


Style evolution in Western choral music: A corpus-based strategy

Benjamin Henzel¹ , Meinard Müller² and Christof Weiß¹

¹Center for Artificial Intelligence and Data Science (CAIDAS), Julius-Maximilians-Universität Würzburg, Germany and

²International Audio Laboratories Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Research Article

Cite this article: Henzel Benjamin, Meinard Müller and Christof Weiß. 2025. "Style evolution in Western choral music: a corpus-based strategy" *Computational Humanities Research*, 1:e13, <https://doi.org/10.1017/chr.2025.10016>

Received: 31 January 2025

Revised: 9 July 2025

Accepted: 23 September 2025

Keywords:

computational musicology; corpus analysis;
cultural evolution; tonal analysis

Corresponding author:

Benjamin Henzel;

Email: benjamin.henzel@uni-wuerzburg.de

Abstract

This article introduces a strategy for the large-scale corpus analysis of music audio recordings, aimed at identifying long-term trends and testing hypotheses regarding the repertoire represented in a given corpus. Our approach centers on computing evolution curves (ECs), which map style-relevant features, such as musical complexity, onto historical timelines. Unlike traditional approaches that rely on sheet music, we use audio recordings, leveraging their widespread availability and the performance nuances they capture. We also emphasize the benefits of pitch-class features based on deep learning, which improve the robustness and accuracy of tonal complexity measures compared to traditional signal processing methods. Addressing the frequent lack of exact work dates (year of composition) in historical corpora, we propose a heuristic method that aligns works with timelines using composers' life dates. This method effectively preserves historical trends with minimal deviation compared to using actual work dates, as validated against available metadata from the Carus Audio Corpus, which spans 450 years of choral and sacred music and contains 5,729 tracks with detailed metadata. We demonstrate the utility of our strategy through case studies of this corpus, showing how ECs provide insights into stylistic developments that confirm expectations from musicology, thus highlighting the potential of computational studies in this field. For example, we observe a steady increase in tonal complexity from the Renaissance through the Baroque period, stable complexity levels in the 19th and 20th centuries, and consistently higher complexity in minor-key works compared to major-key works. Our visualizations also reveal that vocal music was more complex than instrumental music in the 18th century, but less complex in the 20th century. Finally, we conduct comparative analyses of individual composers, exploring how historical and biographical contexts may have influenced their works. Our findings highlight the potential of this strategy for computational corpus studies in musicological research.

Plain Language Summary

In this study, we use computer-based methods to analyze long-term changes in Western music through audio recordings. Unlike traditional studies that rely on sheet music, our approach takes advantage of the accessibility of audio recordings, which also capture performance details that are missing in written music. Our main technical contribution is the introduction of evolution curves, which track changes in musical features such as the tonal complexity over time. We generate these curves using data-driven deep-learning models to analyze the audio, providing a more accurate and reliable method than established signal processing techniques. To handle the challenge of missing work dates, we estimate when the music was created based on the composers' life spans. We test our method using the Carus Audio Corpus, a real-world dataset with over 5,700 tracks spanning 450 years of mostly sacred and choral music. Our analysis reveals trends such as an increase in tonal complexity from the Renaissance to the Baroque period, differences between vocal and instrumental music, and how composers like Schütz, Bach and Mendelssohn adapted their styles over time. This study highlights the benefits of computational methods by objectively testing hypotheses about musical trends, rather than simply offering new musicological insights. In summary, we show how computational tools can analyze large music collections and uncover long-term changes in musical styles, offering a fresh perspective on music history through data and technology.

Introduction

As digitization progresses, more and more comprehensive archives of cultural data become available.¹ In combination with the further development of computational methods, such archives provide promising opportunities for quantitative analyses and large-scale corpus studies in

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



¹This article is an extension of Weiß and Müller (2023), A. Sela, F. Jannidis, and I. Romanowska, Eds., ser. CEUR Workshop Proceedings, vol. 3558, CEUR-WS.org, 2023, pp. 687–702.

computational humanities. Within the context of computational musicology, this article presents a strategy for analyzing style evolution in a corpus of Western choral music. As the central idea of this strategy, we project style-relevant features of musical works onto a historical time axis resulting in *evolution curves* (ECs), which let us study the repertoire over multiple centuries of music history.

Such endeavors rely on the availability of suitable data, which – in the case of music – 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 Bellmann (2012), Koops, Volk, and Bas de Haas (2015), Moss et al. (2019), Nakamura and Kaneko (2019), Temperley (1997), and White (2013), such data are often difficult to obtain in a digital format. Manual creation of digital symbolic scores is tedious, and the automated conversion of graphical sheet music to symbolic scores known as optical music recognition (Calvo-Zaragoza, Hajič Jr, and Pacha 2020) as well as the conversion of audio recordings into symbolic scores known as automatic music transcription (Benetos et al. 2019) often lead to unsatisfactory results, thus requiring labor-intensive post-processing.

To efficiently scale up computational music analyses, corpus-based studies have also been approached directly using raw data such as scanned sheet music images (Rodríguez Zivic, Shiffres, and Cecchi 2013; Viro 2011) or audio recordings (Abeßer et al. 2017; Mauch et al. 2015; Scherbaum, Müller, and Rosenzweig 2017; Weiß et al. 2018; Weiß et al. 2019).

In this article, we conduct a case study of corpus analysis based on audio recordings, using a dataset provided by the *Carus-Verlag*,² a German music publisher specializing in choral and sacred music (Weiß and Müller 2023). 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 5,700 tracks) with a rich set of detailed and well-curated metadata, including information about work dates (i. e., the year of composition), composer dates, instrumentation, singing language, key and other annotations.³ The Carus Audio Corpus (CAC) covers a large time span of about 450 years, which allows for analyzing the development of Western choral music over several centuries.

Performing corpus analyses on audio recordings requires advanced computational techniques that convert the data into semantically meaningful representations, which can be easily interpreted by music experts. Traditional methods rely on signal processing (SP) to measure activations of the 12 chromatic pitch classes (pitches without octave information and enharmonic equivalence, see the section “Pitch-class concept”) in audio recordings. However, these techniques have limitations due to the complex nature of audio data, which encodes not only score-related aspects but also performance nuances and acoustic phenomena. In recent years, deep-learning (DL) methods have significantly improved upon these limitations (Bittner et al. 2017; Korzeniowski and Widmer 2016; Weiß et al. 2021b; Weiß and Müller 2024; Weiß and Peeters 2021; Weiß and Peeters 2022). As a contribution to

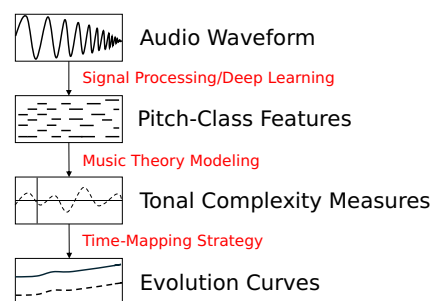


Figure 1. Computational strategy for deriving evolution curves on tonal complexity from music audio recordings.

this article, we use a recently proposed DL approach (Weiß and Peeters 2022) for estimating pitch-class activations, trained on aligned score–audio pairs. From these pitch-class representations, we compute style-relevant measures, such as *tonal complexity* (Weiß and Müller 2014), which has been applied in analyses of jazz (Weiß et al. 2018) and Western classical music (Weiß et al. 2019). While this term is technically defined, it corresponds to an intuitive understanding of tonal complexity, particularly in terms of its behavior over the course of a musical work (see the section “Measuring tonal complexity”).

Beyond individual works, our strategy allows us to compute these measures for large-scale corpora like the CAC, enabling the study of music evolution across broad historical time periods. To achieve this, we project the tonal complexity measures of each work onto a historical timeline using work date annotations, resulting in ECs that capture the stylistic evolution of Western choral music (see Figure 1). However, this projection requires exact work dates for each composition, which are hard to acquire. In many cases, as for the CAC, this is not just a problem of incomplete data, but also of incomplete *knowledge*, that is, the exact years of composition (*work dates*) are uncertain or entirely unknown. To address this issue, Kase, Sobotková, and Hermánková (2023) recently proposed to use date ranges to derive probabilistic distributions for historical dates, which may be a promising method to apply to our corpus in the future. Weiß et al. (2019) proposed a workaround to compute ECs using *composer dates* as a proxy, projecting unknown works onto temporal windows within a composer’s life span. This approach allows us to include all items in the corpus, even those with missing work dates or incomplete metadata, enabling the analysis of the entire CAC. To simplify, we assume that composers remain at a stable level of tonal complexity throughout their lives. However, this assumption has not been verified and is an area for further investigation in this work.

This article extends our previous CHR conference paper (Weiß and Müller 2023), where we introduced the CAC and validated a heuristic strategy for approximating work count curves (WCCs) and ECs using a Tukey window (Weiß et al. 2019). In this study, we significantly enhance that work in several important ways. First, following Weiß and Müller (2023), we optimize the Tukey window parameters by comparing the approximation curves with reference curves derived from the work dates in the CAC. Second, going substantially beyond this work, we replace the traditional SP approach with a DL method for extracting pitch-class features from the CAC’s audio recordings, demonstrating the advantages of this improved technique with a visual example. These enhanced representations lead to more reliable ECs, enabling us to revisit the hypotheses from Weiß and Müller (2023). Moreover, we expand

²<https://www.carus-verlag.com/en/>

³Since the audio recordings are commercial releases, we cannot publish the dataset. However, detailed information about individual recordings along with 30s audio examples is provided at the publisher’s website (<https://www.carus-verlag.com/en/>).

the experiments by testing new hypotheses regarding (1) the relationship between the length of a work (or movement) and its tonal complexity and (2) the stylistic evolution of individual composers. To illustrate this, we plot the relative complexity deviation of composers across their lifetimes and conduct case studies on Heinrich Schütz, Johann Sebastian Bach and Felix Mendelssohn Bartholdy, examining how their music evolved in response to their personal biographies and historical contexts.

The remainder of this article is structured as follows. In the section “The Carus Audio Corpus,” we introduce the CAC used for our experiments. The section “Pitch-class representations and tonal complexity” describes the methods for estimating pitch classes from audio recordings, improvements brought by DL techniques and the tonal complexity measure. In the section “Strategies for computing evolution curves,” we discuss the computational pipeline used to derive ECs. The section “Hypothesis testing using evolution curves” revisits the hypotheses studied in previous work (Weiß and Müller 2023). In the section “Case studies: complexity deviation and stylistic trends of individual composers,” we investigate the style evolution of individual composers through three case studies. The section “Conclusion” concludes the article.

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 chants, madrigals and motets of the Renaissance, to contemporary choral music, and works for jazz and pop choir.⁴ Carus offers scholarly critical music editions of the most important oratorios, masses and cantatas in music history, oriented toward 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 CAC⁵ comprises the majority of the Carus CD releases (as of 2019), totaling 7,115 tracks corresponding to individual works (for one-movement works) or movements (for multi-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’s songs and Christmas songs. After this, 5,729 tracks (movements) remain belonging to 2,401 different works with a total duration of 320:42:33 (hh:mm:ss). While on average, a work consists of 2.4 movements, 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. Statistics of the Carus Audio Corpus (CAC) and its annotations

Annotation type	No. of works
–All–	2,401
Work date	1,153
Instrumentation	1,956
• instrumental	200
• vocal	1,756
• choral	1,392
• solo	364
Key	1,166
• major	673
• minor	348
• other	145

Note: All numbers refer to full works (not individual movements).

Table 1 provides statistics of the CAC’s annotations at the work level. Roughly half of the works (1,153 out of 2,401) has annotations regarding the work date. The majority (1,956 out of 2,401) is annotated regarding instrumentation. As expected, there is a strong focus on vocal music (1,756) in general and on choral music specifically (1,392 out of 1,756).⁶ From the perspective of tonal analysis, the availability of key annotations for roughly half of the works (1,166 out of 2,401) is of particular relevance. As one might expect for this repertoire, there is a bias toward major keys as well as a considerable number of other keys (church modes such as Dorian in early works).

As mentioned above, the 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 10 works. Well-known composers like Felix Mendelssohn Bartholdy, Johann Sebastian Bach, or Wolfgang Amadeus Mozart make up a significant part. However, the 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 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 centuries (which is not covered in Weiß et al. (2019)). In the 20th century, however, we find a lower number of works, almost observing a gap around 1950.

The analysis of audio recordings – as provided by the CAC – offers several advantages over alternatives such as the analysis of sheet music or symbolic data. Although the automatic conversion of sheet music into a symbolic representation (OMR) has made some progress in recent years, perfect annotations by automatic methods are still rare. Audio recordings, on the other hand, are well suited for being processed by computational methods and offer a quick alternative for the collection of large amounts of data. Another advantage is that audio recordings contain additional

⁴<https://www.carus-verlag.com/en/ueber-carus/>

⁵This corpus has been made available to us for research purposes based on a collaborative project, see <https://www.audiolabs-erlangen.de/fau/professor/mueller/projects/anchor>.

⁶Please 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 toward choral works.

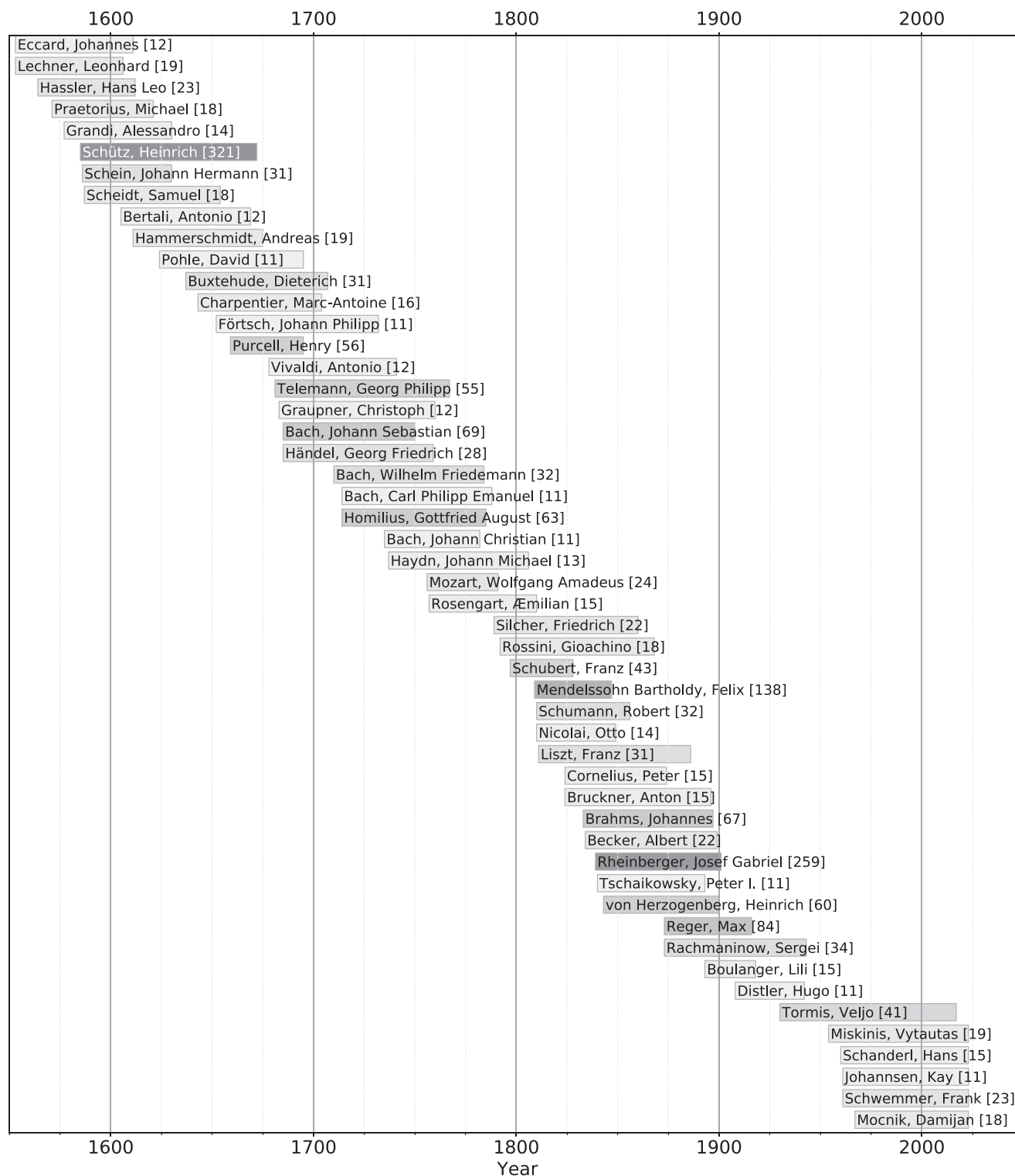


Figure 2. Historical view of the CAC considering all composers with at least 10 works. The number of works by each composer is indicated in square brackets and encoded by the darkness of the bars.

performance-related information. While composers had certain ideas about the sound of their compositions, and professionals may pick up these ideas by reading a score, they are often not explicitly reflected in the sheet music. For example, the timbral characteristics of different instruments or the audibility of solo parts are not indicated in a score, but are immediately apparent to the listener

of an audio recording. Although the usage of audio recordings for computational analysis has other limitations, which we will discuss in the section “Pitch-class representations and tonal complexity,” examining them allows us to take advantage of these factors. In this regard, the CAC offers exceptional coverage and quality, making it ideal for a case study to demonstrate our proposed strategy.

Pitch-class representations and tonal complexity

To study aspects of style evolution within the CAC, we consider as a central measure the tonal complexity (Weiß and Müller 2014), which quantifies the richness and variability of harmonic structures. Derived from pitch-class representations, which reflect the energy distribution across the 12 chromatic pitch classes, tonal complexity offers insights into both local (chord-level) and global (modulation-level) stylistic properties. In this study, pitch-class features are extracted directly from audio recordings, leveraging their broad availability. SP techniques, such as the Constant-Q Transform and other time–frequency transforms, facilitate the extraction of pitch-class representations from audio. DL methods are able to substantially improve such representations, largely overcoming challenges such as noise, percussive components, overtones or timbral variations, thus improving robustness and accuracy. These advancements lead to more reliable complexity measures and enable large-scale, data-driven studies of musical evolution, bridging computational methods with musicological insights. This section introduces the concept of pitch classes (see the section “Pitch-class concept”), details their computation from audio recordings using SP (see the section “Computation from audio recordings”), and explores how DL enhances their reliability (see the section “Enhancements using deep learning”). From these features, we derive a tonal complexity measure (see the section “Measuring tonal complexity”) and examine its relationship to musical factors such as the movement length (see the section “Tonal complexity and movement length”).

Pitch-class concept

In order to perform analysis on large corpora of audio recordings, we need a suitable mid-level representation that is not only semantically meaningful, but also serves as a basis for the computation of further musically informed features such as tonal complexity. By estimating the occurrence of the 12 chromatic pitch classes (pitches without octave information and enharmonic equivalence, i. e., C, C♯, D, ...) in audio recordings, we obtain an appropriate and pragmatic solution to these requirements. First of all, pitches are (at least for most of the time) foundational to Western classical music. While we lose information by reducing to pitch classes, we obtain robust features that carry most of the necessary information about chords and harmonies (on a local level) as well as keys and

modulations (on a global level). In the next step, we can extract this information by computing subsequent measures such as tonal complexity.

Computation from audio recordings

For several years, researchers have applied SP techniques to audio signals in order to measure the relative energy of the occurring pitch classes, resulting in so-called chroma features. As discussed in Weiß et al. (2021a) and Weiß and Peeters (2021), this approach has drawbacks. When applied to real music audio recordings, these features quickly reach their limits due to audio-related properties such as overtones and vibrato and become noisy, thus limiting the interpretability of the subsequent analysis. This is especially evident in vocal music, where characteristics such as a strong vibrato are particularly pronounced. Another problem is that the relative energy of a pitch class affects how it is represented in the resulting chroma features rather than detecting their presence independently of the dynamics. Although several researchers (Klapuri2008; Mauch and Dixon 2010; Müller and Ewert 2010) have proposed improved SP methods to minimize certain drawbacks for specific tasks in the field of music information retrieval, this has often resulted in a trade-off that worsens other properties of the pitch-class features based on SP techniques.

Enhancements using deep learning

In recent years, DL methods have shown substantial progress toward overcoming these problems. For this reason, in this article, we make use of a recently proposed machine-learning approach (Weiß et al. 2021a; Weiß and Peeters 2022) – a deep convolutional neural network with residual connection (ResNet) of medium size (4.8M parameters). The network takes as input a music-specific spectral audio representation (harmonic constant-Q transform, HCQT, see Bittner et al. (2017)) with a resolution of roughly 43 frames per second.

Following Weiß and Müller (2024), we train this network on several datasets comprising various instrumentations and styles. Table 2 provides a detailed description of the training datasets, which comprise a substantial amount of chamber music as well as some choir and orchestra music, thus encompassing all relevant instrumentations present in the CAC (see the section “The Carus Audio Corpus” for details). As training task, the network needs to

Table 2. Training datasets for the DL method to predict pitch-class activations from audio, following (Weiß and Müller 2024)

ID	Dataset name	Style / Instrumentation	Pitch annotation strategy	Mix tracks	Works	Versions	hh:mm
MuN	MusicNet (Thickstun, Harchaoui, and Kakade 2017)	Chamber music (piano, strings, winds)	Aligned scores	330	306	1 up to 3	34:08
SWD	Schubert Winterreise (Weiß et al. 2021b)	Chamber music (piano, solo voice)	Aligned scores	216	24	9	10:50
Tri	TRIOS (Fritsch and Plumbley 2013)	Chamber music (piano, strings, winds)	Multi-track	5	5	1	0:03
B10	Bach10 (Duan, Pardo, and Zhang 2010)	Chamber music (violin, winds)	Multi-track	10	10	1	0:06
PhA	PHENICX-Anechoic (Miron et al. 2016)	Symphonic (orchestra)	Multi-track	4	4	1	0:10
CSD	Choral Singing Dataset (Cuesta et al. 2018)	A cappella (choir)	MIDI-guided performance	3	3	1	0:07

Note: For work cycles, we count each part/movement as a work. We further report the number of versions (performances) per work for multi-version datasets.

predict the activation of the 12 chromatic pitch classes for each audio frame. The annotations used for training are derived from symbolic scores aligned to the individual performances. Applying the trained network to the audio data of the CAC (corresponding to the cross-dataset experiment by Weiß and Müller (2024)), we obtain a 12-dimensional vector of pitch-class activations for each frame. For further details on the network and training process, we refer to Weiß et al. (2021a) and Weiß and Peeters (2022).

In Figure 3, we show a short example from the CAC, representing J. G. Rheinberger's *Abendlied* (op. 69, no. 3), a well-known six-part a cappella composition, in a recording by Frieder Bernius and the Kammerchor Stuttgart.⁷ Specifically, we look at the first five measures, which correspond to the first 20 seconds of the audio recording. Comparing the two chromagrams reveals that the latter, derived from DL, appears to be smoother, as pitch classes that are sustained in the piece are represented as continuous boxes. One potential explanation for this observation is that our improved pitch-class features exhibit increased stability, as energy fluctuations exert less influence on the estimated pitch classes. Consequently, the presence of such blocks is more likely due to this enhanced stability. Most importantly, the vibrato of the voices, which leads to shifting frequencies and thus to fluctuations between neighboring pitch classes, has no effect on the DL-based pitch-class features.

Another observation is that some pitch classes are much clearer in the second chromagram, while they are barely visible in the first one. In particular, the pitch classes D, G, A and B in measures 3 and 4 are almost only visible in the chromagram based on DL. We explain this by the fact that SP techniques measure the relative energy of the pitch classes, which means that softly sung pitches can hardly be measured. In this case, there are two reasons: First, third intervals (over the root) are traditionally sung softer to increase the harmonic stability of chords, and second, due to the structure of the harmonic series, thirds are even less pronounced in the spectrum than octaves and fifths. It is therefore understandable that the B in measure 4 is sung by two voices (Soprano 2 and Tenor 2), but as a third interval only in such a subtle form that it is barely visible in the upper chromagram. The same applies to the pitches D and A in measure 4, which cannot prevail over the dominant C and F in the other voices.

Finally, we notice that for the initial silence in the recording, the pitch-class features based on SP are noisy, which is not the case with the DL-based ones. Since there is no perfect silence in a real recording, SP methods measure the noisy distribution of the remaining audio contents. In contrast, the DL model has learned when there is no presence of a pitch class, that is, when there is silence.

Measuring tonal complexity

Musical complexity is a highly relevant (yet vague and multifaceted) notion for analysis, which has been approached by various researchers. Streich (2007) investigated several aspects of complexity regarding acoustic, timbral or rhythmic properties. Concerning tonality, several authors (Di Giorgi, Zanoni, and Sarti 2017; Mauch and Levy 2011; Streich 2007) have focused on sequential complexity including chord sequences (Di Giorgi, Zanoni, and Sarti 2017). In contrast, Weiß and Müller (2014) introduced tonal complexity

measures that locally describe distributions of energy across the 12 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 4), thus roughly indicating a *degree of dissonance*. Weiß and Müller (2014) showed such measures to have high correspondence to an intuitive understanding of tonal complexity over the course of an individual work, verified on a set of chords as well as for segments of Beethoven's piano sonatas. Averaging such complexity features over many works provides meaningful and stable results regarding long-term trends, which has been demonstrated by a large-scale study of musical evolution in classical music (Weiß et al. 2019) and jazz (Weiß et al. 2018). In a similar fashion, White (2023) successfully applied a complexity measure based on chord entropy to the Yale - Classical Archives Corpus (YCAC).

Following Weiß et al. (2019, Figure 6), we select a geometric complexity measure that accounts for the harmonic relationship between pitch classes across the circle of fifths and is capable of describing the pitch-class content on various temporal levels (fifth-width complexity, see Weiß and Müller (2014)). We now summarize the definition of this measure encoded by the function $\Gamma: \mathbb{R}^{12} \rightarrow [0, 1]$. We first extract 12-dimensional chroma vectors from the audio recording using the DL-method described in the section "Enhancements using deep learning." To enable comparability with the 10 Hz chroma features used in Weiß and Müller (2023), we downsample the DL-based pitch-class predictions by a factor of 4, yielding a resolution of roughly 10.7 Hz. As a result, we obtain chroma vectors $\mathbf{c} = (c_0, c_1, \dots, c_{11})^T \in \mathbb{R}^{12}$ with positive entries ($c_n \geq 0$), which we normalized with respect to the ℓ^1 -norm ($\sum_{n=0}^{11} c_n = 1$). The entries c_n with $n \in [0 : 11]$ indicate the salience or energy of the 12 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 $\mathbf{c}^{\text{fifth}}$:

$$c_n^{\text{fifth}} = c_{(n \cdot 7) \bmod 12}. \quad (1)$$

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

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

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

$$\Gamma(\mathbf{c}) = \sqrt{1 - |\mathbf{r}(\mathbf{c})|}. \quad (3)$$

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 4 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

⁷This performance is part of the CAC, and an audio preview is available at https://www.carusmedia.com/images-intern/medien/80/8333600/8333600_09s.tl_014.mp3.

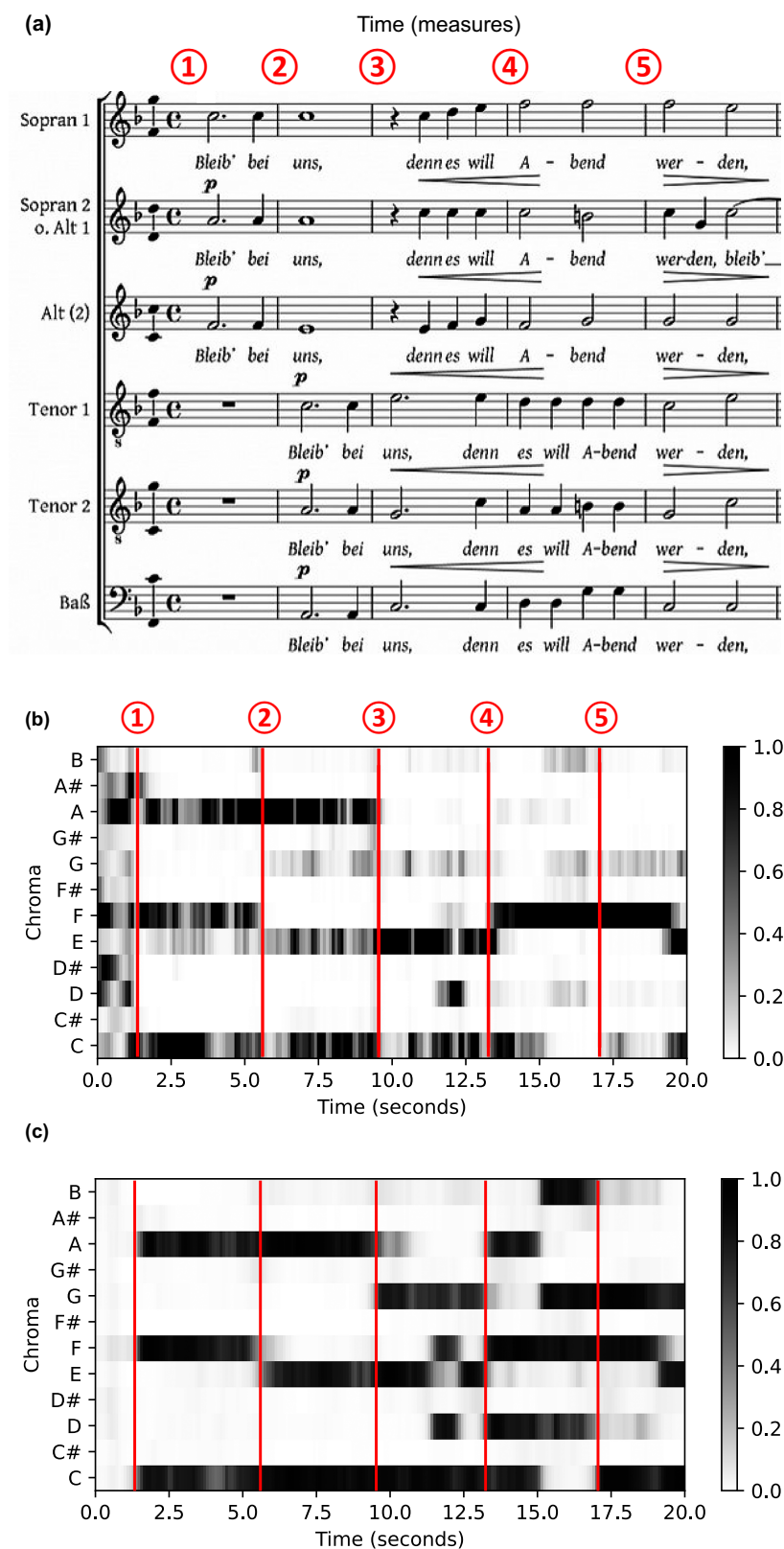


Figure 3. Example pitch-class features for an excerpt of Rheinberger's *Abendlied*, op. 69, no. 3, from the CAC: (a) Score excerpt. (b) Pitch-class features based on SP. (c) Pitch-class activations computed with DL.

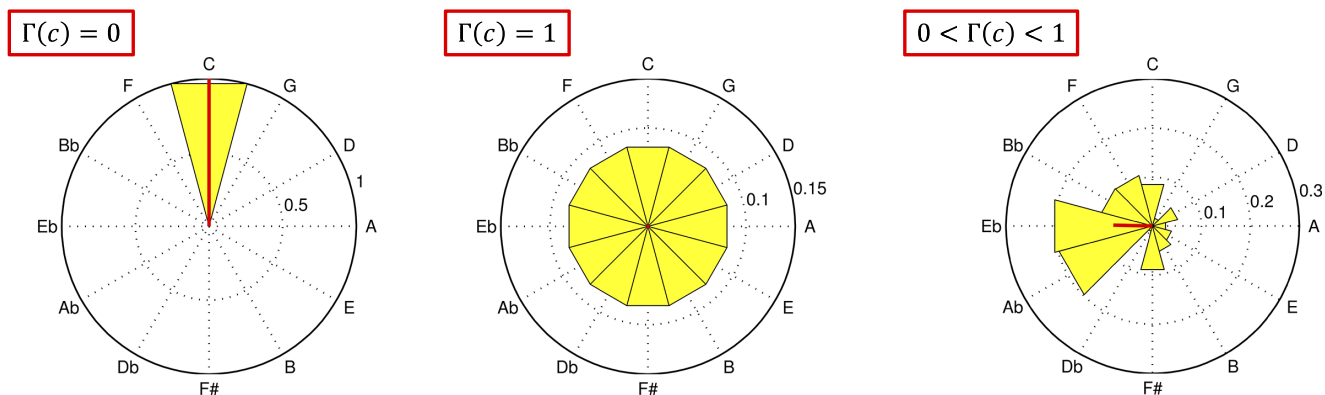


Figure 4. Complexity measure Γ based on the circle of fifths. Values for a sparse chroma vector (left), a flat chroma vector (middle), and a more realistic chroma vector (right) are shown. The red arrows denote the resultant vectors (figure from Weiß et al. 2018).

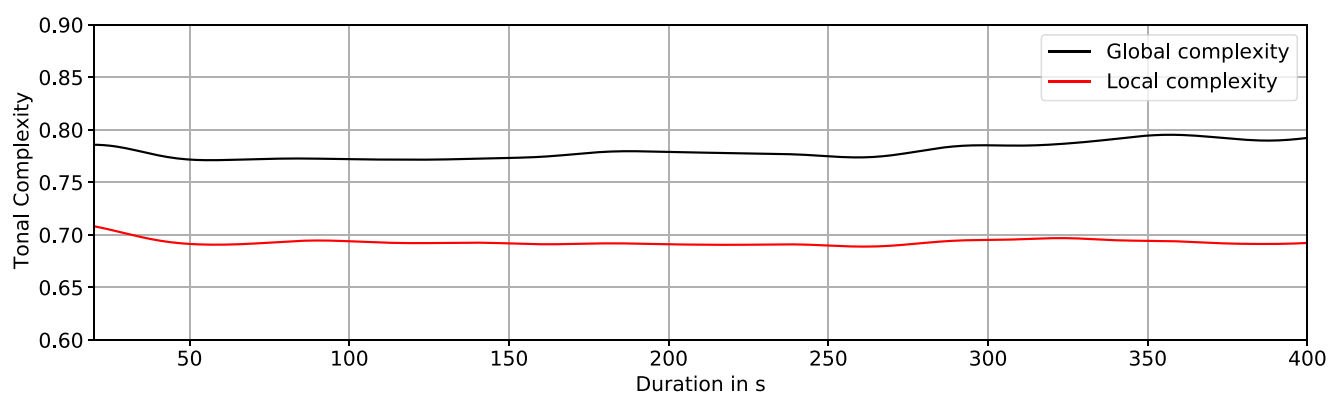


Figure 5. Relationship between the duration of audio recordings and their corresponding tonal complexity.

measure Γ_{local} by calculating $\Gamma(\mathbf{c})$ for all 10.7 Hz chroma vectors \mathbf{c} (i. e., 10.7 chroma vectors per second on average) 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.

From a musicological perspective, this measure of tonal complexity provides a robust approach to capturing tonal and stylistic changes. We can visualize tonal complexity measures to track these changes over a range of values relative to a time axis, whether over the course of a single movement or over a historical timeline based on many works.

Tonal complexity and movement length

Musical notions such as complexity and dissonance are important means for structuring musical compositions, allowing for building up tension and release. As a consequence, longer works or movements may exhibit different characteristics regarding their tonal complexity. Moreover, key changes and modulations crucially influence the global pitch class distribution. Longer movements typically modulate more often and to tonally more distant keys, which may result in a higher global complexity. To better understand such effects on the tonal complexity measure, we now examine the relationship between the length of a movement (audio duration in seconds) and its tonal complexity. We study this relationship for the entire CAC comprising 5,729 tracks with an

average recording length of about 200 seconds. We exclude tracks with durations below 20 seconds and above 400 seconds since such durations occur sparsely, which leads to unstable results as individual tracks have a strong influence on the curve.

Figure 5 shows the results of this experiment. We make two observations: Global complexity slightly increases as the duration of an audio recording increases. This confirms our assumption that, the longer a track or a movement is, the more modulations, and thus, different pitch classes can occur, leading to higher global complexity values. This explanation is also consistent with our local complexity curve, which does not increase with a longer duration. We expect such a behavior, since a longer duration should have no substantial influence on the chord types and harmonic intervals used, which is what we capture by the local complexity. Although the increase in global complexity is relatively small, this could be a characteristic to consider in future studies. For example, complexity values could be normalized according to the length of the corresponding work.

Strategies for computing evolution curves

We now outline the strategy for computing ECs as done in Weiß et al. (2019) and Weiß and Müller (2023). In an ideal case, we can do this based on annotated *work dates*. However, since the CAC provides work dates only for a fraction of works, we additionally apply an approximation strategy as in Weiß et al. (2019), which is based on *composer dates* (birth and death year). We validate

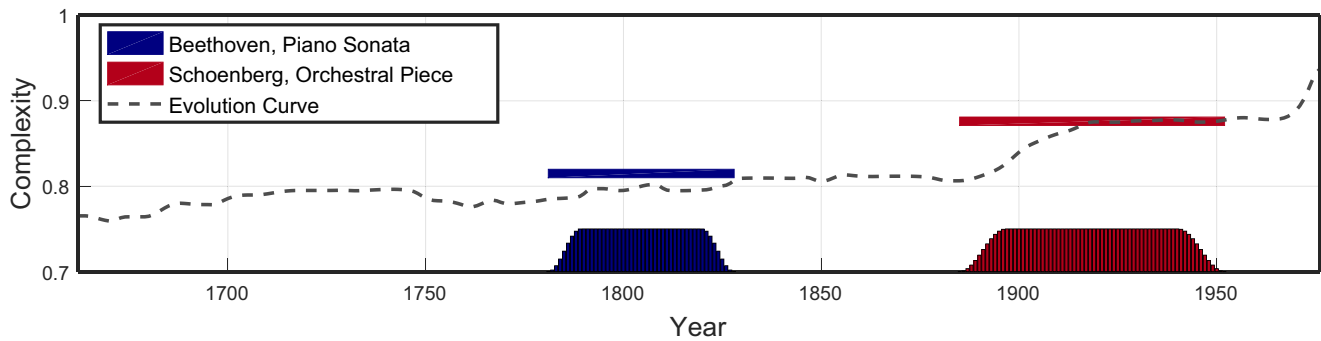


Figure 6. Approximating evolution of tonal complexity based on composer dates (figure from Weiß et al. 2019).

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 the CAC. In the following, we simplify all temporal information by only considering the respective year.

Computation of evolution 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. Using the work date annotations in the CAC, we can project the compositions, or more precisely, the complexity measures computed from the works, onto historical timelines. This allows the identification of long-term trends, limited only by the content, coverage and representativeness of a corpus. Based on the resulting ECs, we can then test hypotheses on topics related to global developments that occur over long periods of time. By relying on the CAC, our pitch-class estimation technique and our tonal complexity measure, ECs thus provide valuable insights into the style evolution of Western choral music (see Figure 1). Note that we use a rather broad definition of *evolution*, referring to the development of a composer’s style over a period of time.

Handling missing data: approximation strategies

Collecting reliable work date annotations for larger corpora requires a substantial amount of manual research. Moreover, such information is unknown or in doubt for quite a number of works. Even if one knows all work dates, it may become difficult to create a dataset with a balanced coverage of all years due to inherent imbalances in the distribution.

Because of such problems, Weiß et al. (2019) adopted a pragmatic approach by projecting works onto the historical time axis based on *composer dates*, that is, 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 and usually more reliable. They proposed an approximation of work counts over the course of a composer’s life. For this distribution, they assumed that a typical composer starts composing not before a certain (fixed) age given by $a_{\text{start}} \in \mathbb{N}$ years, with $a_{\text{start}} = 10$, as in Weiß et al. (2019).

For the remaining years (ages) $[a_{\text{start}} : a_{\text{death}}] := \{a_{\text{start}}, a_{\text{start}} + 1, \dots, a_{\text{death}}\}$, they computed a roughly flat distribution with smooth edges. To this end, they used a so-called *Tukey window* (or *tapered cosine window*) $w : \mathbb{N} \rightarrow \mathbb{R}$ with parameter $\alpha \in \mathbb{R}$:

$$w(n) = \begin{cases} 0.5 \left(1 - \cos \left(\frac{2\pi n}{\alpha N} \right) \right), & 0 \leq n < \frac{\alpha N}{2}, \\ 1, & \frac{\alpha N}{2} \leq n \leq \frac{N}{2}, \\ w(N - n), & \frac{N}{2} < n \leq N, \end{cases} \quad (4)$$

with $n \in [0 : N]$ and $N = a_{\text{death}} - a_{\text{start}}$ being the window length. In Weiß et al. (2019), the parameters were heuristically chosen to a start age of $a_{\text{start}} = 10$ and a Tukey parameter of $\alpha = 0.35$. Figure 6 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 WCC, which indicates the coverage of a certain year with works in the dataset. 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.

In Weiß et al. (2019), 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} , that is, 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 7a). To counteract the effect of imbalanced composition ages, we slightly smooth this curve by convolution with a 5-year kernel $\mathbf{k} = (0.1, 0.2, 0.4, 0.2, 0.1)^T$. Since composers have died at different ages, the blue 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 7a). For each value of $a_{\text{start}} \in [0 : 24]$, we fit the Tukey parameter α (see Eq. (4)) 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 7a), which is slightly higher than the value $a_{\text{start}} = 10$ used in Weiß et al. (2019).

Using $a_{\text{start}} = 13$, we now fit the window parameters for the remaining years, that is, the interval $[a_{\text{start}} : a_{\text{death}}]$. To counteract the effects of different overall ages, we normalize the overall ages

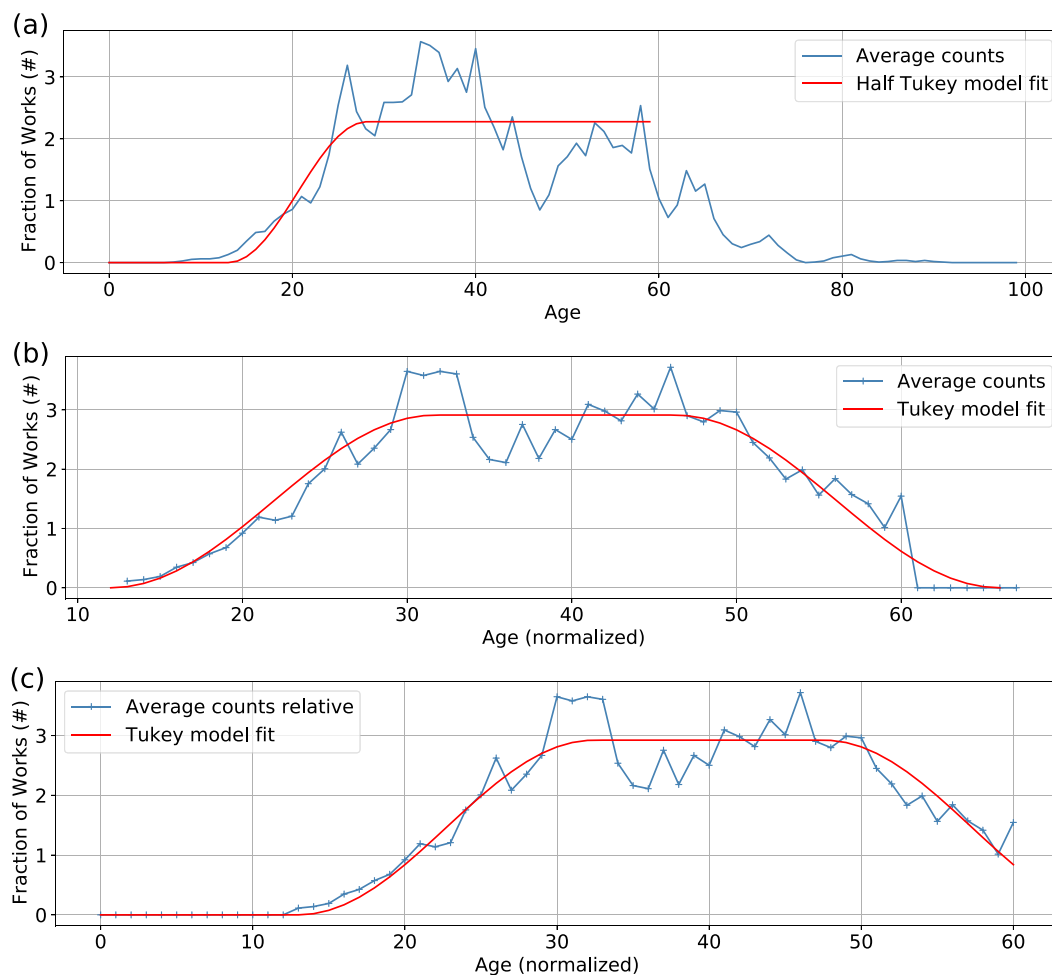


Figure 7. Curve fitting procedure to determine the optimal window parameters (a) Partial curve fit to determine the optimal start composing age $a_{\text{start}} = 13$. (b) Fit to determine optimal parameters N_{end} and α for the Tukey window w . (c) Resulting full window.

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 7b). 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 Weiß et al. (2019). The fitted curve is shown in red in Figure 7b.

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 7c. 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 WCC. To this end, we first compute the reference curve using the work date annotations for 1,153 works that have these annotations. We post-process the curve with an average filter of length 15 years (red curve in Figure 8). We then compare this reference curve with our approximation curve based on composer dates and our optimized Tukey window (blue curve in Figure 8). 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 Weiß et al. (2019), that is, $\alpha = 0.35$, $a_{\text{start}} = 10$, 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.

Impact of approximation strategies on evolution curves

To show the impact of our approximation strategy on the computation of ECs, we now apply them to our measurement of tonal complexity as defined in the section “Measuring tonal complexity.” 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 WCC for normalization. We obtain an EC that indicates the average complexity of the works 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).

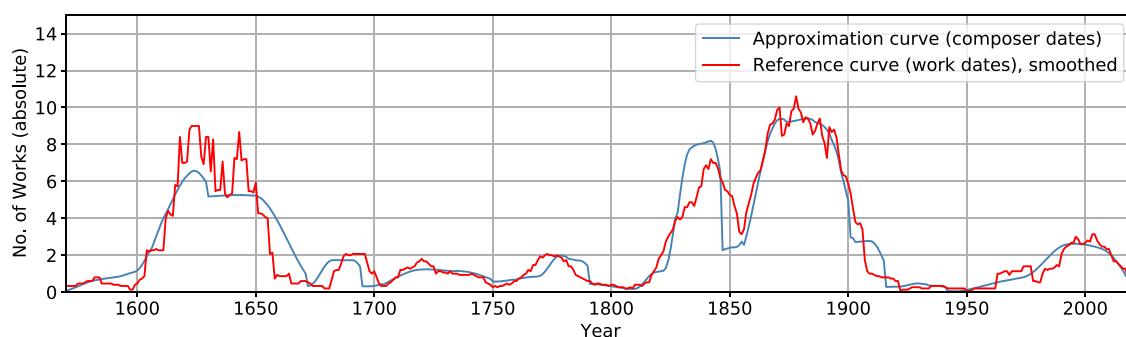


Figure 8. Work count curves based on composer dates (approximation curve, blue) and based on work dates (reference curve, red), respectively.

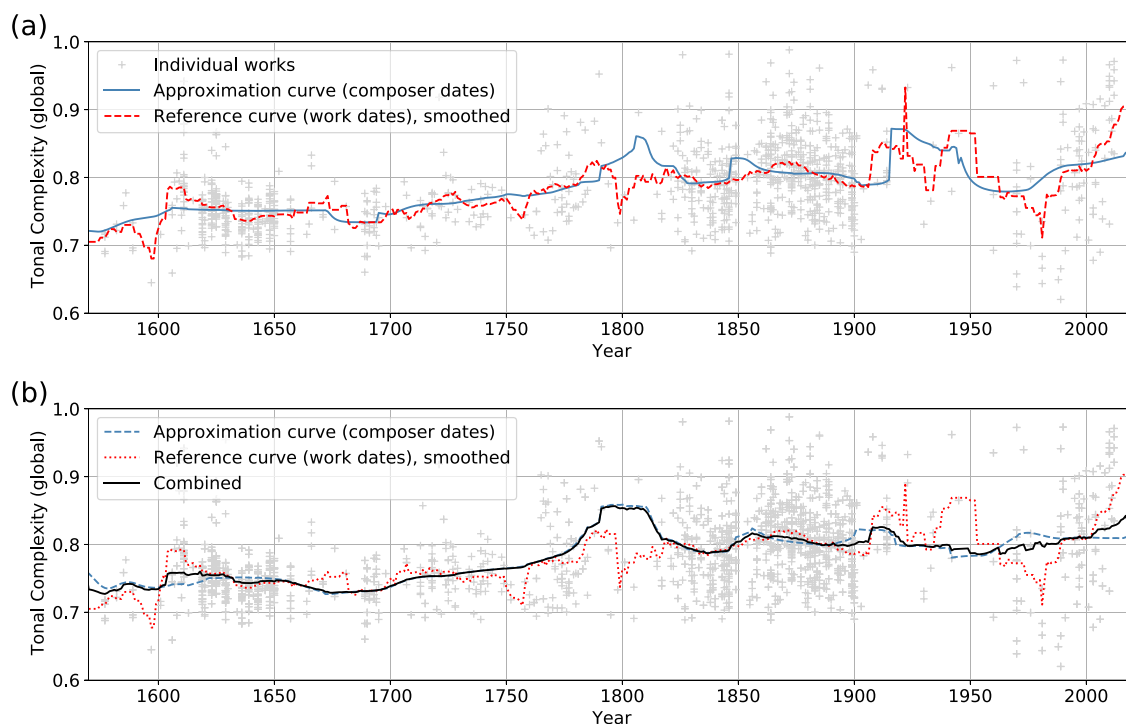


Figure 9. ECs for the global complexity. (a) Comparing ECs based on the subset D_{work} computed as approximation curve using composer dates (blue) and reference curve using work dates (red). (b) Combined EC for the global complexity in D (black) computed using work dates for D_{work} (red) and composer dates for D_{comp} (blue). Original complexity values for works are shown as gray crosses.

Denoting our full dataset as D , we first consider the subset $D_{\text{work}} \subset D$ comprising all works with available work date annotations (1,153 works in total). Figure 9a 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 WCCs (Figure 8), the approximation is still good but the deviations are slightly higher. However, we observe such deviations only in regions where only few works contribute, for example, 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 $D_{\text{work}} \subset D$ (1,153 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 9b). For the subset $D_{\text{comp}} \subset D$ (1,248 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 9b). The combined EC is shown as the black curve in Figure 9b. We observe a stabilized curve where minor outliers are removed (e. g., around the years 1700, 1760 or 1920) while not losing the interesting trends. In periods with a higher number of work dates, the values of the combined curve lie between the two computed curves due to the smoothing effect of the Tukey window. If there are only a few work dates in a certain period (e. g., around 1800), the combined curve will closely follow the approximation curve.

Hypothesis testing using evolution curves

We now apply this strategy for investigating the evolution of tonal complexity in Western choral music, for comparing the results to those obtained for instrumental music in Weiß et al. (2019), and for testing three musicological hypotheses. To this end, we use our mixed approach for computing different variants of the combined EC, always relying on the dataset D , that is, the full CAC. In contrast to Weiß and Müller (2023), we compute our ECs from DL-based

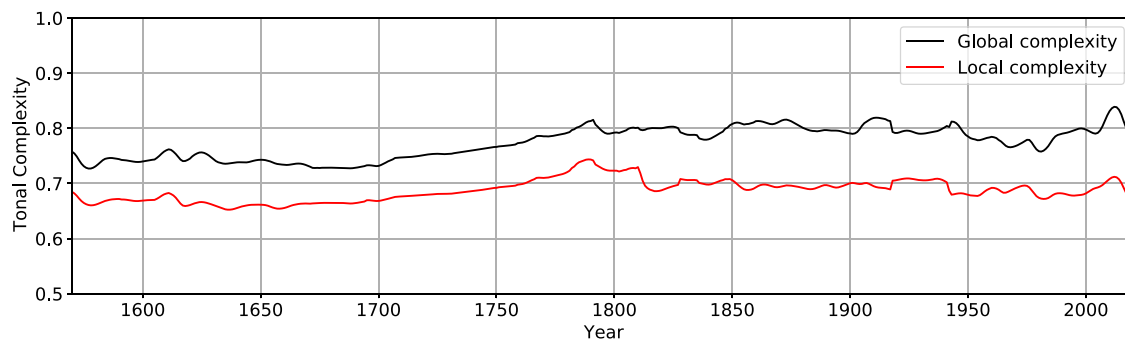


Figure 10. Comparing ECs for global and local complexity.

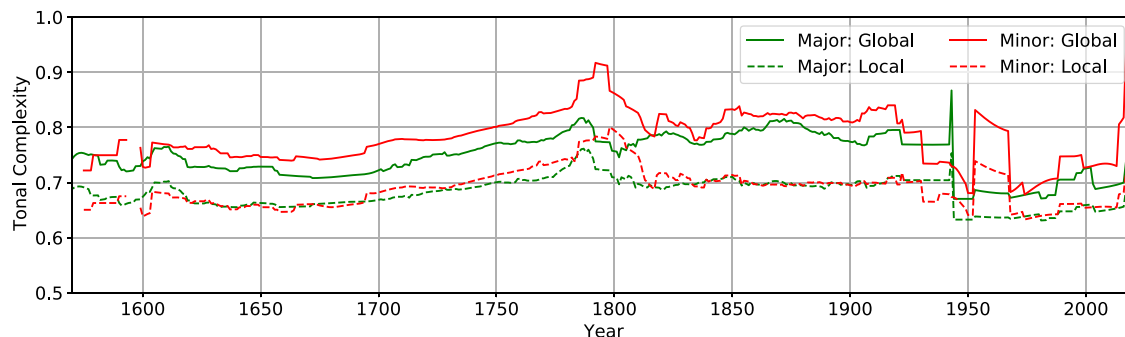


Figure 11. Comparing ECs for global and local complexity separated into major and minor keys.

pitch-class features in order to obtain more robust results. We systematically compare the influence of the feature type (SP vs. DL) in the section “Vocal vs. instrumental music.”

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 10). Looking at the global EC (black), we observe an increase in complexity over the course of the 17th and 18th centuries. Interestingly, we do not observe any drop around 1750, in contrast to Weiß et al. (2019) 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 Weiß et al. (2019) is not visible for the CAC. Even more remarkably, the 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 Schönberg 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. In the section “Vocal vs. instrumental music,” we will discuss such differences between vocal and instrumental music in detail.

Revisiting hypotheses on tonal evolution

Global vs. 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 by Weiß et al. (2019) especially within the 19th century, where the local complexity (referring to the complexity of, e. g., chords) was fairly stable while the global complexity (referring to the complexity of modulations across the whole movement) was clearly increasing. For the CAC, we do not observe such a behavior. Comparing ECs for global and local complexity (Figure 10), we mostly observe a parallel evolution. The distance between 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 occurring within a movement, which can have an influence on the complexity as shown in the section “Tonal complexity and movement length.”

Major vs. minor keys

Our second hypothesis is based on the observation that minor keys are expected to be more complex since they usually exhibit more chromatic inflections as compared to major keys. To this end, we consider the data subset with key annotations (major and minor keys) and compute an EC for each of these two modes (Figure 11). For the global complexity (solid lines), both curves follow a similar trend. However, we see a consistent offset of the minor curve (red) over the major curve (green). This confirms our hypothesis that minor keys use a larger set of pitch classes and, thus, are tonally more complex. For the local complexities (dashed curves), we do not observe this offset in most cases. Only in the 18th century do we see that the difference between major and minor keys is also represented on the local level. Moreover, for the 20th century, we see some fluctuating behavior, which is due to the 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).

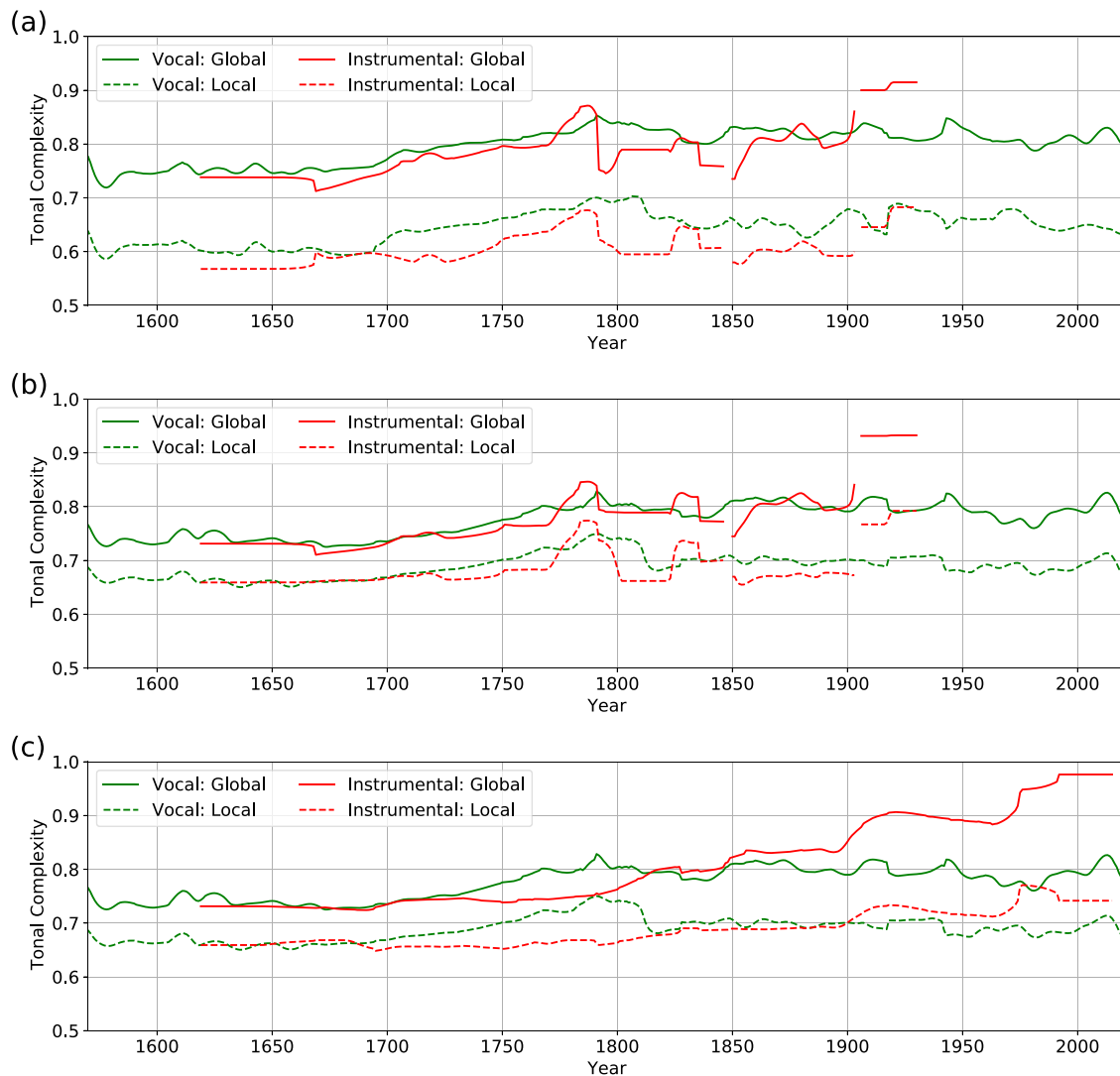


Figure 12. Comparing ECs for global and local complexity separated into vocal and instrumental music based (a) on the CAC based on SP, (b) on the CAC based on DL and (c) on the combined corpus (CAC + CrossEra) based on DL.

Vocal vs. 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 in many cases 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 12).

We first discuss Figure 12a, which is repeated from Weiß and Müller (2023) and shows ECs relying on SP-based pitch-class features. Here, we observe that vocal music seems to be more complex than instrumental music for most time periods, especially for the local complexity (dashed lines). This observation contradicts our hypothesis. Regarding this observation, we already suspected in Weiß and Müller (2023) that the behavior of SP-based complexity measures differs substantially for vocal and instrumental music due to the influence of vocal artifacts (vibrato, bends/slides, overall lower pitch stability) on the features. To this end, we now compare these analyses with our updated ones based on DL.

Figure 12b shows these newly computed ECs. Here, we observe two main differences to Figure 12a: First, the local complexity curve is much closer to the global one, which is mainly due to higher complexity values at the local level. We suspect this effect to be caused by the activations of each pitch class predicted by the DL method in contrast to the energy measured by the SP approach. Since the activations are supposed to be similar for all active pitch classes, they are distributed more equally over the circle of fifths, leading to a higher complexity for a given chord. Our second observation is that the difference between vocal and instrumental music is smaller when using the DL approach. In particular, in the 17th and 18th centuries, where our data coverage is good, their complexity differs only marginally. This supports our assumption that the main difference between vocal and instrumental music observed in Figure 12a is an artifact of SP-based pitch-class features applied to vocal music recordings with their special properties.

We therefore rely on the DL-based EC for the further discussion. As a second challenge in testing this hypothesis, we noticed in Weiß and Müller (2023) the shortage of instrumental works in the CAC (compare Table 1). To improve this situation, we additionally consider an instrumental corpus for our EC, the CrossEra dataset

analyzed in Weiß et al. (2019).⁸ As the CAC, CrossEra is also partially annotated with work dates, for 1,751 of the 2,000 audio tracks. We project the complexity values from the instrumental works onto the timeline and compute new ECs for the combined corpus (7,729 audio tracks in total).

Figure 12c shows the resulting ECs, which we now use to discuss our hypothesis regarding the relationship of vocal and instrumental music. We observe three different situations in specific time spans: First, from 1625–1725, instrumental and vocal music exhibit similar complexity values. Second, from 1725–1810, vocal music appears to be more complex than instrumental music. Third, from 1850 onwards, we find the reversed situation, with instrumental music increasingly exceeding vocal music in complexity. With slight differences, these trends are observable for both local and global complexity measures.

There are several remarks on these observations. For the period 1725–1810, the observed higher complexity of vocal music contradicts our hypothesis. This time span largely coincides with typical definitions of the “classical period,” which is characterized by the upcoming Galant style as well as lighter, clearer textures. This applies especially to instrumental music, such as the arising sonata form used in piano sonatas, symphonies and chamber music. Sacred music, in contrast, is known to have changed less during this period while still showing considerably greater use of chromaticism, dissonance and counterpoint. Note that the hypothesized influence of the string quartet cannot be observed from the ECs in Figure 12c, since neither CAC nor CrossEra contain a considerable amount of quartets. In the later 19th century, we observe instrumental music to be more complex, which may be due to new compositional practices. Instrumental music of all genres became more complex during the 19th century, including longer works and movements, while in sacred music, many works were still expected to be of limited duration due to their liturgical purpose (e.g., masses). Moreover, the second half of the 19th century saw the rise of the Cecilian Movement, which demanded for a restricted use of instruments in sacred music, the reconsideration of Gregorian chant and early vocal polyphony, and a sparse use of chromaticism and late Romantic harmony. In the 20th century, we observe an enormous contrast between the complexity of vocal and instrumental music. We relate this observation in part to the nature of our corpora. For this time span, the CrossEra corpus contains many works by modernist composers such as Shostakovich or Schönberg and his students, who crucially advanced the use of tonality up to dodecaphonic techniques. In contrast, the CAC does not focus on complex atonal works, but instead considers works of more traditional tonality, which can be more easily performed by amateur choirs. This relates to the differences in the composition and performance of instrumental and vocal music. Singing an atonal piece is substantially more challenging than playing it on an instrument since singers have to imagine each note's pitch beforehand, whereas finger positions (and other playing aspects) on instruments provide a strong guidance. While these aspects demand for a more detailed investigation, we nevertheless assume that a clearer, more tonal compositional language, which is required to keep the singing parts manageable, was a characteristic of many vocal compositions in the 20th century, in contrast to the situation in instrumental music of that time.

⁸As for the CAC, the CrossEra dataset is based on commercial recordings. Pitch-class features derived from these recordings and further details on the corpus can be found at <https://www.audiolabs-erlangen.de/resources/MIR/cross-era>.

Case studies: complexity deviation and stylistic trends of individual composers

By using our approximation strategy from the section “Handling missing data: approximation strategies,” we distribute the undated works roughly equally over each composer's life span. This procedure implicitly assumes that the tonal complexity remains at the same level throughout the composer's life. In this section, we want to test this assumption and examine the stylistic evolution of individual composers (see the section “Individual composers' evolution curves”) by comparing the individual ECs of three composers (Heinrich Schütz, Johann Sebastian Bach and Franz Liszt). To extend on these examinations, we present a method for computing a normalized complexity deviation curve with regard to a composer's average complexity (see the section “Complexity deviation”). At the example of three case studies, we then investigate whether we can correlate major shifts in the tonal complexity of compositions to historical contexts and biographical events of the respective composers. To this end, we chose three composers with sufficient works and annotated work dates to produce meaningful results: Heinrich Schütz (see the section “Heinrich Schütz”), Johann Sebastian Bach (see the section “Johann Sebastian Bach”) and Felix Mendelssohn Bartholdy (see the section “Felix Mendelssohn Bartholdy”). Finally, we summarize the main findings of these case studies and critically reflect on our method (see the section “Summary and reflections”).

Individual composers' evolution curves

As indicated by Figure 2, the CAC exhibits a considerable imbalance regarding the number of works per composer. Since we chose every work to have an equal influence on our ECs, this leads to a greater importance of composers with many works in the corpus. Moreover, some of these frequent composers, such as Heinrich Schütz, Johann Sebastian Bach or Max Reger (all in the top 5 regarding work count), clearly stand out among their contemporaries regarding their use of tonality. Altogether, this may result in a crucial influence of such composers on the EC. To investigate these effects, we separate the works by three well-known composers from the CAC and recompute the EC from Figure 10, while adding the individual ECs for Heinrich Schütz, Johann Sebastian Bach and Franz Liszt for reference in color. Figure 13 shows this comparative analysis.

In the case of Heinrich Schütz, we find that throughout his life, his works are globally more complex than those of his contemporaries. On a local level, however, the complexity values seem to be similarly high. Johann Sebastian Bach is considered today as one of the greatest composers of all time, in part because of his sacred music compositions. However, many of his works have survived only in manuscript, and their work dates are in many cases a matter of estimation or even impossible (Geck 2011). Our mapping technique helps us to distribute his compositions over his lifetime and thus to include them in the ECs. Due to the low amount of work dates, we should still treat the exact course of his curve with caution. In terms of tonal complexity, we still observe that Bach is an outlier, as both curves associated with his compositions are above the curves of his contemporaries. Our computed ECs impressively confirm the common description of Bach's compositional style being ahead of its time.

Finally, we look at Franz Liszt. He is best known for his complex piano pieces, but also wrote orchestral pieces, sacred music and other vocal music, the last two of which are contained in the

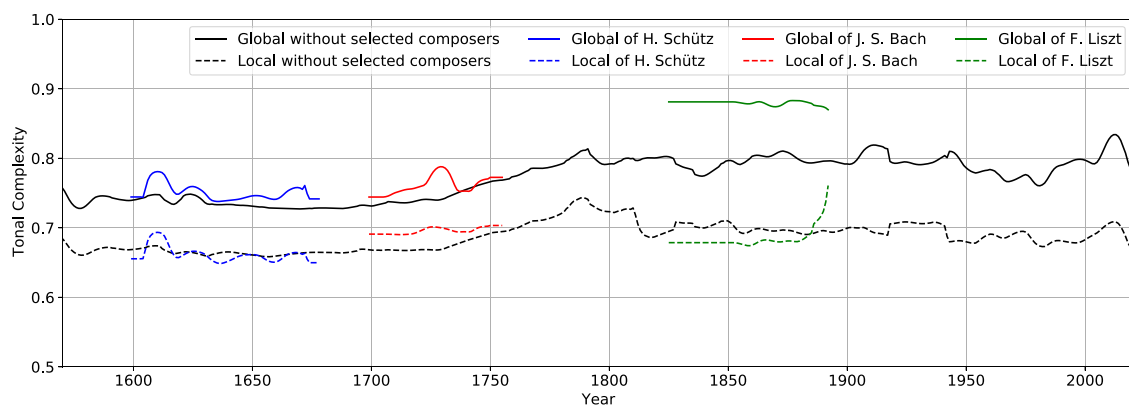


Figure 13. Comparing ECs for global and local complexity with three individual composers separated.

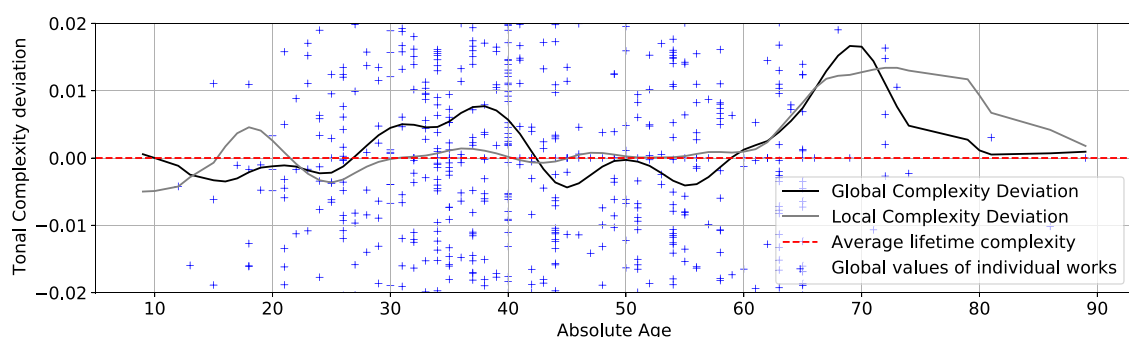


Figure 14. Visualizing the average complexity deviation of all composers in the CAC.

CAC. For Liszt – as compared to the general evolution – we find tremendously higher values of global complexity, but lower local complexity values. Liszt seems to be an outstanding composer in terms of his tonal style, which corresponds to his perception nowadays. He composed in a rather modern way, using numerous modulations and complex chords. This explains the very high global values but would also let us expect high local values. However, his sacred music is not known for overly complex chords with strong dissonances, which may be a reason for the low local values. On the contrary, some of his contemporaries rejected his sacred music, which Liszt deliberately kept simple in the style of Palestrina's compositions and Gregorian chants, in the spirit of the Cecilian reform movement (Hansen 1985). The sharp rise in the local curve toward the end is due to a single piece (*Salve Regina* [S 66] in 1885).

All in all, we cannot confirm the assumption that a composer's complexity remains at one level. We looked at the complexity curves of three individual composers and compared them with the overall ECs. Each composer is a different case: Schütz's global complexity values were above those of his contemporaries, Bach wrote more complex pieces overall compared to his contemporaries, and Liszt showed a more individual curve with very high global but low local values. However, all three composers showed individual stylistic developments. While our time-mapping strategy may not be suitable for a closer look at individual composers, we also observe that their influence on the overall ECs is small, even though Schütz and Bach are among the composers with the most works in the corpus. Even a strong outlier such as Liszt's local curve has no significant influence on the full corpus' EC. We therefore argue that the chosen time-mapping technique is a viable option for the

examination of large-scale corpora with data distributed over a large time span.

Complexity deviation

The changes observed in the section "Individual composers' evolution curves" over the life span of composers suggest that composers develop and adapt their compositional styles. This leads to the question of whether this development occurs gradually, abruptly or in several phases, and if such developments are comparable between different composers. We now want to investigate such questions by studying the average development of composers in the CAC in terms of tonal complexity. To this end, we normalize the complexity values of each work with respect to the average complexity of the respective composer at a certain age (*absolute age*). Since the goal is to study changes over the life of a composer, we exclude works without annotated work dates (1,153 works remaining). We denote the resulting curve as the *complexity deviation*.

Figure 14 shows this complexity deviation on the corpus level (aggregated over all composers) along with the normalized global values of individual works (blue crosses). Although the values themselves are small and the amount of data outside the age range between 20 and 60 is limited, we observe two interesting trends. On average, the works of composers younger than 40 were globally more complex than their average, and the works of composers between 40 and 60 years were globally less complex than their average. Around the age of 40, the curve even suggests the observation of a kind of "midlife crisis." A possible reason for these observations could be that composers developed their style and experimented

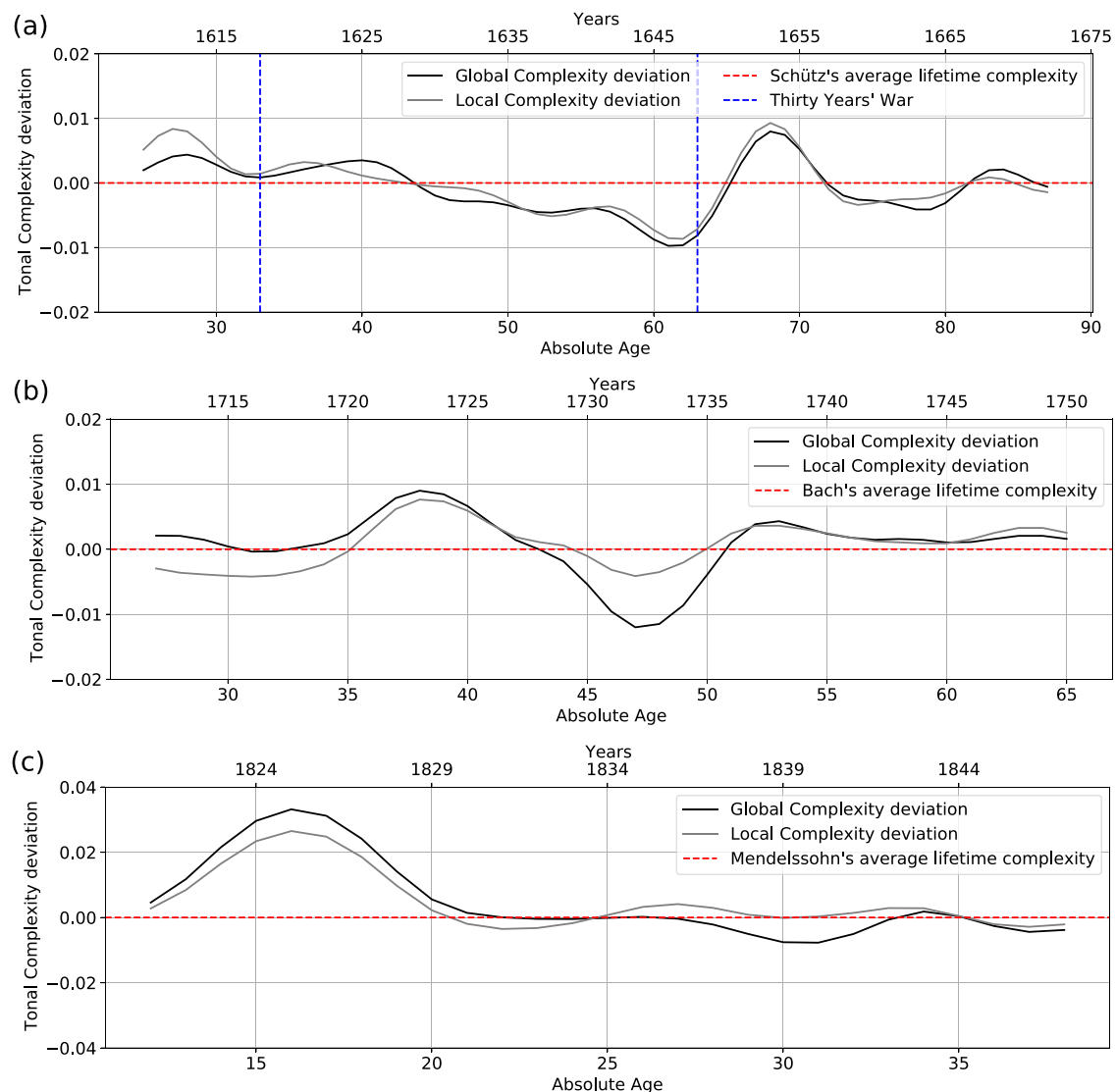


Figure 15. Visualizing the complexity deviation of (a) H. Schütz, (b) J. S. Bach and (c) F. Mendelssohn Bartholdy.

more when they were younger, until they got older and into fixed positions, where their creative freedom was reduced. However, due to the limited amount of data, this explanation remains somewhat speculative.

Heinrich Schütz

Since the CAC contains all 321 works (398 tracks) by Heinrich Schütz (1585–1672) and 256 of them have an annotated work date, we want to look at his individual development in Figure 15a. Interestingly, we notice that the complexity deviation is mostly the same on a global and a local level. Generally, we first observe positive complexity deviation values until the age of 40 in 1625, but then a steady decline until the age of 62. His values then recover briefly and peak at the age of 68 in 1653. Finally, both curves remain just below his average until a brief rebound just after his 80th birthday in 1665.

To understand this behavior, it is worth taking a look at his biography (Breig 2021; Heinemann 2005). Beginning in 1614, Schütz served as court conductor at the Electoral Court in Dresden, where he was in charge of the court chapel including the appointment of

musicians as well as their accommodation and payment. However, Saxony's involvement in the 'Thirty Years' War led to a shortage of money from around 1623. In 1625, Schütz and his employed musicians turned to the Elector for the first time because of the lack of salary payments. In the following years, the Dresden court chapel declined, and by 1639 Schütz had only 10 members left. It was not until the end of the 1640s that the chapel began to recover, and in 1650 a great celebration was held to mark the Peace of Westphalia. During this time, Schütz also published new collections of works, such as his *Geistliche Chormusik*.

Comparing this biographical information with our complexity curves, some parallels can be observed. With the decline of Dresden court music, the tonal complexity of Schütz' works also declined. At the end of the 'Thirty Years' war, the financial situation improved, which coincides with the rising complexity of his works composed at that time. It seems possible that Schütz had to adapt his compositions to such circumstances like a small instrumentation, and that we can detect these stylistic changes with our methods. Schütz did not undergo a typical stylistic development, but had to adapt his works to external circumstances. An argument against this, however, is that he also traveled to Italy and Denmark

during the war and composed for festivities there. He was therefore not exclusively bound to the conditions in Dresden. Also, while the Carus-Verlag produced recordings of Schütz's complete works, his theatrical works composed for festive events at several courts have not survived. We therefore need to be aware that our results are only valid for Schütz's sacred oeuvre.

Johann Sebastian Bach

We now look at the complexity deviation of Johann Sebastian Bach (1685–1750), which we show in Figure 15b. The complexity values of his works start close to their life-time average, then reach their peak when Bach was 38 years old. We then observe a substantial drop to values below the average (between the ages of 45 and 50), before the curve lingers just above the average for the last 15 years of his life. However, only 43 of his works are included here, as only these have a work date annotation.

Once again, we relate the curves to Bach's biographical data (Breig 2016; Geck 2011). In the early years of his career (1708–1723) he worked in Weimar and Köthen, but only a few cantatas in the CAC date from this period. In 1723, Bach began to work in Leipzig, where he was responsible for the music in four churches and often used his own compositions, especially in the early years. Fittingly, the complexity of his works is greatest during this period. He composed motets such as *Komm, Jesu, komm* (BWV 229) and *Jesu, meine Freude* (BWV 227) in 1723 and the cantatas *Nun komm, der Heiden Heiland* (BWV 661) in 1724 and *Erhalt uns, Herr, bei deinem Wort* (BWV 126) in 1725, all of which have relatively high complexity values. The years 1729 to 1739 are considered the second phase of Bach's work in Leipzig. They were marked by a dispute with the city council, as Bach complained about too few singers, too little money and musically untalented pupils. As with Schütz, these difficulties seem to be reflected in the complexity of his works. Bach's *Magnificat in D* (BWV 243) in 1733, his *Christmas Oratorio* (BWV 248) in 1734, and his *Ascension Oratorio* (BWV 11) in 1735 all date from this period and, while they remain popular to this day, they are below his average in tonal complexity. The values finally recover around 1737, when he was 52 years old, for the final phase in Leipzig until 1750 with several masses, including the well-known *Mass in B minor* (BWV 232) in 1749.

Felix Mendelssohn Bartholdy

We finally discuss the complexity deviation curve for Felix Mendelssohn Bartholdy (1809–1847), shown in Figure 15c. His complexity values peak already around the age of 16 years. From the age of 20 onwards, we mostly observe values around his average, with a small negative episode around the age of 31. To contextualize the high complexity values at young age, we may look at the composer's biography (Krummacher and Wehner 2018; Schwingenstein 1994). Mendelssohn grew up in a middle-class family that recognized and strongly encouraged the musical talent of him and his siblings at an early age. The highly complex works were composed at a time when his family regularly organized so-called *Sonntagsmusiken*, where the children performed together with professional musicians. In this context, under guidance of his teacher Carl Friedrich Zelter, Mendelssohn was able to express himself freely as a composer, which may relate to the above-average complexity of his early works. Other well-known works he composed during his lifetime, such as the oratorios *Paulus* (op. 36) in 1831 and *Elias* (op. 70) in 1846, were close to his average in complexity. The small decrease around the age of 30 is due to

several polyphonic songs that he composed around 1840. Here, too, we were able to include only a fraction of the works that still shape the overall image of the composer today. For example, many vocal works, such as songs and operas, are not part of the CAC, as well as all instrumental music, such as symphonies, piano music and chamber music.

Summary and reflections

Overall, our findings show that our computational strategy – calculating DL-based pitch-class activations and then computing ECs of tonal complexity – is suitable for detecting trends in the style evolution of Western choral music, also with regard to individual composers. While we have to reject the simplifying assumption of a constant complexity across a composer's life, our previous analysis in Figure 9 shows that the influence of such individual developments on the EC is limited on a global scale, and thus our methods are still viable for corpus studies of long-term trends. With our novel method of complexity deviation curves, we are able to relate shifts in the complexity of three exemplary composers to their historical contexts and biographies. In the Appendix, we show such analysis for further composers.

Reflecting on our results, we want to acknowledge a few limiting factors to our observations: First, we can only include a fraction of the composers' works because we require work dates in order to calculate the composer's age at the time of composition. For our final case studies, however, we selected three individual composers with sufficient works and annotated work dates to produce meaningful results regarding their compositional style in vocal music. Second, some composers such as Bach did not always write new works from scratch. Instead, he sometimes reworked older compositions or put them together to form a new work. For example, we could argue that parts of his *Christmas Oratorio* (BWV 248) should be assigned to an earlier year because they were re-used from older works such as cantatas. Finally, such kinds of findings can only reveal *correlations* between historical or biographical circumstances and musical properties such as the tonal complexity. Whether there is a *causal effect*, e. g. a historical event inspiring or forcing a composer to adopt his or her style in a certain way, remains subject of further in-depth musicological studies. Such studies, however, can be stimulated and motivated by quantitative observations as the ones made here.

Conclusion

In this article, we presented a strategy for conducting large-scale corpus studies using music audio recordings. By applying a deep learning-based technique to estimate pitch-class activations from these recordings, we were able to calculate tonal complexity measures relating to the circle of fifths. We then projected these measures onto a historical timeline to generate ECs, using work date annotations from the CAC. In cases where work dates were missing, we revisited and further validated a heuristic approach for approximating ECs by using composer life dates, thus enriching our analysis with additional data.

Building on previous work by Weiß and Müller (2023), we demonstrated the practical value of this strategy through several case studies. We compared our results with previous studies on other corpora and revisited hypotheses about the evolution of tonal complexity in Western choral music. Refining our pitch-class features improved the accuracy of our findings, particularly

in distinguishing between global and local complexity, major and minor keys and vocal versus instrumental music.

Our analysis of individual composers revealed that their tonal complexity is not constant throughout their careers, a finding that prompted us to propose a complexity deviation curve. By normalizing ECs for each composer based on the annotated work dates, we were able to examine how their style evolved in relation to historical and biographical factors. Exemplary case studies for Heinrich Schütz, Johann Sebastian Bach and Felix Mendelssohn Bartholdy highlighted the value of this approach for exploring the relationship between compositional style and historical context.

While our case studies are inherently limited by the available data, our results demonstrate that the proposed methodology produces insights of musicological relevance. These findings underline the significant potential of computational approaches for corpus studies, enabling interdisciplinary research at the intersection of computer science and historical musicology. Despite the interpretative challenges posed by some of our approximations, we conclude that the methods presented here offer a valuable tool for future studies in this field.

Acknowledgments. The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Fraunhofer Institute for Integrated Circuits IIS. We cordially thank the Carus-Verlag Stuttgart (Johannes Graulich and Ester Petri) for enabling the study of the CAC.

Data availability statement. Since the audio recordings are commercial releases, we cannot publish the CAC. However, detailed information and preview files for most of the individual audio recordings are available on the publisher's website (<https://www.carus-verlag.com/en/>).

Disclosure of use of AI tools. No artificial intelligence (AI) tool was employed in this research.

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Author contributions. Conceptualization and methodology: B.H., M.M., C.W.; Data curation, implementation and visualization: B.H., C.W.; Writing original draft: B.H., M.M., C.W. All authors approved the final submitted draft.

Funding statement. This work was funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) within the Emmy Noether Junior Research Group on *Computational Analysis of Music Audio Recordings: A Cross-Version Approach* (DFG WE 6611/3-1, Grant No. 531250483) and the DFG Reinhart Koselleck-Projekt (DFG MU 2686/15-1, Grant No. 500643750).

Competing interests. The authors declare none.

References

Abeßer, Jakob, Klaus Frieler, Estefanía Cano, Martin Pfeleiderer, and Wolf-Georg Zaddach. 2017. "Score-Informed Analysis of Tuning, Intonation, Pitch Modulation, and Dynamics in Jazz Solos." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, no. 1: 168–77. <https://doi.org/10.1109/TASLP.2016.2627186>.

Bellmann, Héctor G. 2012. "Categorization of Tonal Music Style: A Quantitative Investigation." PhD diss., Griffith University.

Benetos, Emmanouil, Simon Dixon, Zhiyao Duan, and Sebastian Ewert. 2019. "Automatic Music Transcription: An Overview." *IEEE Signal Processing Magazine* 36, no. 1: 20–30. <https://doi.org/10.1109/MSP.2018.2869928>.

Bittner, Rachel M., Brian McFee, Justin Salamon, Peter Li, and Juan P. Bello. 2017. "Deep Salience Representations for F0 Tracking in Polyphonic Music." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 63–70. Suzhou, China. <https://doi.org/10.5281/zenodo.1417937>.

Breig, Werner. 2016. "Bach, Johann Sebastian." In *MGG Online*, edited by Laurenz Lütteken. New York, Kassel, and Stuttgart: RILM / Bärenreiter / Metzler. <https://www.mgg-online.com/mgg/stable/11672>.

Breig, Werner. 2021. "Schütz, Heinrich." In *MGG Online*, edited by Laurenz Lütteken. New York, Kassel, and Stuttgart: RILM / Bärenreiter / Metzler. <https://www.mgg-online.com/mgg/stable/400567>.

Calvo-Zaragoza, Jorge, Jan Hajič Jr., and Alexander Pacha. 2020. "Understanding Optical Music Recognition." *ACM Computing Surveys* 53, no. 4: 77:1–77:35. <https://doi.org/10.1145/3397499>.

Cuesta, Helena, Emilia Gómez, Agustín Martorell, and Felipe Loáiciga. 2018. "Analysis of Intonation in Unison Choir Singing." In *Proceedings of the International Conference of Music Perception and Cognition (ICMPC)*, 125–130. Graz, Austria.

Di Giorgi, Bruno, Simon Dixon, Massimiliano Zononi, and Augusto Sarti. 2017. "A Data-Driven Model of Tonal Chord Sequence Complexity." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, no. 11: 2237–50. <https://doi.org/10.1109/TASLP.2017.2756443>.

Duan, Zhiyao, Bryan Pardo, and Changshui Zhang. 2010. "Multiple Fundamental Frequency Estimation by Modeling Spectral Peaks and Non-peak Regions." *IEEE Transactions on Audio, Speech, and Language Processing* 18, no. 8: 2121–33.

Fritsch, Joachim, and Mark D. Plumbley. 2013. "Score Informed Audio Source Separation Using Constrained Nonnegative Matrix Factorization and Score Synthesis." In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 888–891. Vancouver, Canada, May.

Geck, Martin. 2011. *Johann Sebastian Bach*, 6th ed., Rowohlt-Taschenbuch-Verlag, Reinbek bei Hamburg.

Hansen, Bernhard. 1985. "Liszt, Franz Ritter von." In *Neue Deutsche Biographie* 14, 701–3. Historische Kommission bei der Bayerischen Akademie der Wissenschaften. München, Germany. <https://www.deutsche-biographie.de/pnd118573527.html#ndbcontent>.

Heinemann, Michael. 2005. *Heinrich Schütz*, 2nd ed., Rowohlt, Reinbek bei Hamburg.

Kase, Vojtech, Adéla Sobotková, and Petra Hermánková. 2023. "Modeling Temporal Uncertainty in Historical Datasets." In *Proceedings of the Computational Humanities Research Conference (CHR)*, 3558: 413–25. CEUR Workshop Proceedings. Paris, France.

Klapuri, Anssi P. 2008. "Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model." *IEEE Transactions on Audio, Speech, and Language Processing* 16, no. 2: 255–66.

Koops, Hendrik Vincent, Anja Volk, and W. Bas de Haas. 2015. "Corpus-Based Rhythmic Pattern Analysis of Ragtime Syncopation." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 483–9. Málaga, Spain.

Korzeniowski, Filip, and Gerhard Widmer. 2016. "Feature Learning for Chord Recognition: the Deep Chroma Extractor." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 37–43. New York City, NY, USA. <https://doi.org/10.5281/zenodo.1416314>.

Krummacker, Friedhelm, and Ralf Wehner. 2018. "Mendelssohn Bartholdy, Felix (Jacob Ludwig)." In *MGG Online*, edited by Laurenz Lütteken. New York, Kassel, and Stuttgart: RILM / Bärenreiter / Metzler. <https://www.mgg-online.com/mgg/stable/51034>.

Mauch, Matthias, and Simon Dixon. 2010. "Approximate Note Transcription for the Improved Identification of Difficult Chords." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 135–40. Utrecht, The Netherlands.

Mauch, Matthias, and Mark Levy. 2011. "Structural Change on Multiple Time Scales as a Correlate of Musical Complexity." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 489–94. Miami, FL, USA.

Mauch, Matthias, Robert M. MacCallum, Mark Levy, and Armand M. Leroi. 2015. "The Evolution of Popular Music: USA 1960–2010." *Royal Society Open Science* 2, no. 5: 1–10. <https://doi.org/10.1098/rsos.150081>.

Miron, Marius, Julio J. Carabias-Orti, Juan J. Bosch, Emilia Gómez, and Jordi Janer. 2016. "Score-Informed Source Separation for Multichannel Orchestral Recordings." *Journal of Electrical and Computer Engineering* 2016: 8363507:1–8363507:19.

- Moss, Fabian C., Markus Neuwirth, Daniel Harasim, and Martin Rohrmeier. 2019. "Statistical Characteristics of Tonal Harmony: a Corpus Study of Beethoven's String Quartets." *PLoS One* 14, no. 6: 1–16. <https://doi.org/10.1371/journal.pone.0217242>.
- Müller, Meinard, and Sebastian Ewert. 2010. "Towards Timbre-Invariant Audio Features for Harmony-Based Music." *IEEE Transactions on Audio, Speech, and Language Processing* 18, no. 3: 649–62.
- Nakamura, Eita, and Kunihiro Kaneko. 2019. "Statistical Evolutionary Laws in Music Styles." *Scientific Reports* 9, no. 1: 15993:1–11. <https://doi.org/10.1038/s41598-019-52380-6>.
- Rodriguez Zivic, H. Pablo, Favio Shifres, and Guillermo A. Cecchi. 2013. "Perceptual Basis of Evolving Western Musical Styles." *Proceedings of the National Academy of Sciences* 110, no. 24: 10034–8. <https://doi.org/10.1073/pnas.1222336110>.
- Scherbaum, Frank, Meinard Müller, and Sebastian Rosenzweig. 2017. "Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966." In *Proceedings of the International Workshop on Folk Music Analysis (FMA)*, 29–36. Málaga, Spain.
- Schwingenstein, Christoph. 1994. "Mendelssohn Bartholdy, Felix." In *Neue Deutsche Biographie* 17, 53–58. Historische Kommission bei der Bayerischen Akademie der Wissenschaften. München, Germany. <https://www.deutsche-biographie.de/pnd118580779.html#ndbcontent>.
- Streich, Sebastian. 2007. "Music Complexity: A Multi-Faceted Description of Audio Content." PhD diss., University Pompeu Fabra.
- Temperley, David. 1997. "An Algorithm for Harmonic Analysis." *Music Perception: An Interdisciplinary Journal* 15, no. 1: 31–68.
- Thickstun, John, Zaïd Harchaoui, and Sham M. Kakade. 2017. "Learning Features of Music from Scratch." In *Proceedings of the International Conference on Learning Representations (ICLR)*. Toulon, France.
- Viro, Vladimir. 2011. "Peachnote: music Score Search and Analysis Platform." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 359–62. Miami, FL, USA.
- Weiße, Christof, Stefan Balke, Jakob Abeßer, and Meinard Müller. 2018. "Computational Corpus Analysis: A Case Study on Jazz Solos." In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, 416–423. Paris, France. <https://doi.org/10.5281/zenodo.1492439>.
- Weiße, Christof, Matthias Mauch, Simon Dixon, and Meinard Müller. 2019. "Investigating Style Evolution of Western Classical Music: A Computational Approach." *Musicae Scientiae* 23, no. 4: 486–507. <https://doi.org/10.1177/1029864918757595>.
- Weiße, Christof, and Meinard Müller. 2014. "Quantifying and Visualizing Tonal Complexity." In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, 184–7. Berlin, Germany.
- Weiße, Christof, and Meinard Müller. 2023. "Studying Tonal Evolution of Western Choral Music: A Corpus-Based Strategy." In *Proceedings of the Computational Humanities Research Conference (CHR)*, 3558: 687–702. CEUR Workshop Proceedings.
- Weiße, Christof, and Meinard Müller. 2024. "From Music Scores to Audio Recordings: Deep Pitch-Class Representations for Measuring Tonal Structures." *Journal on Computing and Cultural Heritage* (New York, NY, USA) 17, no. 3 (July). 45: 1–19. <https://doi.org/10.1145/3659103>.
- Weiße, Christof, and Geoffroy Peeters. 2021. "Training Deep Pitch-Class Representations with a Multi-Label CTC Loss." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 754–61.
- Weiße, Christof, and Geoffroy Peeters. 2022. "Comparing Deep Models and Evaluation Strategies for Multi-Pitch Estimation in Music Recordings." *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 30: 2814–27. <https://doi.org/10.1109/TASLP.2022.3200547>.
- Weiße, Christof, Frank Zalkow, Viora Arifi-Müller, Meinard Müller, Hendrik Vincent Kooops, Anja Volk, and Harald Grohgan. 2021a. "Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis." *ACM Journal on Computing and Cultural Heritage (JOCCH)* 14, no. 2: 25: 1–18. Association for Computing Machinery. New York, NY, United States. <https://doi.org/10.1145/3429743>.
- Weiße, Christof, Johannes Zeitler, Tim Zunner, Florian Schuberth, and Meinard Müller. 2021b. "Learning Pitch-Class Representations from Score–Audio Pairs of Classical Music." In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 746–53. Online.
- White, Christopher W. 2013. "Some Statistical Properties of Tonality, 1650–1900." PhD diss., Yale University.
- White, Christopher W. 2023. *The Music in the Data: corpus Analysis, Music Analysis, and Tonal Traditions*. Routledge. New York City, NY.

Appendix

In this appendix, we present some figures that go beyond the scope of the main article while still providing interesting results. First, we used additional annotations in the CAC to conduct further studies on the general stylistic evolution of Western choral music (as in the section "Hypothesis testing using evolution curves"). Figure A1 shows separate ECs computed for different languages of the sung texts. Here, we found three main categories: German, Latin and grouped less represented languages such as English, Italian and Russian. Figure A2 presents separate ECs for the text source of sacred music in particular, and distributed the works with annotated Bible verses according to the Old and New Testaments.

Second, we computed additional complexity deviation curves (as in the section "Case studies: complexity deviation and stylistic trends of individual composers") for other individual composers. For this, we focused on relatively well-known composers from different time periods: Wolfgang Amadeus Mozart (Figure A3), Franz Liszt (Figure A4), Johannes Brahms (Figure A5), Josef Gabriel Rheinberger (Figure A6), Max Reger (Figure A7) and Veljo Tormis (Figure A8).

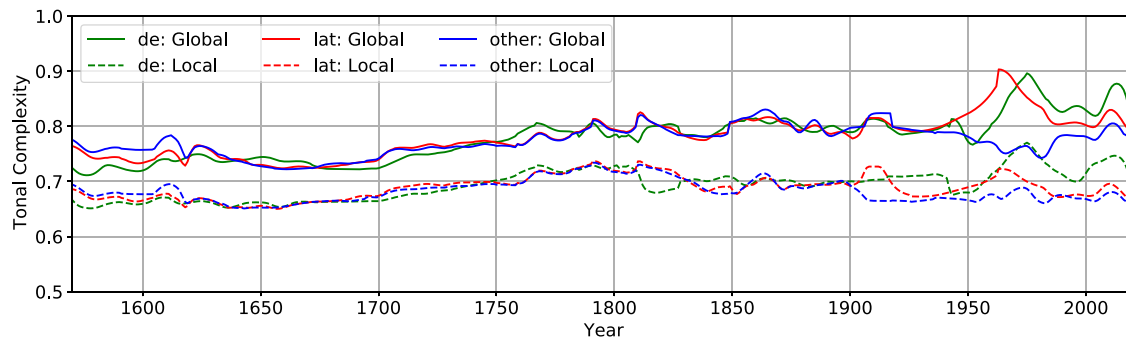


Figure A1. Comparing ECs for global and local complexity separated by texts' language (German, Latin or other).

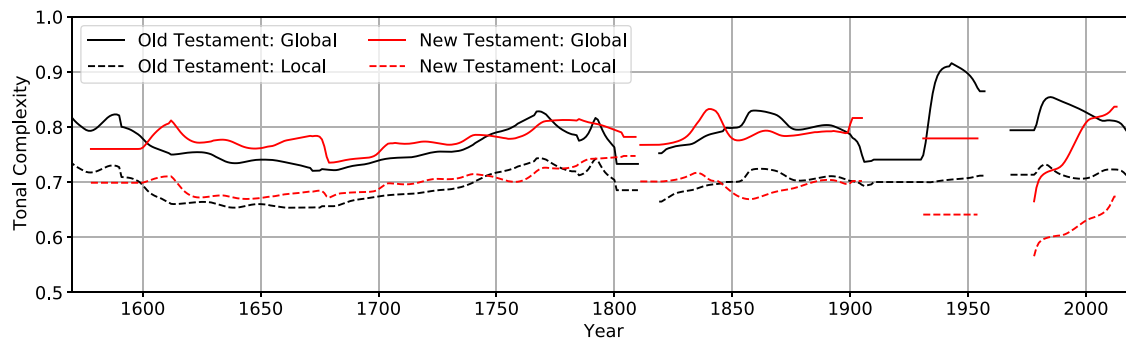


Figure A2. Comparing ECs for global and local complexity separated by text from the Old Testament and the New Testament.

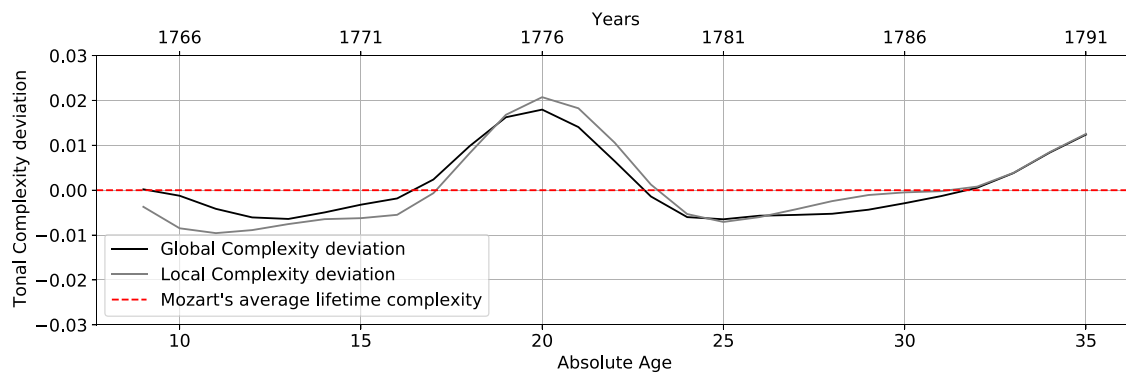


Figure A3. Visualizing the complexity deviation of W. A. Mozart.

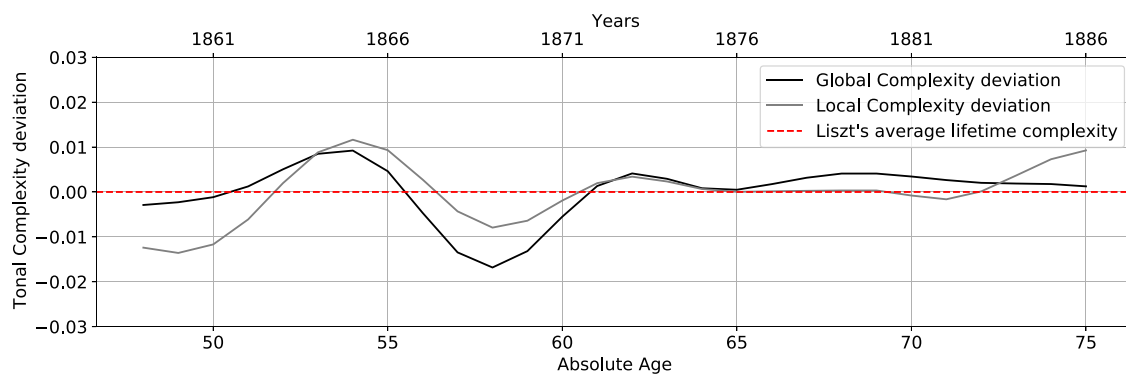


Figure A4. Visualizing the complexity deviation of F. Liszt.

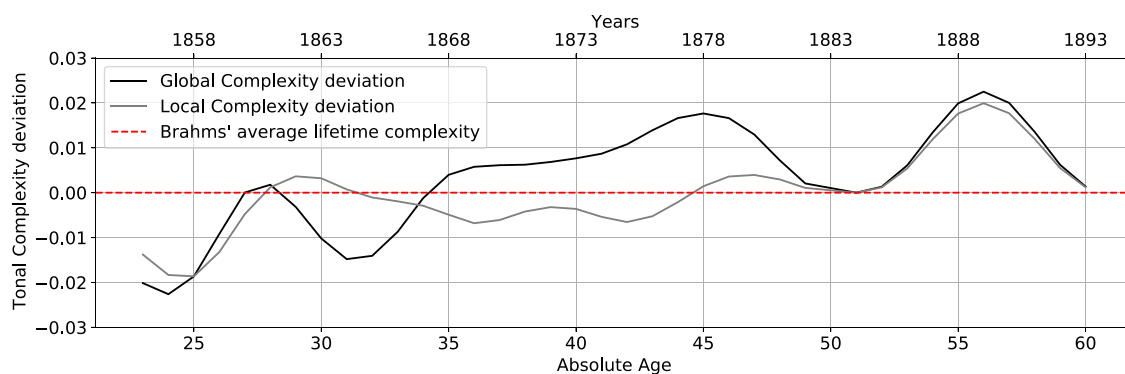


Figure A5. Visualizing the complexity deviation of J. Brahms.

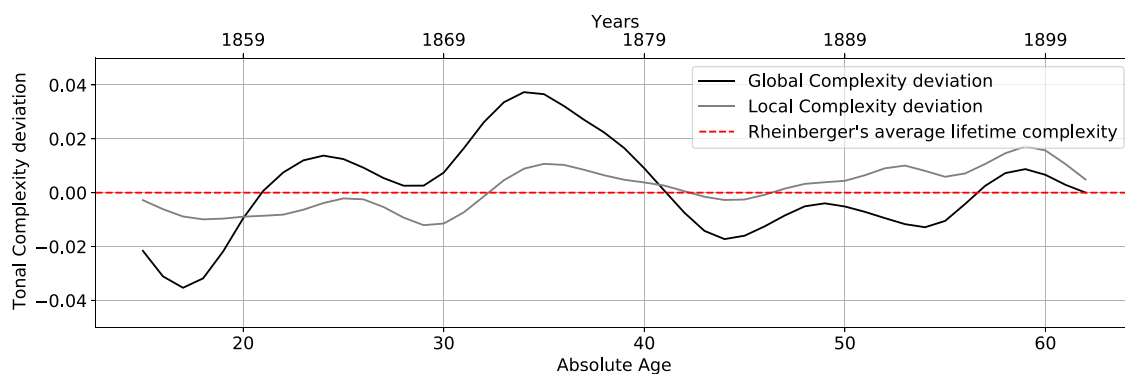


Figure A6. Visualizing the complexity deviation of J. G. Rheinberger.

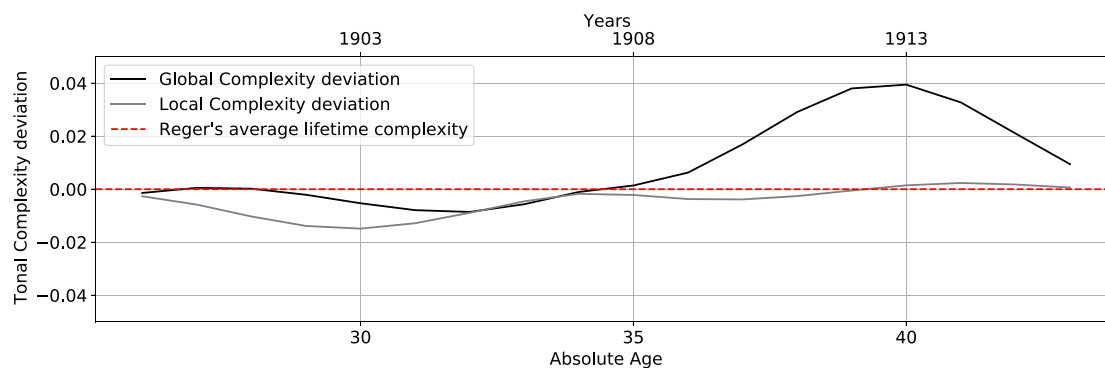


Figure A7. Visualizing the complexity deviation of M. Reger.

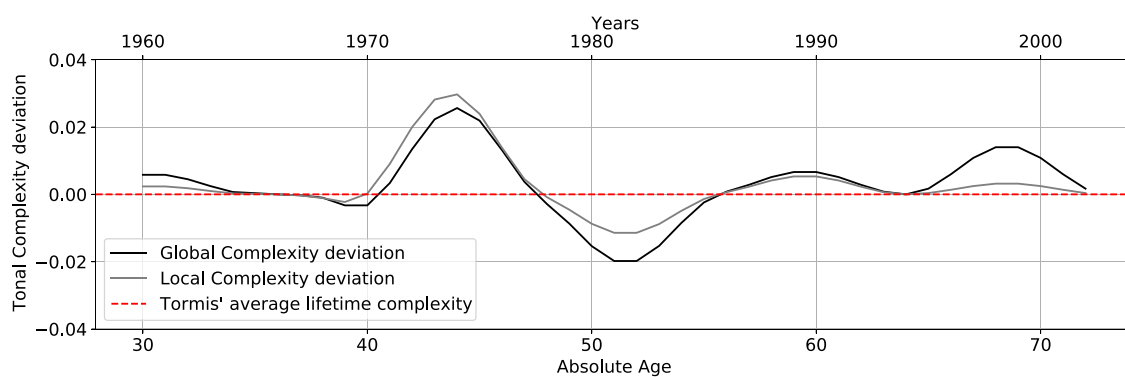


Figure A8. Visualizing the complexity deviation of V. Tormis.