Investigating Style Evolution of Western Classical Music: A Computational Approach

Christof Weiß[1], Matthias Mauch[2], Simon Dixon[2], and Meinard Müller[1]

Abstract
In musicology, there has been a long debate about a meaningful partitioning and description of music history regarding composition styles. Particularly, concepts of historical periods have been criticized since they cannot account for the continuous and interwoven evolution of style. To systematically study this evolution, large corpora are necessary suggesting the use of computational strategies. This paper presents such strategies and experiments relying on a dataset of 2000 audio recordings, which cover more than 300 years of music history. From the recordings, we extract different tonal features. We propose a method to visualize these features over the course of history using evolution curves. With the curves, we re-trace hypotheses concerning the evolution of chord transitions, intervals, and tonal complexity. Furthermore, we perform unsupervised clustering of recordings across composition years, individual pieces, and composers. In these studies, we found independent evidence of historical periods that broadly agrees with traditional views as well as recent data-driven experiments. This shows that computational experiments can provide novel insights into the evolution of styles.

Keywords
Computational Musicology, Music Information Retrieval, Tonal Audio Features, Style Analysis, Composer Style, Corpus Analysis

Introduction

Western art music style steadily evolved over centuries. Musicologists commonly agree that this evolution proceeded in several phases rather than in a linear fashion (Pascall, 2001). Some of these phases exhibit a certain homogeneity with respect to stylistic aspects. This is why a

[1]International Audio Laboratories Erlangen, Germany
[2]Centre for Digital Music, Queen Mary University of London, UK

Corresponding author:
Christof Weiß, International Audio Laboratories Erlangen, Am Wolfsmantel 33, 91058 Erlangen, Germany.
Email: christof.weiss@audiolabs-erlangen.de

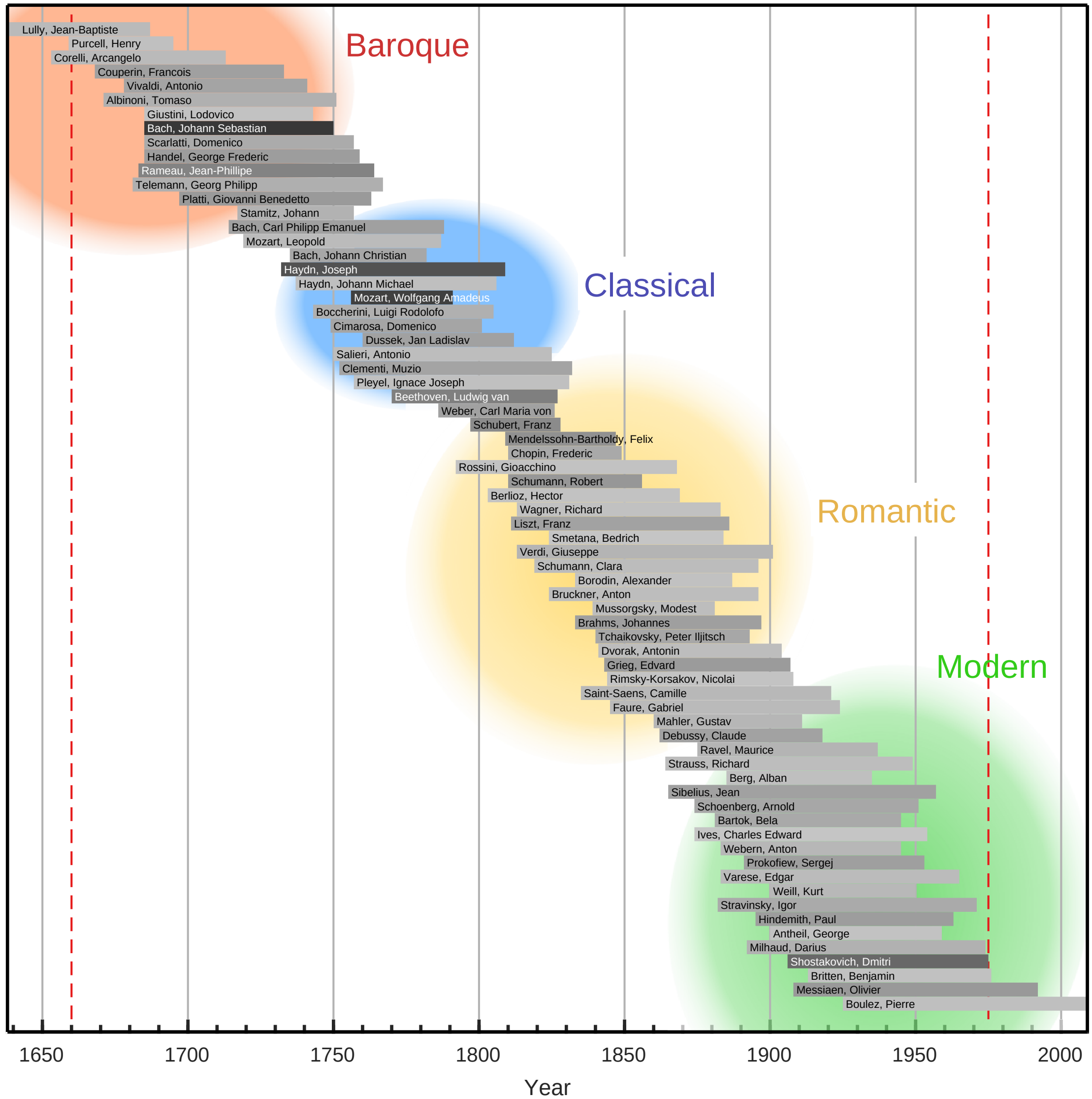

Figure 1. Overview of the composers in the dataset. A box corresponds to the composer's lifetime. Darker boxes indicate that more pieces by a composer are considered in the dataset (e. g., for J. S. Bach).

categorization of music according to *historical periods* or *eras*—as indicated by the clouds in Figure 1)—has been a "customary method" in musicology (Frank, 1955, p. 1). Until today, these categories' names serve as important terminology and "basis for discussion" (Godt, 1984, p. 38) for describing musical style in the historical context.

Nevertheless, a categorization into a few historical periods cannot reflect the complex structure of musical style's continuous and interlaced evolution (Webster, 2004; Clarke, 1956). Long transitional phases, parallel or contrasting trends, bifurcations due to esthetic controversies,[1] as well as slow but steady changes in musical style defy a classification using such simple categories. On closer examination, stylistic similarity of pieces does not necessarily imply temporal proximity of their composition dates (Frank, 1955). The geographical context adds another layer of complexity to the overall picture. Composer styles can be influenced by local folk culture or particular social conditions. The balance between a composer's *personal style* and a time-related *contemporary style* or *epochal style* has also changed over the course of music history (Pascall, 2001). Furthermore, even individual composers have not always written in a homogeneous style throughout their life. Beethoven or Schoenberg are only two examples of this observation.

Because of such reasons, musicologists have criticized models of historical periods for decades. Nowadays, analyzing the style of individual composers or small regional groups is the preferred approach in musicology (Webster, 2004). Adler and Strunk (1934) suggest three definitions of style relating to *time*, *place*, and *author*. They describe the time-related categorization as the "essence of independent style-criticism" while regarding author identification to be "style-criticism in its highest form," which, however, "sometimes turns on subordinate details." This indicates that the detailed analysis of individual composers often lacks the possibility of generalization and does not provide an overview of larger time spans. To obtain such an overview, which allows for identifying stylistically homogeneous phases as well as phases of change,[2] one needs to consider a broad variety of pieces covering both composer-specific aspects such as lifetime or place of residence as well as musical aspects such as instrumentation, key, tempo, or genre.

In order to account for this variety, one needs datasets of several hundreds or thousands of pieces where manual inspection is impractical. To make a corpus-based analysis feasible, computational approaches are required. These approaches often rely on statistical methods (Fucks & Lauter, 1965; Bellmann, 2012; White, 2013; Rodriguez Zivic et al., 2013) and, therefore, allow for analyzing style characteristics within a corpus in an objective and unbiased fashion. As a technical prerequesite, the musical pieces have to be accessible in a computer-readable format. Musicologists typically choose a symbolic score representation such as MusicXML (Good, 2006) or MEI (Pugin et al., 2012). In practice, the availability of symbolic scores in high quality is a major limitation when compiling a dataset. Manual creation of scores is very time-consuming and current systems for Optical Music Recognition (OMR) do not yet show adequate performance (Byrd & Simonsen, 2015). As a consequence, studies on manually curated symbolic scores employ rather small datasets such as the study by Bellmann (2012), who analyzed 297 piano pieces by 27 composers.[3] Some researchers accept the loss caused by limited OMR performance and hope to achieve meaningful analysis results when averaging over a large dataset of uncorrected OMR output. Using this strategy, Rodriguez Zivic et al. (2013) presented a promising study relying on the Peachnote corpus.[4] They calculated statistics of melodic intervals mapped to composition years and subsequently clustered the year-wise features resulting in cluster boundaries roughly at the years 1765, 1825, and 1895.

Another option are MIDI files, which are available in large numbers for classical music. Similarly to scanned sheet music, however, the quality of available MIDI files is heterogeneous since many files contain errors and the encoding is often not consistent. Furthermore, the selection is biased—in particular, orchestral pieces or works by less popular composers are sometimes hard to find. Using a limited set of 19 popular composers, White (2013, Chapter 3) presented an interesting study on 5000 MIDI files.[5] Based on chord progression statistics, he found that composers and composer groups "tend to cluster in ways that conform to our intuitions about stylistic traditions and compositional schools" (White, 2013, p. 176).

As an alternative to using scanned sheet music or MIDI files, one may consider audio recordings of musical pieces. For the typical classical music repertoire, a high number of such recordings are easily available. Though capturing a specific interpretation, a recording better corresponds to the "sonic reality" of a musical piece than a score representation does. To analyze such recordings, one needs to apply audio processing tools as developed in the field of Music Information Retrieval (MIR). These algorithms are often error-prone and do not reach a high level of specificity regarding human analytical concepts. In particular, note objects as specified by a musical score are not given explictly and, thus, are hard to extract from a recording (Benetos et al., 2013). Nevertheless, several studies (Izmirli, 2009; Sheh & Ellis, 2003; Weiß & Müller, 2014) have shown that suitable audio features can capture meaningful information that correlates to music theory.

In this paper, we present several experiments for such an audio-based style analysis. To this end, we compiled a dataset of 2000 music recordings by 70 composers covering more than 300 years of music history (see Figure 1). We choose a number of audio features that may be capable of describing style characteristics of the music. To achieve a certain invariance to the instrumentation, we focus on features capturing harmonic and tonal aspects. More specifically, our features describe the presence of chord progression types and harmonic interval types as well as the tonal complexity. Restricting to harmony does not provide a comprehensive description of musical style since, for instance, melody or rhythm capture further important aspects. Nevertheless, our results show that tonal features alone can provide a meaningful description and lead to interesting insights. Furthemore, rhythmic and melodic characteristics can have an influence on our features and, thus, are implicitly captured to a certain degree.

As one main contribution of this paper, we propose a novel visualization technique. For these *evolution curves*, we project the piece-wise feature values onto the historical timeline using the composers' lifetime. We show several such curves in order to investigate tonal properties of our data in a statistical way. Performing aggregation and clustering with unsupervised techniques[6]—i. e., without incorporating any prior information about stylistic similiarity—, we analyze the evolution of musical styles regarding composition years, individual pieces, and composers. We found interesting coherences that widely agree with traditional views as well as other data-driven experiments. Even though the choices of data (pieces) and methods (features) have crucial influence on the results and these choices are also subjective, our investigations generally demonstrate how computational strategies can contribute to the understanding of musical style and its evolution from a quantitative and objective perspective.

The remainder of the paper is organized as follows. First, we describe our dataset. Second, we explain the main aspects of our computational procedure including the extraction and temporal aggregation of audio features as well as our strategy of computing evolution curves. Third, we present such evolution curves for different types of features and discuss musicological implications. Finally, we conduct analyses and clustering experiments for investigating the stylistic relationships regarding years, pieces, and composers. The main findings of this work rely on the first author's dissertation (Weiß, 2017, Chapter 7).

Dataset

In this study, we consider the typical repertoire of Western classical music. Thus, we put special emphasis on composers whose works frequently appear in concerts and on classical radio programs. For example, we include a relatively large number of works by popular composers such as J. S. Bach or W. A. Mozart. At the same time, we try to ensure a certain variety and diversity regarding other aspects (countries, composers, musical forms, keys, tempi, etc.). Following such principles, we compiled a dataset of 2000 music recordings[7] from 70 different composers covering more than 300 years of music history.[8] Figure 1 provides a visualization of the dataset with respect to the composers' lifetime. The darkness of the "lifetime boxes" indicates the number of recordings contained in the dataset by the respective composer. We strived towards a homogeneous coverage of the timeline with composers. The years before 1660 and after 1975 were ignored for the further analysis since less than three composers contribute here.

To avoid effects due to timbral characteristics, we balanced our dataset regarding the instrumentation by including each 1000 pieces for piano and orchestra. To avoid timbral particuliarities within the piano data, we only selected piano recordings performed on the modern grand piano (also for keyboard pieces from the 17th and 18th century, where we did not include any harpsichord recordings). Moreover, the orchestral data neither includes works featuring vocal parts nor solo concertos. We took care of a certain diversity among each composer's works by considering various musical forms (e. g., sonatas, variations, suites, symphonies, symphonic poems, or overtures). Furthermore, the dataset exhibits a mixture of time signatures, tempi, keys, and modes (major/minor). For most aspects—such as tempo and time signature—, we obtained this variety by including all movements of a work cycle or multi-movement work. However, the selection is not *systematically* balanced regarding all of these characteristics. Instead, we prioritized balancing the instrumentations in order to avoid biases caused by audio-related effects. Beyond this, we put special emphasis on the coverage of the timeline and on the regional balance of the composers' countries of residence. Since our experiments rely on statistical procedures, we ensured a certain size of the dataset (2000 pieces) and, therefore, could not achieve perfect balance regarding all aspects. A systematical investigation of principles for data compilation and their influence on experimental results is beyond the scope of this paper and should be addressed in future work.

The recordings originate from commercial audio CDs. To allow reproduction of our experiments and to provide detailed insight into the content, we published a list of the recordings along with annotations and audio features extracted from these recordings.[9]

Computational Methods

Overview

The computational analysis of music recordings is a young field of research. Extracting score-like information from audio—referred to as *automatic music transcription*—is a complex problem

Table 1. Overview of interval and complexity features. The interval features rely on local NNLS chroma features (10 Hz). For the tonal complexity, we considered four different time resolutions.

Feature	Description
F_1	Interval Category 1 (minor second / major seventh)
F_2	Interval Category 2 (major second / minor seventh)
F_3	Interval Category 3 (minor third / major sixth)
F_4	Interval Category 4 (major third / minor sixth)
F_5	Interval Category 5 (perfect fourth / perfect fifth)
F_6	Interval Category 6 (tritone)
F_7	Tonal Complexity Global (full movement)
F_8	Tonal Complexity Medium (10 s)
F_9	Tonal Complexity Medium (500 ms)
F_{10}	Tonal Complexity Local (100 ms)

where state-of-the-art systems do not show satisfactory performance in most scenarios (Benetos et al., 2013). In particular, the output of such systems does not provide a reliable basis for applying methods developed for score analysis. Nevertheless, some analysis tasks can be approached without the need of explicit information such as note events. Instead, *semantic mid-level representations* can be used, which can be directly computed from the audio recordings while allowing for human interpretation.

Feature Extraction

For tonal analysis, chroma features have turned out to be useful mid-level representations. These representations indicate the distribution of spectral energy over the twelve chromatic pitch classes (Müller, 2015, Chapter 3) and robustly capture tonal information of music recordings. Several advanced chroma extraction methods were proposed in order to improve the timbre invariance of chroma features (Gómez & Herrera, 2004; Lee, 2006; Müller & Ewert, 2010). For our studies, we rely on a chroma feature type that reduces the influence of overtones using a Nonnegative Least Squares (NNLS) algorithm (Mauch & Dixon, 2010a).[10] The chroma features computed for our experiment locally correspond to 100 ms of audio (feature resolution of 10 Hz). We provide details on the feature extraction in Section S1 of the Supplemental Material Online (SMO) section.

On the basis of such chroma features, researchers developed algorithms for analysis tasks such as global key detection (van de Par et al., 2006; Papadopoulos & Peeters, 2012), local key detection (Sapp, 2005; Papadopoulos & Peeters, 2012), or chord recognition (Sheh & Ellis, 2003; Mauch & Dixon, 2010b; Cho & Bello, 2014). In this paper, we rely on similar algorithms extracting various types of tonal features. To account for different aspects of tonality, we consider 65 features, which we refer to as $F_1, \ldots, F_{65}$. Tables 1 and 2 outline some of these features.

The first type of features serves to quantify the presence of different *harmonic intervals* within the local analysis segments. Since chroma features refer to the level of *pitch classes*, we can only discriminate six different interval types when ignoring the octave and the unison. The system of these *interval categories* (IC) was developed for style analysis in the context of the pitch class set theory (Honingh et al., 2009). Based on local NNLS chroma features, we calculate six interval features as proposed in (Weiß et al., 2014). We denote these features with $F_1, \ldots, F_6$.

Table 2. Overview of root note transition features. The arrows denote the direction of the root note interval ($\nearrow$ = upwards, $\searrow$ = downwards). Transitions by complementary intervals in opposite direction belong to the same category. Δ indicates the interval size in semitones.

Feature	Interval	Δ	Complementary	Δ	Quality
–	Perfect unison	0	Perfect octave $\searrow$	−12	None
F_{11}	Minor second $\nearrow$	+1	Major seventh $\searrow$	−11	Authentic
F_{12}	Major second $\nearrow$	+2	Minor seventh $\searrow$	−10	Authentic
F_{13}	Minor third $\nearrow$	+3	Major sixth $\searrow$	−9	Plagal
F_{14}	Major third $\nearrow$	+4	Minor sixth $\searrow$	−8	Plagal
F_{15}	Perfect fourth $\nearrow$	+5	Perfect fifth $\searrow$	−7	Authentic
F_{16}	Augmented fourth $\nearrow$	+6	Diminished fifth $\searrow$	−6	None
F_{17}	Perfect fifth $\nearrow$	+7	Perfect fourth $\searrow$	−5	Plagal
F_{18}	Minor sixth $\nearrow$	+8	Major third $\searrow$	−4	Authentic
F_{19}	Major sixth $\nearrow$	+9	Minor third $\searrow$	−3	Authentic
F_{20}	Minor seventh $\nearrow$	+10	Major second $\searrow$	−2	Plagal
F_{21}	Major seventh $\nearrow$	+11	Minor second $\searrow$	−1	Plagal
–	Perfect octave $\nearrow$	+12	Perfect unison	0	None

For example, F_1 corresponds to minor second or major seventh intervals (IC1) and F_2 denotes major second and minor seventh intervals (IC2); see Table 1 for an overview. Due to the fine temporal resolution (100 ms), the features mainly describe harmonic intervals (simultaneously played notes). At note transitions, the segmentation procedure can lead to blurry features. More detailed information on the feature computation can be found in Section S2 of the SMO.

Next, we consider the more abstract notion of *tonal complexity*. In MIR, several approaches have been proposed for measuring tonal complexity from audio data (Streich, 2006; Honingh & Bod, 2010). In this paper, we rely on a feature variant presented in (Weiß & Müller, 2014), which can be computed directly from chroma representations. These features turned out to be useful for style classification of classical music recordings (Weiß & Müller, 2015). In particular, we consider the fifth-based complexity feature, which measures the spread of the pitch class content around the circle of fifths. Flat distributions of pitch classes result in high complexity values. Since tonal complexity refers to different time scales (chords, segments, or full movements), we calculate four features $F_7, \ldots, F_{10}$ based on different temporal resolutions of the chromagram (local features with 100 ms resolution, two intermediate resolutions of 500 ms and 10 s, and a global histogram). In Section S3 of the SMO, we explain the feature computation in more detail. Figure 2 shows the complexity features for two pieces.

We further look at *chord transitions* to capture sequential properties. For estimating the chords, we use the public algorithm Chordino.[11] This method relies on NNLS chroma features and incorporates Hidden Markov Models for concurrently estimating and smoothing the chord labels (Mauch & Dixon, 2010a). In Section S4 of the SMO, we report the parameter settings and chord types used in this work. Motivated by music theory concepts (Gárdonyi & Nordhoff, 2002), we only consider the *relative root note distance* between the chords. To this end, we reduce the chord estimates by only retaining the root note information of the chords (see Figure 3). We count the occurrence of different intervals between these root notes for all pairs of chord symbols.

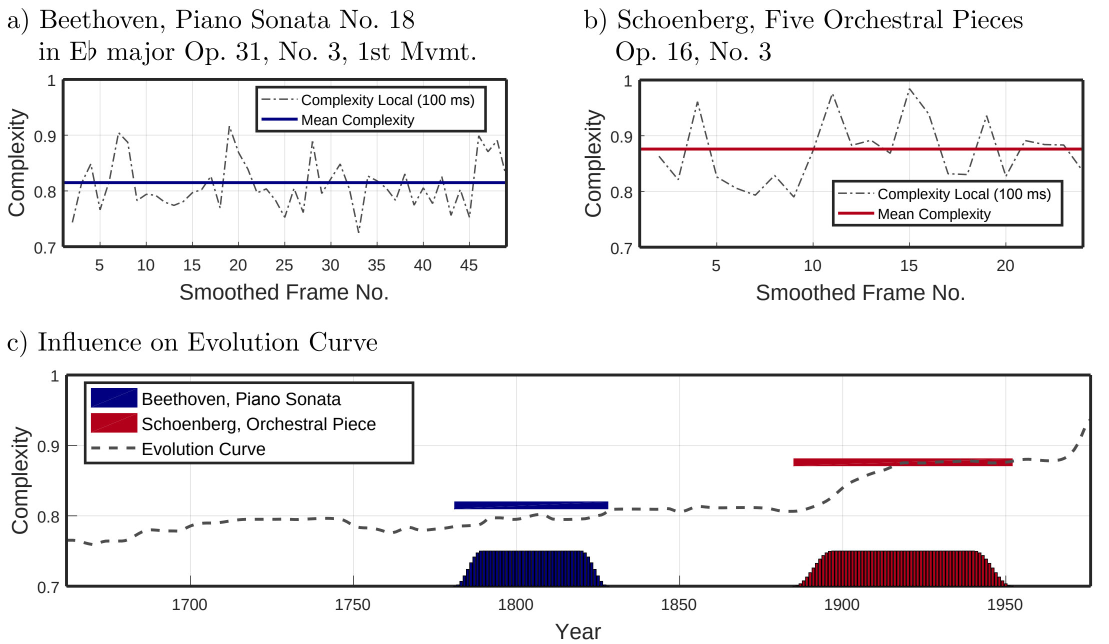

Figure 2. Temporal aggregation and evolution curve. For two pieces by Beethoven (a) and Schoenberg (b), we compute the mid-scale complexity feature (10 s) and average over the piece (colored line). Figure c) shows the projection of these features onto the timeline using the composers' lifetime.

Since the root note information refers to the pitch class level (no octave information), we can discriminate only twelve types of steps as given in Table 2. For example, the root transition C→A can be described by a minor third downwards (m3↘) or by a major sixth upwards (M6↗)—the complementary interval in opposite direction. Since we have a temporal order, we can discriminate between the directions of a given interval here. For example, C→A (m3↘) belongs to a different category than A→C (m3↗). Ignoring self-transitions of root notes (such as C major → C minor), we end up with eleven different features referred to as $F_{11}, \ldots, F_{21}$. For the later experiments, we account for specific chord transitions by looking at the chord types. Only counting transitions from a major chord to another major chord ($maj \rightarrow maj$), we obtain the features $F_{22}, \ldots, F_{32}$ referring to the eleven root note intervals. Similarly, we consider the combinations $maj \rightarrow min$ ($F_{33}, \ldots, F_{43}$), $min \rightarrow maj$ ($F_{44}, \ldots, F_{54}$), and $min \rightarrow min$ ($F_{55}, \ldots, F_{65}$).

An automatic chord estimation system is not free of errors. Moreover, the chosen selection of chord types may not be suitable for all musical styles in the dataset. For atonal pieces, a specific "measurement error" may be characteristic rather than a semantically meaningful output. Nevertheless, we expect certain tendencies to occur since we look at a large number of works and, thus, the "measurement noise" may get smoothed out in the global view. Moreover, errors concerning the chord *types* do not affect our experiments since we only consider the chords' *root notes* and their transitions.

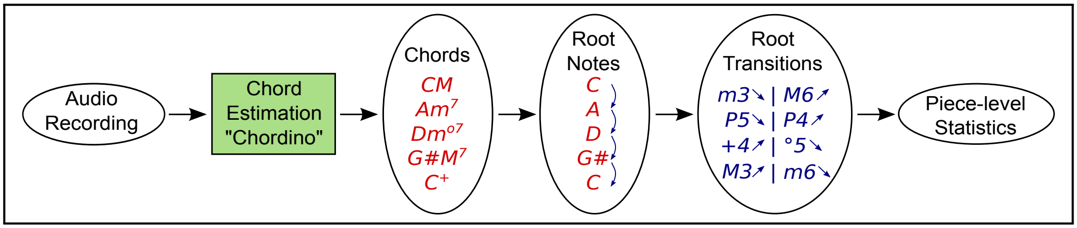

Figure 3. Estimation of root note transitions. In this schematic overview, we show the processing pipeline for estimating statistics on root note transitions. First, we reduce the output of the chord estimator by only considering the root notes (without octave information). From this root note sequence, we calculate interval statistics according to the categories presented in Table 2.

Temporal Aggregation

The experiments in this paper are based on on a comparison of entire musical pieces. For this reason, we need movement-level descriptors rather than local ones. To obtain piece-wise features, we simply average the local feature values over each recording. Averaging provides an easily interpretable summary even though higher-order statistics such as the variance might lead to a more detailed description. As for the chord transitions, we divide the counts of every root note transition by the total number of chord transitions in a piece in order to obtain relative values. In the following, our feature symbols $F_1, \ldots, F_{65}$ always refer to these globally averaged values. Thus, each feature has exactly one value per piece. Figure 2 shows the global mean value along with the local values for one of the complexity features.

Evolution Curves

For analyzing musical styles in their historical context, the composition dates of the pieces in our dataset are of major interest. Compiling all this information requires a huge effort. For many works, the composition year is unknown or in doubt. If we had all the composition dates at hand, it would constitute a difficult task to find an equal amount of works for all years while balancing the dataset regarding other aspects (such as instrumentation, key, or tempo). For these reasons, we pursue a pragmatic approach where we project the works of a composer onto a timeline using the composer's lifetime. As an approximation, we use a roughly flat distribution with smooth edges (a Tukey window with parameter $\alpha = .35$) while excluding the first ten years of the lifetime. Figure 2c shows the distribution for Beethoven and Schoenberg.

Subsequently, we apply this projection strategy to all 2000 pieces in our dataset. For a given feature, we obtain an *evolution curve* (EC), which shows the average value of the piece-wise values over the timeline. Thereby, each piece contributes to that part of the timeline which corresponds to the composer's lifetime as indicated by our distribution. Within this procedure, all pieces are given an equal weight.[12] The dashed line in Figure 2c shows the EC for the complexity feature F_9. The projection strategy of our EC is rather simplistic, and it is obvious that one cannot resolve details of style evolution in this way. For example, the assumption of stylistic homogeneity over a composer's lifetime is often violated. Here, one may think of composers with several "creative periods" such as Schoenberg, who developed from late Romantic style to dodecaphony in several steps. In our study, however, we are interested in a rather "global" view and look at the overall tendencies. For this reason, we assume that the simplifications of the EC does not have a crucial impact when analyzing the general trends over centuries. With this procedure, the pieces in our dataset are distributed in an approximately equal fashion over the timeline. For the EC, we consider the span 1660–1975 as indicated by the red dashed lines in Figure 2.[13]

Feature Aggregation

Since it is hard to obtain an overview of our 65 feature dimensions, aggregation of several features to a new one-dimensional feature F^* can be useful. Such an aggregation can be a linear combination or a ratio of selected features where the individual features F_n can obtain different weights w_n. Moreover, there are aggregation techniques that automatically determine these weights with respect to some optimization criterion. One example is Principal Component Analysis (PCA), see (Pearson, 1901). Hereby, the first principal component points to the direction of maximal variance and, thus, contains the highest amount of information that can be expressed in one dimension. With increasing number, the components contain less variance. Later, we will use PCA for aggregating features as well as for analyzing the variance of the initial features in the EC. Section S5 of the SMO gives mathematical details for calculating the aggregated features.

Style Analysis Using Evolution Curves

Analysis of Chord Transitions

A comprehensive analysis of musical style has to reflect a wide range of different aspects and musical parameters. According to LaRue (1962), we can find style indicators in the domains of sound, form, rhythm, melody, and harmony. The situation is complex because of a high interdependency of these categories. Apart from the sound with its "psychological firstness" (LaRue, 1962, p. 92), researchers consider harmony as important and notice "clear conventions of harmonic behavior" within a period (LaRue, 1992, p. 39). Belaiev (1930, p. 375) stresses the importance of "chordal combinations" and harmonies in general for defining a style. Other theorists focus on more specific aspects of harmony but discuss these issues along with their stylistic meaning (Gárdonyi & Nordhoff, 2002; de la Motte, 1976/1991). In addition to this, harmony as a musical dimension is—to a certain degree—independent from timbral properties such as the instrumentation.

For these reasons, our study focuses on tonal and harmonic characteristics. We consider several types of tonal audio features as described in the previous section. Relying on these features, we want to investigate and re-trace hypotheses regarding tonal aspects of musical style. To this end, we first look at a categorization scheme for chord transitions proposed by Bárdos (1961), taken up by Gárdonyi and Nordhoff (2002). This concept is an extension of the well-known distinction of cadences into the *plagal* type with an ascending perfect fifth (or descending perfect fourth) between the chords' root notes and the *authentic* type with a descending (falling) perfect fifth. According to Bárdos' extension, authentic transitions comprise root note transitions of descending fifth and third intervals as well as ascending second (descending seventh) intervals. Plagal transitions are of opposite direction (see Table 2). These qualities only refer to *pitch classes* and are independent from any octave inversion. Thus, transitions by complementary intervals

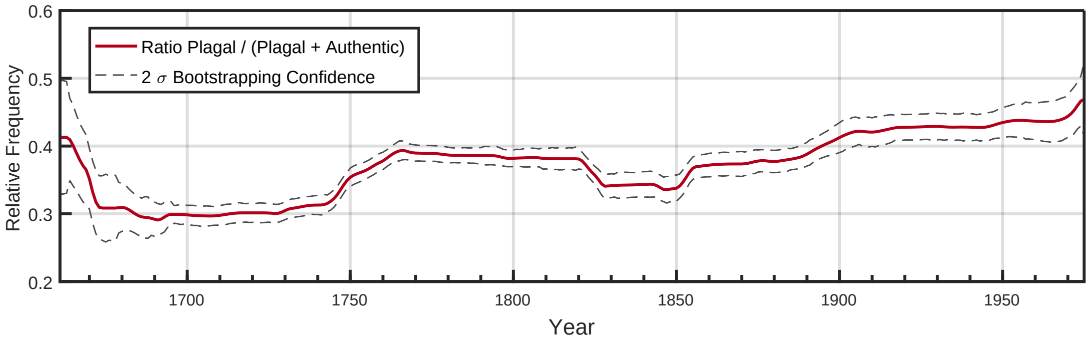

Figure 4. Evolution curve for the ratio of plagal chord transitions. The red curve displays the amount of plagal transitions compared to the total amount of plagal and authentic ones (ignoring tritone and self-transitions). The dashed error lines are calculated with a bootstrapping procedure.

in the opposite direction belong to the same category.[14] According to Gárdonyi and Nordhoff (2002), the quantitative relation between authentic and plagal transitions constitutes a useful criterion for discriminating musical styles. They claim modal harmony of the 17th century to exhibit a higher ratio of plagal transitions compared to 18th century harmony. During the 19th century, plagal transitions play an important role again (Gárdonyi & Nordhoff, 2002, p. 133).

Motivated by such hypotheses, we estimate for each recording the plagal transition occurrences by summing up the features $F_{13}, F_{14}, F_{17}, F_{20}$, and F_{21}. Similarly, we estimate the authentic transition occurrences by summing up $F_{11}, F_{12}, F_{15}, F_{18}$, and F_{19} (Table 2). We aggregate these two quantities by calculating the ratio of plagal transition occurrences to the sum of plagal and authentic transition occurrences. We then compute an EC projecting this ratio onto the timeline. Figure 4 shows the resulting EC along with confidence intervals obtained from a so-called bootstrapping procedure (Efron, 1992). The proportion of plagal transitions considerably changes over the years—from around 0.3 up to almost 0.5. Overall, we always find a lower number of plagal transitions compared to authentic ones (ratio < 0.5). This points to a high importance of chord progressions such as authentic cadences or "circle of fifths" sequences which are typical for a "functional" or "progressive" concept of harmony. Around the year 1750, we find an increase of the ratio. Around this year, the contribution of several Baroque composers disappears (J. S. Bach, Handel, and others). We conclude that the dominance of authentic transitions constitutes a criterion to discriminate late Baroque from Classical style. Between the years 1820–1850, we find a decrease of plagal transitions. In this period, works by R. Schumann and Mendelssohn contribute, among others. We speculate that the new popularity of the Baroque music in this time influenced the style of these composers.[15] Interestingly, this observation is contradictory to Gárdonyi and Nordhoff (2002), who let us expect an increase of plagal transitions in the 19th century. During the 20th century, the ratio gradually comes closer to 0.5 (equal presence of plagal and authentic transitions). This confirms our expectation of a random-like chord estimation or "measurement error," leading to an equal distribution of chord transition types. Overall, the proposed analysis technique allows for testing an existing hypothesis on a style-relevant harmonic

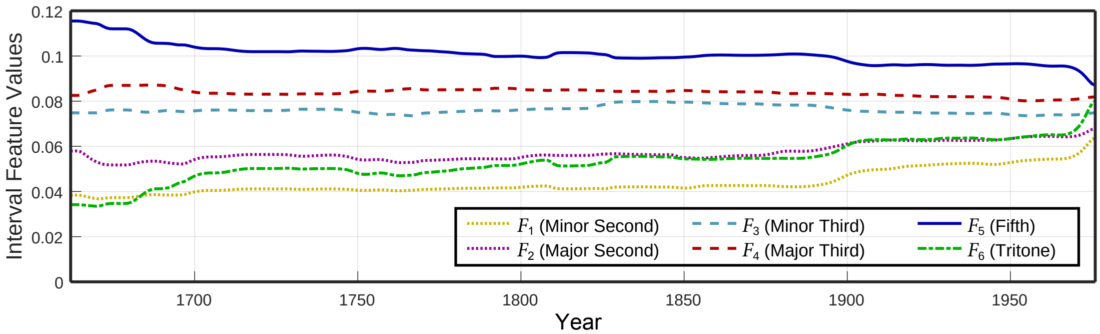

Figure 5. Interval category features distributed over the years. For the interval features, inversion (complementary intervals) cannot be resolved. For example, "Minor Third" also describes a major sixth.

phenomenon, which we could verify in partial. For detailed results showing the relevance of individual chord transitions and types, we refer to (Weiß, 2017, p. 125ff.).

Analysis of Interval Types

To analyze further aspects of tonality, we consider the measurement of interval categories (ICs), which constitutes an established analysis method (Honingh et al., 2009). Inspired by the ICs, we calculate our interval features $F_1, \ldots, F_6$ (see Table 1). Since we use a fine temporal resolution (100 ms), the features mainly refer to simultaneously sounding intervals (harmonic intervals). Figure 5 shows the ECs for these features. We observe a prominent role of the feature F_5 corresponding to perfect fifth and fourth intervals. During the 20th century, F_5 decreases and the values of the interval classes become more similar. In the 20th century, the "dissonant" categories represented by F_1 (semitone), F_2 (whole tone), and F_6 (tritone) are more frequent. We expect such a behavior since 20th century composers typically use more dissonant chords. Fucks and Lauter (1965) found similar results when statistically analyzing instrumental (violin, flute) and vocal parts based on symbolic data. They observed a prominent role of the major seventh and the minor ninth intervals—both corresponding to our F_1—in works by Schoenberg and Webern.

Analysis of Tonal Complexity

Next, we visualize measures for tonal complexity (Weiß & Müller, 2014). As described in the previous section, we calculate the complexity features $F_7, \ldots, F_{10}$ based on different chroma resolutions. We average the values and compute ECs shown in Figure 6. For all temporal resolutions, we find a general increase with time. After 1750, the complexity features decrease. This supports the composers' demand for more "simplicity" at that time, which musicologists often claim to be a paradigm for the beginning of the Classical period. During the 19th century, *global* complexity increases, whereas *local* complexity stays approximately constant. We assume that this effect originates from an increasing use of modulations—leading to a flatter global chroma histogram—whereas local structures such as chords remain less complex. This relationship changes towards the 20th century, where we observe a strong increase of complexity

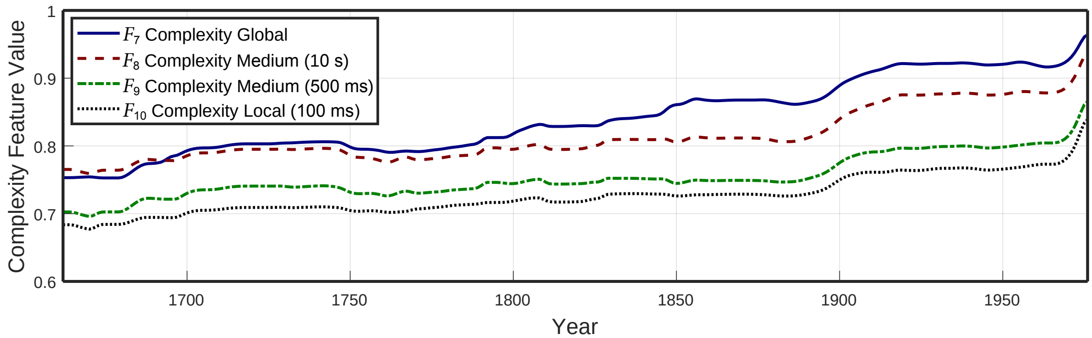

Figure 6. Tonal complexity features (lower plot) distributed over the years. The complexity features relate to different temporal resolutions of the underlying chroma features.

for all temporal scales. For the 20th century, we also find *locally* complex phenomena such as highly dissonant chords, which mainly stem from pieces by Schoenberg, Webern, and others.[16]

Style Analysis Using Data Mining Techniques

Analysis and Clustering Regarding Years

In the previous section, we directly investigated the evolution of tonal features using ECs. We showed that, at first glance, some of the observed phenomena are in accordance with hypotheses from historical musicology and music theory. We now apply data mining techniques such as feature aggregation and clustering in order to analyze the similiarity of music recordings across pieces, composers, and composition years. Assuming that our features capture some style-relevant aspects, the results of unsupervised learning strategies can provide interesting arguments for discussing the existence and borders of historical periods. These experiments are inspired by Mauch et al. (2015), who investigated the history of popular music using suitable audio features.

First, we want to focus on chord transition statistics. To this end, we individually consider the root note transition features $F_{11}, \ldots, F_{21}$, which we project onto the years with our EC method. To the eleven ECs, we perform feature aggregation (PCA) in order to analyze the importance of the individual transitions.[17] We obtain the aggregated features $F_1^*, \ldots, F_{11}^*$ (PCA scores). Furthemore, we obtain the weight vectors or *loadings* $\mathbf{w}^1, \ldots, \mathbf{w}^{11}$. The vector components indicate how much the initial features contribute to each new feature. Figure 7 shows ECs for the first three aggregated features, Table 3 lists the corresponding weights. In Figure 7, F_1^* decreases over time, capturing the difference between early periods and modern styles. Looking at the weight vector $\mathbf{w}^1$ in Table 3, we find the largest entries for the perfect fifth transitions with an emphasis on the authentic one (0.871). All components have negative signs except for perfect fifth and major second transitions—the most important transitions in tonal music.[18] Thus, F_1^* describes the presence of these "tonal transitions" in relation to all others. From 1850 on, other transitions become more frequent leading to a smaller value of F_1^*. Concerning the second component F_2^*, the corresponding weight vector $\mathbf{w}^2$ also has large values for the perfect fifth transitions but, with different signs. The plagal fifth transition has a large positive coefficient

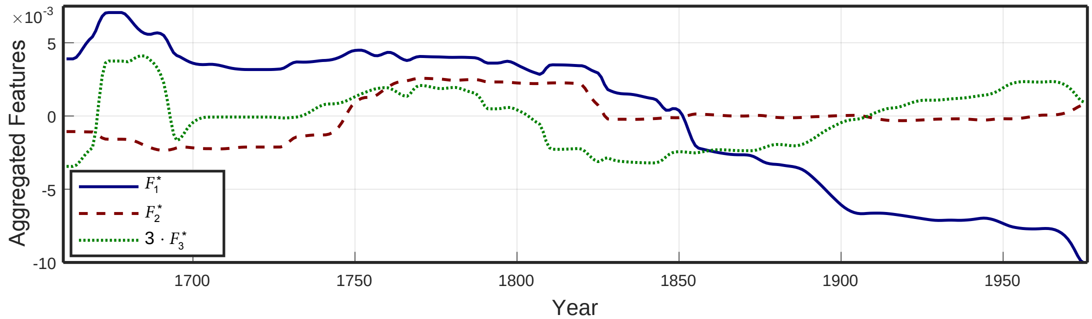

Figure 7. Aggregated features obtained from root note transitions. We display ECs for the aggregated features F_1^*, F_2^*, and F_3^* obtained from the root note transition features $F_{11}, \dots, F_{21}$. To better recognize the small component F_3^*, we multiplied it with the factor 3.

Table 3. Principal component weights for root note transitions. We re-ordered the vector entries according to plagal and authentic categories.

Feature	Interval	Δ	$\mathbf{w}^1$	$\mathbf{w}^2$	$\mathbf{w}^3$	Quality
F_{16}	Tritone ↗ | ↘	± 6	-0.138	-0.178	-0.045	None
F_{21}	Minor second ↘	-1	-0.127	-0.159	-0.012	Plagal
F_{20}	Major second ↘	-2	0.038	-0.155	0.358	Plagal
F_{13}	Minor third ↗	$+3$	-0.139	-0.039	-0.136	Plagal
F_{14}	Major third ↗	$+4$	-0.121	0.068	-0.330	Plagal
F_{17}	Perfect fifth ↗	$+7$	0.325	0.715	0.407	Plagal
F_{15}	Perfect fifth ↘	-7	0.871	-0.202	-0.418	Authentic
F_{18}	Major third ↘	-4	-0.114	-0.039	-0.250	Authentic
F_{19}	Minor third ↘	-3	-0.081	-0.125	-0.021	Authentic
F_{12}	Major second ↗	$+2$	0.199	-0.579	0.576	Authentic
F_{11}	Minor second ↗	$+1$	-0.082	-0.095	-0.087	Authentic

(0.715) whereas all authentic transitions (including the authentic fifth and second transitions) have negative coefficients. This means that F_2^* describes some kind of *difference between plagal and authentic transitions*. Looking at Figure 7, we see that F_2^* mainly distinguishes the Classical period (about 1750–1820) from the other years. In our opinion, this is a fascinating result since it stems from an *unsupervised* transformation of the transition features—without using any pre-knowledge from music theory. The EC in Figure 4, in contrast, is based on a *manual* grouping of chord transitions into plagal and authentic. We conclude that the relation between plagal and authentic transitions indeed constitutes an important style marker.

We now extend these analyses to the interval features $F_1, \dots, F_6$ and the complexity features $F_7, \dots, F_{10}$.[19] Similarly to the previous experiment, we denote the aggregated features by $G_1^*, \dots, G_{10}^*$ where G_1^* is the first principal component. The corresponding weight vectors are denoted as $\mathbf{v}^1, \dots, \mathbf{v}^{10}$. In Figure 8, we show ECs for the aggregated features. Table 4 lists the entries of the associated weight vectors. The first component G_1^* increases over the years and particularly marks the stylistic change at about 1900. Looking at the entries of $\mathbf{v}^1$ in Table 4, we see that most features have a similar absolute weight, which is an effect of the standardization.

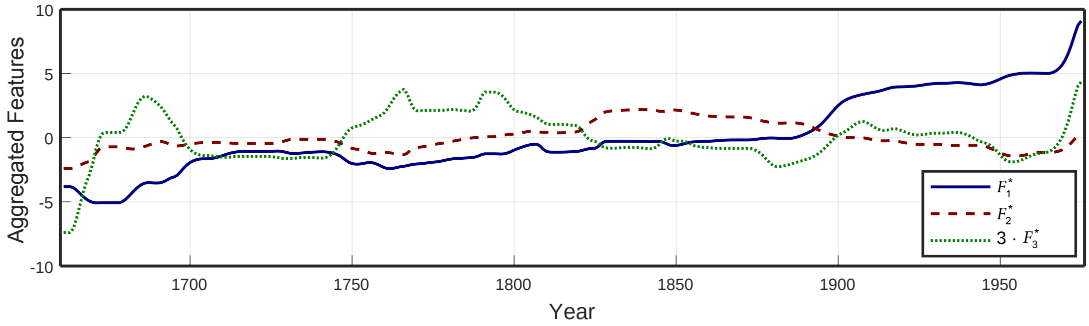

Figure 8. Aggregated features obtained from interval and complexity features. We display ECs for the aggregated features G_1^*, G_2^*, and G_3^* obtained from interval features $F_1, \ldots, F_6$ and complexity features $F_7, \ldots, F_{10}$. To improve visual recognition, we re-scaled the third component G_3^* with the factor 3.

Table 4. Principal component weights for interval and complexity features.

Feature	Feature type	$\mathbf{v}^1$	$\mathbf{v}^2$	$\mathbf{v}^3$
F_1	Interval Cat. 1 (minor second / major seventh)	0.341	−0.140	0.081
F_2	Interval Cat. 2 (major second / minor seventh)	0.334	−0.128	−0.287
F_3	Interval Cat. 3 (minor third / major sixth)	−0.087	0.881	−0.363
F_4	Interval Cat. 4 (major third / minor sixth)	−0.292	0.204	0.739
F_5	Interval Cat. 5 (perfect fourth / perfect fifth)	−0.310	−0.265	−0.424
F_6	Interval Cat. 6 (tritone)	0.336	0.197	0.149
F_7	Complexity Global (full movement)	0.335	0.174	−0.047
F_8	Complexity Mid-Scale (10 s)	0.344	−0.031	0.009
F_9	Complexity Mid-Scale (500 ms)	0.347	0.011	0.132
F_{10}	Complexity Local (100 ms)	0.344	0.077	0.110

The entries for the complexity features have positive sign indicating a correlation between G_1^* and tonal complexity, which increases over the years. The entries of $\mathbf{v}^1$ for the interval features support this assumption: Dissonant interval features (F_1, F_2, and F_6) have positive sign whereas consonant interval features (F_3, F_4, and F_5) have negative sign. Looking at the weight vector $\mathbf{v}^2$, the second feature G_2^* describes the relation between thirds (in particular, minor thirds with a weight of 0.881) and other intervals such as perfect fifths (F_5 with negative sign). Figure 8 shows that this component mainly discriminates the Romantic period (about 1825–1890) from the other years. We conclude that chords with many third intervals such as seventh or ninth chords are important for Romantic style. The positive coefficient of the tritone in $\mathbf{v}^2$ indicates an important role of diminished chords and dominant seventh chords.

We saw that chord transition statistics, interval, and complexity features may capture different aspects of style. In the following, we combine all feature types. To add more detailed information about chord transitions, we also consider specific root note transitions with respect to the chord types (major / minor type chords).[20] As before, we perform PCA based on all features $F_1, \ldots, F_{65}$ applying prior standardization. We obtain aggregated features denoted by $H_1^*, \ldots, H_{65}^*$. Based on the components H_1^*, H_2^*, and H_3^*, we automatically partition the years into segments using the unsupervised K-means clustering algorithm (MacQueen, 1967). Since the choice of K (number

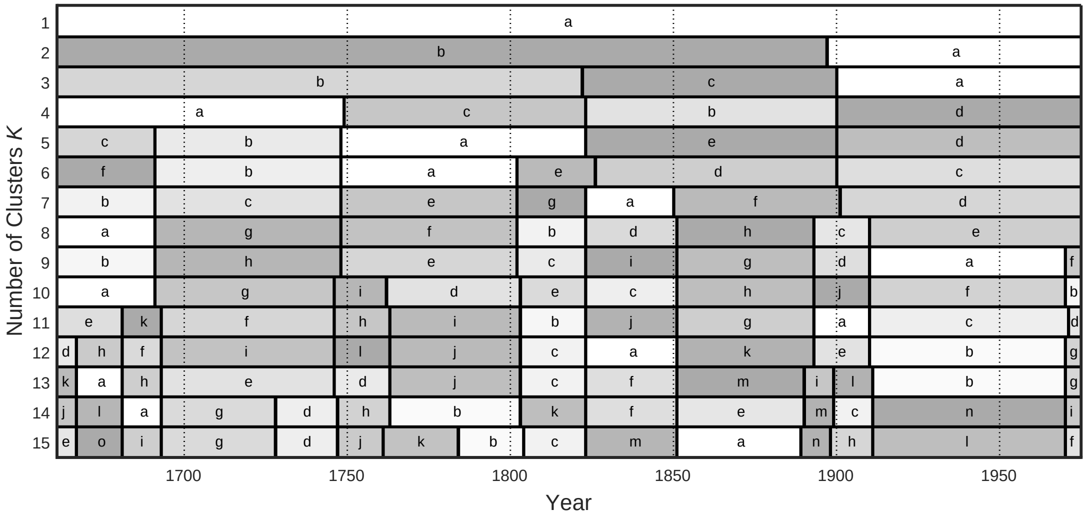

Figure 9. Clustering result for a combination of features. Based on the first three principal components from all features, we plot the cluster assignment of the years for different numbers of clusters K.

of clusters) is crucial for the result, we perform clustering for different values of K (Figure 9). We observe several stable cluster boundaries and repeating clusters, which occur for different values of K. In particular, the years 1750 and 1900 seem to play a major role for separating clusters. The boundary at 1900 bifurcates into two boundaries for $K \geq 8$. Furthermore, a boundary at 1820 seems to be important. The Baroque period splits at about 1700 for $K \geq 5$. Using $K \geq 6$, we find at least one "intermediate period" between the Classical and Romantic eras. As we mentioned before, Rodriguez Zivic et al. (2013) performed a similar clustering of years based on melodic interval statistics from sheet music data.[21] Similar to our results, they obtained stable boundaries at the years 1760, 1825, and 1895. This agreement is remarkable since the approaches crucially differ from each other. First, Rodriguez Zivic et al. use graphical scores whereas our experiment relies on audio recordings. Second, they investigate melodic descriptors where we focus on tonality. Third, the datasets are very different. We conclude that these clustering methods uncover some historical trends in musical style evolution—even though both approaches are based on various simplifications and may suffer from errors in the feature extraction step.[22]

Clustering Individual Pieces

In the introduction, we discussed the inhomogeneity and complexity of style evolution. From this point of view, our procedure—averaging all works over a year—constitutes a coarse and simplified approach. To better account for this inhomogeneity, we perform clustering using a different setting. We consider all 65 features for each of the 2000 pieces *individually* (no EC). On the resulting feature matrix, we perform PCA (after standardization). Based on the three principal components, we apply K-means clustering algorithm and then assign every piece in the dataset to one of the K clusters. We use a value of $K = 5$.[23] We then compute ECs for the

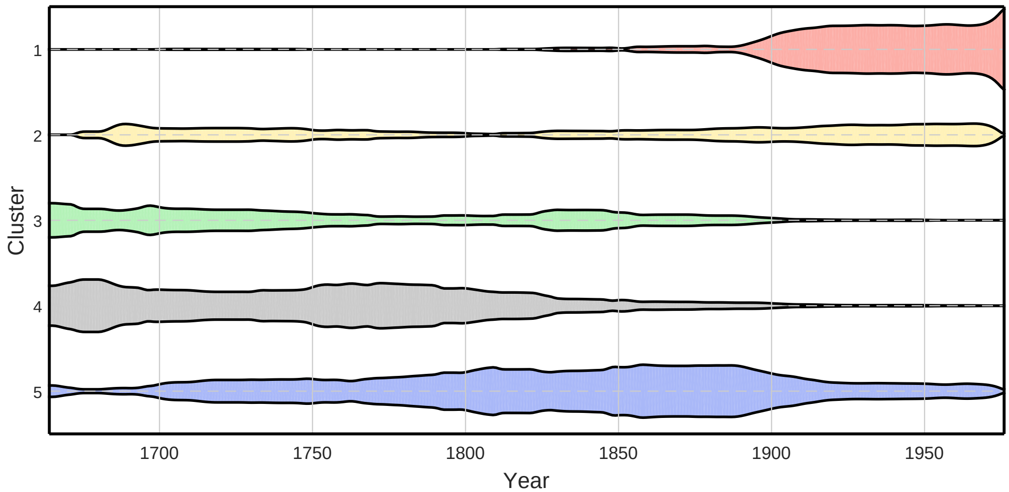

Figure 10. K-means clustering of individual pieces with $K = 5$. For each year, the fraction of pieces belonging to a cluster is indicated by the width of the respective spindle.

resulting *cluster assignments*. In Figure 10, we plot the resulting curves as spindle plots describing the fraction of pieces belonging to each cluster over the years. Compared to the previous section, the results are less clear. Cluster 1 exhibits the most extreme distribution. This cluster gradually builds up during the 19th century and plays an important role in the 20th century. We assume that this cluster is mostly characterized by atonal pieces. In the 20th century, Cluster 5 is also present, which is the most prominent cluster throughout the 19th century. The presence of Cluster 1 and Cluster 5 during the years 1910–1960 may reflect the parallelism of styles during this time. For example, Romantic pieces by Strauss and dodecaphonic pieces by Schoenberg simultaneously contribute here. Cluster 2 obtains a flat distribution over the years and, thus, is hard to interpret ("noise cluster"). Clusters 3 and 4 seem to mostly describe 17th and 18th century pieces and slowly disappear after 1850. Here, Cluster 3 is slightly more prominent for the Baroque time and contributes less to the years 1750–1820 (Classical period). This experiment shows that the situation is much less distinct when clustering pieces *before* mapping to years. The individuality of pieces appears to be stronger than the stylistic homogeneity of a period. To study this homogeneity, we show in the SMO (Section S6) an analysis of diversity over time.

Clustering Composers

Finally, we analyze the stylistic relationships between individual composers. For each of the 70 composers, we average chord transition, interval, and complexity features over all pieces by the respective composer. On the resulting feature matrix, we perform PCA followed by K-means clustering ($K = 5$) on the first three principal components. Figure 11 shows the resulting cluster assignments. Widely, composers with a similar lifetime belong to the same cluster. This points towards a fundamental relation between historical and stylistic periods. For example, Cluster 1 (green) comprises most of the Baroque composers. Single composers appear as outliers to this

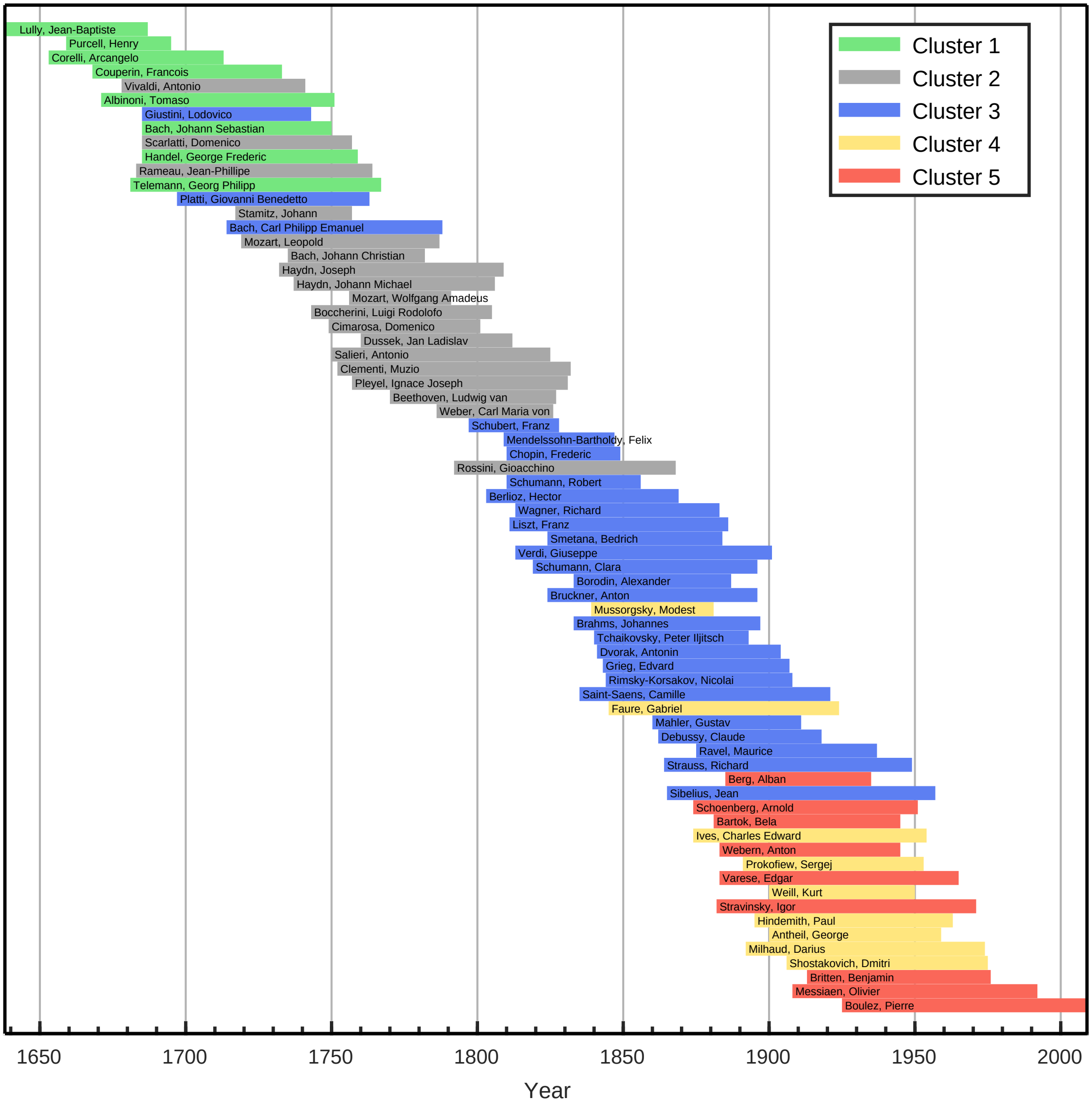

Figure 11. K-means clustering of composers with $K = 5$. The color indicates the cluster assignments.

simple partitioning. For example, Vivaldi and Scarlatti are assigned to the "Classical" group. C. P. E. Bach was assigned to the "Romantic" Cluster 3 (blue). This may be an interesting observation since some musicologists point to such a connection: "[C. P. E.] Bach's career coincided with the transition between Baroque and Classical styles, even heralding the Romantic" (Schulenberg, 2014, p. 6). Other pre-classical composers such as Stamitz or J. C. Bach are assigned to the "Classical" Cluster 2 (gray). For the change at about 1820, we find a clear separation. Beethoven, von Weber, and Rossini constitute the last Classical representatives whereas Schubert and Mendelssohn are assigned to the Romantic cluster. For the 20th century, we find two parallel clusters. Cluster 5 (red) comprises the avantgarde of that time with composers such as Schoenberg, Webern, Varèse, Bartók, or Boulez. Cluster 4 (yellow), the other modern cluster, contains composers with a moderately modern style such as Prokofiev and Shostakovich. The assignment of Mussorgsky and Faure to this cluster is rather surprising since most of the late romantic composers (Mahler, Strauss) as well as the impressionists (Debussy, Ravel) are assigned to the Romantic cluster. This kind of unexpected observations could serve as an inspiration for musicological research. Looking at these clustering results, we may arrive at a similar conclusion as White (2013) drew from his MIDI-based studies: "Although stylistic proximity was found to correlate to chronology, it also seems that stylistic norms can best be represented as groups of composers whose time periods often overlap" (White, 2013, p. 177).

Conclusion

In this paper, we presented computational methods and experiments for analyzing the evolution of Western classical music styles in a historical context. From a dataset comprising 2000 audio recordings of piano and orchestral music, we extracted different tonal features. Projecting the features onto the timeline in evolution curves, we could verify musicological hypotheses regarding chord transitions, interval types, and tonal complexity. This shows that audio-based strategies can be useful tools for analyzing musical pieces not only individually but also in a larger context. Using automated feature aggregation, tonal complexity as well as the ratio of plagal and authentic transitions arised as style markers in an unsupervised fashion. This shows the benefits of computational methods for obtaining insights that are not based on existing theories. Such experiments may serve as a source of inspiration for music research. Clustering the recordings across composers and composition years, we independently observed stable periods and boundaries in accordance with traditional views as well as recent data-driven experiments. In contrast, first clustering individual pieces and then projecting the assignments onto the timeline produced less clear results. This observation suggests that style evolution is complex and that the individuality of pieces is stronger than the stylistic homogeneity within a period. Averaging over many works by a composer seems to balance out individual pieces' characteristics and, thus, helps to uncover the composer's style. Our study pointed out how such fundamental questions might be approached using computational methods. Even though the possibilities of audio-based analysis are limited, meaningful descriptors relating to music theory can be successfully extracted from recordings. Musicological hypotheses can be used to set up and refine analysis methods with a "human in the loop." This enables corpus studies in a novel order of magnitude and, thus, has the potential to open up a new dimension for musicological research.

Acknowledgements

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. C. W. thanks

Queen Mary University of London for an extended research stay. Parts of this work were carried out at Fraunhofer Institute for Digital Media Technology IDMT in Ilmenau, Germany.

Funding

This work was supported by the German Research Foundation (DFG) within the project „Computergestützte Analyse harmonischer Strukturen“ (DFG MU 2686/7-1). Furthermore, C. W. received support by the Foundation of German Business (sdw) including the funding of a research stay in London.

Supplemental material

Tables and figures/audio files with the index "S" are available as Supplemental Online Material, which can be found attached to the online version of this article at http://msx.sagepub.com. Click on the hyperlink "Supplemental material" to view the additional files.

Notes

1. One example is the conflict about programmatic music during the 19th century.
2. For example, we think of the transition phase between late Baroque and pre-classical style at about 1730–1760.
3. "Building the database was heavily time-consuming, particularly on account of the limitations of the software needed to convert the image to digital and remove errors created by the process" (Bellmann, 2012, p. 255).
4. http://www.peachnote.com. This dataset contains statistics of melodic and harmonic progressions for individual composition years obtained from scanned sheet music with OMR techniques (Viro, 2011).
5. The MIDI files stem from the commercial platform http://www.classicalarchives.com/
6. Unsupervised learning strategies serve to find structure in unlabeled data.
7. For multi-movement works or work cycles, we count every movement as a piece/work in the dataset.
8. Parts of this dataset (1600 pieces) served as evaluation scenario for classification into four historical periods (Baroque, Classical, Romantic, Modern) published in (Weiß et al., 2014; Weiß & Müller, 2015; Weiß, 2017).
9. http://www.audiolabs-erlangen.de/resources/MIR/cross-era
10. This algorithm is published as a vamp plugin under http://isophonics.net/nnls-chroma
11. http://isophonics.net/nnls-chroma
12. Thus, a composer with more works in the dataset has a stronger influence on the EC. We decided for this weighting since otherwise—giving equal weight to all *composers*—the pieces by less prominent composers would have a disproportionate effect on the EC.
13. For the years before 1660 and after 1975, less than three composers contribute to the year-wise analysis. Thus, the EC may be heavily biased towards the pieces of individual composers.
14. Because of enharmonic equivalence in the features, we cannot assign the tritone transition (six semitones) to one of these categories (the tritone could be mapped to an augmented fourth or to a diminished fifth interval).
15. For example, many treatises on music history consider the performance of J. S. Bach's "St. Matthew Passion" conducted by Mendelssohn in 1829 as an important event.

16. For studying the complexity regarding individual composers' works, we refer to the dissertation (Weiß, 2017).
17. As for normalization, we first subtract from each row its mean value. For features of different type, a division of each row's values by the standard deviation would also be necessary. Since we have features of similar type, we do not divide by the standard deviation in order to maintain the overall influence of each chord transition type.
18. These transitions appear in typical chord progressions such as cadences (II-V-I, IV-V-I), pendula (I-V-I, I-IV-I), or sequences (I-V-VI-III-IV-I-IV-V, and the circle-of-fifths sequence), vgl. (Gárdonyi & Nordhoff, 2002; Roig-Francolí, 2011).
19. Again, we normalize the rows by subtracting their mean value before performing PCA. Furthermore, we standardize the rows so that the features values lie in the same range across all feature types.
20. In the dissertation (Weiß, 2017, p. 128), a detailed analysis of root note transitions can be found.
21. Though Rodriguez Zivic et al. (2013) know the composition dates—in contrast to our scenario—, the results are comparable to some degree since they use a smoothing window of ten years in order to balance out local outliers in the clustering results.
22. Among others, these weaknesses comprise the imperfect mapping of pieces to years, pitch and duration identification errors in OMR, the influence of overtones or vibrato on the chromagrams and, resulting from these, erroneous estimation of melodic shapes, interval types, chords and chord progressions.
23. For the composer clustering in the next section, $K = 5$ arised as optimial using the silhouette score, a method to estimate the quality of a clustering result. To enable comparability, we used the same value in this section.

References

Adler, G., & Strunk, W. O. (1934). Style-criticism. *Musical Quarterly*, *20*(2), 172–176.

Bárdos, L. (1961). *Modális harmóniák (Modal harmonies)*. Budapest, Hungary: Ed. Zenemukiadó.

Belaiev, V. (1930). The signs of style in music. *Musical Quarterly*, *16*(3), 366–377.

Bellmann, H. G. (2012). *Categorization of tonal music style: A quantitative investigation* (Doctoral dissertation). Griffith University, Brisbane, Australia.

Benetos, E., Dixon, S., Giannoulis, D., Kirchhoff, H., & Klapuri, A. (2013). Automatic music transcription: Challenges and future directions. *Journal of Intelligent Information Systems*, *41*(3), 407–434.

Byrd, D., & Simonsen, J. G. (2015). Towards a standard testbed for optical music recognition: Definitions, metrics, and page images. *Journal of New Music Research*, *44*(3), 169–195.

Cho, T., & Bello, J. P. (2014). On the relative importance of individual components of chord recognition systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *22*(2), 477–492.

Clarke, H. L. (1956). Toward a musical periodization of music. *Journal of the American Musicological Society*, *9*(1), 25–30.

de la Motte, D. (1991). *The study of harmony: A historical perspective* (J. L. Prater, Trans.). Dubuque, IA: William C. Brown. (Original work published 1976)

Efron, B. (1992). Bootstrap methods: Another look at the jackknife. In *Breakthroughs in statistics* (pp. 569–593). New York, NY: Springer.

Frank, P. L. (1955). Historical or stylistic periods? *Journal of Aesthetics and Art Criticism*, *13*(4), 451–457.

Fucks, W., & Lauter, J. (1965). *Exaktwissenschaftliche Musikanalyse.* Köln and Opladen, Germany: Westdeutscher Verlag.

Gárdonyi, Z., & Nordhoff, H. (2002). *Harmonik* (2nd ed.). Wolfenbüttel, Germany: Möseler.

Godt, I. (1984). Style periods of music history considered analytically. *College Music Symposium*, *24*(1), 33–48.

Gómez, E., & Herrera, P. (2004). Estimating the tonality of polyphonic audio files: Cognitive versus machine learning modelling strategies. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 92–95). Barcelona, Spain.

Good, M. (2006). Lessons from the adoption of MusicXML as an interchange standard. In *Proceedings of XML.* Boston, MA.

Honingh, A., & Bod, R. (2010). Pitch class set categories as analysis tools for degrees of tonality. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 459–464). Utrecht, The Netherlands.

Honingh, A., Weyde, T., & Conklin, D. (2009). Sequential association rules in atonal music. In *Mathematics and computation in music (MCM)* (pp. 130–138). Berlin and Heidelberg, Germany: Springer.

Izmirli, Ö. (2009). Tonal-atonal classification of music audio using diffusion maps. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 687–691). Kobe, Japan.

LaRue, J. (1962). On style analysis. *Journal of Music Theory*, *6*(1), 91–107.

LaRue, J. (1992). *Guidelines for style analysis.* Sterling Heights, MI: Harmonie Park Press.

Lee, K. (2006). Automatic chord recognition from audio using enhanced pitch class profile. In *Proceedings of the International Computer Music Conference (ICMC)* (pp. 306–311). New Orleans, LA.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Berkeley and Los Angeles, CA.

Mauch, M., & Dixon, S. (2010a). Approximate note transcription for the improved identification of difficult chords. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 135–140). Utrecht, The Netherlands.

Mauch, M., & Dixon, S. (2010b). Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(6), 1280–1289.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science*, *2*(5).

Müller, M. (2015). *Fundamentals of music processing.* Cham, Switzerland: Springer.

Müller, M., & Ewert, S. (2010). Towards timbre-invariant audio features for harmony-based music. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(3), 649–662.

Papadopoulos, H., & Peeters, G. (2012). Local key estimation from an audio signal relying on harmonic and metrical structures. *IEEE Transactions on Audio, Speech, and Language*

Processing, *20*(4), 1297–1312.
Pascall, R. (2001). Style. In D. Root (Ed.), *Grove music online: Oxford music online.* Oxford, UK: Oxford University Press.
Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine and Journal of Science*, *2*(11), 559–572.
Pugin, L., Kepper, J., Roland, P., Hartwig, M., & Hankinson, A. (2012). Separating presentation and content in MEI. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 505–510). Porto, Portugal.
Rodriguez Zivic, P. H., Shifres, F., & Cecchi, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences*, *110*(24), 10034–10038.
Roig-Francolí, M. A. (2011). *Harmony in context.* New York, NY: McGraw-Hill.
Sapp, C. S. (2005). Visual hierarchical key analysis. *ACM Computers in Entertainment*, *3*(4), 1–19.
Schulenberg, D. (2014). *The music of Carl Philipp Emanuel Bach.* Rochester, NY: University of Rochester Press.
Sheh, A., & Ellis, D. P. W. (2003). Chord segmentation and recognition using EM-trained hidden Markov models. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 185–191). Baltimore, MD.
Streich, S. (2006). *Music complexity: A multi-faceted description of audio content* (Doctoral dissertation). Universitat Pompeu Fabra, Barcelona, Spain.
van de Par, S., McKinney, M. F., & Redert, A. (2006). Musical key extraction from audio using profile training. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 328–329). Victoria, Canada.
Viro, V. (2011). Peachnote: Music score search and analysis platform. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 359–362). Miami, FL.
Webster, J. (2004). The eighteenth century as a music-historical period? *Eighteenth Century Music*, *1*(1), 47–60.
Weiß, C. (2017). *Computational methods for tonality-based style analysis of classical music audio recordings* (Doctoral dissertation). Ilmenau University of Technology, Ilmenau, Germany.
Weiß, C., Mauch, M., & Dixon, S. (2014). Timbre-invariant audio features for style analysis of classical music. In *Proceedings of the Joint International Computer Music Conference (ICMC) and Sound and Music Computing Conference (SMC)* (pp. 1461–1468). Athens, Greece.
Weiß, C., & Müller, M. (2014). Quantifying and visualizing tonal complexity. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)* (pp. 184–187). Berlin, Germany.
Weiß, C., & Müller, M. (2015). Tonal complexity features for style classification of classical music. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (pp. 688–692). Brisbane, Australia.
White, C. W. (2013). *Some statistical properties of tonality, 1650-1900* (Doctoral dissertation). Yale University, New Haven, CT.