Studying Tonal Evolution of Western Choral Music: A Corpus-Based Strategy

Christof Weiß[1,*], Meinard Müller[2]

[1]Center for Artificial Intelligence and Data Science (CAIDAS), Universität Würzburg, Germany
[2]International Audio Laboratories Erlangen, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany

Abstract
The availability of large digital music archives combined with significant advances in computational analysis methods have enabled novel strategies for musicological corpus studies. This includes approaches based on audio recordings, which are available in large quantities for different musical works and styles. In this paper, we take up such an audio-based approach for studying the tonal complexity of music and its evolution over centuries. In particular, we examine the tonal evolution of Western choral and sacred music exploiting a novel audio corpus (5773 tracks) with a rich set of annotations. The data stems from one of the world's leading music publisher for choral music, the Carus-Verlag, which is specialized on scholarly-critical sheet music editions of this repertoire and also runs an own record label. Based on this corpus, we revisit a heuristic strategy that exploits composer life dates to approximate work count curves over the years, validate this approximation strategy, and optimize its parameters using the reference composition years annotated in the Carus dataset. We then apply this strategy to derive evolution curves from the full Carus dataset. We compare the results to a study based on a purely instrumental dataset and test three hypotheses on tonal evolution, namely that (1) global complexity increases faster than local complexity, that (2) major keys are tonally more complex than minor keys, and that (3) instrumental music is more complex than vocal music. The results provide interesting insights into the choral music repertoire and suggest that well-curated publisher data constitutes a valuable resource for the computational humanities.

Keywords
Computational Musicology, Corpus Analysis, Musical Style Evolution, Tonal Analysis

1. Introduction

As digitization progresses, more and more comprehensive archives of cultural data become available. In combination with the further development of analysis algorithms, such archives provide promising opportunities for quantitative analyses and large-scale corpus studies in computational humanities. This also applies to music data, which exists in a variety of styles and digital data types, including graphical sheet music, symbolic (i. e., machine-readable) scores, and audio recordings. While symbolic scores, which explicitly encode musical symbols, usually allow for the most detailed analyses (as in [1, 2, 3, 4, 5, 6]), such data is hard to ac-

CHR 2023: Computational Humanities Research Conference, December 6–8, 2023, Paris, France
*Corresponding author.
christof.weiss@uni-wuerzburg.de (C. Weiß); meinard.mueller@audiolabs-erlangen.de (M. Müller)
https://go.uniwue.de/ch (C. Weiß); https://www.audiolabs-erlangen.de/fau/professor/mueller (M. Müller)
0000-0003-2143-4679 (C. Weiß); 0000-0001-6062-7524 (M. Müller)

CEUR Workshop Proceedings (CEUR-WS.org)

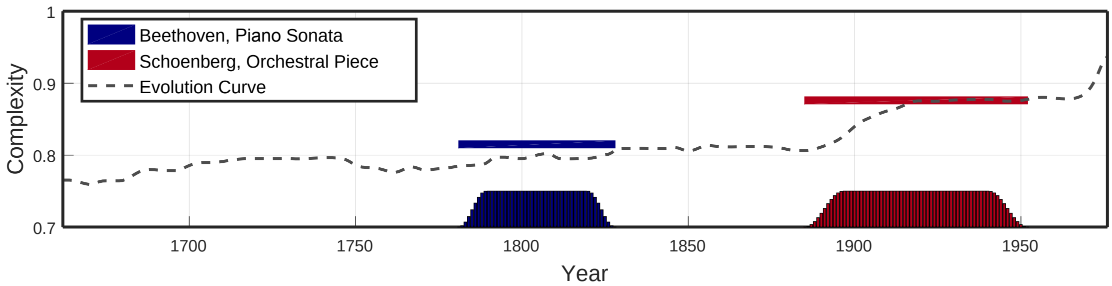

Figure 1: Approximating evolution of tonal complexity based on composer dates (figure from [15]).

quire. Manual creation of symbolic data is tedious, and automated conversion of graphical sheet music to symbolic scores known as optical music recognition (OMR) [7] or automatic music transcription (AMT) for converting audio recordings to symbolic scores [8] often lead to unsatisfactory results, thus requiring labour-intensive post-processing.

For efficiently scaling up computational music analyses, corpus-based studies have also been approached directly based on raw data such as sheet music images [9, 10] or audio recordings [11, 12, 13, 14, 15]. This requires advanced computational techniques that convert the data into semantically meaningful representations that can be directly interpreted by music experts. An example for such a representation is the measurement of tonal complexity [16], which has been applied for corpus analyses of jazz [14] and Western classical music [15] based on pitch-class representations (chroma) of audio recordings.

Beyond the computational tools, comprehensive and carefully curated datasets are essential for conducting corpus analyses [17]. While some well-annotated public datasets of limited size and scope are available (see e. g., [18, 19, 20, 21]), a good coverage of a larger repertoire is required to draw more general conclusions. However, annotations and historical metadata is often hard to acquire for large corpora. In previous work [15], we made an attempt to compile a diverse, medium-sized dataset (*Cross-Era*) of 2000 classical music recordings (piano and orchestral music) spanning roughly 350 years of Western music history. Since this dataset did not contain any fine-grained annotations of composition years (*work dates*), we proposed a workaround to map tonal analysis results onto a historical time axis ("evolution curves", compare Figure 1) based on composers' lifetime (*composer dates*). Until now, this simplifying approach has not been systematically tested on any dataset with composition year annotations.

In this paper, we approach this problem by studying the distribution of work dates over the lifetime of a composer. To this end, we consider the data repository of the *Carus Verlag*,[1] a German music publisher specializing in choral and sacred music. Carus produces high-quality editions conforming to a historical-critical standard, also employing leading musicologists with comprehensive expertise on their repertoire. Since Carus is also active as a record label releasing reference recordings of their own editions, their repository comprises a large number of audio recordings (more than 7000) with a rich set of detailed and well-curated metadata, includ-

[1]https://www.carus-verlag.com/en/

ing information about work dates, composer dates, instrumentation, singing language, key, and other annotations.[2]

Based on the Carus audio corpus (CAC), we make the following contributions in this paper. First, we revisit the heuristic strategy for approximating work count curves and evolution curves based on Tukey windows (Figure 1) proposed in [15]. We systematically validate this strategy and optimize the Tukey window parameters by comparing the approximation curves with reference curves derived from the work dates annotated in the CAC. As an exemplary application, we then consider the measurement of tonal complexity as proposed in [16], visualized over composer dates and work dates, respectively. Second, using this strategy, we perform multiple analyses regarding the tonal evolution. In contrast to [15], the CAC allows us to go beyond instrumental music and focus on vocal/choir and sacred music instead. Moreover, we consider a substantially extended time span of 450 years in CAC (as opposed to roughly 300 in [15]). Finally, the detailed annotations in the CAC allow for testing different hypotheses about the tonal complexity of Western (vocal) music, i. e.: (1) Global complexity increases earlier than local complexity. (2) Major keys are tonally more complex than minor keys. (3) Instrumental music is more complex than vocal music. The computed evolution curves provide interesting insights regarding these questions and indicate that well-curated publisher data can be of high value for the computational humanities.

The remainder of this paper is organized as follows: Section 2 presents information and statistics of the CAC. Section 3 deals with the approximation of work count curves and evolution curves, tests the validity of this strategy, and determine optimal parameters based on the reference annotations in CAC. In Section 4, we use this strategy to compute evolution curves on tonal complexity and to test three hyptheses on tonal evolution. Section 5 concludes the paper. Further related work is discussed in the respective sections.

2. The Carus Audio Corpus

The Carus-Verlag, founded near Stuttgart, Germany, in 1972 is a family business focusing on vocal and sacred music. Their sheet music editions include around 45,000 works (most of them vocal compositions) and reflect the development of five centuries of choral music, ranging from Gregorian chant, madrigals, and motets of the Renaissance, to contemporary choral music, and works for jazz and pop choir.[3] Carus offers scholarly-critical music editions of the most important oratorios, masses, and cantatas in music history, oriented towards historically informed performance practice. Being also active as a record label, Carus releases reference recordings based on their own editions. A core mission of the company is to help amateur and semi-professional choirs to improve their skills. To this end, digital tools such as the *Carus music* app have been created.

The CAC[4] comprises the majority of the Carus CD releases (as of 2019), totalling 7115 tracks corresponding to individual works (for one-movement works) or movements (for multi-

[2]Since the audio recordings are commercial releases, we cannot publish the dataset. However, detailed information about individual recordings is provided at the publisher's website (https://www.carus-verlag.com/en/).
[3]https://www.carus-verlag.com/en/ueber-carus/
[4]This corpus has been made available to us for research puposes based on a collaborative project.

Table 1
Statistics of the Carus audio corpus and its annotations. All numbers refer to full works (not individual movements).

Annotation type	*No. of works*
–All–	2409
Work date	1151
Instrumentation	1964
• instrumental	200
• vocal	1764
• choral	1400
• solo	364
Key	1166
• major	673
• minor	348
• other	145

movement works and work cycles). Since we want to focus on original art music compositions, we perform a first cleaning step where we remove works without composer, works without composer life dates, arrangements, pop music, children songs, and christmas songs. After this, 5773 tracks (movements) remain belonging to 2409 different works with a total duration of 389:52:20 (hh:mm:ss). On average, a work has 2.4 movements and a duration of 9:43 (mm:ss). However, we note that the number of movements per work is highly unbalanced, with many one-movement works on the one hand and many large-scale works (oratorios, passions, etc.) with more than 30 movements on the other hand. In the following, we present all statistics and analysis results at the *work level*, where information such as key or instrumentation always refer to the overarching work (note that e. g., a mass in C minor for choir and orchestra may also contain individual movements in other keys and instrumentations).

Table 1 provides statistics over the CAC's annotations at the work level. Roughly half of the works (1151 out of 2409) has annotations regarding the year of composition (work date). The majority (1964 out of 2409) is annotated regarding instrumentation. As expected, there is a strong focus on vocal music (1764) in general and on choral music specifically (1400 out of 1764).[5] From the perspective of tonal analysis, the availability of key annotations for roughly half of the works (1166 out of 2409) is of particular relevance. As one might expect for this repertoire, there is a bias towards major keys as well as a considerable number of other keys (church modes such as dorian in early works).

As mentioned above, CAC spans roughly 450 years, covering the period from about 1570–2020. In total, the works stem from 234 different composers. Figure 2 shows a historical view on the composer dates for composers with at least five works. Well-known composers like Felix Mendelssohn Bartholdy, Johann Sebastian Bach, or Wolfgang Amadeus Mozart make up a significant part. However, CAC also comprises less known composers such as Heinrich Schütz (featuring the complete edition) or Max Reger. Carus even makes great efforts to bring almost

[5]Please note that, due to the work-related annotations, individual solo vocal movements (e. g., an aria) within a choir work (e. g., an oratorio) are counted towards choral works.

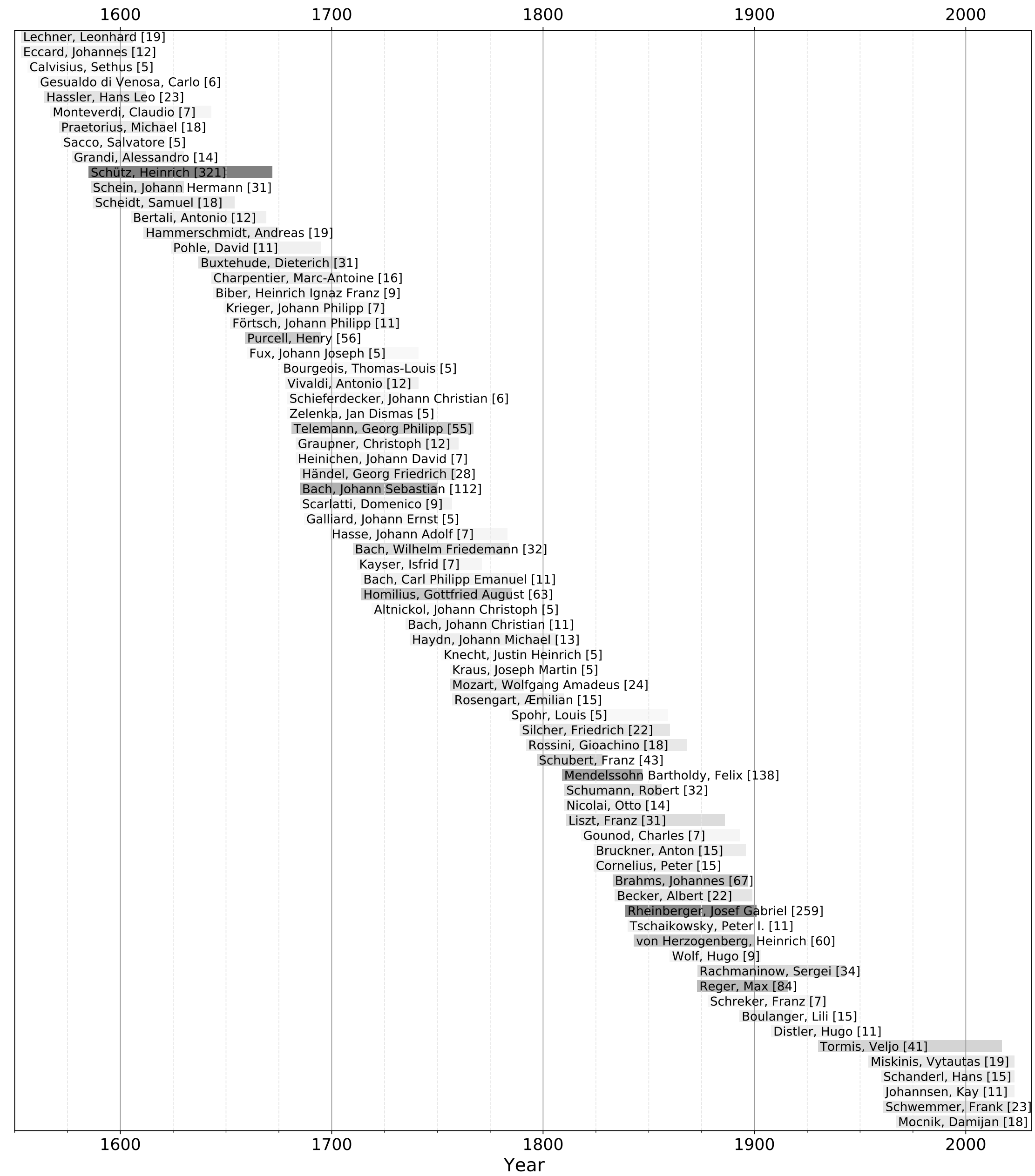

Figure 2: Historical view of CAC considering all composers with at least five works. The number of works by each composer is indicated in square brackets and encoded by the darkness of the bars.

forgotten works by Gottfried August Homilius or Josef Gabriel Rheinberger back into the focus of the German choir scene and beyond. A particular interesting fact is the good coverage of the late 15th and 16th century (which is not covered in [15]). In the 20th century, however, we find a lower number of works, almost observing a gap around 1950.

3. Approximation Strategy for Work Count Curves

We now outline the approximation of work count curves and the strategy for computing evolution curves as done in [15]. We then validate this strategy and optimize the involved parameters by comparing approximation curves based on composer dates to the reference curves based on true work dates using the annotations in CAC. In the following, we simplify all temporal information by only considering the respective year.

3.1. Work count curves

To analyze musical styles in their historical context, one ideally has information about the true *work dates*, which we assume to be the year $t_{\text{work}} \in \mathbb{N}$, where a composition was completed. Musical styles may evolve rapidly, and composing is subject to trends and influenced by other composers, the taste of audiences, or extra-musical stimuli such as political events. One might think of composers with several "creative periods," such as Ludwig van Beethoven or Arnold Schönberg. However, collecting reliable work date annotations for larger datasets requires a substantial amount of manual research, and this information is unknown or in doubt for quite a number of works. Even if one knows all composition dates, it becomes difficult to create a dataset with a balanced coverage of all years.

Because of such problems, we adopted in previous work [15] a pragmatic approach by projecting works onto the historical time axis based on *composer dates*, i. e., the information on birth year t_{birth}, death year t_{death}, and overall age $a_{\text{death}} = t_{\text{death}} - t_{\text{birth}}$, which is considerably faster to acquire. We proposed an approximation of work counts over the course of a composer's life. For this distribution, we assumed that a typical composer starts composing not before a certain (fixed) age given by $a_{\text{start}} \in \mathbb{N}$ years (with $a_{\text{start}} = 10$ in [15]). For the remaining years (ages) $[a_{\text{start}} : a_{\text{death}}] := \{a_{\text{start}}, a_{\text{start}} + 1, \ldots, a_{\text{death}}\}$, we computed a roughly flat distribution with smooth edges. To this end, we used a so-called *Tukey window* (or *tapered cosine window*) $w : \mathbb{N} \rightarrow \mathbb{R}$ with parameter $\alpha \in \mathbb{R}$:

$$w(n) = \begin{cases} 0.5\left(1 - \cos\left(\frac{2\pi n}{\alpha N}\right)\right), & 0 \leq n < \frac{\alpha N}{2} \\ 1, & \frac{\alpha N}{2} \leq n \leq \frac{N}{2} \\ w(N - n), & \frac{N}{2} < n \leq N \end{cases} \tag{1}$$

with $n = [0 : N]$ and $N = a_{\text{death}} - a_{\text{start}}$ being the window length. In [15], the parameters were heuristically chosen to a start age of $a_{\text{start}} = 10$ and a Tukey parameter of $\alpha = 0.35$. Figure 1 shows the resulting distribution for Beethoven and Schönberg. The total distribution is then amplitude-normalized to $\sum_n w(n) = 1$ and weighted with the total number of works by a composer in the dataset, resulting in a so-called *work count curve* (WCC). That way, each work contributes to the part of the time axis that corresponds to the composer's lifetime, as indicated in the distribution. This means that a composer with more works in the dataset will have a greater influence on the WCC.

3.2. Validating and optimizing the approximation strategy

In [15], the Tukey window w and its parameters were chosen heuristically without any further validation since work date annotations were not available for the dataset used. The CAC contains such annotations for roughly half of the works (compare Table 1). Using these annotations, we now validate the approximation strategy and search for optimal values of the parameters α and a_{start}. We do this in a stepwise fashion: First, we determine the start age a_{start}, i. e., the age at which we expect an average composer to start composing. To this end, we calculate the percentage of all works that were composed at a specific absolute age in years (blue curve in Figure 3a). To counteract the effect of imbalanced composition ages, we slightly smooth this curve by convolution with a 5-year kernel $\mathbf{k} = (0.1, 0.2, 0.4, 0.2, 0.1)^{\text{T}}$. Since composers have died at different ages, the red curve slowly decreases after an age of approximately 60. We then define a half Tukey window for the range $[a_{\text{start}} : 60]$ preceded by zeros (red curve in Figure 3a). For each value of $a_{\text{start}} \in [0 : 24]$, we fit the Tukey parameter α (see Eq. (1)) as well as a magnitude scaling factor using non-linear least squares. We obtain a minimal squared distance (Euclidean distance) between the curve and the half-Tukey approximation at $a_{\text{start}} = 13$ (compare Figure 3a), which is slightly higher than the value $a_{\text{start}} = 10$ used in [15].

Using $a_{\text{start}} = 13$, we now fit the window parameters for the remaining years, i. e., the interval $[a_{\text{start}} : a_{\text{death}}]$. To counteract the effects of different overall ages, we normalize the overall ages from $[a_{\text{start}} : a_{\text{death}}]$ to $[a_{\text{start}} : 60]$ by interpolating work dates accordingly followed by smoothing with the kernel $\mathbf{k}$ (blue curve in Figure 3b). Since the curve ends steeper than it begins, we allow the fitted Tukey window to cover a range $[a_{\text{start}} : 60 + a_{\text{add}}]$ (the additional years will be set to zero later). With the same fitting strategy as above (non-linear least squares), we then find an optimal value of $a_{\text{add}} = 6$. For the Tukey parameter α, we determine the optimal value to $\alpha = 0.72$, which is considerably larger than the value of $\alpha = 0.35$ used in [15]. The fitted curve is shown in red in Figure 3b.

We finally set the curve to zero for all ages > 60 and normalize the window weights such that the total weight amounts to 1. The resulting curve is shown in Figure 3c. For a given composer with final age a_{death}, we then re-normalize this window length back from the range $[a_{\text{start}} : 60]$ to $[a_{\text{start}} : a_{\text{death}}]$ by suitable interpolation.

With these optimized window parameters, we now validate the approximation strategy for the work count curve. To this end, we first compute the reference curve using the work date annotations for 1151 works that have these annotations. We post-process the curve with an average filter of length 15 years (red curve in Figure 4). We then compare this reference curve with our approximation curve based on composer dates and our optimized Tukey window (blue curve in Figure 4). Overall, the approximation seems to be suitable. In some periods (e. g., around 1680), the approximation curve is ahead, for others (e. g., at 1770), it lags behind the reference curve. In a quantitative comparison, we measure an Euclidean distance of 0.046 (averaged per year). In contrast, when using the parameters of [15], i. e., , $\alpha = 0.35$, $a_{\text{start}} = 10$, and $a_{\text{add}} = 0$, we measure an average distance of 0.068. We conclude that the approximation based on Tukey windows is a suitable strategy to compensate for missing work date annotations.

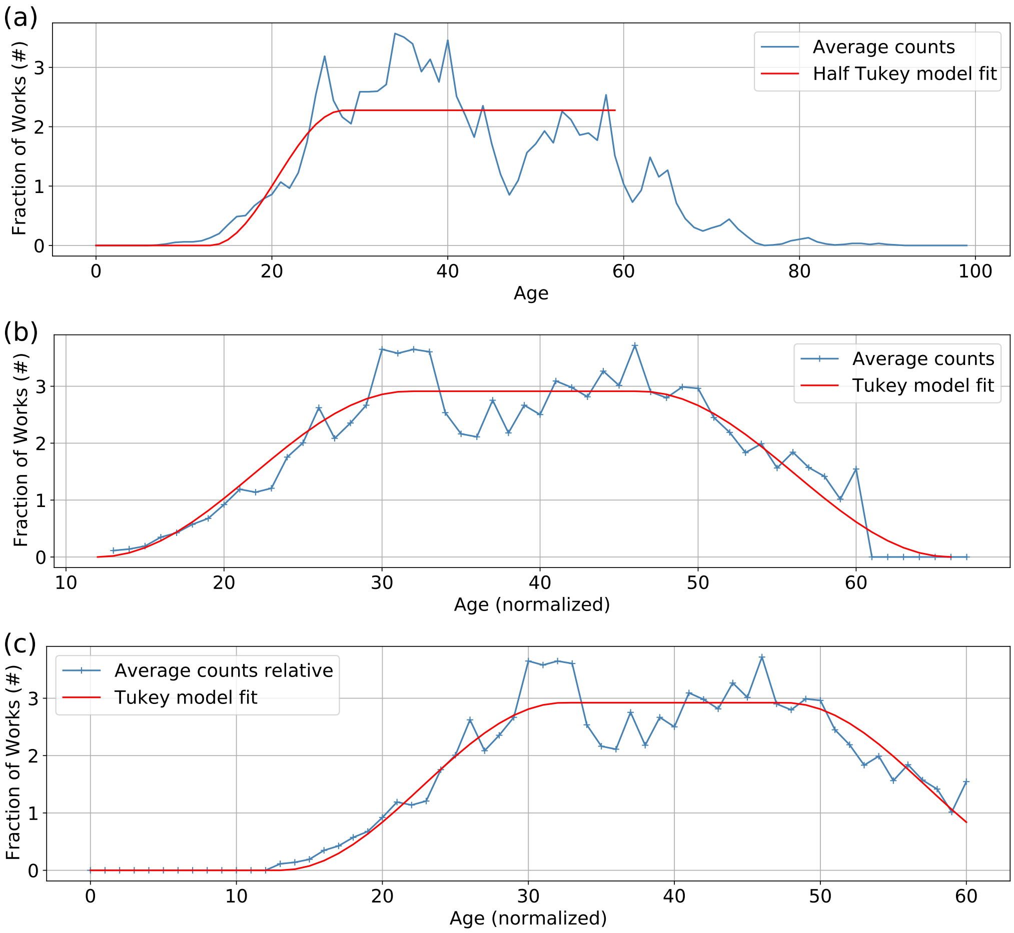

Figure 3: Curve fitting procedure to determine the optimal window parameters (a) Partial curve fit to determine the optimal start composing age $a_{\text{start}} = 13$. (b) Fit to determine optimal parameters N_{end} and α for the Tukey window w. (c) Resulting full window.

4. Studying the Evolution of Tonal Complexity

With the validated strategy, we now investigate the tonal evolution of choral music in the CAC. First, we summarize the computational approach for measuring tonal complexity from audio recordings. Then, we compare the results to the study in [15] and then use our evolution curves to test three common hypotheses about the repertoire.

4.1. Measuring tonal complexity

We now revisit the measurement of tonal complexity from audio recordings as performed in [15]. First, we discuss related work regarding complexity. Then, we present the method applied here, closely following [14].

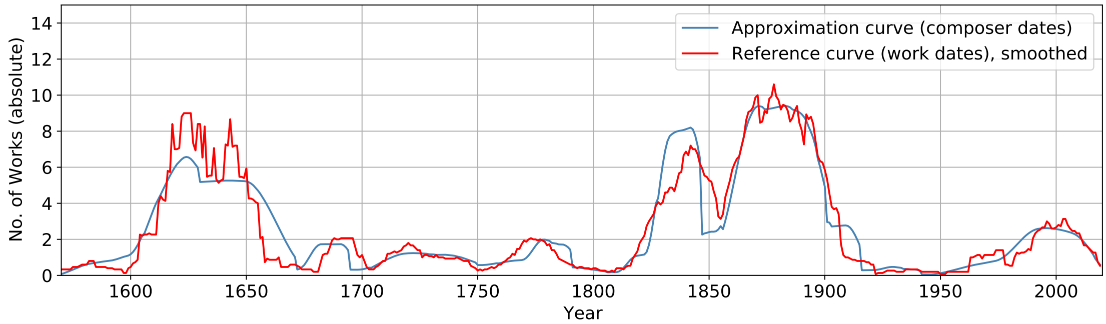

Figure 4: Work count curves based on composer dates (approximation curve, blue) and based on work dates (reference curve, red), respectively.

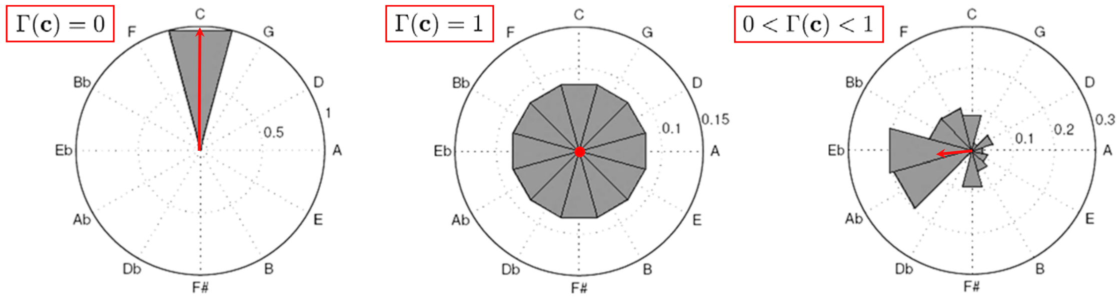

Figure 5: Complexity measure Γ based on the circle of fifths. Values for a sparse chroma vector (left), a flat chroma vector (middle), and a more realistic chroma vector (right) are shown. The red arrows denote the resultant vectors (figure from [14]).

Musical complexity is a highly relevant (yet vague and multi-faceted) notion for analysis, which has been approached by various researchers. In [22], several aspects of complexity regarding acoustic, timbral, or rhythmic properties were investigated. Concerning tonality, several authors [22, 23, 24] have focused on sequential complexity including chord sequences [23]. In contrast, we introduced in [16] tonal complexity measures that locally describe distributions of energy across the twelve chromatic pitch classes used in the Western tonal system. As one principle, these measures quantify the variety of pitch classes used such that flat distributions (e. g., chromatic clusters) result in high complexity values while sharp distributions (e. g., single notes) result in low ones (see Figure 5), thus indicating an average degree of dissonance. Such features have shown good correspondence to an intuitive understanding of tonal complexity over the course of an individual work, which we have verified on a set of chords as well as for segments of Beethoven's piano sonatas [16]. Averaging such complexity features over many works provides meaningful and stable results, which has been demonstrated by a large-scale study of musical evolution in classical music [15] and jazz [14].

Following [15, Fig. 6], we select a geometric complexity measure that accounts for the harmonic relationship between pitch classes and is capable of describing the pitch-class content on various temporal levels (fifth-width complexity, see [16]). We now summarize the definition of this measure encoded by the function $\Gamma : \mathbb{R}^{12} \rightarrow [0, 1]$. First, we extract a chroma representation from the audio data using the filter-bank method presented in [25], with a resolution of 10 Hz (ten chroma vectors per second). As a result, we obtain chroma vectors $\mathbf{c} = (c_0, c_1, \ldots, c_{11})^{\mathrm{T}} \in \mathbb{R}^{12}$ with positive entries ($c_n \geq 0$) normalized with respect to the ℓ^1-norm $\left(\sum_{n=0}^{11} c_n = 1\right)$. The entries c_n with $n \in [0:11]$ indicate the salience or energy of the twelve pitch classes C, C♯, ..., B, respectively. Because of octave invariance, the features are of a cyclic nature (a transposition results in a cyclic shift).

For computing the complexity $\Gamma(\mathbf{c}) \in [0,1]$, we map the chroma features onto the circle of fifth. To this end, we first re-order the chroma values according to perfect fifth intervals (having a size of 7 semitones) resulting in the vector $\mathbf{c}^{\text{fifth}}$:

$$c_n^{\text{fifth}} = c_{(n \cdot 7) \bmod 12}. \tag{2}$$

Based on the reordered vector $\mathbf{c}^{\text{fifth}}$, we compute circular statistics using the resultant vector $\mathbf{r}(\mathbf{c})$:

$$\mathbf{r}(\mathbf{c}) = \tfrac{1}{N} \textstyle\sum_{n=0}^{N-1} c_n^{\text{fifth}} \exp\left(\tfrac{2\pi i n \pi}{12}\right). \tag{3}$$

Then, the complexity $\Gamma(\mathbf{c})$ relates to the inverse length of $\mathbf{r}(\mathbf{c})$ and is defined as:

$$\Gamma(\mathbf{c}) = \sqrt{1 - |\mathbf{r}(\mathbf{c})|}. \tag{4}$$

This measure corresponds to the angular deviation (the circular equivalent to the standard deviation) and describes the spread of the pitch classes around the circle of fifths. Figure 5 illustrates the definition of the complexity feature and the resultant vector $\mathbf{r}(\mathbf{c})$ (in red) showing examples for three input chroma vectors $\mathbf{c}$. For a sparse vector (left), the complexity is minimal ($\Gamma(\mathbf{c}) = 0$). For a flat vector (middle), we obtain maximal complexity ($\Gamma(\mathbf{c}) = 1$). Other chroma vectors yield intermediate complexity values ($0 < \Gamma(\mathbf{c}) < 1$).

Finally, we note that there are different strategies of aggregation to track-wise (i. e., movement-wise) values. First, we define a local measure Γ_{local} by calculating $\Gamma(\mathbf{c})$ for all 10 Hz chroma vectors $\mathbf{c}$ (i. e., ten chroma vectors per second) and then averaging over these features. Second, we first compute a global chroma statistics by averaging and ℓ_1-normalizing the features and then calculating a single complexity value Γ_{global} for each movement. Aggregation to works is then done by averaging over the complexity values for all movements.

4.2. Evolution curves

In Section 3, we have studied the total number of works in CAC over the course of the years (work count curves) using the work dates or our approximation strategy based on composer dates. As an example for a quantitative analysis, we now apply these strategies to our measurements of tonal complexity as defined in Section 4.1. For the approximation curves, we again use the window parameters as determined above. For the reference curves, we use a 15-year average filter for smoothing.

While the windows for each work were weighted with the value of 1 to account for the total number of works, we now use the complexity value Γ of the respective work for weighting. We sum up all weighted windows and divide by the respective work count curve for normalization. We obtain a so-called *evolution curve* (EC) that indicates the average complexity of the works

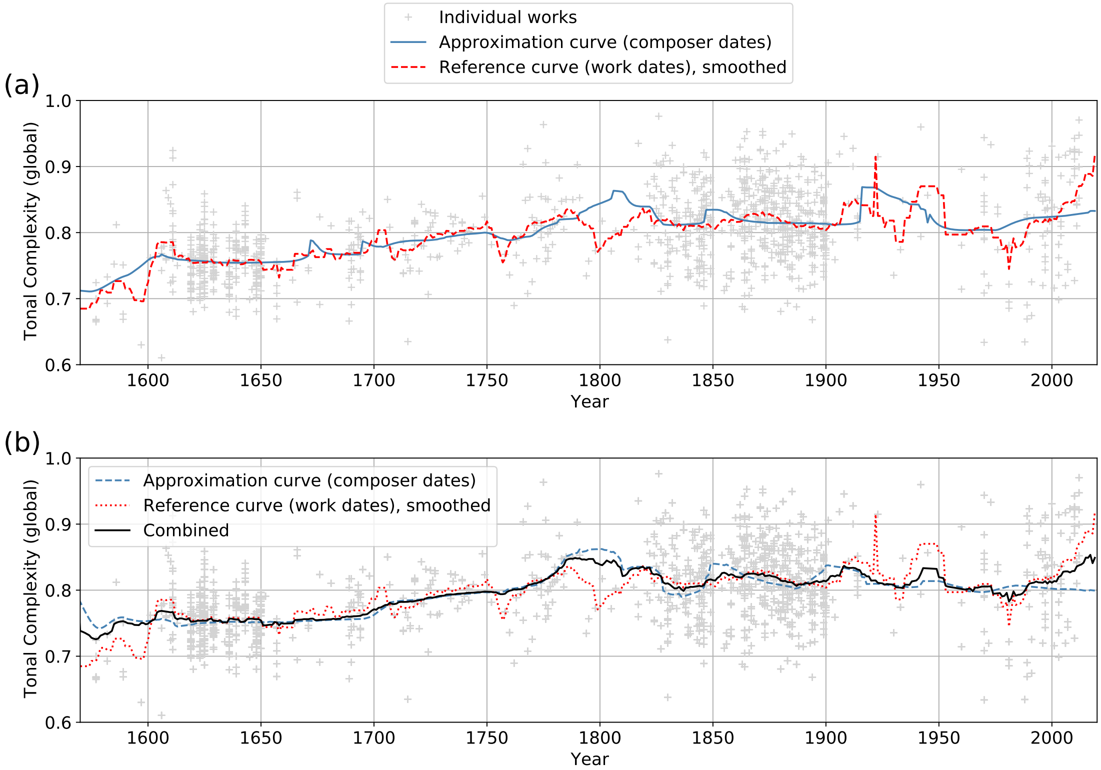

Figure 6: ECs for the global complexity. (a) Comparing ECs based on the subset D_{work} computed as approximation curve using composer dates (blue) and reference curve using work dates (red). (b) Combined EC for the global complexity in D (black) computed using work dates for D_{work} (red) and composer dates for D_{comp} (blue). Original complexity values for works are shown as gray crosses.

along the historical time axis. That way, each work contributes to the part of the time axis that corresponds to its work date (for the reference curve) or its composer's life dates (for the approximation curve).

Denoting our full dataset as D, we first consider the subset $D_{\text{work}} \subset D$ comprising all works with available work date annotations (1151 works in total). Figure 6a shows the resulting EC for the global complexity both as approximation curve (blue) and reference curve (red), together with the individual works' complexity values (gray crosses). Compared to the work count curves (Figure 4), the approximation is still good but the deviations are slightly higher. However, we observe such deviations only in regions where only few works contribute, e. g., around the years 1600, 1750, 1800, or 1920–1950. As long as there is sufficient coverage of works/composers, the approximation curve closely resembles the reference curve.

Based on this finding, we now analyze the full dataset D applying a combined strategy: For the subset $D_{\text{work}} \subset D$ (1151 works), we make use of the work date annotations and map them directly to the time axis (smoothed as above) as done for the reference curves (red curve in Figure 6b. For the subset $D_{\text{comp}} \subset D$ (1258 works), which contains the works without work date annotations, we use the mapping based on our optimized Tukey windows as done for the approximation curves (blue curve in Figure 6b). The resulting combined EC is shown as the black curve in Figure 6b. We observe a stabilized curve where minor outliers are removed (e. g., around the years 1700, 1760, or 1920) while not loosing the interesting trends.

4.3. Three hypotheses on tonal evolution

We now apply this mixed strategy for investigating the evolution of the tonal complexity in CAC, for comparing the results to those in [15], and for testing three musicological hypotheses. To this end, we use our mixed approach for computing various variants of the combined EC, always using the full dataset D.

Comparison to related work. We start with two of the combined ECs, one based on the local complexity Γ_{local} and the other based on the global complexity Γ_{global}, respectively (Figure 7). Looking at the global EC (black), we observe an increase in complexity over the course of the 17th and 18th century. Interestingly, we do not observe any drop around 1750, in contrast to [15] where the demand for more "simplicity" after the Baroque era was clearly visible (however, this trend is supported by a small number of works available for the period around 1800). On the other hand, the increase during the 19th century observed in [15] is not visible for CAC. Even more remarkably, CAC does not show any major increase in complexity during the 20th century. The modernism in tonality, pushed by expressionist and dodecaphonic composers such as Arnold Schoenberg or Igor Stravinsky, does not seem to be reflected in choral music to the same degree. This could be based on different stylistics trends in choral music, but also be a property of the CAC, where complex atonal works might not be in the focus since they are hard to be performed by amateur choirs.

Global versus local complexity. We now test different hypotheses starting with the assumption that the global complexity evolves independently from the local one. This behavior was observed in [15] especially within the 19th century, where the local complexity (refering to the complexity of e. g., chords) was fairly stable while the global complex (refering to the complexity of modulations across the whole piece) was clearly increasing. For CAC, we do not observe such a behavior. Comparing ECs for global and local complexity, we mostly observe a parallel evolution. The distance beween the curves only marginally increases after 1820. A possible reason might be the typical movement length, which can be considerably higher in instrumental works such as string quartets or symphonies, as opposed to the shorter movements of oratorios or masses. This shorter length might restrict the number and tonal distance of modulations occuring within a movement. However, this hypothesis needs further investigation.

Major versus minor keys. Our second hypothesis is based on the observation that minor keys usually exhibit more chromatic inflections as compared to major keys. To this end, we consider the data subset with key annotations (major and minor) and compute an EC for each of them (Figure 8). For the global complexity (solid lines), both curves follow a similar trend. However, we see a small but consistent offset of the minor curve (red) over the major curve (green). This confirms our hypothesis that minor keys use a larger pitch-class range and, thus, are tonally more complex. For the local complexities (dashed curves), we do not observe this offset. Moreover, for the 20th century, we see some fluctuating behavior, which is due to the

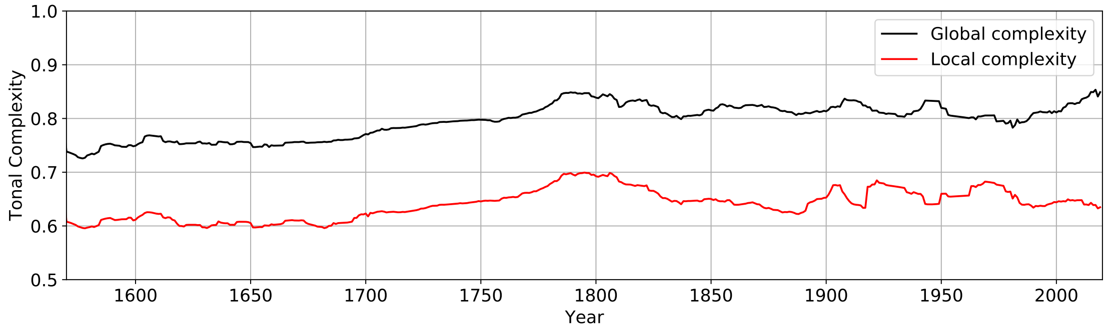

Figure 7: Comparing ECs for global and local complexity.

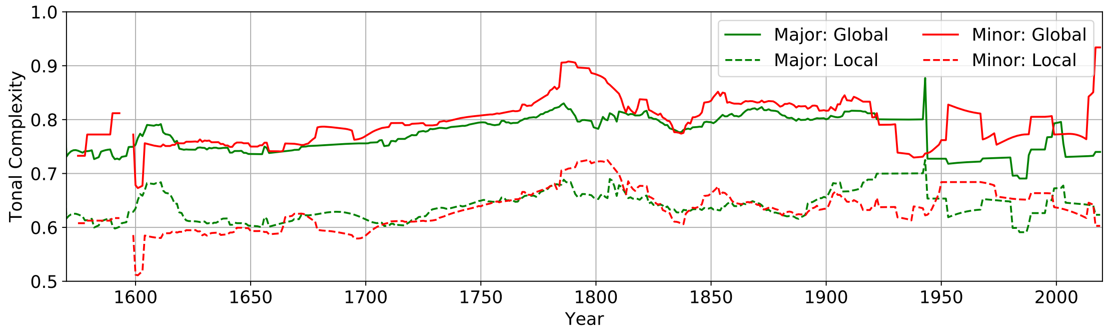

Figure 8: Comparing ECs for global and local complexity separated into major and minor keys.

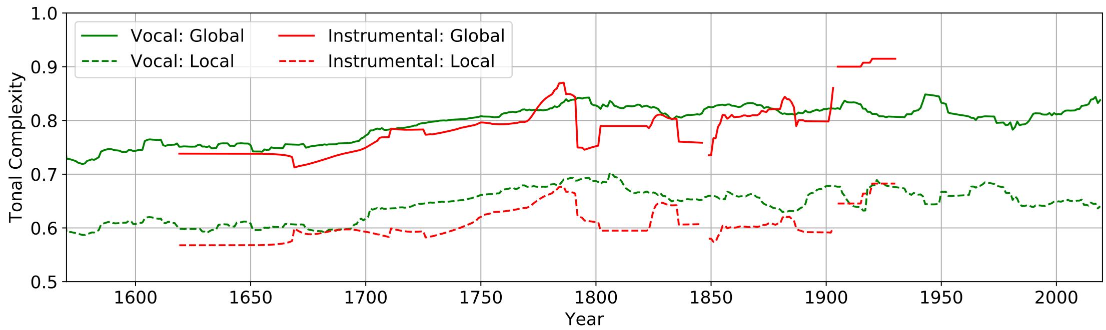

Figure 9: Comparing ECs for global and local complexity separated into vocal and instrumental music.

fact that there is little data with key annotations for that period (for atonal and free tonal music, key is often not a relevant concept).

Vocal versus instrumental music. Next, we investigate the hypothesis that instrumental music is more complex than vocal music. We expect such behavior since vocal compositions need to account for the higher difficulty in producing pitches when singing, especially for large and complex intervals. Moreover, musicologists often claim that compositional "revolutions" were often happening in compact instrumental settings such as the string quartet. To test our hypothesis, we use the instrumentation annotations and compute a vocal as well as an instrumental EC (Figure 9). As a downside of CAC, we find an unbalanced situation (compare

Table 1), resulting in a small number of works available for the instrumental EC. Nevertheless, we observe a clear tendency that contradicts our hypothesis: Vocal music seems to be more complex than instrumental music for most time periods. In particular, the offset is large for the local complexity (dashed lines). However, we suspect a technical reason for this behavior. Our chroma features are based on a signal processing approach, which maps frequency components extracted from audio recordings to the twelve chroma bands. When dealing with recorded vocal music, this process often leads to substantial artifacts since pitch stability is much lower than for instruments and effects such as vibrato, portamento, or typical deviations from the twelve-tone equal temperament (pure tuning) substantially blur the chromagrams. This can lead to quasi-chromatic artifacts that may push the complexity measurements even locally.

5. Conclusions and Future Work

In this paper, we considered an approach for studying the tonal evolution of music based on partially annotated corpora of music audio recordings. As our first contribution, we revisited and validated a strategy for computing work count curves, compensating for missing work composition dates using heuristics based on composer life dates as an approximation. To this end, we exploited the novel Carus audio corpus (CAC), which contains work date annotations for a substantial part of the works. We showed that a good choice of the parameters helps to minimize the deviations of the approximation curve from the reference curve. On this basis, we performed a combined approach for computing evolution curves that map musical features onto the time axis. This strategy allowed us to compare the CAC with previous studies and to test three hypotheses on tonal complexity in this repertoire. In our future work, we plan to substantially extend, improve, and deepen these studies. In particular, we want to investigate potential technical reasons for higher complexity measurements in choral music. To this end, more recent chroma extraction strategies based on deep neural networks are of high potential since they have shown to be successful for deriving tonal information from vocal recordings by reducing typical artifacts [26]. Beyond that, a combination of the analysis based on CAC with other datasets such as the one in [15] will provide better insights into the evolution of tonal music and allow for testing further hypotheses. Finally, this paper also aimed for providing high-level insights into the CAC. While the analyses revealed that a very good coverage of the time period under investigation is crucial for obtaining reliable results and that additional data might be beneficial for some periods, we see a high potential of such well-curated publisher datasets for studies in computational musicology and beyond.

Acknowledgments

We cordially thank the Carus-Verlag Stuttgart (Johannes Graulich and Ester Petri) for enabling the study of the Carus audio corpus. This work was supported by the German Research Foundation within the project “Computational Analysis of Harmonic Structures” (DFG MU 2686/7-2). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

References

[1] D. Temperley, An algorithm for harmonic analysis, Music Perception: An Interdisciplinary Journal 15 (1997) 31–68.

[2] H. G. Bellmann, Categorization of Tonal Music Style: A Quantitative Investigation, Ph.D. thesis, Griffith University, Brisbane, Australia, 2012.

[3] F. C. Moss, M. Neuwirth, D. Harasim, M. Rohrmeier, Statistical characteristics of tonal harmony: A corpus study of beethovenś string quartets, PLOS ONE 14 (2019). doi:10.1371/journal.pone.0217242.

[4] H. V. Koops, A. Volk, W. B. de Haas, Corpus-based rhythmic pattern analysis of ragtime syncopation, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Málaga, Spain, 2015, pp. 483–489.

[5] C. W. White, Some Statistical Properties of Tonality, 1650-1900, Ph.D. thesis, Yale University, New Haven, Connecticut, USA, 2013.

[6] E. Nakamura, K. Kaneko, Statistical evolutionary laws in music styles, Scientific Reports 9 (2019). doi:10.1038/s41598-019-52380-6.

[7] J. Calvo-Zaragoza, J. Hajič Jr., A. Pacha, Understanding optical music recognition, ACM Computing Surveys 53 (2020). doi:10.1145/3397499.

[8] E. Benetos, S. Dixon, Z. Duan, S. Ewert, Automatic music transcription: An overview, IEEE Signal Processing Magazine 36 (2019) 20–30. doi:10.1109/MSP.2018.2869928.

[9] V. Viro, Peachnote: Music score search and analysis platform, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Miami, Florida, USA, 2011, pp. 359–362.

[10] P. H. Rodriguez Zivic, F. Shifres, G. A. Cecchi, Perceptual basis of evolving Western musical styles, Proceedings of the National Academy of Sciences 110 (2013) 10034–10038. doi:10.1073/pnas.1222336110.

[11] F. Scherbaum, M. Müller, S. Rosenzweig, Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966, in: Proceedings of the International Workshop on Folk Music Analysis (FMA), Málaga, Spain, 2017, pp. 29–36.

[12] J. Abeßer, K. Frieler, E. Cano, M. Pfleiderer, W. Zaddach, Score-informed analysis of tuning, intonation, pitch modulation, and dynamics in jazz solos, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (2017) 168–177. URL: https://doi.org/10.1109/TASLP.2016.2627186. doi:10.1109/TASLP.2016.2627186.

[13] M. Mauch, R. M. MacCallum, M. Levy, A. M. Leroi, The evolution of popular music: USA 1960–2010, Royal Society Open Science 2 (2015). doi:10.1098/rsos.150081.

[14] C. Weiß, S. Balke, J. Abeßer, M. Müller, Computational corpus analysis: A case study on jazz solos, in: Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR), Paris, France, 2018, pp. 416–423. doi:10.5281/zenodo.1492439.

[15] C. Weiß, M. Mauch, S. Dixon, M. Müller, Investigating style evolution of Western classical music: A computational approach, Musicae Scientiae 23 (2019) 486–507. doi:10.1177/1029864918757595.

[16] C. Weiß, M. Müller, Quantifying and visualizing tonal complexity, in: Proceedings of the Conference on Interdisciplinary Musicology (CIM), Berlin, Germany, 2014, pp. 184–187.

[17] X. Serra, Creating research corpora for the computational study of music: The case of the CompMusic project, in: Proceedings of the AES International Conference on Semantic Audio, London, UK, 2014.
[18] M. Neuwirth, D. Harasim, F. C. Moss, M. Rohrmeier, The Annotated Beethoven Corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets, Frontiers in Digital Humanities 5 (2018) 16.
[19] J. Hentschel, M. Neuwirth, M. Rohrmeier, The annotated mozart sonatas: Score, harmony, and cadence, Transactions of the Intational Society for Music Information Retrieval (TISMIR) 4 (2021) 67–80. doi:10.5334/tismir.63.
[20] S. Rosenzweig, F. Scherbaum, D. Shugliashvili, V. Arifi-Müller, M. Müller, Erkomaishvili Dataset: A curated corpus of traditional Georgian vocal music for computational musicology, Transactions of the International Society for Music Information Retrieval (TISMIR) 3 (2020) 31–41. doi:10.5334/tismir.44.
[21] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, X. Serra, Corpora for music information research in Indian art music, in: Proceedings of the Joint Conference 40th International Computer Music Conference (ICMC) and 11th Sound and Music Computing Conference (SMC), Athens, Greece, 2014.
[22] S. Streich, Music Complexity: A Multi-Faceted Description of Audio Content, Ph.D. thesis, University Pompeu Fabra, Barcelona, Spain, 2007.
[23] B. Di Giorgi, S. Dixon, M. Zanoni, A. Sarti, A data-driven model of tonal chord sequence complexity, IEEE/ACM Transactions on Audio, Speech, and Language Processing 25 (2017) 2237–2250. doi:10.1109/TASLP.2017.2756443.
[24] M. Mauch, M. Levy, Structural change on multiple time scales as a correlate of musical complexity, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Miami, Florida, USA, 2011, pp. 489–494.
[25] M. Müller, F. Kurth, M. Clausen, Audio matching via chroma-based statistical features, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), London, UK, 2005, pp. 288–295. doi:10.5281/zenodo.1416800.
[26] C. Weiß, G. Peeters, Training deep pitch-class representations with a multi-label CTC loss, in: Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Online, 2021, pp. 754–761.