

---

## DATASET

# The Annotated Mozart Sonatas: Score, Harmony, and Cadence

Johannes Hentschel\*, Markus Neuwirth\*<sup>†</sup> and Martin Rohrmeier\*

---

This article describes a new expert-labelled dataset featuring harmonic, phrase, and cadence analyses of all piano sonatas by W.A. Mozart. The dataset draws on the DCML standard for harmonic annotation and is being published adopting the FAIR principles of Open Science. The annotations have been verified using a data triangulation procedure which is presented as an alternative approach to handling annotator subjectivity. This procedure is suited for ensuring consistency, within the dataset and beyond, despite the high level of analytical detail afforded by the employed harmonic annotation syntax. The harmony labels also encode contextual information and are therefore suited for investigating music theoretical questions related to tonal harmony and the harmonic makeup of cadences in the classical style. Apart from providing basic statistical analyses characterizing the dataset, its music theoretical potential is illustrated by two preliminary experiments, one on the terminal harmonies of cadences and the other on the relation between performance durations and harmonic density. Furthermore, particular features can be selected to produce more coarse-grained training data, for example for chord detection algorithms that require less analytical detail. Facilitating the dataset's reusability, it comes with a Python script that allows researchers to easily access various representations of the data tailored to their particular needs.

---

**Keywords:** expert-annotated dataset; tonal harmony; cadence; classical style; piano music

---

## 1. Introduction

Polyphonic music is typically characterized by its harmonic makeup. The study of (tonal) harmony thus occupies a prominent position in musicological research. Owing to the growing availability of machine-readable datasets of harmonic analyses (see Section 2), harmony can now be examined across different styles and periods using advanced computational and empirical methods (e.g., Quinn and Mavromatis, 2011; Broze and Shanahan, 2013; Temperley and de Clercq, 2013; Jacoby et al., 2015; Chen and Su, 2018; White and Quinn, 2018; Moss et al., 2019). The main contribution of this paper is the introduction and description of a new corpus of annotated scores under the FAIR principles of Open Science (Wilkinson et al., 2016). The corpus consists of the digital scores of all 18 piano sonatas (1774–1789) by Wolfgang Amadé Mozart which have been annotated by music theory experts on three levels: harmony, cadences, and phrases. The chord and phrase labels follow the DCML harmonic annotation standard,<sup>1</sup> whereas the cadence labels are based on the typology and definitions by Caplin (2004) and Rohrmeier and Neuwirth (2015). The data is provided not only as annotated scores but also as feature matrices representing notes, bars, and

annotation labels. These representations, together with a script for easy and versatile data access, enable detailed note-level analyses of a prominent sample of tonal harmony from the so-called classical era. The data is being published under a Creative Commons License and is available at [github.com/DCMLab/mozart\\_piano\\_sonatas](https://github.com/DCMLab/mozart_piano_sonatas).

## 2. Related Work and Motivation

### 2.1 Harmonic Datasets

Among the growing number of datasets featuring analyses of harmony, one of the most influential is the Kostka-Payne Corpus<sup>2</sup> compiled by David Temperley (2009). This dataset has been used, among other things, to support a particular theory of harmonic syntax (Temperley, 2011), as a ground truth for automated harmonic analysis (e.g., Pardo and Birmingham, 2002), and for estimating the abstract harmonic categories underlying surface chords in a Hidden Markov Model (White and Quinn, 2018). There are, however, two main drawbacks inherent to this dataset: first, although it covers Western tonal music from ca. 1750 to 1900, it is very small, involving only 919 labels. Second, this dataset is derived from musical excerpts taken from a well-known music theory textbook and hence may corroborate a particular theoretical standpoint (that one may or may not agree with).

Further (and more recent) datasets enabling researchers to study classical harmony are the Joseph Haydn Harmonic

---

\* École Polytechnique Fédérale de Lausanne, CH

<sup>†</sup> Anton Bruckner University, Linz, AT

Corresponding author: Johannes Hentschel  
([johannes.hentschel@epfl.ch](mailto:johannes.hentschel@epfl.ch))

Analysis Annotations Dataset,<sup>3</sup> the Annotated Beethoven Corpus,<sup>4</sup> the Beethoven Piano Sonata with Functional Harmony dataset,<sup>5</sup> and the TAVERN corpus of harmonically annotated theme-and-variation movements.<sup>6</sup> All these datasets have in common that they are bound to the oeuvres of particular (and prominent) composers.

In addition, there are several medium-sized datasets that allow scholars to examine harmony in Rock and Pop idioms. The largest among them is the McGill Billboard corpus<sup>7</sup> that consists of 743 transcriptions of popular music in the US between 1958 and 1991 and has been used, for instance, by Burgoyne et al. (2013) and Gauvin (2015). Similarly broad in scope is the Rolling Stone 200 Corpus,<sup>8</sup> which is based on Rolling Stone magazine's list of the "500 Greatest Songs of All Time" and has been analyzed by Temperley and de Clercq (2013). More specific Rock/Pop idioms have been studied using the Harte Beatles corpus.<sup>9</sup> A recent study by Moss et al. (2020) deals with harmony in 295 pieces in the Choro Songbook, providing the first quantitative style analysis of an idiom that emerged in Brazil at the end of the 19th century.

Apart from symbolic datasets of harmonic analyses, there are also numerous datasets of audio recordings that have been used for inferring harmonic characteristics. For instance, Zalkow et al. (2017) explore the notoriously complex tonal harmony in Richard Wagner's Ring cycle, relying on chroma features. Mauch et al. (2015) study the stylistic evolution of Pop music between 1960 and 2010, drawing on harmony as a prominent feature. Weiß et al. (2019) examine, among other things, the evolution of harmonic progressions on an even larger time-scale of several hundred years. Audio-based studies of harmony have the obvious advantage that they can, in principle, consider massive amounts of data since time-consuming human annotations play a smaller role compared to symbolic corpora. The flipside is, however, that they do not reach the level of detail and context-sensitivity that may be desirable from a music theoretical point of view.

## 2.2 Cadence Datasets

Within the domain of modal and tonal harmony, cadences act as salient structural patterns used to achieve closure on multiple hierarchical levels. Harmonically, these patterns are organized in different temporal phases: initial tonic  $\Rightarrow$  predominant  $\Rightarrow$  dominant  $\Rightarrow$  tonic. Each of these stages can be realized by selecting one or more chords fulfilling the proper functional criteria. Apart from the harmonic content, the closural effect of cadences crucially depends on further criteria, in particular structural voice-leading patterns (bass and soprano), the (hyper)metrical placement of a cadence, and its positioning in the form.

There is a growing number of theoretical and computational studies focusing specifically on the use of cadences in the "classical" repertoire (e.g., Caplin, 2004; Duane, 2019; Ito, 2014; van Kranenburg and Karsdorp, 2014; Neuwirth and Bergé, 2015; Sears, 2017a; Sears et al., 2018). However, the automated detection of cadences in music scores (Bigo et al., 2018) is as of yet not accurately feasible, mainly for two reasons.

First, there is the lack of formally concise definitions of cadence. Cadences are seen to emerge from coordinated activities of harmony, voice-leading, rhythm, and meter that are difficult to disentangle and have therefore been accounted for from a schema-theoretical perspective (e.g., Temperley, 2004; Gjerdingen, 2007).

Second, there are as of yet only few (mainly small) datasets available that contain cadence labels (e.g., Ito, 2014; van Kranenburg and Karsdorp, 2014; Duane, 2019; Sears, 2017a). For instance, Sears et al. (2018) provide and explore a comparatively small dataset of 270 cadence tokens in 50 selected string quartet expositions (1771–1803) from Joseph Haydn's oeuvre. As a result, analysts can examine the use of cadence only within one particular formal context, namely sonata form.

Similarly, Duane and Jakubowski (2018) confine themselves to exploring cadences in first-movement string quartet expositions (apart from Haydn also by Mozart and Beethoven). Two annotators were involved in creating this dataset; the intersection of the annotations were used for learning cadential categories in both supervised and unsupervised contexts (based on scale-degree distributions).

Note that none of these cadence datasets are accompanied by explicit and exhaustive harmonic analyses. Both issues, size and analytical richness, are tackled by the cadence dataset introduced in the present paper, which is not only much larger than previous datasets but is also supplemented by detailed harmonic annotations. As a result, it constitutes a valuable resource for investigating the complex interplay of structural components and has the potential of advancing our understanding of the harmonic nature of cadences. Further, it invites systematic comparison with the above-mentioned datasets.

## 3. Creation of the Dataset

### 3.1 Digital Scores

The dataset comprises the scores of the 54 sonata movements according to the *Neue Mozart-Ausgabe* (NMA) (Plath and Rehm, 1986) in the uncompressed XML format of the open source notation software MuseScore 3, thus providing an alternative to Craig Sapp's *\*\*kern* edition of the *Alte Mozart-Ausgabe* (AMA).<sup>10</sup> Compared to the AMA, the NMA introduces the additional sonata K. 533/494 and reflects modern critical edition practices. The MuseScore format combines the advantages of a free and easy-to-use software, i.e. consistently typesetting the scores across platforms and systems, with those of a dialect-free XML encoding that affords robust parsing and text-based version control.

As a starting point, existing encodings of the sonatas were collected and converted from various sources and file formats. Since conversion between formats tends to be lossy, we relied, where possible, on existing transcripts in MuseScore format by Lucas Mossman<sup>11</sup> and in Capella format by Tobias Schölkopf<sup>12</sup> (which can also be processed with MuseScore). For the remaining sonatas we converted Craig Sapp's digital edition to musicXML and then to MuseScore format. The only missing sonata, K. 533/494, was typeset by Tom Schreyer specially for this edition.

The converted files were corrected by the professional transcription company *tunescrivers.com* to make them conform to the *Neue Mozart Ausgabe* with respect to pitch, rhythm, articulation, dynamics, and bar numbering. Thus, the scores' content conforms in many respects to a modern authoritative edition.

### 3.2 General Annotation Principles

#### 3.2.1 Contextual Information and Granularity

The harmonic analyses included in this dataset encode expert knowledge of professional music theorists in a string-based format following a pre defined syntax (see Subsection 3.4). This syntax has been designed such that it is as close as possible to the conventional Roman numeral notation used in many theory textbooks, while being applicable to a broad variety of musical styles (e.g. Baroque, Romanticism, and Jazz) and allowing for a high level of detail to be encoded (e.g. nonchord tones; see Subsection 3.4.2). It is self-evident, however, that a larger vocabulary of chord symbols entails higher analytical contingency (the number of technically correct alternative interpretations) and therefore enhances the potential for inter-annotator disagreement. In the remainder of this section, we will make a case for the possibility of encoding a large variety of harmonic interpretations. The problem of analytical consistency is addressed in Section 4.

Harmonic analysis as taught through prominent textbooks (e.g., Kostka et al., 2013; Laitz, 2015; Clendinning and Marvin, 2016; Aldwell et al., 2011) does not involve merely labelling vertical sonorities in relation to roots and keys; rather, it is heavily informed by recognition of horizontal structures (e.g., suspensions, neighboring motions, sequential patterns, and voice-leading schemata). Although Roman numerals represent vertical entities (chords), they offer the possibility to encode, to a certain extent, horizontal aspects and hence the chord's context, for example by taking into account suspensions, neighbors, or pedal points.

As an example of the horizontal (contextual) aspects that are frequently encoded in Roman numeral analyses, take the excerpt in **Figure 1**. In mm. 19–24 it was decided to account for the melodic context—a horizontally shifted upper voice—by viewing the first sonority of each bar as a suspension chord. For instance, the first sonority in m. 20 is interpreted as a chord with root  $vi$  (E) where the fifth is suspended by a sixth, rather than as the first inversion of a C-major triad,  $IV6$ . Bar 22, beat 1 shows a more intricate case: in this specific context, the fifth (A4), though nominally

a chord tone, is better to be interpreted as a suspension (indicated by the arrow-like  $v$ ) of the fourth suspension that is in turn part of the cadential 64 chord. Furthermore, the  $I[$  part of the label in m. 25 marks the beginning of a pedal tone G which continues until the closing  $]$  in m. 32 (not shown). Finally, the example shows that—depending on a researcher's needs—this rather fine-grained way of annotating can also be evaluated on a more coarse-grained level. For example, disregarding all suspensions (within curved parentheses) and grouping labels by their numerals would produce the underlying progression  $ii6 IV V7 vi ii6 V I$ ; grouping labels by the bass notes they implicitly express would result in  $C D E C D G$  (or  $4 5 6 4 5 1$ , when expressed as scale degrees of G major).

#### 3.2.2 Analytic Consistency

Any analytical system crucially depends on the underlying criteria. In the case of a historically grown and wide-spread system such as Roman numeral analysis, the criteria that annotators apply would likely depend on their musical training and may therefore differ. At present, the task of setting up a universally valid, formalized set of analytical criteria that would lead any domain expert to the same Roman numeral analysis regardless of their musical training and of the musical style at hand, is still out of reach. We therefore abandon the idea of one incontrovertible harmonic ground-truth analysis and instead opt for solutions that reflect a consensus between at least two experts under a shared set of guidelines.<sup>13</sup> The guidelines underline the importance of being consistent with one's own analytical decisions throughout a piece, while reviewers are required to ensure the analytical consistency with other annotated pieces and corpora (compare Section 4).

#### 3.2.3 Encoding Temporal Positions

Every annotation label in this dataset is attached to a position in the corresponding score, whether it was created in MuseScore (harmony and phrases) or in an external file (cadences). Temporal positions encoded in XML differ, however, from how musicians and musicologists generally refer to them. In the first place, the positions need to be immediately comprehensible for humans who conventionally use measure numbers (MN) and beats, the latter depending on the given time signature. For a machine, however, MNs are not always sufficient: the same MN may be comprised of several `<measure>` nodes in an XML encoding, as is the case for divided measures (e.g. **Figure 2a**) or first and second endings (e.g. **Figure 2b**).

The figure shows a musical score excerpt for measures 17-25. The top staff is the treble clef and the bottom staff is the bass clef. The key signature has one sharp (F#). The time signature is 4/4. The score includes dynamic markings *p* and *f*. Below the bass line, Roman numeral annotations are provided for each measure:  $ii^6$ ,  $IV^p$ ,  $V^{(6)} V^7$ ,  $vi^{(6)} vi$ ,  $ii^{(6)} ii^6$ ,  $V^{(6v^5)} V^{(6)}$ ,  $V^{(4)} V$ ,  $I^{(7)} I$ , and  $I[V^7/IV\{$ .

**Figure 1:** Measures 17–25 of the final movement of K. 283. The example shows an annotated score excerpt employing the DCML syntax as displayed in MuseScore 3.

(a) Due to the placement of the repetition signs and key signature change, MN 64 is divided. It consists of MCs 70 and 71.

(b) First and second ending: Both represent a manifestation of MN 96 but have two different MCs, 104 and 105, and even different lengths, due to the beginning of the repetition with an upbeat.

**Figure 2:** Examples from the third movement of K. 331 where measure numbers (MN) and measure counts (MC) diverge.

In the present context, the issue of score addressability plays an important role for two reasons. First, we want to present the various facets of the dataset in a uniform way that allows for correctly joining and interrelating them (for an example, see Subsection 5.2.1). Second, we want to enable users of our dataset to automatically remove and add sets of annotations from and to MuseScore files (e.g., inserting the cadence labels that in the first place were not contained in the scores themselves).<sup>14</sup> Both cases require a temporal encoding that unequivocally references positions in the score's XML structure. Therefore, we express every position as a tuple  $(MC, onset)$  where MC (measure count) represents a running count of `<measure>` nodes (independent of length and time signature) which always starts at 1 (this corresponds to the bar counts displayed in MuseScore's status bar). Consequently, the `onset` part of a position is given as distance from the MC's first event, measured in fractions of a whole note. In other words, the three A's at the beginning of **Figure 2a**'s MN 64 have onset 0, as do the A4 on beat 2 and the grace note B4; beat 2 of MN 65 has onset 1/4.  $(MC, onset)$  tuples unambiguously reference positions within XML-encoded scores and can be easily converted to human-readable positions  $(MN, beat)$  which depend on time signatures and conventions. In the case of the cadence annotations, which were created using the latter convention (see Subsection 3.3), the beats were converted into fractions of a whole note, which allowed the `ms3` parsing library (Hentschel and Rohrmeier, 2020) to infer the  $(MC, onset)$  positions that are now stored with them.

Note that by considering the time intervals between positions, any set of positions in a given score also represents a segmentation of it. In that sense, the current dataset also provides various score segmentations, depending on which musical features researchers may want to include in a set:

- key regions (derived from harmony annotations; for an example, see **Figure 3**)
- segments between cadences (cadence annotations)
- phrases (phrase annotations)
- harmonies (harmony annotations)
- rhythmic layers (particular selection of note onsets)

### 3.3 Annotating Cadences

The cadence annotations included in this dataset were created in a tabular format, as tab-separated values (TSV). Using a simple text editor, the annotator encoded each label jointly with the corresponding temporal positions. Note that the cadence labels were prepared by the second author independently of, and without considering, the harmonic annotations described in Subsection 3.4. Informal harmonic analysis was but one component guiding the annotation of cadences, accompanied by consideration of melodic, contrapuntal, and (hyper)metric information.

Taking these structural dimensions into account, the cadence analyses adopt a typology that is based on recent music theoretical work (e.g., Caplin, 2004; Neuwirth and Bergé, 2015), operating with five labels. Two main cadence types are distinguished: *authentic* (perfect and imperfect, i.e., PAC and IAC) and *half cadences* (HC). The two core strategies for avoiding cadential closure have been labelled as *deceptive* and *evaded cadences* (i.e., DC and EC).

Note that the typology used here is tailored to the classical style and hence differs somewhat from those proposed in prominent textbooks (e.g., Kostka et al., 2013; Laitz, 2015; Clendinning and Marvin, 2016; Aldwell et al., 2011). Most importantly, we do not assume *plagal* and *contrapuntal cadences* to be genuine cadence types. For more details on this typology, the reader may wish to consult the README file.

### 3.4 Harmony and Phrase Annotations

Using notation software (such as MuseScore) currently provides the most comfortable way of creating, displaying, and manipulating annotations within a single framework, dispensing with a tedious and error-prone manual encoding of the label positions within a score. MuseScore's chord symbol functionality, for example, allows music theorists to quickly navigate and annotate even if they are not particularly computer-savvy (see the example in **Figure 1**). Using this functionality, three music theorists (Uli Kneisel, 42 movements; Tal Soker, 8 movements; and Adrian Nagel, 4 movements) created the harmonic annotations in this dataset, following the syntax and annotation guidelines developed at the Digital and Cognitive Musicology Lab (DCML). This syntax can

be automatically validated (see Section 4) and encodes a whole range of key aspects of the Roman numeral chord nomenclature, which is the de facto standard for the harmonic analysis of Western tonal music. The data verification was performed directly in MuseScore, and the most recent version of the harmony labels is included in the MuseScore files.

The DCML harmonic annotation standard<sup>15</sup> consists of a plain text syntax that allows for a highly detailed encoding of harmonic interpretations. The harmonic analyses in our dataset provide information on properties of keys, the relation of chords to a given key, chord types, chordal roots, chord tones, non-chord tones, harmonic motion over pedal points, and musical phrases. The features that the standard encodes are listed in **Table 1** and explained in the remainder of this subsection. Their combinations follow a predefined syntax that can be parsed using a regular expression.

### 3.4.1 Encoding Tonal Hierarchy

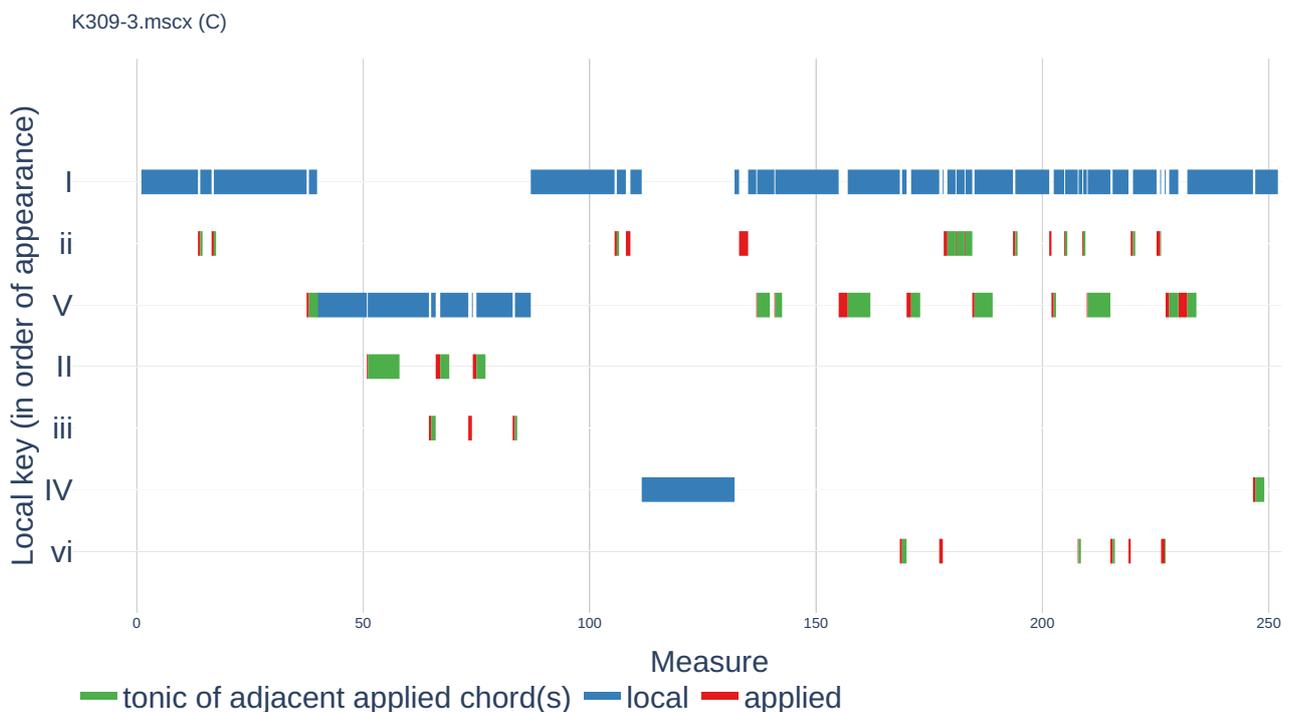
The way in which chord features are expressed in the DCML harmony annotation standard largely follows music theoretical conventions. One of these conventions is the analysis of chords in terms of a tonal hierarchy. From bottom to top, every chord is expressed relative to a local key, and a local key is in turn expressed relative to a global key in terms of Roman numerals. The resulting local key segments are visualized in **Figure 3** as blue lines. In addition, this chart shows the next lower level of the tonal hierarchy, namely the one introduced by applied chords (in red, often subsumed under the term chord borrowing), such as secondary dominants. Direct adjacency of the local tonic that the label applies to is shown in green.

### 3.4.2 Encoding Chord and Non-Chord Tones

Drawing on a slightly modified Roman-numeral nomenclature, the harmony labels encode chord tones, especially the root, as well as its exact chord type (e.g., diminished triad, major seventh chord) and inversion. Moreover, the DCML standard offers the possibility of annotating non-chord tones such as suspensions, additions, and

**Table 1:** Features encoded in the DCML standard. RN stands for Roman numeral (with uppercase and lowercase numerals distinguishing between major and minor). <NA> designates null values which may encode chord information as well (e.g., the lack of an inversion symbol indicating a root-position triad).

Feature	Encoding	Examples
Global key	Name.	Ab.I, g#.i
Local key	RN.	v.i, bVII.I
Chordal root	RN	I, bII, #vii
Chord type	<NA>, +, o, %, M	viio, IV+
Chord inversion	<NA>, 6, 64, 7, 65, 43, 2	I6, ii%65
Replacing interval(s)	()	V(64), i(#74)
Added interval(s)	(+)	I(+6), V(+b9+4)
Lower-level reference	/RN	V7/V, #viiio/ii
Phrase boundary	{, }, }	V}, I6{
Pedal point	RN[ ]	I[V7/IV IV I]



**Figure 3:** Gantt chart representing the disposition of local keys in the final movement of K. 309. Blue lines show which measures are in which local key; red lines show temporary tonicizations and interrupt the blue line; green lines mark presence of the temporarily tonicized key’s tonic. [interactive]

alterations. Such non-chord tones are expressed as Arabic numbers enclosed in parentheses, each representing an interval above the root as it would occur naturally in the given scale. Several examples have been mentioned above in Subsection 3.2.1.

#### 3.4.3 Phrase Annotations

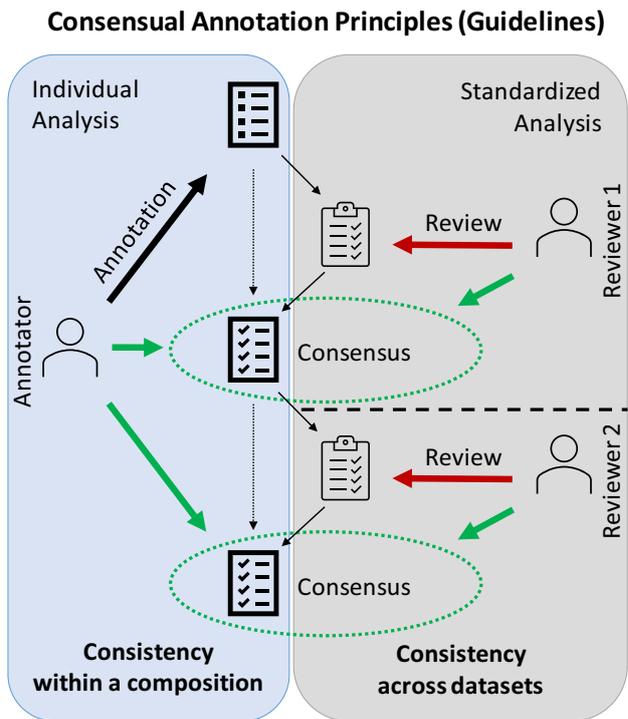
The DCML harmony annotation standard can be enriched with a very simple syntax for determining phrase boundaries. It uses the symbols { and } which can be attached to the end of any harmony label or else stand alone. { marks the beginning of a musical phrase and } its ending (e.g., **Figure 1**, mm. 24f.). The decoupling of beginnings and endings allows annotators to distinguish between (a) cases where two phrases are linked by a small transitory unit which is part of neither and (b) cases of phrase interlocking, where the endpoint of a phrase is also the beginning of the next, annotated as }{ (see, for instance, Caplin, 2001). These annotations have been added to the MuseScore files by Adrian Nagel in a separate annotation step, relying on cadential and textural cues.

## 4. Data Validation and Verification

In this context, we use the word “valid” for data that is syntactically correct, be it valid XML in the case of scores, or valid strings according to the employed annotation standards and their syntax. The validity of scores is guaranteed by the fact that they can be opened and displayed with the current version of MuseScore 3 without throwing warnings and that they have been successfully parsed with Python’s parsing library BeautifulSoup. The validity of cadence labels becomes evident by checking all labels with respect to the predefined vocabulary of the five cadence types. The harmony and phrase annotations have been validated using a regular expression.

By “verification” we refer to the process of checking data for semantic correctness (i.e., music theoretical plausibility). In the case of the scores themselves, the main criterion of this process was the correspondence with the *Neue Mozart-Ausgabe* in terms of content (but not, for example, in page layout). As laid out in Subsection 3.1, this criterion has been ensured by professional typesetters.

When it comes to the annotations in our dataset, we rely on two criteria: (A) Every label has to represent a consensus between at least two theory experts as to which choice best satisfies the annotation guidelines, and (B) analytical decisions need to be consistent within at least one movement. The annotations were verified twice by the first two authors in exchange with the annotators, thus leading to a consensus between three experts. A schematic diagram of the process is shown in **Figure 4**. Each of the two reviewers checked the entire set of sonata movements. The suggested changes reflected either the correction of an objectively wrong label (e.g., in terms of the notes it represents), the suggestion of a different harmonic interpretation (both pertaining to criterion A), or the correction of an analytical inconsistency (criterion B). For every movement, the changes were then shared with the respective annotator who could either agree or object to each suggestion. The latter case would lead to an exchange of arguments supporting or contradicting each of the two solutions, which would



**Figure 4:** Data triangulation scheme for verifying a set of expert annotations for a particular composition. Annotator and reviewer(s) share the goal of reaching a consensus on a set of annotation labels that best represents the structural properties of a composition given the predefined annotation principles (guidelines). Consensus is reached through discussions between annotator and one or several reviewers. Taking the common guidelines into account, annotators ensure analytic consistency within a composition by defending their own analytical choices, while reviewers base their suggestions and arguments on how these guidelines have previously been realized across datasets.

eventually result in a consensus on which label best reflects the discussed aspects, or in the decision to maintain both solutions as alternatives. The procedure is based on the idea of triangulation as a means of data verification (e.g., Flick, 2018) and ideally leads to a consistent, high-quality set of annotations (see the discussion in Subsection 6.2).

**Figure 4** reveals the procedure’s similarity to the widespread git-flow branching model<sup>16</sup> and can indeed be put into practice using a remote Git repository.

## 5. Characteristics of the Dataset and Use Cases

### 5.1 Basic Statistical Properties

The dataset consists of expert annotations of all 18 piano sonatas by W.A. Mozart with three movements each. It contains roughly 104,500 notes distributed over 7,500 measures, 15,000 harmony labels (466 types), and 1,100 cadence labels (5 types). **Figure 5** shows the distribution of pitch class counts over the whole corpus ordered on the line of fifths, which remarkably conforms to the shape of an almost perfect bell curve centered around G and D. This tallies with previous findings suggesting that the line of fifths is one of the prevailing topological structures underlying musical pitch space (Moss, 2019; Temperley, 2000).

Distinguishing between harmony labels occurring in major (12,700) and minor (2,300) key segments,

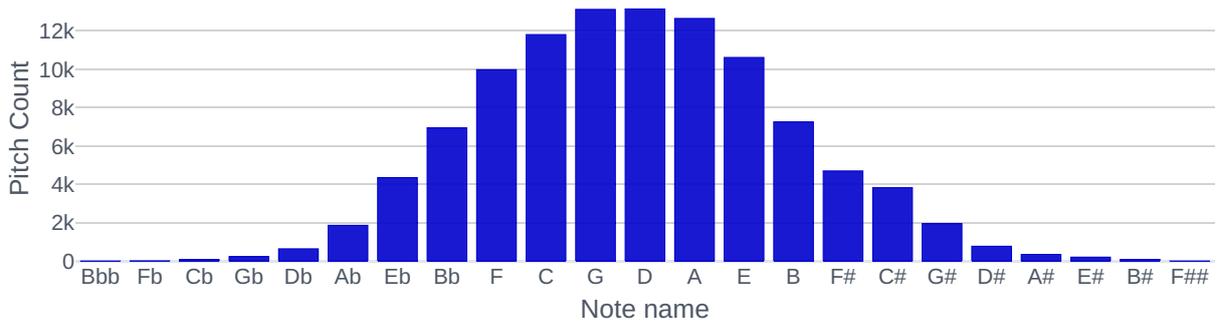
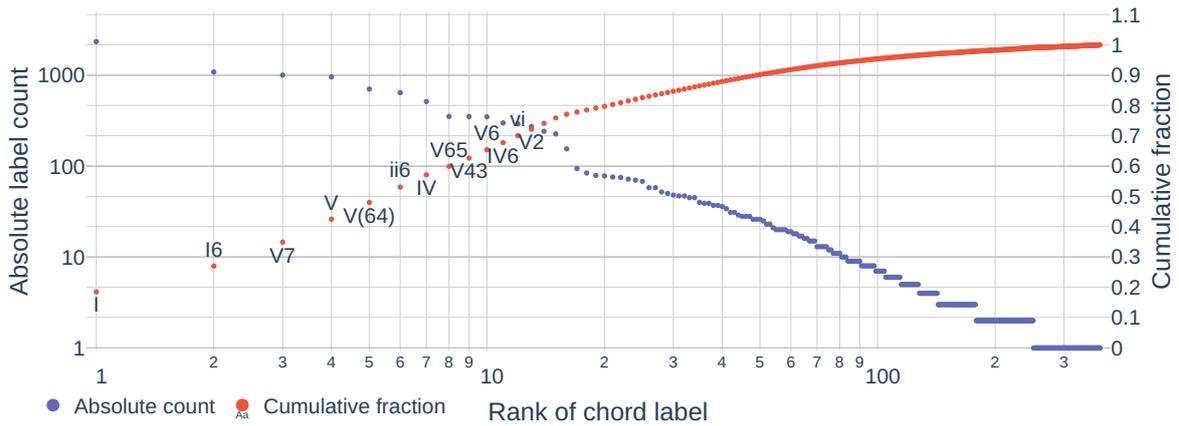
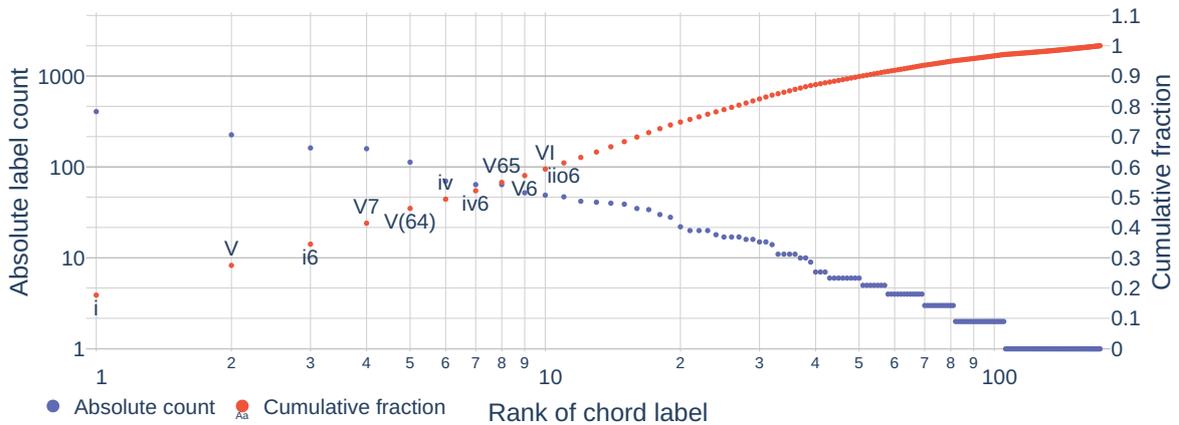


Figure 5: Distribution of pitch classes over the dataset. [interactive]



(a) Major



(b) Minor

Figure 6: Unigram statistics for all (a) major [interactive] and (b) minor segments [interactive], ordered by rank. Blue markers show absolute counts, red markers the cumulated token fraction of the current and all previous ranks.

Figure 6 plots the number of chord tokens for every chord type (blue markers), as well as the cumulative fraction of the current and all previous ranks with respect to all tokens (red markers), ordered by rank. The plots show that, out of the 466 different chord types (out of which 79 are shared between major and minor segments), relatively few make up for a large portion of all tokens. Both in major and minor keys, the first five ranks are taken up by the labels I/i, I6/i6, V, V7, and V(64), which together make up for 48.0% in major and 46.3% in minor, while 75.0% of all tokens are covered by the top 15 (major) and top 21 (minor) types. The decrease of label counts with increasing rank roughly shows the shape of a decaying power law, which has been found to be a consistent pattern for frequency distributions of chord labels, pitch class collections, and

timbres in Western tonal music of the last three centuries (Zanette, 2006; Rohrmeier and Cross, 2008; Serrà et al., 2012; Moss, 2019), as well as for words in corpora of natural language and for many domains other than music and language (Mandelbrot, 1953; Piantadosi, 2014).

Table 2 shows the distribution of cadence labels over the entire dataset and compares it to the one evaluated by Sears et al. (2018), which consists of cadence labels for 50 sonata-form expositions selected from Joseph Haydn’s string quartets. The two distributions have a very high positive correlation ( $r(3) = .975, p = .0017$ ). The fact that Mozart and Haydn were Austrian contemporaries, with their works being interrelated in multiple ways (e.g., Klauk and Kleinertz, 2016), invites further investigation into whether these cadence distributions are representative of

(a) genre-bound forms only, (b) these two composers' entire oeuvres, (c) the compositional practice of a particular era in Vienna, or even (d) a larger historico-geographical space.

The heatmaps in **Figure 8** show the relative bigram frequencies for the top 25 label types in major on the

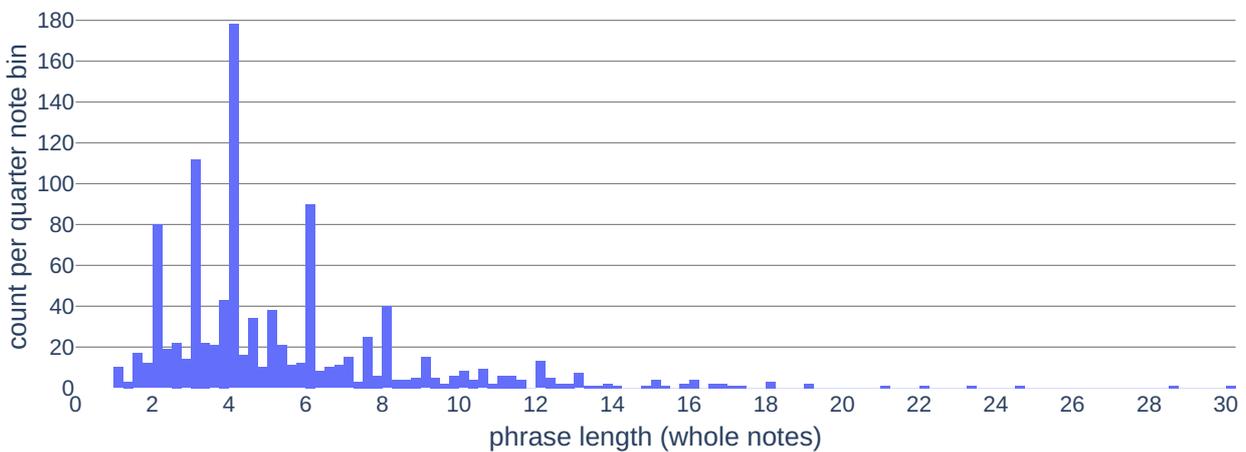
left, and in minor on the right. Black bars indicate the normalized entropy of a given label's distribution. For example, the  $\vee 2$  has a rather low entropy because in the vast majority of cases (79.1% in major and 85.7% in minor), it proceeds to the same chord,  $\text{I}6$  and  $\text{i}6$  respectively.

The chart in **Figure 9** shows the distribution of the five cadence types over the 54 sonata movements. The quantitative prevalence of the PAC and HC shown there is also evident when considering the across-movement distribution of the labels. All movements contain at least two cadence types; eleven of them contain as a minimal prerequisite only PACs and HCs. In 16 movements, this core is complemented by one of the three remaining types (either EC, DC, or IAC) and only five movements make use of all cadence types (K. 281, II; 284, II; 309, III; 310, III; and 533, III).

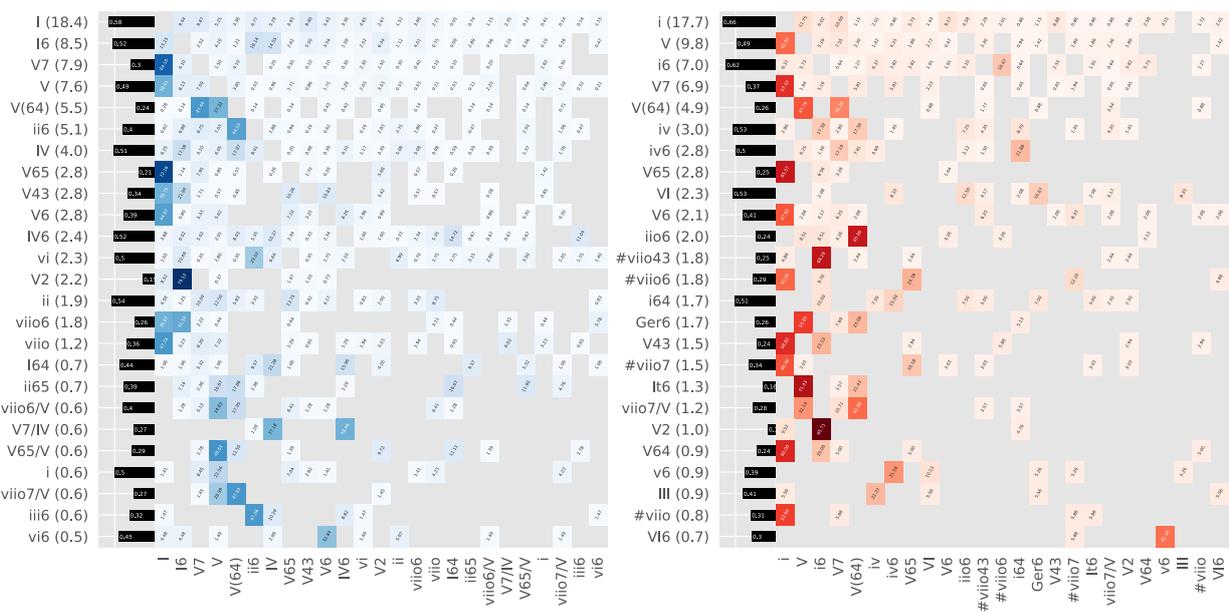
The distribution of phrase lengths is shown in **Figure 7**. It displays peaks for time-spans that are 2, 3, 4, 6, and 8 whole notes in length, and no phrase is shorter than a 4/4 measure.

**Table 2:** Comparison of this dataset's cadence distribution with that of a similar dataset based on 50 string quartet expositions by Joseph Haydn (Sears et al., 2018).

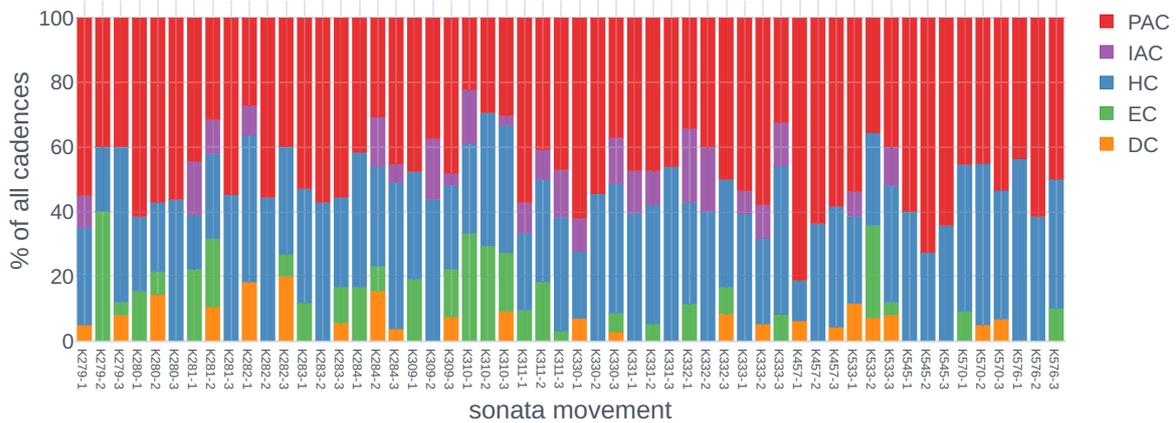
type	Mozart	%	Haydn	%
PAC	517	46.9	122	49.8
HC	398	36.1	84	34.3
EC	81	7.3	11	4.5
IAC	69	6.3	9	3.7
DC	38	3.4	19	7.8
sum	1103		245	



**Figure 7:** Histogram with bin size of a quarter note, showing the distribution of phrase lengths. [interactive]



**Figure 8:** Heatmaps showing the relative bigram frequencies (as percentages) for the top 25 chord types of all major (left, blue) and minor (right, red) segments. The black bars show each label's entropy, and the values behind the single chords indicate their relative (unigram) frequency.



**Figure 9:** Distribution of cadence labels over each of the 54 sonata movements. PAC/IAC: Perfect/Imperfect Authentic Cadence; HC: Half Cadence; EC: Evaded Cadence; DC: Deceptive Cadence. [interactive]

## 5.2 Preliminary Experiments

### 5.2.1 Terminal Harmonies of Cadences

In a first attempt to combine the cadence and harmony annotations contained in the present dataset, we address the basic question of what harmonies the various cadence types end on, and whether the finding conforms to what one would expect based on textbook knowledge (see Subsection 3.3). Since each cadence label marks the endpoint of a cadence, this task can be accomplished by joining the two sets of annotations together and looking up, for every cadence label, the corresponding harmony label. The results in **Table 3** are largely unsurprising: perfect and imperfect authentic cadences invariably end on tonic chords (overwhelmingly major), half cadences on dominant chords (almost exclusively root-position and without a seventh dissonance), and deceptive cadences on chords that have scale degree 6 as a bass note (most frequently carrying the theoretically expected  $\text{vi}$  chord). Only the terminal chords of evaded cadences show a much greater variety than what would be expected (cf. Schmalfeldt, 1992): while 46.9% of ECs resemble authentic cadences in that they end on a root-position tonic, 23.5% do not end on a tonic chord at all (but on 12 alternative chords). The high proportion of first-inversion tonic endings (29.7%) in ECs is largely in keeping with prior theoretical expectations.

Following up on this finding, it could be examined whether PACs on the one hand and IACs, HCs, and failed cadences (DC and EC) on the other differ primarily with respect to their endings (e.g., Caplin, 2004) or also with regard to their entire harmonic makeup. Related to that is the question of the extent to which cadence types can be predicted based on harmonic information alone, or conversely, whether cadence tokens may be harmonically similar across types (e.g., measured by the amount of shared harmonic vocabulary or information theoretical measures).

Since our dataset combines harmonic and cadence annotations, it can be used to determine potential similarities of cadence instances across the conventional types on the basis of multiple harmonic features, thus enabling scholars to scrutinize the results proposed by Sears (2017b) and Bigo et al. (2018) considering a much larger dataset.

**Table 3:** The respective frequencies for the harmonies ending of each of the five cadence types (i.e., the harmonic labels coinciding with the cadence labels). For example, 87.6% of all perfect authentic cadences end on a major tonic chord, I.

Cadence Type	Harmony	Count	%
PAC (517)	I	453	87.6
	i	64	12.4
HC (398)	V	392	98.5
	V6	3	0.75
	V7	3	0.75
EC (81)	I	32	39.5
	I6	22	27.2
	i	4	4.9
	#vii <sup>o</sup> 43/ii	4	4.9
	IV6	3	3.7
	#vii <sup>o</sup> 65/ii	2	2.5
	i6	2	2.5
	I(7)	2	2.5
	V65	2	2.5
	#vii <sup>o</sup> 43/vi	1	1.2
	#vii <sup>o</sup> /ii	1	1.2
	IV	1	1.2
	#vii <sup>o</sup> 64	1	1.2
V43	1	1.2	
vii <sup>o</sup> 2	1	1.2	
#vii <sup>o</sup> 43	1	1.2	
#vii <sup>o</sup> 65/iv	1	1.2	
IAC (69)	I	62	89.9
	i	7	10.1
DC (38)	vi	25	65.8
	vii <sup>o</sup> 6/V	5	13.2
	VI	4	10.5
	V43/V	2	5.3
	Ger6	1	2.6
	vi(6)	1	2.6

### 5.2.2 Performance Durations and Harmonic Density

Having shown in the previous section how two types of annotations (harmony and cadence) can profitably be combined, this section exemplifies how symbolic annotations can be put in relation to audio data, e.g., the harmonic density of a recording. To that purpose, we correlate the average harmonic densities (measured in labels per minute) of the sonata movements to their tempos (or musical densities, measured in quarter beats per minute). Both measures require actual performance durations for which we use aggregates. The underlying question is whether the speed at which harmony changes in a given piece correlates positively with its overall tempo. An alternative outcome could be that, on the contrary, the harmonic density remains within a certain range, covarying with different factors.

Our second experiment started off with two main assumptions, namely that

- all measures within the same movement which have the same time signature have the same real-time duration and can be used to approximately calculate the tempo of a piece;<sup>17</sup>
- every harmony label in this dataset represents a change of harmony in the music, and the harmonic density of a piece can be expressed by averaging its label count over its typical performance duration.

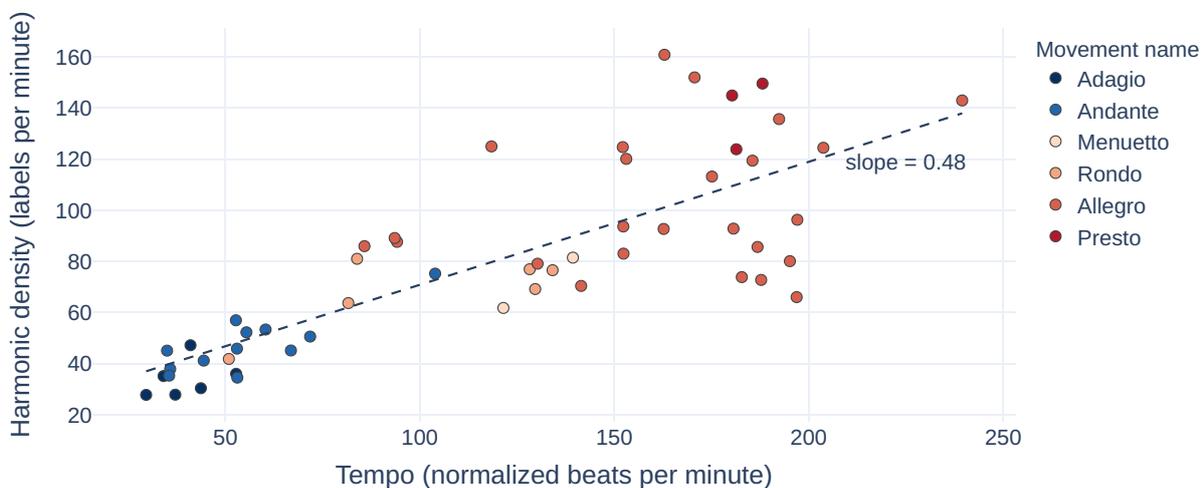
The median performance durations for the 54 sonata movements were retrieved from the Spotify API. To that end, six complete recordings were selected<sup>18</sup> because their filenames could automatically be matched to the sonata movements and because their durations showed no missing values. Then, the harmony labels had to be unfolded according to the repeat structures of the individual movements so that their counts would reflect the actual musical chronology. With this data, the harmonic density was calculated by averaging the resulting label count for every movement over its median performance duration.

In order to approximate the tempo of every movement, we averaged the length of every score over the same

performance durations. Considering the diverse time signatures and (unfolded) measure counts—ranging from 40 bars (K. 332, II) to 550 bars (the *Presto* movement of K. 283)—we opted for a uniform representation of the score lengths expressed as the number of quarter notes that can be fit into one entire rendition (including repetitions), in order to normalize the different measure lengths represented by the various time signatures. Consequently, we call the unit of the computed tempos ‘bpm’ (beats per minute), although we are dealing with a constant beat size rather than with beats in the metrical sense.

The plot in **Figure 10** suggests a strong correlation ( $r(52) = .80$ ,  $p = 4.25e-13$ ) between the two sets of values. Normalizing both label counts and beat counts by the same performance durations reveals a clear trend for faster movements to change harmony more quickly than slower movements. The cluster in the lower left part, consisting mainly of blue markers, contains 16 out of the 18 mostly slow middle movements of the corpus (the remaining two are minuets (*Menuetto*)), which generally have shorter scores but equal or longer performance times than outer movements. But the cluster is also set apart vertically, which seems to suggest that slower harmonic changes are characteristic for the harmony of *Adagio* and *Andante* movements.

This approach invites a couple of improvements and ideas for future work. For example, instead of using only the durations from complete sets of recordings of the 18 sonatas, one could opt for a random sampling approach to evaluate more performances and to therefore produce more robust statistics showing the typicality of a given duration. Also, some of the initial assumptions might have to be revisited. For example, the extreme outlier suggesting a tempo of 239 quarter notes per minute is due to the fact that for this particular piece—the first movement of K. 533/494—there seems to be a convention among pianists to repeat the first part of the piece, but not the second (as the score would suggest), which of course reduces the performance duration. Further investigations might also refine the idea of what makes for a change in harmony. As shown in Subsection 3.2.1, labels can be grouped into larger units, thus reducing the label counts. Furthermore, it might



**Figure 10:** Harmonic densities of all sonata movements plotted over their respective tempos and color coded by their movement names. Both density and tempo depend on the median performance duration. [interactive]

prove beneficial to apply linear mixed effects models in order to evaluate the effects of such a treatment. This would also be a useful approach for quantifying confounding factors such as human psychology which might lead annotators to analyze slow movements differently, or for shedding light on hidden patterns occurring in the disposition of harmonic densities between adjacent movements.

## 6. Discussion

### 6.1 Heterogeneous Data in a Unified and FAIR Format

The publication of this dataset adheres to the FAIR principles of Open Science, given that the data and associated metadata is

- *findable*, because it has been published online, and entered into various data registries;
- *accessible*, namely by granting unrestricted access to the repository through an Open Access publication with an attached digital object identifier (DOI);
- *interoperable*, because it uses text-based formats (TSV and XML) exclusively and presents its various facets in a unified tabular format;
- and *reusable* since the accompanying Python script allows to flexibly select, join, and transform parts of it, and because the metadata has been enhanced with identifiers from Wikidata,<sup>19</sup> VIAF,<sup>20</sup> MusicBrainz,<sup>21</sup> and IMSLP.<sup>22</sup>

One limitation of the dataset is that the scores come in a single format. Although MuseScore can be used to export the scores to musicXML, such exports cannot be consistently imported, currently not even by MuseScore itself. In the future, it would be desirable to further increase reusability by publishing additional validated files in widely used formats, such as musicXML or MEI. Without the availability of lossless conversion tools, however, this would immensely increase curatory efforts. Nonetheless, providing score information in a tabular format that can be produced using a publicly available parser can be viewed as a valuable alternative.

### 6.2 An Alternative Procedure for Verifying Expert Annotations

Since the advent of crowdsourcing platforms such as MTurk (Buhrmester et al., 2011), the quality assessment of subjective annotations is a much-researched topic (e.g., Nguyen et al., 2016; Kutlu et al., 2020). Two persisting problems, however, are (1) the opacity of the analytical criteria employed by crowd annotators (especially when using disparate chord vocabularies) and (2) the question of how to assess the quality of annotation sets in which many labels do not coincide (for example in the case of diverging analytical granularities, see Subsection 3.2.1). In Section 4, we therefore introduced an alternative way of ensuring the quality of annotations. It is based on the ideas of a standardized chord vocabulary, transparency and consistency of annotation principles across annotators and pieces, and achieving analytical consensus between two or more experts. The facts that both annotators and reviewers rely on the same (publicly available) annotation

guidelines and that their names are known, situate the annotated Mozart sonatas within the best practices of the Open Science philosophy (Vicente-Saez and Martinez-Fuentes, 2018).

### 6.3 Toward More Music-Theoretically Informed Annotations

Most existing chord annotation standards are tailored to the description of vertical sonorities, be it those relying on traditional Roman numeral analysis (e.g., Huron, 2020; Temperley and de Clercq, 2013; Cambouropoulos, 2016; White and Quinn, 2016; Chen and Su, 2018; Tymoczko et al., 2019), or on absolute (“guitar”) chords (e.g., Burgoyne et al., 2011; Harte, 2010; Broze and Shanahan, 2013; Choi et al., 2016). With this publication we want to make a case for harmonic annotations that try to overcome some of the shortcomings of the solely vertical perspective by including voice-leading and other horizontal contexts such as suspensions, retardations, neighbouring motions, and organ points (see Subsection 3.2.1). The DCML harmonic annotation standard provides experts with a more expressive syntax, allowing them to consistently encode contrapuntal sequences, schemata, and other voice-leading techniques that heavily inform harmonic analysis. Similarly, the cadence typology used for our corpus depends on many more criteria than those explained above (see Subsection 3.3), including voice-leading, (hyper)meter, and form. Cadence annotators trained in the outlined “Caplinian” tradition likely apply these criteria implicitly in their analyses, but it requires the quantitative study of a large dataset to shed light on them empirically. Since harmonic progressions and voice-leading patterns can be realized in an intractable multitude of musical surfaces (for the case of cadences, see Rohrmeier and Neuwirth (2015)), it is difficult to define or enumerate annotation criteria and rules beforehand without falling back to ad-hoc principles. Large datasets of annotations reflecting the intricate decisions and intuitions of expert analysts therefore represent an important step toward the development of comprehensive formalized models of music and their application in the field of MIR.

## Notes

<sup>1</sup> [github.com/DCMLab/standards](https://github.com/DCMLab/standards)

<sup>2</sup> [davidtemperley.com/kp-stats](https://davidtemperley.com/kp-stats)

<sup>3</sup> [github.com/napulen/haydn\\_op20\\_harm](https://github.com/napulen/haydn_op20_harm)

<sup>4</sup> [github.com/DCMLab/ABC](https://github.com/DCMLab/ABC)

<sup>5</sup> [github.com/Tsung-Ping/functional-harmony](https://github.com/Tsung-Ping/functional-harmony)

<sup>6</sup> [getTAVERN.org](https://getTAVERN.org)

<sup>7</sup> [ddmal.music.mcgill.ca/research/The\\_McGill\\_Billboard\\_Project\\_\(Chord\\_Analysis\\_Dataset\)](https://ddmal.music.mcgill.ca/research/The_McGill_Billboard_Project_(Chord_Analysis_Dataset))

<sup>8</sup> [rockcorpus.midside.com/](https://rockcorpus.midside.com/)

<sup>9</sup> [isophonics.net/content/reference-annotations-beatles](https://isophonics.net/content/reference-annotations-beatles)

<sup>10</sup> [github.com/craigsapp/mozart-piano-sonatas](https://github.com/craigsapp/mozart-piano-sonatas)

<sup>11</sup> [musescore.com/lukemossman](https://musescore.com/lukemossman)

<sup>12</sup> [tobis-notenarchiv.de](https://tobis-notenarchiv.de)

- <sup>13</sup> <https://dcmlab.github.io/standards/tutorial>
- <sup>14</sup> This can be achieved with the *ms3* parsing library for Python.
- <sup>15</sup> [github.com/DCMLab/standards](https://github.com/DCMLab/standards)
- <sup>16</sup> <http://nvie.com/posts/a-successful-git-branching-model/>
- <sup>17</sup> There are only two time signature changes, and each one of them occurs in the last variation of each of the two variation movements K. 284, III and K. 331, I.
- <sup>18</sup> Daniel Barenboim (2013, Warner), Marta Deyanova (1996, Nimbus), Michael Endres (2011, Oehms), Roberte Mamou (2015, Ligia), Siegfried Mauser (2014, Celestial Harmonies), and Fazıl Say (2016, Warner).
- <sup>19</sup> [wikidata.org](https://www.wikidata.org/)
- <sup>20</sup> [viaf.org](https://viaf.org/)
- <sup>21</sup> [musicbrainz.org](https://www.musicbrainz.org/)
- <sup>22</sup> [imslp.org](https://www.imslp.org/)

### Acknowledgements

This publication is possible thanks to funding by the Swiss National Science Foundation (grant no. 182811); the VolkswagenStiftung; and Mr. Claude Latour for supporting this research through the Latour Chair in Digital Musicology at EPFL.

### Competing Interests

The authors have no competing interests to declare.

### References

- Aldwell, E., Schachter, C., and Cadwallader, A. C.** (2011). *Harmony & Voice Leading*. Schirmer/Cengage Learning, Boston, MA, 4 edition.
- Bigo, L., Feisthauer, L., Giraud, M., and Levé, F.** (2018). Relevance of musical features for cadence detection. In *19th International Society for Music Information Retrieval Conference, Paris*.
- Broze, Y., and Shanahan, D.** (2013). Diachronic changes in jazz harmony: A cognitive perspective. *Music Perception: An Interdisciplinary Journal*, 31(1): 32–45. DOI: <https://doi.org/10.1525/mp.2013.31.1.32>
- Buhrmester, M., Kwang, T., and Gosling, S. D.** (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, 6(1): 3–5. DOI: <https://doi.org/10.1177/1745691610393980>
- Burgoyne, J. A., Wild, J., and Fujinaga, I.** (2011). An expert ground-truth set for audio chord recognition and music analysis. *12th International Society for Music Information Retrieval Conference*, pages 633–638.
- Burgoyne, J. A., Wild, J., and Fujinaga, I.** (2013). Compositional data analysis of harmonic structures in popular music. In Yust, J., Wild, J., and Burgoyne, J. A., editors, *Mathematics and Computation in Music*, volume 7937, pages 52–63. Springer Berlin Heidelberg, Berlin, Heidelberg. DOI: [https://doi.org/10.1007/978-3-642-39357-0\\_4](https://doi.org/10.1007/978-3-642-39357-0_4)
- Cambouropoulos, E.** (2016). The harmonic musical surface and two novel chord representation schemes. In Meredith, D., editor, *Computational Music Analysis*, pages 31–56. Springer, New York. DOI: [https://doi.org/10.1007/978-3-319-25931-4\\_2](https://doi.org/10.1007/978-3-319-25931-4_2)
- Caplin, W. E.** (2001). *Classical Form: A Theory of Formal Functions for the Instrumental Music of Haydn, Mozart, and Beethoven*. Oxford University Press, New York.
- Caplin, W. E.** (2004). The classical cadence: Conceptions and misconceptions. *Journal of the American Musicological Society*, 57(1): 51–118. DOI: <https://doi.org/10.1525/jams.2004.57.1.51>
- Chen, T.-P., and Su, L.** (2018). Functional harmony recognition of symbolic music data with multitask recurrent neural networks. *19th International Society for Music Information Retrieval Conference*, pages 90–97.
- Choi, K., Fazekas, G., and Sandler, M.** (2016). Textbased LSTM networks for automatic music composition. *arXiv preprint arxiv:1604.05358*.
- Clendinning, J. P., and Marvin, E. W.** (2016). *The Musician's Guide to Theory and Analysis*. WW. Norton & Company, New York; London, third edition edition.
- Duane, B.** (2019). Melodic patterns and tonal cadences: Bayesian learning of cadential categories from contrapuntal information. *Journal of New Music Research*, 48(3): 197–216. DOI: <https://doi.org/10.1080/09298215.2019.1607396>
- Duane, B., and Jakubowski, J.** (2018). Harmonic clusters and tonal cadences: Bayesian learning without chord identification. *Journal of New Music Research*, 47(2): 143–165. DOI: <https://doi.org/10.1080/09298215.2017.1410181>
- Flick, U.** (2018). Triangulation. In Denzin, N. K. and Lincoln, Y. S., editors, *The SAGE Handbook of Qualitative Research*, pages 444–461. SAGE, Los Angeles, fifth edition edition. DOI: <https://doi.org/10.2307/j.ctvddzffm.4>
- Gauvin, H. L.** (2015). “The Times They Were A-Changin’”: A database-driven approach to the evolution of musical syntax in popular music from the 1960s. *Empirical Musicology Review*, 10(3): 215–238. DOI: <https://doi.org/10.18061/emr.v10i3.4467>
- Gjerdingen, R. O.** (2007). *Music in the Galant Style*. Oxford University Press, New York.
- Harte, C.** (2010). *Towards Automatic Extraction of Harmony Information from Music Signals*. PhD thesis, Queen Mary University of London, London.
- Hentschel, J., and Rohrmeier, M.** (2020). Creating and evaluating an annotated corpus using the library *ms3*. In *Digital Music Research Network One-day Workshop, Queen Mary University of London, United Kingdom*. [youtu.be/UBY3wuIS4wc](https://youtu.be/UBY3wuIS4wc).
- Huron, D.** (2020). \*\*harm representation for Western functional harmony. <https://www.humdrum.org/rep/harm>.
- Ito, J. P.** (2014). Koch's metrical theory and Mozart's music: A corpus study. *Music Perception: An Interdisciplinary Journal*, 31(3): 205–222. DOI: <https://doi.org/10.1525/mp.2014.31.3.205>
- Jacoby, N., Tishby, N., and Tymoczko, D.** (2015). An information theoretic approach to chord categorization and functional harmony. *Journal of New*

- Music Research*, 44(3): 219–244. DOI: <https://doi.org/10.1080/09298215.2015.1036888>
- Klauk, S., and Kleinertz, R.** (2016). Mozart's Italianate response to Haydn's Opus 33. *Music and Letters*, 97(4): 575–621. DOI: <https://doi.org/10.1093/ml/gcw102>
- Kostka, S., Payne, D., and Almén, B.** (2013). *Tonal Harmony, with an Introduction to Twentieth-Century Music*. McGraw-Hill, New York, 7 edition.
- Kutlu, M., McDonnell, T., Lease, M., and Elsayed, T.** (2020). Annotator rationales for labeling tasks in crowdsourcing. *Journal of Artificial Intelligence Research*, 69: 143–189. DOI: <https://doi.org/10.1613/jair.1.12012>
- Laitz, S. G.** (2015). *The Complete Musician: An Integrated Approach to Theory, Analysis and Listening*. Oxford University Press, New York, 4 edition.
- Mandelbrot, B.** (1953). An informational theory of the statistical structure of language. In Jackson, W., editor, *Communication Theory*, pages 486–502. Butterworths Scientific Publications, London.
- Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M.** (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, 2(5): 150081. DOI: <https://doi.org/10.1098/rsos.150081>
- Moss, F. C.** (2019). *Transitions of Tonality: A Model-Based Corpus Study*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
- Moss, F. C., Neuwirth, M., Harasim, D., and Rohrmeier, M.** (2019). Statistical characteristics of tonal harmony: A corpus study of Beethoven's string quartets. *PLoS One*, 14(6). DOI: <https://doi.org/10.1371/journal.pone.0217242>
- Moss, F. C., Souza, W. F., and Rohrmeier, M.** (2020). Harmony and form in Brazilian Choro: A corpus-driven approach to musical style analysis. *Journal of New Music Research*, pages 1–22. DOI: <https://doi.org/10.1080/09298215.2020.1797109>
- Neuwirth, M., and Bergé, P., editors.** (2015). *What is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*. Leuven University Press, Leuven. DOI: <https://doi.org/10.2307/j.ctt14jxt45>
- Nguyen, A. T., Halpern, M., Wallace, B. C., and Lease, M.** (2016). Probabilistic modeling for crowdsourcing partially-subjective ratings. *4th AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, pages 149–158.
- Pardo, B., and Birmingham, W. P.** (2002). Algorithms for chordal analysis. *Computer Music Journal*, 26(2): 27–49. DOI: <https://doi.org/10.1162/014892602760137167>
- Piantadosi, S. T.** (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21: 1112–1130. DOI: <https://doi.org/10.3758/s13423-014-0585-6>
- Plath, W., and Rehm, W., editors.** (1986). *Klaversonaten*, volume 1–2. Bärenreiter, Kassel, Neue Mozart-Ausgabe IX/25 edition.
- Quinn, I., and Mavromatis, P.** (2011). Voice-leading prototypes and harmonic function in two chorale corpora. In *Mathematics and Computation in Music: Third International Conference*, Lecture notes in computer science, pages 230–240. Springer, Heidelberg. DOI: [https://doi.org/10.1007/978-3-642-21590-2\\_18](https://doi.org/10.1007/978-3-642-21590-2_18)
- Rohrmeier, M., and Cross, I.** (2008). Statistical properties of tonal harmony in Bach's chorales. In *Proceedings of the 10th International Conference on Music Perception and Cognition*, pages 619–627. Hokkaido University Sapporo, Japan.
- Rohrmeier, M., and Neuwirth, M.** (2015). Towards a syntax of the Classical cadence. In Neuwirth, M. and Bergé, P., editors, *What Is a Cadence? Theoretical and Analytical Perspectives on Cadences in the Classical Repertoire*, pages 285–336. Leuven University Press, Leuven. DOI: <https://doi.org/10.2307/j.ctt14jxt45.12>
- Schmalfeldt, J.** (1992). Cadential processes: The evaded cadence and the “One More Time” technique. *Journal of Musicological Research*, 12(1–2): 1–52. DOI: <https://doi.org/10.1080/01411899208574658>
- Sears, D. R. W.** (2017a). *The Classical Cadence as a Closing Schema: Learning, Memory, and Perception*. PhD thesis, McGill University, Montreal, Canada.
- Sears, D. R. W.** (2017b). Family resemblance and the classical cadence typology: Classification using phylogenetic trees. *Psychology*, 4: 328–350.
- Sears, D. R. W., Pearce, M. T., Caplin, W. E., and McAdams, S.** (2018). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research*, 47(1): 29–52. DOI: <https://doi.org/10.1080/09298215.2017.1367010>
- Serrà, J., Corral, Á., Boguñá, M., Haro, M., and Arcos, J. L.** (2012). Measuring the evolution of contemporary Western popular music. *Scientific Reports*, 2(521). DOI: <https://doi.org/10.1038/srep00521>
- Temperley, D.** (2000). The line of fifths. *Music Analysis*, 19(3): 289–319. DOI: <https://doi.org/10.1111/1468-2249.00122>
- Temperley, D.** (2004). *The Cognition of Basic Musical Structures*. The MIT Press, Cambridge, MA.
- Temperley, D.** (2009). A unified probabilistic model for polyphonic music analysis. *Journal of New Music Research*, 38(1): 3–18. DOI: <https://doi.org/10.1080/09298210902928495>
- Temperley, D.** (2011). Composition, perception, and Schenkerian theory. *Music Theory Spectrum*, 33(2): 146–168. DOI: <https://doi.org/10.1525/mts.2011.33.2.146>
- Temperley, D., and de Clercq, T.** (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, 42(3): 187–204. DOI: <https://doi.org/10.1080/09298215.2013.788039>
- Tymoczko, D., Gotham, M., Cuthbert, M. S., and Ariza, C.** (2019). The RomanText format: A flexible and standard method for representing roman numeral analyses. *20th International Society for Music Information Retrieval Conference*, pages 123–129.
- van Kranenburg, P., and Karsdorp, F.** (2014). Cadence detection in Western traditional stanzaic songs using melodic and textual features. *15th International Society for Music Information Retrieval Conference*, pages 391–396.

- Vicente-Saez, R.,** and **Martinez-Fuentes, C.** (2018). Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88: 428–436. DOI: <https://doi.org/10.1016/j.jbusres.2017.12.043>
- Weiß, C., Mauch, M., Dixon, S.,** and **Müller, M.** (2019). Investigating style evolution of Western classical music: A computational approach. *Musicae Scientiae*, 23(4): 486–507. DOI: <https://doi.org/10.1177/1029864918757595>
- White, C. W.,** and **Quinn, I.** (2016). The Yale-Classical Archives Corpus. *Empirical Musicology Review*, 11(1): 50. DOI: <https://doi.org/10.18061/emr.v11i1.4958>
- White, C. W.,** and **Quinn, I.** (2018). Chord context and harmonic function in tonal music. *Music Theory Spectrum*, 40(2): 314–335. DOI: <https://doi.org/10.1093/mts/mty021>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.** (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1). DOI: <https://doi.org/10.1038/sdata.2016.18>
- Zalkow, F., Weiß, C.,** and **Müller, M.** (2017). Exploring tonal-dramatic relationships in Richard Wagner's Ring cycle. In *18th International Society for Music Information Retrieval Conference*, pages 642–648, Suzhou, China.
- Zanette, D. H.** (2006). Zipf's Law and the creation of musical context. *Musicae Scientiae*, 10(1): 3–18. DOI: <https://doi.org/10.1177/102986490601000101>

**How to cite this article:** Hentschel, J., Neuwirth, M., & Rohrmeier, M. (2021). The Annotated Mozart Sonatas: Score, Harmony, and Cadence. *Transactions of the International Society for Music Information Retrieval*, 4(1), pp. 67–80. DOI: <https://doi.org/10.5334/tismir.63>

**Submitted:** 30 April 2020

**Accepted:** 14 April 2021

**Published:** 19 May 2021

**Copyright:** © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

]u[

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

**OPEN ACCESS** 