sonata movement is also distributed along the line of fifths but it exhibits a considerably wider spread that is roughly symmetric about the most frequent tone A (the tonic of the dominant key). Furthermore, the pitch classes of the tonic D-minor and the dominant A-major triads occupy most probability mass, resulting in the major third axis (F–A–C♯) being more pronounced than in Bach's piece. Finally, the pitch classes in Liszt's piece are most widely distributed, covering a range from F♭ to D♯♯ (24 fifths). In contrast to the other two pieces, the particularly wide distribution of pitch classes in this piece reflects the fact that it juxtaposes harmonically distant keys (in particular through mediantic relations), such as F♯, D, and B♭. These key relations are not reflected in the line-of-fifths topology but are explicitly modeled by the major and minor third axes of the Tonnetz, which suggests the



**(a)** Johann Sebastian Bach, Prelude in C major, BWV 846 (1722, Baroque).

**(b)** Ludwig van Beethoven, Sonata op. 31, no. 2 in D minor 'Tempest', 1st mov. (1802, Classical).

**(c)** Franz Liszt, *Bénédiction de Dieu dans la Solitude*, S. 173/3 (1853, Romantic).

**Figure 3:** Tonal pitch-class distributions plotted as heatmaps on the Tonnetz (darker colors correspond to higher probabilities). The plots were generated with the Python library *pitchplots* (Moss et al., 2019).

Tonnetz to be a more suitable representation for this kind of extended tonality. We will discuss the tonal structures of these three pieces in more detail in Section 5.3.

### *3.3 Moving on the Tonnetz*

A central idea of the TDM is that the generation of tones is associated to certain motions on the Tonnetz. Specifically, the Tonnetz captures harmonic relations, so that moving on the Tonnetz corresponds to typical harmonic changes occurring in a piece. These changes happen on several time scales, from large scale modulations between different sections of a piece, to harmonic progressions, and polyphony on the surface level. Motion on the Tonnetz thus reflects the deep structural dynamics of a piece.

Different paths through the Tonnetz express different harmonic interpretations of the involved tones. This distinction of harmonic functions expressed in different paths on the Tonnetz is an explicit part of our model.

For instance, the tone E can be reached from the tone C by ascending four perfect fifths or by ascending a major third. In the first case, E is conceived to be the perfect fifth of A, which is the perfect fifth of D, which is the perfect fifth of G, which, ultimately, is the perfect fifth of C, reflecting a recursive application of applied dominants, while in the second case, the tone E stands in a direct major third relation to C. We take this distinction as expressing two different interpretations of the harmonic function of E relative to C. That is, we assume that it corresponds to a difference on the cognitive level, because it relates the tones C and E in two different ways.

In some contexts, this might be reflected in different tunings and ways of hearing, as ascending four perfect fifths leads to the Pythagorean major third E, while the E reached by ascending a major third corresponds to the major third E in just intonation. However, since these different paths from C to E are indistinguishable in a notated score, our cognitive model does *not* require them to be physically distinct.

In the TDM, we assume that different paths have different probabilities of occurrence, that is, that some paths are more likely than others. For instance, we assume that different steps occur with different probabilities and that paths cannot have an arbitrary length. The step probabilities and the path-length distribution are explicit components of the TDM (Section 4). Altogether, these aspects reflect different ways of hearing tonal relations, as well as different stylistic characteristics of tonal music in general and of concrete pieces more specifically.

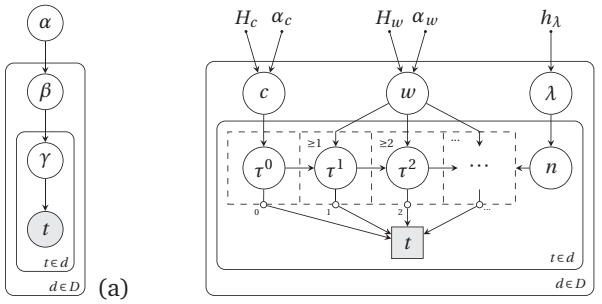
### *3.4 The Tonal Origin*

Tonal music is fundamentally characterized by the existence of one or more tonal centers that are related to each other in a hierarchical manner, for instance, by modulating through different local keys (Schenker, 1935; Schoenberg, 1969; Lerdahl and Jackendoff, 1983; Rohrmeier, 2011; Koelsch et al., 2013; Rohrmeier, 2020). In the TDM, we therefore introduce the concept of the *tonal origin* and assume that a unique tonal origin exists for each piece (see Section 4.2). Intuitively, the tonal origin corresponds to the single pitch class that is best suited to explain all tones in the piece by starting at this point on the Tonnetz and taking a small number of primary interval steps. There are two important caveats to keep in mind for interpreting the tonal origin: 1) Assuming a single tonal origin does *not* mean that the piece cannot modulate between different keys having their own respective local centers. 2) The tonal origin in our model is *not* necessarily equivalent to the tonic of the global key. Please see Appendix C for a more detailed discussion of these two points.

## 4. The Tonal Diffusion Model

The purpose of the TDM is to model the intuitions presented in Section 3 formally and derive quantitative measures that capture the diffusion of probability mass from the tonal origin along the different axes of the Tonnetz. To this end, we define an explicit generative model, in which each tone in a piece is generated by starting at the tonal origin and taking a number of steps on the Tonnetz. As described in Section 3.3, there are many possible paths connecting two tones on the Tonnetz, which correspond to different cognitive interpretations of their harmonic relation. In our model, these different derivations are treated as a latent representation, which is marginalized out to determine the overall probability of reaching a particular tone.

In the general formulation of our model, we include the possibility of multiple tonal origins, and allow for an




corpus-level parameters ($\alpha$)
$H_c, \alpha_c, H_w, \alpha_w, h_\lambda$
piece-level variables ($\beta$)
$c$ : distribution over tonal centers
$w$ : distribution over intervals
$\lambda$ : distribution over path lengths
latent variables ($\gamma$)
$\tau^i$ : intermediate tones
$n$ : number of steps
observed variables
$t$ : tones in a piece (c)

**Figure 4: (a)** General structure of a topic model. **(b)** The Tonal Diffusion Model. We use an extended gate notation (Minka and Winn, 2009): random variables in dashed boxes and edges connected via a gate (empty circles) exist only if the random variable connected to the border ($n$ in this case) fulfills the condition in the top-left corner or next to the gate, respectively. **(c)** Overview of the variables and parameters, see Section 4.1 for more details.

arbitrary set of intervals, as well as a generic path-length distribution. For the evaluation of the model on tonal music, presented in Section 5, we then make the more specific assumptions motivated in Section 3. Specifically, we assume a single tonal origin (Section 3.4) and restrict the allowed interval steps to the set of primary intervals present in the Tonnetz (Section 3.2).

The TDM has the basic structure of a topic model with two nested levels of generation, as shown in **Figure 4a**. On the inner level, each piece (i.e. each document) $d$ is represented as a 'bag of tones' and all tones $t \in d$ are generated independently, conditional on the piece-level variables $\beta$. On the outer level, the corpus $D$ is represented as a 'bag of pieces' and all pieces $d \in D$ are generated independently, conditional on the corpus-level parameters $\alpha$. The tones $t$ of a piece are the observed variables while $\gamma$ represents any latent variables involved in the generation of a single tone $t$.

We extend this basic structure by splitting the corpus-level and piece-level variables into multiple distinct variables with a clear semantics and by replacing the inner generative step for a single tone with a model of the underlying latent cognitive process. The complete model is shown in **Figure 4b** and will now be explained in detail.

### 4.1 Generative Process

Let $\mathcal{T} \equiv \mathbb{Z}$ be the space of all possible TPCs and $\mathcal{I} = \{t - t' \mid t, t' \in \mathcal{T}\} \equiv \mathbb{Z}$ be the space of all TPC intervals (see also **Figure 1**). We assume the following generative process, which corresponds to the graphical model in **Figure 4b**.

For each piece $d$ in a corpus $D$ draw

$$c \sim \mathrm{Dir}(H_c, \alpha_c) \tag{1}$$

$$w \sim \mathrm{Dir}(H_w, \alpha_w), \tag{2}$$

that is, the distribution of tonal origins $c$ is drawn from a Dirichlet process with base distribution $H_c$ over the tonal space $\mathcal{T}$ and concentration parameter $\alpha_c$; and the interval weights $w$ are drawn from a Dirichlet process with base distribution $H_w$ over the interval space $\mathcal{I}$ and concentration parameter $\alpha_w$. Furthermore, the path-length distribution $\lambda$ is drawn from a prior with hyper-parameters $h_\lambda$

$$\lambda \sim p(\lambda \mid h_\lambda), \tag{3}$$

where the prior depends on the specific path-length distribution being used: for a Poisson distribution the conjugate prior is a gamma distribution, for a binomial distribution it is a beta distribution.

For each tone $t$ in a piece $d$ draw

$$n \sim \mathrm{PathLength}(\lambda) \tag{4}$$

$$\tau^0 \sim \mathrm{Categorical}(c) \tag{5}$$

$$\tau^{i+1} \sim \tau^i + \mathrm{Categorical}(w) \tag{6}$$

$$t \leftarrow \tau^n, \tag{7}$$

that is, a number of steps $n$ is drawn from the path-length distribution (Poisson or binomial); a series of latent tones $\tau^0, \ldots, \tau^n$ is generated by first drawing $\tau^0$ from the distribution of tonal origins and then transitioning $n$ times from $\tau^i$ to $\tau^{i+1}$ by adding an interval; finally, the last tone $\tau^n$ of the sequence is observed as $t$.

### 4.2 Inference

For a given corpus $D$ of musical pieces and corpus-level parameters $\alpha \equiv (H_c, \alpha_c, H_w, \alpha_w, h_\lambda)$ we would like to compute the *maximum posterior* (MAP) estimate $\beta^*$ of the piece-level variables $\beta \equiv (c, w, \lambda)$ for each piece $d$,

$$\beta^* = \arg\max_\beta p(\beta \mid d) \tag{8}$$

$$\text{with} \qquad p(\beta \mid d) \propto p(\beta; \alpha) \prod_{t \in d} p(t \mid \beta), \tag{9}$$

where, following Bayes' theorem, $p(\beta;\alpha)$ is the prior over piece-level variables and $p(t|\beta) \equiv p(t|c,w,\lambda)$ is the marginal likelihood for a single tone $t$ in the piece $d$ with the latent variables ($\gamma \equiv \tau^0, \ldots, \tau^n, n$) being marginalized out. Since in our model all tones within a piece are drawn independently and identically distributed, the distribution $p(t|c,w,\lambda)$ needs to be computed only once per piece.

$$p(t\,|\,c,w,\lambda) = \sum_{n=0}^{\infty} p(n\,|\,\lambda) \underbrace{\sum_{\tau^{n-1}} p(\tau^{n}\,|\,\tau^{n-1},w) \overbrace{\sum_{\tau^{n-2}} p(\tau^{n-1}\,|\,\tau^{n-2},w) \cdots \sum_{\tau^{0}} p(\tau^{1}\,|\,\tau^{0},w)\, p(\tau^{0}\,|\,c)}^{p(\tau^{n-1}|\,c,w)}}_{p(\tau^{n}|\,c,w)}, \tag{10}$$

The expansion of the marginal likelihood $p(t|c,w,\lambda)$ is shown in Equation (10), where $p(n|\lambda)$ is the distribution over path lengths, $p(\tau^i|\tau^{i-1},w)$ are the transition probabilities in the latent cognitive process, and $p(\tau^0|c)$ is the initial distribution corresponding to possible tonal origins in the piece. To compute $p(t|c,w,\lambda)$, we have to explicitly marginalize out the latent variables $(\gamma \equiv \tau^0,\ldots,\tau^n,n)$.

The marginal likelihood $p(t|c,w,\lambda)$ has a recursive structure, which becomes apparent in Equation (10) by highlighting the intermediate terms $p(\tau^i|c,w)$, where the subset of latent variables $\tau^0,\ldots,\tau^{i-1}$ has already been marginalized out: the latent cognitive process is a Markov chain in the tonal space $\mathcal{T}$ and the marginal distribution $p(\tau^i|c,w)$ at step $i$ of the process can be recursively computed via dynamic programming. The path length $n$ can be marginalized out by interleaving the dynamic programming updates with weighted increments to the final distribution $p(t|c,w,\lambda)$, resulting in the procedure shown in **Algorithm 1**. Being able to compute the marginal likelihood $p(t|c,w,\lambda)$ allows to optimize the piece-level variables $c$, $w$ and $\lambda$ to find their MAP estimate.

For our evaluation, we make three additional assumptions that are specific to tonal music and well-established in music theory (see Section 3): 1) we assume that only a single tonal origin per piece exists and that all tones are a priori equally likely to become the tonal origin; 2) we restrict the number of allowed interval steps to the six primary intervals present in the Tonnetz; 3) we assume a uniform prior over the path-length variable $\lambda$. These assumptions facilitate inference (see Appendix D for technical details) and imply that computing a MAP estimate becomes equivalent to a *maximum likelihood* (ML) estimate.

We infer values for the piece-level variables $\beta$ by maximizing their posterior probability or (equivalently, due to using uniform priors) minimizing the negative data log-likelihood, cross-entropy, or *Kullback-Leibler divergence* (KLD). The model was implemented in PyTorch (Paszke et al., 2019) and optimization was done in two nested procedures: 1) The values for $w$ and $\lambda$ were optimized via gradient decent using the Adam optimizer (Kingma and Ba, 2015). 2) For specific values of the weight and path-length variables, $w$ and $\lambda$, we chose the tonal origin $c$ that minimizes the KLD (gradients are propagated through this step using PyTorch's automatic differentiation functionality).

**Algorithm 1** Computing the marginal likelihood $p(t|c,w,\lambda)$ by explicitly marginalizing out the latent variables $\gamma \equiv (\tau^0,\ldots,\tau^n,n)$. In practice, the infinite loop over $n$ is terminated by summing over $p(n|\lambda)$ and stopping when this cumulative probability approaches 1 (with tolerance $10^{-5}$).

**Input: Output:** $c,w,\lambda$

**Output:** $p(t|c,w,\lambda)$

1: $\mathbf{u} \leftarrow \mathbf{0}$ *# initialize output distribution*
2: $\mathbf{v} \leftarrow p(\tau^0|c)$ *# initialize intermediate distribution*
3: $\mathbf{M} \leftarrow p(\tau^i|\tau,w)$ *# initialize transition matrix*
4: **for** $n \in \{0,1,\ldots\}$ **do**
5: $\quad\mathbf{u} \leftarrow \mathbf{u} + p(n|\lambda) \cdot \mathbf{v}$ *# update, marginalizing out n*
6: $\quad\mathbf{v} \leftarrow \mathbf{M}\mathbf{v}$ *# transition, marginalizing out* $\tau^i$
7: **end for**
8: **return u**

## 5. Results and Discussion

We evaluate the TDM in two ways: first, we introduce several baseline models (Section 5.1) and perform a quantitative comparison (Section 5.2) on a corpus of 248 pieces from different historical epochs. Second, we perform a detailed qualitative analysis (Section 5.3) on three exemplary pieces, inspecting the inferred parameters and discussing our musical interpretation of these results.[1]

### *5.1 Baseline Models*

We introduce several baseline models to verify whether the structural assumptions incorporated in our model effectively improve performance. In particular, we are interested in validating the impact of the Tonnetz topology and the assumed latent process on model performance.

The two static baseline models, *Static (1 Profile)* and *Static (2 Profiles)*, do not incorporate any topological information except for transposition invariance. These models also lack the semantic interpretability aimed for with the TDM and most closely resemble the categorical representation of PCDs, as used in template-based key finders or the model of Hu and Saul (2009b). The line-of-fifths model, *TDM (Binomial, 1D)*, is a reduced version of the TDM that uses the line of fifths but not the full Tonnetz topology. To evaluate the influence of the path-length distribution we also include *TDM (Poisson)*, a version of the full *TDM* with a Poisson path-length distribution, in addition to the best-performing *TDM (Binomial)* with a binomial path-length distribution.

#### 5.1.1 Static Model (1 Profile)

The static baseline model consists of a single global PCD, that is, a fixed template or profile. The model has an individual parameter for each tonal pitch class; together these parameters determine the shape of the profile and they are trained via gradient descent on the entire corpus.

The profile is individually matched to each piece (during training and for evaluation) by shifting it along the line of fifths. The model thus has a large number of continuous corpus-level parameters (one for each tonal pitch class) but only a single discrete piece-specific variable (the transposition). The optimal profile corresponds to the purple line in **Figures 6a–6c**.

#### 5.1.2 Static Model (2 Profiles)

This model is identical to the static model described above but comprises two different static PCDs. Matching with a specific piece is done by choosing the best profile *and* transposition. It thus has two discrete piece-specific variables, the transposition (as above) and the binary profile class. This model is conceptually very similar to conventional template-based keyfinding algorithms with two profiles, which are usually assumed to correspond to the major and minor modes – a questionable assumption that will be discussed below. The optimal profiles correspond to the blue line in **Figure 6a** (first profile) and **Figure 6b** and **6c** (second profile).

#### 5.1.3 Line-of-Fifths TDM (Binomial, 1D)

To specifically test the impact of the Tonnetz topology, we introduce a reduced version of the TDM that uses only the line-of-fifths topology. This model is identical to the full TDM (with binomial path-length distribution), except that it only uses fifth steps (no major or minor thirds). The model has three piece-specific variables (one for weighting +P5 against –P5 and two for the binomial distribution) and produces bell-shaped PCDs on the line of fifths, including Gaussian and skewed bell shapes.

### *5.2 Corpus Evaluation*

For the quantitative evaluation of the TDM, we used a corpus of 248 pieces that are representative for the Baroque, Classical, and Romantic era: all preludes and fugues from Bach's *Well-Tempered Clavier* Vol. I & II (96 pieces), all movements from Beethoven's piano sonatas (101 pieces), and 51 pieces by Liszt, including etudes, pieces from the *Années de Pèlerinage*, and the B-minor Sonata.

All models were trained on the entire corpus to minimize the *Kullback-Leibler divergence* (KLD) between the model and empirical PCD (or, equivalently, minimize the cross-entropy, or maximize the data likelihood). To train the two static models with corpus-level parameters, the three composers were weighted equally: each piece inversely proportional to the total number of pieces of the respective composer. The results are shown in **Figure 5**.

As expected, the static model with a single global PCD performs worst. The inferred profile can be seen in the detailed analysis of the single pieces in **Figure 6**. It is bimodal and can be understood as a combined major-minor profile with additional long tails along the line of fifths. For pieces in major mode, the lower mode coincides with the tonic and dominant pitch class, while the upper mode covers the major third. For pieces in minor mode the reverse is true, with the upper mode corresponding to the tonic and dominant, while the lower mode covers the minor third. This single profile can be interpreted as the best attempt to match all pieces in the corpus to a single profile.

Also not surprisingly, the static model with two profiles performs considerably better. Notably, for the Bach pieces this model performs as well as the TDM (Poisson)

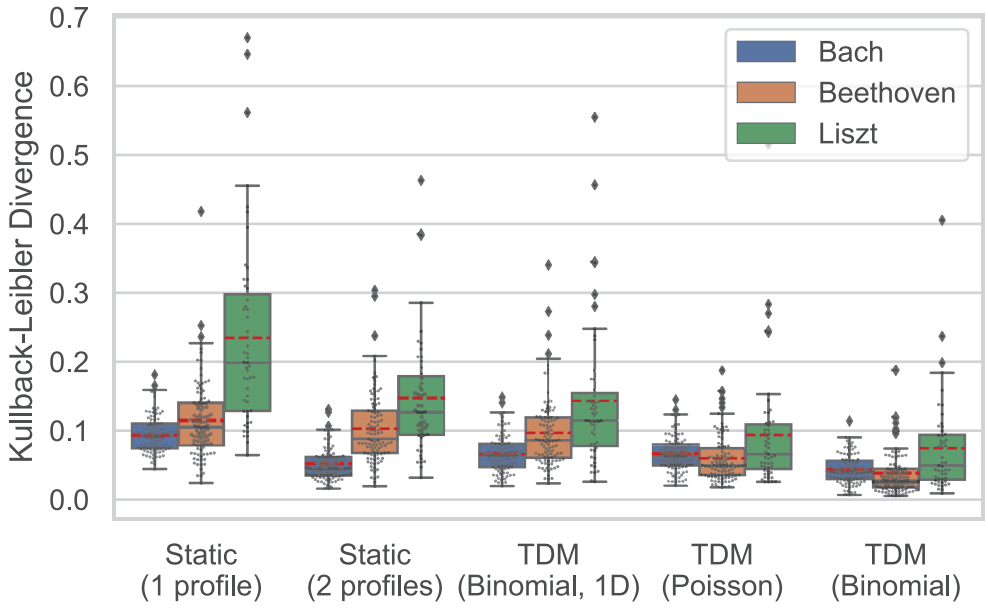



**Figure 5:** Comparison of model performance for composers from different historical epochs: Bach (Baroque), Beethoven (Classical), Liszt (Romantic). The box plots indicate quartiles (25%, 50%, 75%), whiskers extend 1.5 times the inter-quartile range (IQR). The mean is indicated as a dashed red line. Single pieces are shown as black dots in a swarm plot, outliers as larger diamonds.

and almost as well as the TDM (Binomial). Presumably, the main reason for the strong performance of the static model on Bach's pieces is the fact these pieces are typically confined to a relatively well-defined range on the line of fifths, after which the PCD quickly drops to zero. This shape does not correspond to the smooth decay assumed in the TDM, which is more typically observed in the pieces by Beethoven. In contrast to the relatively good performance for Bach's pieces, for Beethoven and Liszt the static two-profile model performs even worse than the reduced line-of-fifth TDM (Binomial, 1D).

Inspecting the two inferred profiles in **Figure 6** reveals an interesting property: instead of learning a major and a minor profile (as one might have expected), one of the profiles is almost identical to the compound major-minor profile of the single-profile model (just with shorter tails), while the second profile has three modes that form an augmented triad. The number of pieces associated to the respective profiles are 96/0 (Bach), 88/13 (Beethoven), 8/43 (Liszt). This suggests that reducing the tonality of the Classical and Romantic era to only a major and a minor mode might not always be appropriate. The profiles could rather be interpreted as a 'conventional diatonic' profile and an 'extended tonality' profile. In fact, when trained on only the Bach pieces (results not included), the static two-profile model learns a major and minor profile, outperforms the TDM, and displays a key classification accuracy of 97%. Our results thus question established key classifiers by showing that their accuracy is very high only in a relatively narrow musical repertoire. In conclusion, the static models strongly overfit on the corpus being used, which may or may not be acceptable, depending on the application. In contrast, note that the TDM has more piece-specific degrees of freedom than the static model but no corpus-level parameters; it is thus immune to this kind of overfitting on the corpus level.

Finally, we compare the performance of the different versions of the TDM. For the line-of-fifths based TDM (Binomial, 1D), we observe a decreasing performance from Bach to Beethoven and Liszt. This corresponds to the music-theoretic insight that Bach's harmony is predominantly fifths-based, while Beethoven incorporates an increasing amount of third-based harmonic progressions, which is yet again extended by Liszt. Beethoven's pieces therefore tend to be multimodal on the line of fifths – but not on the Tonnetz. In contrast, some of Liszt's pieces are multimodal even on the Tonnetz and tend to be fragmented on the line of fifths.

This is also strongly reflected in the different performances of the full TDMs, which both show the best results for Beethoven. However, the reason for the decrease in performance for Bach and Liszt as compared to Beethoven are presumably different.

As mentioned above, Bach's pieces tend to have PCDs with a relatively sharp decay, which conflicts with the smooth decay of the TDM, more commonly found in Beethoven's pieces. On the other hand, the mentioned fact that (due to the extended tonality) Liszt's pieces tend to be multimodal even on the Tonnetz, means that, even though they have a smooth decay, they may not be well captured by the TDM. A slight modification of the TDM to allow for multiple tonal origins, might also allow to model this kind of extended tonality appropriately.

### *5.3 Case Studies*

We chose three exemplary pieces with the goal to evaluate whether the TDM is able to capture characteristics of the respective piece and historical period. We used the same three pieces for which the empirical PCDs are shown in **Figure 3**, that is,

1. Baroque: Johann Sebastian Bach, Prelude in C major, BWV 846 (1722),
2. Classical: Ludwig van Beethoven, Sonata op. 31, no. 2 in D minor 'Tempest', 1st mov. (1802),
3. Late-Romantic: Franz Liszt, *Bénédiction de Dieu dans la Solitude*, S. 200/1 (1882).

As described above, the piece-level variables were determined via MAP/ML estimation. The results are shown in **Figure 6** with the values in brackets indicating the KLD between the empirical and model distributions.

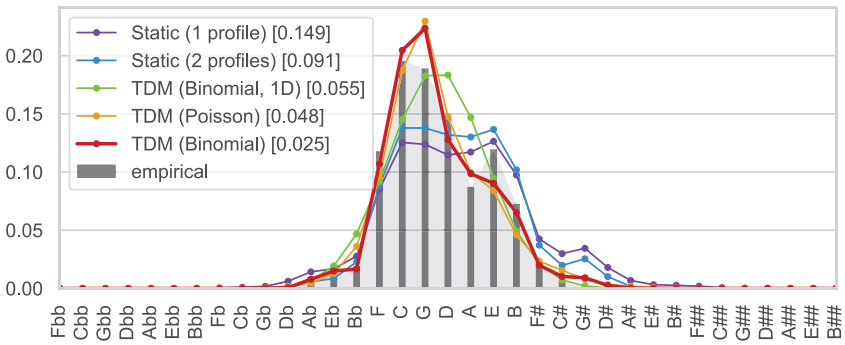
For each of the three example pieces, the MAP estimates for the parameters of the binomial TDM are shown in **Figure 7**. The parameters $\mu$ and $\sigma$, given in the caption, are the mean and standard deviation of the path-length distribution. The probabilities $p_i$ for choosing the different intervals in a particular step are indicated in the box at the bottom right. The tonal origin is displayed in the center of each plot and the lengths of the arrows (also given in their labels) indicate the expected numbers of steps $\mu p_i$ in the six primary interval directions.

#### 5.3.1 Bach: C major Prelude

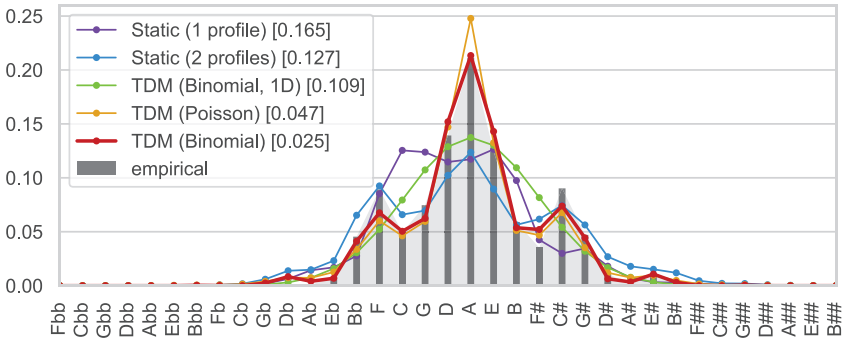
For Bach's prelude (**Figures 3a, 6a** and **7a**), the model was able to capture the PCD mainly extending along the line of fifths as reflected in the high weights for both perfect fifth intervals +P5 and –P5. The stronger weight for the descending fifth ($p_{-\mathrm{P5}} = 0.475$) reflects that the tonal origin G lies a fifth above the most frequent tone and global tonic C. The strong descending fifth is balanced by the combination of the ascending fifth ($p_{+\mathrm{P5}} = 0.288$) and the descending minor third that also ascends along the line of fifths ($p_{-\mathrm{m3}} = 0.137$). Interestingly, explaining the PCD using the global tonic C as the tonal origin (results not shown) is only marginally less accurate with correspondingly changed weights (stronger ascending fifth and a strong ascending major third instead of the descending minor third). Given the lack of temporal information, this ambiguity is musically consistent and reflects the harmonically close relation of G and C (also see Section 3.4). The general importance of the line of fifths for the organization of the tonal material is characteristic of Baroque pieces.

#### 5.3.2 Beethoven: 'Tempest' Sonata

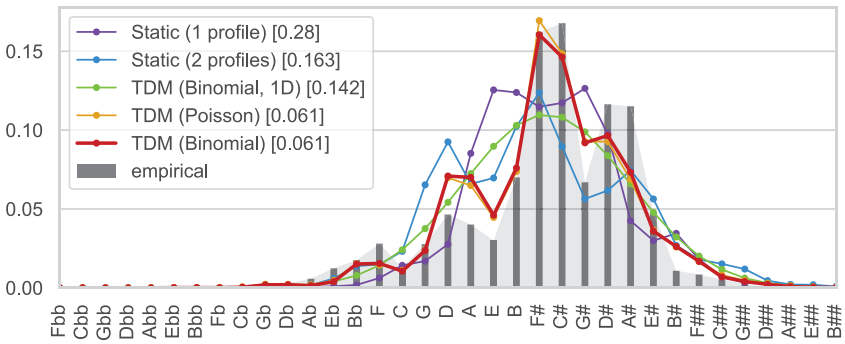
The optimal parameters for Beethoven's Sonata movement (**Figures 3b, 6b** and **7b**) also prioritize the two perfect fifth directions, almost in a symmetrical manner ($p_{+\mathrm{P5}} = 0.331$, $p_{-\mathrm{P5}} = 0.356$). However, as opposed to Bach's piece, a significant amount of the probability




**(a)** Bach, C major Prelude.

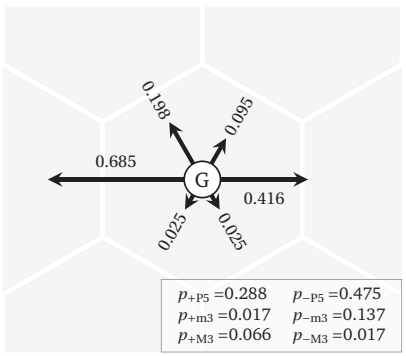



**(b)** Beethoven, 'Tempest' Sonata.

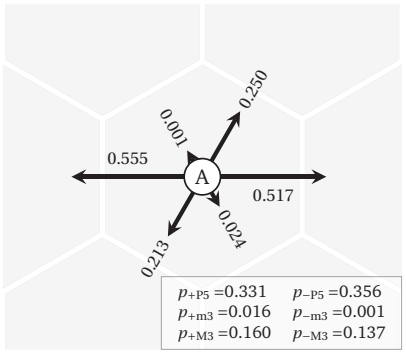



**(c)** Liszt, *Bénédiction de Dieu dans la Solitude*.

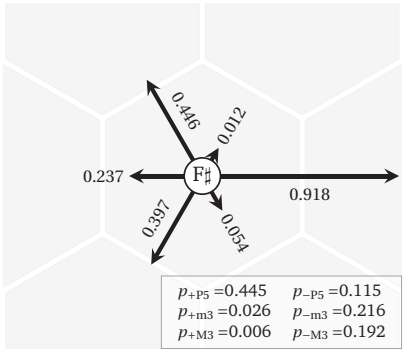
**Figure 6:** Comparison of the empirical PCD (gray bars with shaded background) with different models (colored plots). The corresponding Kullback-Leibler divergence is indicated in square brackets after the model.




**(a)** Bach, C major Prelude ($\mu = 1.44, \sigma = 0.69$).




**(b)** Beethoven, 'Tempest' Sonata ($\mu = 1.56, \sigma = 0.76$).




**(c)** Liszt, *Bénédiction de Dieu dans la Solitude* ($\mu = 2.06, \sigma = 1.25$).

**Figure 7:** Optimal parameters for the TDM (Binomial) model (see text for details).

mass (≈30%) is assigned to the two major third components ($p_{+M3} = 0.160$, $p_{-M3} = 0.137$), again essentially symmetric. This is typical for pieces in the minor mode due to the above-mentioned overall prominence of the (minor) tonic and (major) dominant triads. However, it can also be interpreted as reflecting the stylistic changes in the Classical period that allow for broader ranges of mediantic (i.e. third-based) local key relations, as can also be observed in this movement. The approximate point symmetry of the empirical distribution around the pitch class A and along the three axes of the Tonnetz (see **Figure 3b**) is reflected in similar weights of the ascending and descending intervals along the same axis, as shown in **Figure 6b**. Note that A, the point of symmetry and the model's choice for the tonal origin, in fact is the fifth of the tonic D and the root of the dominant triad. Again, this exemplifies that the tonal origin chosen by the TDM does not need to be the tonic but rather incorporates a more general notion of intervallic relations.

#### 5.3.3 *Liszt*, Bénédiction de Dieu dans la Solitude

The most diverse distribution of pitch classes is found in Liszt's piano piece (**Figures 3c, 6c** and **7c**). The empirical distribution in **Figure 6c** is clearly multimodal on the line of fifths around F♯/C♯, D♯/A♯, and D/A with another smaller mode around F/B♭, likewise pointing towards mediantic relations as in Beethoven's piece. But unlike the Sonata, the distribution is asymmetric. The extended harmonic relations in Liszt's piece are reflected in the model outcome,

which assigns non-zero probability mass to all six primary intervals, with an acceptable overall fit of the estimated distribution (KLD=0.061) as compared to the two previous pieces and the range of KLD values in **Figure 5**.

The model was able to capture a number of important characteristics of this piece. The harmonic structure of the entire piece is fundamentally governed by major third relations. Its three sections (*Moderato*, *Andante*, and *Più sostenuto, quasi Preludio*) stand in the keys F♯ major, D major, and B♭ major, respectively. The major-third relation between those keys also permeates each of the sections on more local levels (e.g. D major and B♭ major passages in the F♯ major sections) and thus governs the harmonic structure of the piece on several hierarchical levels.

Since each of the sections is largely diatonic with some ornamental chromaticism, it is not surprising that also for this Romantic piece the perfect fifths together account for more than 50% of the overall weights, followed by the descending minor and major thirds (0.216 and 0.192, respectively). This entails, for example, that the upper major third A♯ of the frequently used F♯ major triad is largely explained by the combination of an ascending fifth and a descending minor third, while D and B♭ are more directly explained as descending major third steps. In particular, the difference of the model explanations between B♭ and A♯ is meaningful since these tonal pitch classes bear different harmonic implications, which would be lost in a neutral pitch-class representation.

## 6. Conclusion and FutureWork

We presented the *Tonal Diffusion Model* (TDM), a generative probabilistic model for *pitch-class distributions* (PCDs) that incorporates relevant music-theoretic and cognitive insights. As opposed to most existing models for PCDs, the TDM incorporates algebraic and geometric structures from music theory and describes the generation of tones in a piece as a latent cognitive process. It thereby relates the statistical properties of PCDs to musically meaningful and interpretable variables.

The model was evaluated quantitatively on a corpus of 248 pieces showing superior performance to traditional models. Comparing against several baseline models, the positive impact of incorporating the Tonnetz structure was demonstrated. Furthermore, a detailed analysis of three exemplary pieces showed that the TDM is able to capture characteristic properties of these pieces and the respective period.

The TDM is well-suited to study a range of relevant questions in digital musicology and MIR. It may be extended and adapted in multiple ways. For instance, it can be adapted to incorporate more specific assumptions about the underlying cognitive processes and the relevant musical style and it allows for corpus-based studies to investigate historical developments and stylistic differences among a large number of musical pieces. In MIR, the piece-level variables (tonal origin, interval weights, and path-length distribution) can be conceived as a tonal fingerprint of the piece, going beyond the notion of a pitch profile, and thus allowing to determine its tonality in a novel way. The model can be extended to include a larger (or infinite) set of intervals $\mathscr{I}$ by employing a full Dirichlet process prior and the modeled cognitive process can be adapted by using independent path-length distributions for the different intervals. The tonal space can be augmented to include tuning differences and take into account different interpretations for different generation paths. And finally, the model may be generalized to include a time component that takes sequential and syntactic dependencies in musical pieces into account.

The TDM thus provides a novel approach to modeling PCDs in a compact and musically interpretable way, while outperforming existing approaches in terms of accuracy. It may thereby serve as a broad foundation for further developing generative models of PCDs in music and opens up multiple highly promising directions for future research.

## Note

[1] The data and code to reproduce our results can be found at https://github.com/DCMLab/tonal-diffusion-model.

## Additional File

The additional file for this article can be found as follows:

- **Supplementary Material.** Appendix. DOI: https://doi.org/10.5334/tismir.46.s1

## Acknowledgements

This project was partially funded through the Swiss National Science Foundation within the project "Distant Listening – The Development of Harmony over Three Centuries (1700–2000)". Also, this project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 760081 – PMSB. Martin Rohrmeier acknowledges the kind support by Mr Claude Latour through the Latour chair of Digital Musicology at EPFL.

## Competing Interests

The authors have no competing interests to declare.

## References

**Aldwell, E., Schachter, C.,** & **Cadwallader, A.** (2010). *Harmony and Voice Leading*. Cengage Learning, 4th edition.

**Bigo, L.,** & **Andreatta, M.** (2016). Topological Structures in Computer-Aided Music Analysis. In Meredith, D., editor, *Computational Music Analysis*, pages 57–80. Springer, Berlin. DOI: https://doi.org/10.1007/978-3-319-25931-4_3

**Blei, D. M., Ng, A. Y.,** & **Jordan, M. I.** (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

**Chew, E.** (2000). *Towards a Mathematical Model of Tonality*. Doctoral dissertation, Massachussets Institute of Technology, Cambridge, MA.

**Clough, J.,** & **Myerson, G.** (1985). Variety and multiplicity in diatonic systems. *Journal of Music Theory*, *29*(2), 249–270. DOI: https://doi.org/10.2307/843615

**Cohn, R.** (1997). Neo-Riemannian operations, parsimonious trichords, and their "Tonnetz" representations. *Journal of Music Theory*, *41*(1), 1–66. DOI: https://doi.org/10.2307/843761

**Cohn, R.** (2012). *Audacious Euphony: Chromatic Harmony and the Triad's Second Nature*. Oxford University Press, Oxford. DOI: https://doi.org/10.1093/acprof:oso/9780199772698.001.0001

**Douthett, J.,** & **Steinbach, P.** (1998). Parsimonious graphs: A study in parsimony, contextual transformations and modes of limited transposition. *Journal of Music Theory*, *42*(2), 241–263. DOI: https://doi.org/10.2307/843877

**Euler, L.** (1739). *Tentamen novae theoriae musicae ex certissimis harmoniae principiis dilucide expositae*. Ex Typographia Academiae Scientiarum, St. Petersburg.

**Fiore, T. M.,** & **Noll, T.** (2011). Commuting groups and the topos of triads. In Agon, C., Amiot, E., Andreatta, M., Assayag, G., Bresson, J., & Manderau, J., editors, *Mathematics and Computation in Music*, volume 6726 of *Lecture Notes in Artificial Intelligence*. Springer, Berlin. DOI: https://doi.org/10.1007/978-3-642-21590-2_6

**Gárdonyi, Z.,** & **Nordhoff, H.** (2002). *Harmonik*. Möseler Verlag, Wolfenbüttel.

**Griffiths, T., Steyvers, M., Blei, D., & Tenenbaum, J.** (2005). Integrating topics and syntax. *Advances in Neural Information Processing Systems*, *17*, 537–544.

**Haas, B.** (2004). *Die neue Tonalität von Schubert bis Webern: Hören und Analysieren nach Albert Simon*. Florian Noetzel, Wilhelmshaven.

**Harasim, D., Noll, T.,** & **Rohrmeier, M.** (2019). Distant neighbors and interscalar contiguities. In Montiel, M., Gomez-Martin, F., & Agustín-Aquino, O. A., editors, *Mathematics and Computation in Music*, Lecture Notes in Computer Science, pages 172–184. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-21392-3_14

**Harasim, D., Schmidt, S. E.,** & **Rohrmeier, M.** (2016). Bridging scale theory and geometrical approaches to harmony: The voice-leading duality between complementary chords. *Journal of Mathematics and Music*, *10*(3), 193–209. DOI: https://doi.org/10.1080/17459737.2016.1216186

**Hostinský, O.** (1879). *Die Lehre von den musikalischen Klängen: Ein Beitrag zur aesthetischen Begründung der Harmonielehre*. H. Dominicus, Prague.

**Hu, D. J.,** & **Saul, L. K.** (2009a). A probabilistic topic model for music analysis. In *22nd Conference on Neural Information Processing Systems, Workshop on Applications for Topic Models: Text and Beyond*.

**Hu, D. J.,** & **Saul, L. K.** (2009b). A probabilistic topic model for unsupervised learning of musical keyprofiles. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 441–446.

**Huron, D.,** & **Veltman, J.** (2006). A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system. *Empirical Musicology Review*, 1(1). DOI: https://doi.org/10.18061/1811/24072

**Kingma, D. P.,** & **Ba, J.** (2015). Adam: A method for stochastic optimization. In Bengio, Y., & LeCun, Y., editors, *3rd International Conference on Learning Representations*.

**Koelsch, S., Rohrmeier, M., Torrecuso, R.,** & **Jentschke, S.** (2013). Processing of hierarchical syntactic structure in music. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15443–8. DOI: https://doi.org/10.1073/pnas.1300272110

**Krumhansl, C. L.** (1990). *Cognitive Foundations of Musical Pitch*. Oxford University Press, New York.

**Krumhansl, C. L.** (1998). Perceived triad distance: Evidence supporting the psychological reality of neo-Riemannian transformations. *Journal of Music Theory*, *42*(2), 265–281. DOI: https://doi.org/10.2307/843878

**Krumhansl, C. L.,** & **Kessler, E. J.** (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334–368. DOI: https://doi.org/10.1037/0033-295X.89.4.334

**Lerdahl, F.,** & **Jackendoff, R. S.** (1983). *A Generative Theory of Tonal Music*. MIT Press, Cambridge, MA.

**Milne, A. J.,** & **Holland, S.** (2016). Empirically testing Tonnetz, voice-leading, and spectral models of perceived triadic distance. *Journal of Mathematics and Music*, *10*(1), 59–85. DOI: https://doi.org/10.1080/17459737.2016.1152517

**Minka, T.,** & **Winn, J.** (2009). Gates. In *Advances in Neural Information Processing Systems*, pages 1073–1080.

**Moss, F. C.** (2019). *Transitions of Tonality: A Model-Based Corpus Study*. Doctoral dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

**Moss, F. C., Loayza, T.,** & **Rohrmeier, M.** (2019). pitchplots. DOI: https://doi.org/10.5281/zenodo.3265393

**Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J.,** & **Chintala, S.** (2019). PyTorch: An imperative style, highperformance deep learning library. In Wallach, H., Larochelle, H., Beygelzimer, A., dAlché-Buc, F., Fox, E., & Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

**Riemann, H.** (1896). *Dictionary of Music*. Augener, London.

**Rohrmeier, M.** (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, *5*(1), 35–53. DOI: https://doi.org/10.1080/17459737.2011.573676

**Rohrmeier, M.** (2020). The syntax of jazz harmony: Diatonic tonality, phrase structure, and form. *Music Theory and Analysis (MTA)*, *7*(1), 1–63. DOI: https://doi.org/10.11116/MTA.7.1.1

**Schenker, H.** (1935). *Der freie Satz*. Universal Edition, Wien.

**Schoenberg, A.** (1969). *Structural Functions of Harmony*. Norton, New York.

**Selfridge-Field, E.,** editor (1997). *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, Cambridge, MA.

**Steyvers, M.,** & **Griffiths, T.** (2007). Probabilistic topic models. In Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W., editors, *Handbook of Latent Semantic Analysis*, pages 424–440. Lawrence Erlbaum Associates, Mahwah, NJ.

**Temperley, D.** (2000). The line of fifths. *Music Analysis*, *19*(3), 289–319. DOI: https://doi.org/10.1111/1468-2249.00122

**Toiviainen, P.,** & **Krumhansl, C. L.** (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*(6), 741–766. DOI: https://doi.org/10.1068/p3312

**Tymoczko, D.** (2011). *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice*. Oxford University Press, Oxford.

**von Oettingen, A.** (1866). *Harmoniesystem in dualer Entwicklung*. W. Gläser, Dorpat und Leipzig.

**Weber, G.** (1851). *The Theory of Musical Composition, Treated with a View to a Naturally Consecutive Arrangement of Topics*. Messrs Robert Cocks and Co., London.

**How to cite this article:** Lieck, R., Moss, F. C., & Rohrmeier, M. (2020). The Tonal Diffusion Model. *Transactions of the International Society for Music Information Retrieval,* 3(1), pp. 153–164. DOI: https://doi.org/10.5334/tismir.46

**Submitted:** 11 January 2020 **Accepted:** 28 August 2020 **Published:** 16 October 2020



]u[ *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS

Transactions of the International Society for Music Information Retrieval

Lieck, R., Moss, F. C., and Rohrmeier, M. (2020). The Tonal Diffusion Model (Appendix), *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 153–164, DOI: https://doi.org/10.5334/tismir.46

# The Tonal Diffusion Model (Appendix)

Robert Lieck, Fabian C. Moss, and Martin Rohrmeier

## A. Representations of Pitch-Class Distributions

Many large corpora are available in the MIDI format (Huang et al., 2017; Madsen et al., 2007; Rohrmeier and Cross, 2008; White, 2014), possibly inferred from audio (Serrà et al., 2012; Weiß et al., 2018). A major drawback of the MIDI representation is the assumption of twelve equivalence classes, which do not allow for the distinction between enharmonically equivalent pitch classes (Selfridge-Field, 1997). As a consequence, using the MIDI representation implies that many relevant musical relations are obfuscated, such as the difference between enharmonically equivalent notes in dominant seventh and German sixth chords, for instance, (A♭, C, E♭, G♭) versus (A♭, C, E♭, F♯). Algorithms for *pitch spelling* (Bora et al., 2018; Cambouropoulos, 2003; Chew and Chen, 2005; Meredith, 2006; Stoddard et al., 2004) attempt to infer this missing information from the musical context and can to some degree be employed to mitigate this problem. Recent research has also produced datasets in which the tonal spelling of pitch classes is retained, for instance, in the **kern, MusicXML, or MEI formats, which are increasingly being made available online (Freedman, 2014; Fujinaga et al., 2014; Sapp, 2005).

As the large datasets in MIDI format are widely used in the MIR community, this simplified representation has largely carried over to the representation of PCDs. In musical corpus studies, PCDs are often obtained by simply counting the relative frequencies of notes (Temperley and Marvin, 2008; Albrecht and Huron, 2014) resulting in categorical distributions with one separate and independent parameter for each pitch class. They are visualized in a linear chromatic arrangement. This implies three shortcomings: 1) The visual representation in a linear arrangement suggests some ordering and proximity relation (and thus an implicit dependency between the pitch classes) that is not reflected in the categorical distribution. 2) The commonly used chromatically ascending order does not reflect tonal relations well, especially with respect to harmony and key. An arrangement along the line or circle of fifths would be better suited for these relations. 3) The space of pitch classes exhibits an inherent cyclic topology (whether arranged chromatically or along the circle of fifths), which is not reflected in a linear arrangement.

A more recent approach from mathematical music theory is the application of the Discrete Fourier Transform (DFT) to PCDs, which takes the cyclical nature of pitch-class space into account (e.g. Amiot, 2016; Noll, 2019; Quinn, 2006, 2007; Yust, 2019). The Fourier representation of PCDs makes use of the implicit ordering and topology and has some interpretatory value as different coefficients can be associated to different musical properties (e.g. the third coefficient reflects triadicity and the fifth coefficient diatonicity). While in some cases the Fourier representation may yield a somewhat better interpretability than the categorical representation, it still has two drawbacks: 1) Inherent to the Fourier representation is the assumption of a circular pitch-class space, which implies the same shortcomings in expressing musical relations as for the MIDI representation. 2) The Fourier representation is still inflexible when it comes to incorporating additional music-theoretic and cognitive knowledge, making model improvements impossible.

## B. Comparison to Hu and Saul

The only music-specific application of topic models we are aware of is by Hu and Saul (2009). However, while sharing the general topic model structure with our TDM, they target a very different goal and there are several profound differences, which make a direct comparison difficult.

The main difference is that Hu and Saul address the problem of key finding under the standard assumption of two distinct modes (major and minor) with 12 possible values for the tonic (all neutral pitch classes in 12TET). This is contrary to the purpose of our model in two ways. First, we aim to provide a more fine-grained representation of tonality – specifically, one that goes beyond the coarse classification into two discrete modes – by using weighted directions on the Tonnetz. Second, we argue that this should be achieved by providing a more structured and interpretable representation of the PCDs (the weights in our model), while Hu and Saul learn one fixed, unstructured distribution for each mode.

Another difference is that Hu and Saul add one additional layer to the standard topic modeling architecture, resulting in three layers (piece, section, note). As a result, they can infer key labels for different sections of each piece, which provides more detailed information. However, this does not necessarily improve accessibility because these separate key labels still require interpretation by a domain expert and do not compactly represent the global harmonic character of the entire piece (besides again adhering to the simple binary classification).

Finally, given their interest in traditional key finding, Hu and Saul compare their model's accuracy to established template-based key finders. In contrast, as our interest lies in capturing the fine-grained structure of PCDs, we use the KLD to the empirical distribution

(or equivalently the cross-entropy or data likelihood) as a performance measure in our evaluation. Consequently, a direct comparison to Hu and Saul's results is hardly possible. We do, however, include a baseline model that is similar to their model in that it learns two key profiles for the entire corpus. Our evaluation shows that its performance is acceptable on Baroque pieces but strongly decays for pieces from the Classical and Romantic era. Furthermore, the learned profiles indicate that the pervasive assumption of a piece being either in major or minor mode is a coarse simplification, which is justified only in a narrow stylistic range.

## C. The Tonal Origin

There are two important caveats to keep in mind for interpreting the tonal origin. First, assuming a single tonal origin does *not* mean that the piece cannot modulate between different keys having their own respective local centers. Rather, we assume that one pitch class exists, which governs the entire piece and can serve to explain the occurrence of all other pitch classes, among others by taking possible modulations into account. While local modulations may affect the perception of a tonal center (Cuddy and Thompson, 1992; Farbood, 2016), the assumption of a unique global key is generally reasonable for pieces up to the late Romantic epoch, which tend to have a well-defined and unique global key. In fact, our general model definition (see Section 4) also allows for more than one tonal origin and applying the TDM to a corresponding corpus of pieces is highly interesting future work.

The second important caveat is that the tonal origin in our model is *not* necessarily equivalent to the tonic of the global key. While in the hierarchical dependency structure of harmonic relations, the tonic of the global key is the structurally dominating and most stable center, this does not directly translate into the statistical properties of PCDs. The most important confounding factor in this respect is that only the relative frequencies of pitch classes are preserved but not the information on their temporal order. For instance, in a piece (say in the key of C) with a short sub-dominant section (in F major) and a short dominant section (in G major), the global tonic (pitch class C) will generally be the statistically most salient pitch class. However, the PCD of another piece in the key of F major with a short pre-dominant section on the second scale degree (G), and an extended dominant section (C) may look exactly the same. In the second piece, the most salient pitch class will again be C, which here is the tonic of the dominant. This example demonstrates that the global tonic cannot generally be identified based on the PCD alone. Instead, F would be identified as the global tonic of the second piece based on temporal information, such as the piece starting and ending in that key. Some approaches to algorithmic key finding therefore take only the beginnings or endings of pieces into account (e.g. Albrecht and Shanahan, 2013), effectively ignoring most of the tonal material. But precisely this temporal information is not retained in the overall PCDs and we thus cannot expect a model of PCDs to directly relate to these concepts.

The tonal origin in our model therefore has to be conceived as a more general statistical concept. Even though the global tonic is a good candidate in many cases, other pitch classes (especially those that are closely related to the global tonic on the Tonnetz) may be better suited to explain a given PCD.

## D. Specific Assumptions for the Evaluation

For our evaluation, we make several assumptions that are specific to tonal music and well-established in music theory (see Section 3).

Our first assumption is that only a single tonal origin per piece exists and that all tones are a priori equally likely to become the tonal origin. As discussed above (see Section 3.4), this does *not* mean that the piece cannot modulate between different keys. Mathematically, this corresponds to the base distribution $H_c$ being uniform over the tonal space $\mathscr{T}$ and the concentration parameter $\alpha_c$ being equal to zero. As a result, optimizing over $c$ corresponds to finding the single best-matching tonal origin for the given piece. This optimization is further facilitated by the fact that transitions are defined in terms of intervals, which means that shifting the tonal origin simply produces a shifted version of the distribution with the same shape and thus does not require recomputing the distribution via dynamic programming.

Our second assumption consists in restricting the number of allowed intervals (i.e. the number of possible transitions in the cognitive process) to a finite set and assuming a uniform prior over their weights $w$. The Dirichlet process is thus replaced by its finite-dimensional equivalent, a Dirichlet distribution, with all concentration parameters equal to one. Specifically, we use the six primary intervals present in the Tonnetz (see Figure 2) so that the transition probability $p(\tau' \mid \tau, w)$ can be written as

$$p(\tau' \mid \tau, w) = \begin{cases} p_i & \text{if } \tau' - \tau = i \\ 0 & \text{else} \end{cases} \tag{11}$$

where $\tau, \tau' \in \mathscr{T}$ are TPCs on the Tonnetz and $i \in \{+\mathrm{P5}, -\mathrm{P5}, +\mathrm{M3}, -\mathrm{M3}, +\mathrm{m3}, -\mathrm{m3}\} \subset \mathscr{I}$ are the directed primary intervals of a perfect fifth, major third, and minor third, ascending and descending, respectively. This means that the transition probability $p(\tau' \mid \tau, w)$ can only be non-zero if $\tau$ and $\tau'$ are neighboring tones on the Tonnetz. Accordingly, the weights $w$ can be represented as a six-dimensional probability vector $w = (p_{+\mathrm{P5}}, p_{-\mathrm{P5}}, p_{+\mathrm{M3}}, p_{-\mathrm{M3}}, p_{+\mathrm{m3}}, p_{-\mathrm{m3}})$. All other intervals, such as ascending or descending seconds, tritones, or augmented sixths, are expressed as combinations of these primary intervals.

Finally, we assume a uniform prior over the path-length variable $\lambda$, that is, the rate in case of a Poisson distribution and the success probability and number of trials for a binomial distribution. Note that by choosing uniform priors over all piece-level variables $\beta$, the MAP estimate becomes equivalent to a ML estimate.

## References


Albrecht, J. and Shanahan, D. (2013). The Use of Large Corpora to Train a New Type of Key-Finding Algorithm: An Improved Treatment of the Minor Mode. *Music Perception: An Interdisciplinary Journal*, 31(1):59–67.

Albrecht, J. D. and Huron, D. (2014). A Statistical Approach to Tracing the Historical Development of Major and Minor Pitch Distributions, 1400–1750. *Music Perception: An Interdisciplinary Journal*, 31(3):223–243.

Amiot, E. (2016). *Music Through Fourier Space*. Computational Music Science. Springer Nature.

Bora, U., Tekin Tezel, B., and Vahaplar, A. (2018). An Algorithm for Spelling the Pitches of Any Musical Scale. *Information Sciences*, 472:203–222.

Cambouropoulos, E. (2003). Pitch Spelling: A Computational Model. *Music Perception: An Interdisciplinary Journal*, 20(4):411–429.

Chew, E. and Chen, Y.-C. (2005). Real-Time Pitch Spelling Using the Spiral Array. *Computer Music Journal*, 29(2):61–76.

Cuddy, L. L. and Thompson, W. F. (1992). Asymmetry of perceived key movement in chorale sequences: Converging evidence from a probe-tone analysis. *Psychological Research*, 54(2):51–59.

Farbood, M. M. (2016). Memory of a Tonal Center after Modulation. *Music Perception: An Interdisciplinary Journal*, 34(1):71–93.

Freedman, R. (2014). The Renaissance Chanson goes Digital: digitalduchemin.org. *Early Music*, 42(4):567–578.

Fujinaga, I., Hankinson, A., and Cumming, J. E. (2014). Introduction to SIMSSA (single interface for music score searching and analysis). In *Proceedings of the 1st International Workshop on Digital Libraries for Musicology*, DLfM '14, page 1–3, New York, NY, USA. Association for Computing Machinery.

Hu, D. J. and Saul, L. K. (2009). A probabilistic topic model for unsupervised learning of musical key-profiles. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, pages 441–446.

Huang, P., Wilson, M., Mayfield-jones, D., Coneva, V., and Frank, M. (2017). The Evolution of Western Tonality: A Corpus Analysis of 24,000 Songs from 190 Composers over six Centuries. *SocArXiv*.

Madsen, S. T., Widmer, G., and Kepler, J. (2007). Key-Finding with Interval Profiles. In *Proceedings of the International Computer Music Conference (ICMC 07)*, pages 27–31, Copenhagen, Denmark.

Meredith, D. (2006). The *ps13* Pitch Spelling Algorithm. *Journal of New Music Research*, 35(2):121–159.

Noll, T. (2019). Insiders' Choice: Studying Pitch Class Sets Through Their Discrete Fourier Transformations. In *Mathematics and Computation in Music*, pages 371–378. Springer International Publishing.

Quinn, I. (2006). General Equal-Tempered Harmony (Introduction and Part I). *Perspectives of New Music*, 44(2):114–158. Publisher: Perspectives of New Music.

Quinn, I. (2007). General Equal-Tempered Harmony: Parts 2 and 3. *Perspectives of New Music*, 45(1):4–63. Publisher: Perspectives of New Music.

Rohrmeier, M. and Cross, I. (2008). Statistical Properties of Tonal Harmony in Bach's Chorales. In Miyazaki, K., editor, *Proceedings of the 10th International Conference on Music Perception and Cognition*, volume 6, pages 619–627.

Sapp, C. S. (2005). Online Database of Scores in the Humdrum File Format. In *Proceedings of the International Symposium on Music Information Retrieval (ISMIR)*.

Selfridge-Field, E., editor (1997). *Beyond MIDI: The Handbook of Musical Codes*. MIT Press, Cambridge, MA.

Serrà, J., Corral, L., Boguñá, M., Haro, M., and Ll Arcos, J. (2012). Measuring the Evolution of Contemporary Western Popular Music. *Scientific Reports*, 2(521):1–6.

Stoddard, J., Raphael, C., and Utgoff, P. E. (2004). Well-Tempered Spelling: A Key-Invariant Pitch Spelling Algorithm. In *Proceedings of the 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain.

Temperley, D. and Marvin, E. W. (2008). Pitch-Class Distribution and the Identification of Key. *Music Perception: An Interdisciplinary Journal*, 25(3):193–212.

Weiß, C., Mauch, M., and Dixon, S. (2018). Investigating Style Evolution of Western Classical Music: A Computational Approach. *Musicae Scientiae*, 23(4):486–507.

White, C. W. (2014). Changing Styles, Changing Corpora, Changing Tonal Models. *Music Perception: An Interdisciplinary Journal*, 31(3):224–253.

Yust, J. (2019). Stylistic Information in Pitch-Class Distributions. *Journal of New Music Research*, 48(3):217–231.