Computational Audio and Music Analysis

Christof Weiß

https://orcid.org/0000-0003-2143-4679

Abstract With the ongoing digitization, not only textual documents but also other types of media have become available in large quantities. This includes audio recordings comprising three main types of content: speech, environmental sounds (e.g. natural or urban soundscapes), and music. While all may be relevant for theological research, this chapter focuses on using audio recordings for studying sacred music (Computational Musicology). After introducing fundamentals of audio data, we first outline a technique for visualizing the tonal content (local keys and modulations) within a music recording and apply this technique to Bach's *Johannespassion* BWV 245. Second, we demonstrate the potential of audio recordings for corpus analysis. We present an approach for studying the tonal complexity and its evolution over centuries. With this technique, we examine the tonal evolution of sacred music exploiting an annotated audio corpus (5,773 tracks) stemming from a leading music publisher for choral music, the Carus-Verlag Stuttgart.

Keywords Audio Signal Processing, Harmony Analysis, Computational Musicology, Corpus Analysis

1. Audio Data and Applications

Ongoing digitization efforts result in an increasing number of archives and corpora on cultural artifacts. Textual data have been the starting point for the Computational Humanities (CH) by using statistical methods on comprehensive literary texts. Nowadays, further modalities are available in the same vein, including audio (sound) recordings. In contrast to text, raw audio poses a number of challenges: First, due to a considerably larger size (one second of uncompressed stereo audio corresponds to 88,200 16-bit values), audio storage and transmission demands for more resources – a problem that has been addressed by efficient audio coding technology beginning with MP3 audio compression and similar codecs. Second, the computational analysis of audio data requires more elaborate processing techniques. In contrast to text, explicit symbols such as characters or words (in speech) or note events (in music) are not directly accessible from audio. To extract this information, algorithmic solutions have been developed for decades, comprising techniques from engineering (signal processing) and computer science (pattern recognition, machine learning, and nowadays deep learning/AI technology). Central venues for this research are the *International*

Christof Weiß: Computational Audio and Music Analysis, in: Christopher A. Nunn and Frederike van Oorschot (eds.): Compendium of Computational Theology 1. Introducing Digital Humanities to Theology, Heidelberg: heiBOOKS, 2024, S. 127–145. DOI: https://doi.org/10.11588/heibooks.1521.c21944

Conference on Acoustics, Speech, and Signal Processing (ICASSP)[1] or the IEEE/ACM *Transactions on Audio, Speech, and Language Processing.*[2]

In general, audio data contains a mixture of various different sources. Consider, for example, the case of a movie sound track or an audio book, which may comprise speech (by different speakers), background music, as well as diegetic sounds (sound effects, sounds generated by people or objects in the plot, or music played or sung within the plot). The separation of these sources is a major computational challenge (Smaragdis 2004). In many cases, however, we are dealing with clean audio data, covering one of three types of content (speech, environmental sound, music), which we shortly summarize in the following.

Speech data. Since the most fundamental type of audio is spoken language, a large part of audio technology is motivated by applications for inter-personal communication. Consequently, speech processing has driven the development of digital audio technology with research on fundamental time-frequency transforms, specific audio features such as *Mel-frequency cepstral coefficients* (MFCCs), dynamic programming techniques such as *Hidden Markov Models,* and more recently, machine-learning algorithms based on neural networks (Bäckström et al. 2022). A central venue for this research field is the annual INTERSPEECH conference by the *International Speech Communication Association* (ICSA).[3] Motivated by different applications, speech processing comprises a variety of tasks such as speech coding and transmission, speaker identification, speech-to-text transcription, analysis of emotion, prosody, or dialect, or audio forensics. More recently, the generation of coherent speech signals from text or directly from a user query has matured due to tremendous progress in generative deep-learning techniques. Interactive voice assistants are one of the most prominent applications of such technology. For CH, efficient speech-to-text (or automatic speech recognition, ASR) systems (Schneider et al. 2019) are of high interest since they can be used as a preprocessing step for the subsequent application of text-based CH strategies.

Environmental sounds. Besides speech, a second field of study covers the processing of sounds in a more general sense. There is a dedicated research community on the detection and classification of acoustic scenes and events (DCASE),[4] which addresses a variety of sound event detection and acoustic scene classification tasks within the annual DCASE challenge. One prominent application of such technology is wildlife and biodiversity monitoring where, for instance, natural reserves are equipped with microphone devices to capture animal sounds for analyzing the presence of species, e.g.,

1 See https://ieeexplore.ieee.org/xpl/conhome/1000002/all-proceedings (Accessed: 21 June 2024).
2 See https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6570655 (Accessed: 21 June 2024).
3 See https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=6570655 (Accessed: 21 June 2024).
4 See https://dcase.community (Accessed: 21 June 2024).

of birds (Bardeli et al. 2010). Another application is monitoring urban sound scenes for targeting issues such as noise pollution or detecting crime or potential dangers. This has been done, for instance, in a large project in New York City (Bello et al. 2019). Specific to such applications is the demand for low-resource technology since many signals are recorded in parallel over large areas and very long time spans. Thus, (pre-) processing has to be done locally on individual sensor units (edge devices).

Music audio recordings. The third major category is music audio. In general, music data exists in a variety of digital data types including (beyond audio) graphical sheet music or symbolic (i.e., machine-readable) scores, which explicitly encode musical symbols and usually allow for the most detailed analyses (Temperley 1997; Bellmann 2012; White 2013; Nakamura & Kaneko 2019). However, scores are not available for a variety of music traditions and styles including improvised (e.g. organ improvisation in church), electronically generated, or orally transmitted music. Moreover, audio-based approaches allow for studying performance aspects such as the behavior of congregational singing in church. Finally, symbolic scores are hard to acquire since manual creation is time-consuming and automated conversion from sheet music images (*optical music recognition,* OMR, see Calvo-Zaragoza et al. 2020) or audio recordings (*automatic music transcription,* AMT, see Benetos et al. 2019) to symbolic scores remains often unsatisfactory and demands for considerable manual post-processing. For this reason, audio recordings are a promising alternative since they allow for efficiently scaling up computational music analyses to large corpora (Scherbaum et al. 2017; Mauch et al. 2015; Weiß et al. 2018; 2019). This requires advanced computational techniques that convert the data into semantically meaningful representations that can be directly interpreted by music experts. Such technology is developed within an interdisciplinary research community centered around the *International Society for Music Information Retrieval* (ISMIR),[5] which offers an annual conference and a journal.[6] *Music Information Retrieval* (MIR) comprises a variety of tasks and applications including music synchronization, harmony analysis (chord and key detection), beat and tempo tracking, genre and style classification, audio decomposition, and music transcription (Müller 2021). Beyond these audio analysis tasks, music generation tasks and other musical data types play an important role within MIR.

In the following, we focus on the potential of MIR technology for musicological research, which demands for specific datasets and analysis techniques. We present two analytical studies. The first one (Section 2) deals with the visualization of harmony (local keys and scales) to analyze the tonal organization of large-scale musical works considering Johann Sebastian Bach's *Johannespassion* BWV 245 as an example. The second one (Section 3) demonstrates an audio-based corpus analysis of musical style

5 See https://www.ismir.net (Accessed: 21 June 2024).

6 See https://transactions.ismir.net (Accessed: 21 June 2024).

in Western sacred music relying on a dataset by a leading publisher for choral and sacred music, the Carus-Verlag Stuttgart.

2. Visualizing Tonal Structures: A Case Study on Bach's *Johannespassion*

This section presents an algorithmic approach for visualizing tonal information over the course of an audio recording. Following (Weiß & Müller 2021), we introduce basic notions of audio, fundamental processing techniques, and our visualization strategy at the example of the choral No. 22 "Durch Dein Gefängnis" from J.S. Bach's *Johannespassion* BWV 245. We finally apply this technique to the complete *Johannespassion* and show its potential for studying the tonal organization of large-scale works.

2.1 Extracting Spectral Information from Audio

The starting point of audio analysis is an acoustic waveform (also referred to as *signal*), as shown in Fig. 1a for a recording of the Bach choral. In a first step, we perform a spectral analysis (Müller 2021, Chap. 2). For this purpose, we first divide the signal into local time windows *(frames).* The width of the time window (given in seconds) is a critical parameter that has to be adapted to the particular application requirements since there is a trade-off between frequency and time resolution. Within a frame, the salience of different frequencies is calculated, which can be realized, for instance, by the Fourier transform.[7] This time window is now shifted over the signal, so that one receives for each frame a local frequency distribution. This results in a time-frequency representation, a so-called *spectrogram*, which is shown in Fig. 1b for the Bach choral example.

For tonal analysis, we further summarize this spectral information according to musical pitches. To this end, we make the simplifying assumption that the pitch content can be described well enough by the twelve-tone equal-tempered scale. We further assume that pitch-class information (ignoring a pitch's octave) is sufficient for our tonal analysis tasks. Thus, we end up with the twelve chromatic pitch classes c, c♯, d, d♯, …, b. Here, enharmonic differentiation of pitches such as c♯ and d♭ is not possible. For each frame of the spectrogram, the frequency components are aggregated according to these twelve pitch classes. This results in a time-chroma representation

7 Together with the windowing procedure described before, this specific variant is denoted as (discrete) *short-time Fourier transform* (STFT). Other transforms have been developed for specific applications such as the *constant-Q transform* (CQT) for pitch analysis or the *modified discrete cosine transform* (MDCT) for audio coding.

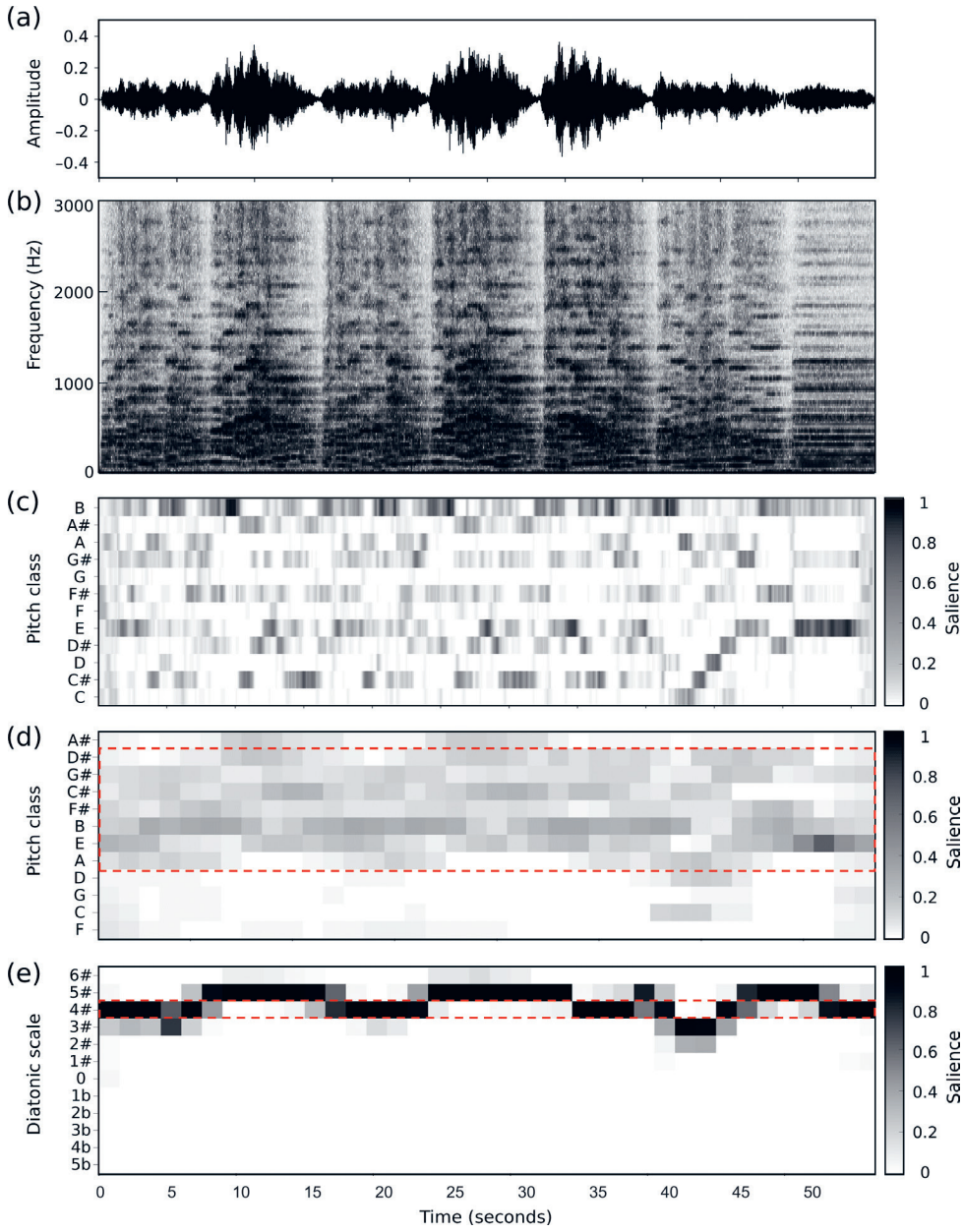

Fig. 1 Choral No. 22 "Durch Dein Gefängnis" from J.S. Bach's *Johannespassion* BWV 245, recording by *The Scholars Baroque Ensemble* (Naxos 1994). (a) Waveform of the audio signal. (b) Spectrogram. (c) Chromagram. (d) Smoothed chromagram, re-ordered according to perfect fifths. (e) Visualization of diatonic scale probabilities.

or chromagram. Fig. 1c shows such a chromagram (time resolution 10 Hz) for the Bach choral example. This representation captures the energy distribution of the audio signal over the twelve chromatic pitch classes over time. Converting music recordings into chromagrams as an intermediate step in the processing chain is a fundamental approach for various MIR tasks such as key estimation, scale analysis, and chord recognition. For details of the underlying signal processing, we refer to Müller (2021).

2.2 Visualizing Diatonic Scales

The observation and measurement of energy distributions in pitch classes provides only limited information for tonally complex polyphonic music. More useful categories for this purpose are intervals, chords, or scales, which have to be described by further processing steps. In the following, we consider the measurement of pitch content according to the twelve diatonic scales. For this purpose, we first smooth the chromagram (by local averaging) in order to account for the coarser time resolution of musical scales (the pitches of a scale do usually not occur within a time span as short as our 100 ms chromagram frames). This is indicated by Fig. 1d, where we chose a filter length of 45 frames (i. e., 4.5 seconds). Second, a chromagram for each of these smoothed frames is compared with binary templates corresponding to the scales. For example, the 0 diatonic (pitch content of C major or A minor, no accidentals) is modeled by a template in which the values for the seven pitch classes c, d, e, f, g, a, h are set to 1 and for the remaining five pitch classes c♯, d♯, f♯, g♯, a♯, to 0. Fig. 1d highlights in red the pitch content of the 4♯ scale (corresponding to E major, the global key of the choral). Due to the similarity of diatonic scales that are related by a perfect fifth (which share six out of seven pitch classes), we organize the scales according to the circle of fifths.

By locally matching the chromagram frames with the twelve different templates, probabilities for the occurrence of these scales over time can be computed. This leads to a generalized time-diatonic representation where the probabilities can be visualized via a grayscale scheme (here, black corresponds to probability 1 and white to probability 0). Fig. 1e shows such a visualization for a recording of the Bach choral. Additional processing steps may be required to better highlight important structures. For example, further temporal smoothing or enhancement of the higher energy values and suppression of low, noise-like values can lead to clearer visual structures. The latter is realized by means of an exponential rescaling of the energy values (similar to the *Softmax* function).

The time-diatonic representation derived from a music recording is initially organized according to physical time steps (seconds). For some applications, such as evaluating the musical form of large-scale works, this can be useful. However, for comparison with score representations or other interpretations of a work (cross-version analysis), it is useful to consider musical time information (e. g., structural bound-

aries or measures/beats). If such information is available, the temporal components of the time-diatonic representation can be musically smoothed in order to obtain, for example, a representation with quarter note resolution.

2.3 Bach Choral Example

Let us now take a closer look at the visualization result for our Bach choral (Fig. 1e). The first part starts in the choral's global key (4♯) and modulates into the upper fifth key or dominant key (5♯) starting at about 8 seconds. An interesting observation is the deviation from the 4♯ scale at roughly 5 seconds, where alterations (here the d in the chord g♯-d-e-h) affect the result. In fact, no d♯ is found in the entire first measure, so a high probability is visible for the 3♯ diatonic as well. This shows that our procedure does not provide an explicit *recognition* of keys, but only describes the local pitch content in terms of diatonic scales. On the basis of the time-diatonic representation, the rough harmonic progression of the Bach choral can be followed conveniently. The beginning phrase (4♯ with modulation to 5♯) is repeated with other lyrics. The choral proceeds with more complex harmonies. Here, the chromatic passage in the bass (text "Unsere Knechtschaft") at about 40 seconds results in several scales obtaining non-zero probabilities. Finally, the choral ends in the global key of E major (4♯). Beyond the easy access to music recordings and their straightforward processing, the direct applicability to audio is of particular advantage since the analysis can be directly linked to the acoustic impression, e.g., by using a running cursor as animation.

2.4 Large-Scale Tonal Visualization of the *Johannespassion*

One of the major benefits of the presented analysis strategy is its scalability. Large-scale works such as operas, oratorios, or symphonies can be visualized in a compact and consistent way, thus allowing for a good overview of their tonal conception – a very important aspect since tonality has been a central means for formally structuring long works. We now demonstrate this by extending our analysis from the one-minute choral to the whole *Johannespassion* BWV 245, which (in the performance by *Scholars Baroque*) amounts to a total playing time of roughly two hours. Fig. 2 shows the result of this analysis, with the whole work being condensed to a single plot.

To avoid micro-fluctuations obscuring the coarse-scale tonal structure, we now opt for a much larger window size when averaging the chroma features before the template matching. As opposed to the more fine-grained analysis of the isolated choral (Section 2.2) with a filter length of 4.5 seconds, we now chose a filter length of 60 seconds. This leads to a suppression of details but enhances the robustness of the method and helps to emphasize the overall tonal structure. Let us now discuss the results. For better readability, Fig. 2 only contains the numbers of the individual movements.

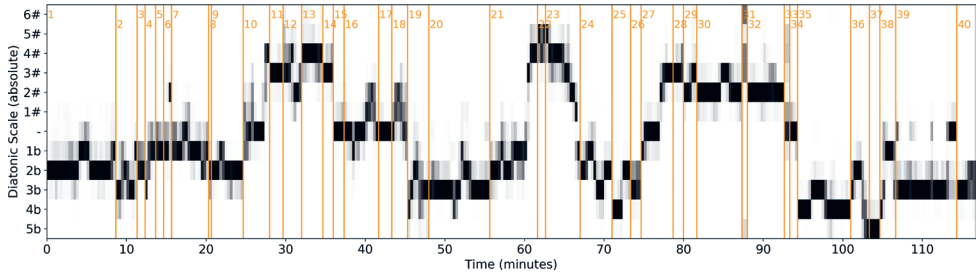

Fig. 2 J.S. Bach's *Johannespassion* BWV 245, complete recording by Scholars Baroque. Diatonic scale visualization for all movements.

The form (instrumentation) and text cues are given in Table 1 for reference. First, we observe a certain local stability. Within a movement and between neighbouring movements, distant modulations are rare, meaning that the next movement usually starts in the same or a closely related key. Second, we identify a certain tonal closure. The work starts in the the −2 diatonic (No. 1, Chorus "Herr, unser Herrscher"in G minor) and ends with two numbers in the −3 diatonic (No. 39, Chorus "Ruht wohl, ihr heiligen Gebeine" in C minor and No. 40, Choral "Ach Herr, laß dein lieb Engelein" in E♭ major.), only a fifth apart. In between, this tonal region is clearly left, with the peak at our choral discussed above, No. 22 "Durch Dein Gefängnis", which is, with its 4♯, about seven fifths apart. Not only is this remarkable since E major has been related to death in the Baroque era – Johann Mattheson ascribed this key a "deadly sadness" ("tödliche Traurigkeit") in his 1713 writing Das *Neu-eröffnete Orchester*. Indeed, the choral is also the center of the passion in several respects. Regarding physical playing time, it occurs right in the middle of the work, after roughly one hour (which is not the beginning of the second part). Moreover, there is a clear symmetry around this choral, which can be nicely observed in our visualization. Numbers 21 and 23 (both a series of recitatives and choruses, which are repeated with different text) include transitions to and from 4♯, respectively. The arias Nos. 20 and 24 each comprise the scales 3♭ and 2♭. Nos. 15–17 and 27 emphasize the 0 diatonic, and so forth. These tonal relationships show the ingenious conception of Bach's *Johannespassion* and his deliberate usage of tonality for structuring the work and for putting emphasis on the theological messages of the passion narrative.

3. Audio-based Corpus Analysis

Scaling analyses to comprehensive works is one major advantage of computational methods. Another highly interesting possibility is to analyze whole corpora of musical works. Here, a corpus may refer to a closed set of works, e.g., all chorals by

Tab. 1 J.S. Bach's *Johannespassion* BWV 245, overview of numbers (movements), forms, and text cues.

No.	Form	Text cue
Parte Prima		
1	Chorus	Herr, unser Herrscher
2	Rezitativ, Chorus	Jesus ging mit seinen Jüngern
3	Choral	O große Lieb
4	Rezitativ	Auf daß das Wort erfüllet würde
5	Choral	Dein Will gescheh
6	Rezitativ	Die Schar aber und der Oberhauptmann
7	Arie	Von den Stricken meiner Sünden
8	Rezitativ	Simon Petrus aber folgete Jesu nach
9	Arie	Ich folge dir gleichfalls
10	Rezitativ	Derselbige Jünger war dem Hohepriester bekannt
11	Choral	Wer hat dich so geschlagen
12	Rezitativ, Chorus	Und Hannas sandte ihn gebunden
13	Arie	Ach, mein Sinn
14	Choral	Petrus, der nicht denkt zurück
Parte Seconda		
15	Choral	Christus, der uns selig macht
16	Rezitativ, Chorus	Da führeten sie Jesum
17	Choral	Ach großer König
18	Rezitativ, Chorus	Da sprach Pilatus zu ihm
19	Arioso	Betrachte, meine Seel
20	Arie	Erwäge, wie sein blutgefärbter Rücken
21	Rezitativ, Chorus	Und die Kriegsknechte flochten eine Krone
22	Choral	Durch dein Gefängnis, Gottes Sohn
23	Rezitativ, Chorus	Die Jüden aber schrieen und sprachen
24	Arie	Eilt, ihr angefochtnen Seelen
25	Rezitativ, Chorus	Allda kreuzigten sie ihn
26	Choral	In meines Herzens Grunde
27	Rezitativ, Chorus	Die Kriegsknechte aber
28	Choral	Er nahm alles wohl in acht
29	Rezitativ	Und von Stund an nahm sie der Jünger
30	Arie	Es ist vollbracht
31	Rezitativ	Und neiget das Haupt
32	Arie	Mein teurer Heiland, laß dich fragen
33	Rezitativ	Und siehe da, der Vorhang im Tempel zerriß
34	Arioso	Mein Herz, in dem die ganze Welt
35	Arie	Zerfließe, mein Herze
36	Rezitativ	Die Jüden aber, dieweil es der Rüsttag war
37	Choral	O hilf, Christe, Gottes Sohn
38	Rezitativ	Darnach bat Pilatum Joseph von Arimathia
39	Chorus	Ruht wohl, ihr heiligen Gebeine
40	Choral	Ach Herr, laß dein lieb Engelein

J.S. Bach or all string quartets by L. van Beethoven. It might, however, also refer to an open subset of a whole time span. To sketch a long-term goal, we envision the analysis of the development of Western sacred music, spanning more than a thousand years from monophonic chant up to today's avantgarde compositions. Following Weiß & Müller (2023), we now present a first step towards such corpus analysis. To this end, we consider an audio dataset provided by a leading publisher of choral and sacred music.

3.1 The Carus Audio Corpus

The Carus-Verlag, founded near Stuttgart, Germany, in 1972 is a family business focusing on vocal and sacred music. Their sheet music editions include around 45,000 works (most of them vocal compositions) and reflect the development of five centuries of choral music, ranging from Gregorian chant, madrigals, and motets of the Renaissance, to contemporary choral music, and works for jazz and pop choir.[8] Carus offers scholarly-critical music editions of the most important oratorios, masses, and cantatas in music history, orientated towards historically informed performance practice. Being also active as a record label, Carus releases reference recordings based on their own editions. The CAC comprises the majority of the Carus CD releases (as of 2019), totalling 7,115 tracks corresponding to individual works or movements (for multi-movement works). Since we want to focus on original art music compositions, we perform a first cleaning step where we remove works without composer, works without composer life dates, arrangements, pop music, children songs, and christmas songs. After this, 5,773 tracks (movements) remain belonging to 2,409 different works with a total duration of 389:52:20 (hh:mm:ss). On average, a work has 2.4 movements and a duration of 9:43 (mm:ss). However, we note that the number of movements per work is highly unbalanced, with many one-movement works on the one hand and many large-scale works (such as the *Johannespassion*) on the other hand. Table 2 provides statistics over the CAC's annotations at the *work level*, where information such as key or instrumentation always refer to the overarching composition. Roughly half of the works has annotations regarding the year of composition (work date). The majority is annotated regarding instrumentation. As said, there is a strong focus on vocal music in general as well as on choral music.[9] From the perspective of tonal analysis, the availability of key annotations for roughly half of the works (1,166 out of 2,409) is of particular relevance. There is a bias towards major keys as well as a considerable number of other keys (church modes). As mentioned above, CAC spans roughly 450 years, covering the period from about 1570–2020. In

8 See https://www.carus-verlag.com/en/ueber-carus (Accessed: 21 June 2024).

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

Tab. 2 Statistics of CAC and its annotations. All numbers refer to full works (not individual movements).

Annotation type	No. of works
– All –	2,409
Work date	1,151
Instrumentation	1,964
instrumental	200
vocal	1,764
choral	1,400
solo	364
Key	1,166
major	673
minor	348
other	145

total, the works stem from 234 different composers. Fig. 3 shows a historical view on the composer dates for composers with at least five works. Well-known composers like F. Mendelssohn-Bartholdy, J.S. Bach, or W.A. Mozart contribute a significant part. However, CAC also comprises less known composers such as H. Schütz or M. Reger. Carus even makes great efforts to bring almost forgotten works by G.A. Homilius or J.G. Rheinberger back into the focus of the choir scene. A particular interesting fact is the good coverage of the late 15th and 16th century. In the 20th century, in contrast, we find a lower number of works, almost observing a gap around 1950.

3.2 Work Count Curves and Approximation Strategy

To analyze musical styles in their historical context, one ideally has information about the true *work dates*, i.e., the year where a composition was completed. Musical styles may evolve rapidly, and composing is subject to trends, being influenced by other composers, the taste of audiences, or extra-musical stimuli. One might think of composers with several *creative periods*, such as L. van Beethoven or A. Schönberg. However, collecting reliable work date annotations for larger datasets requires a substantial amount of manual research, and this information is unknown or in doubt for quite a number of works. Even if one knows all composition dates, it becomes difficult to create a dataset with a balanced coverage of all years. Because of such problems, in our previous work (Weiß et al. 2019), we adopted a pragmatic approach by projecting works onto the historical time axis based on *composer dates*, i.e., the information on birth and death year, which is considerably faster to acquire. To approximation of work counts over the course of a composer's life, we assumed that a composer starts composing not before a certain (fixed) age. For the remaining years (ages), we com-

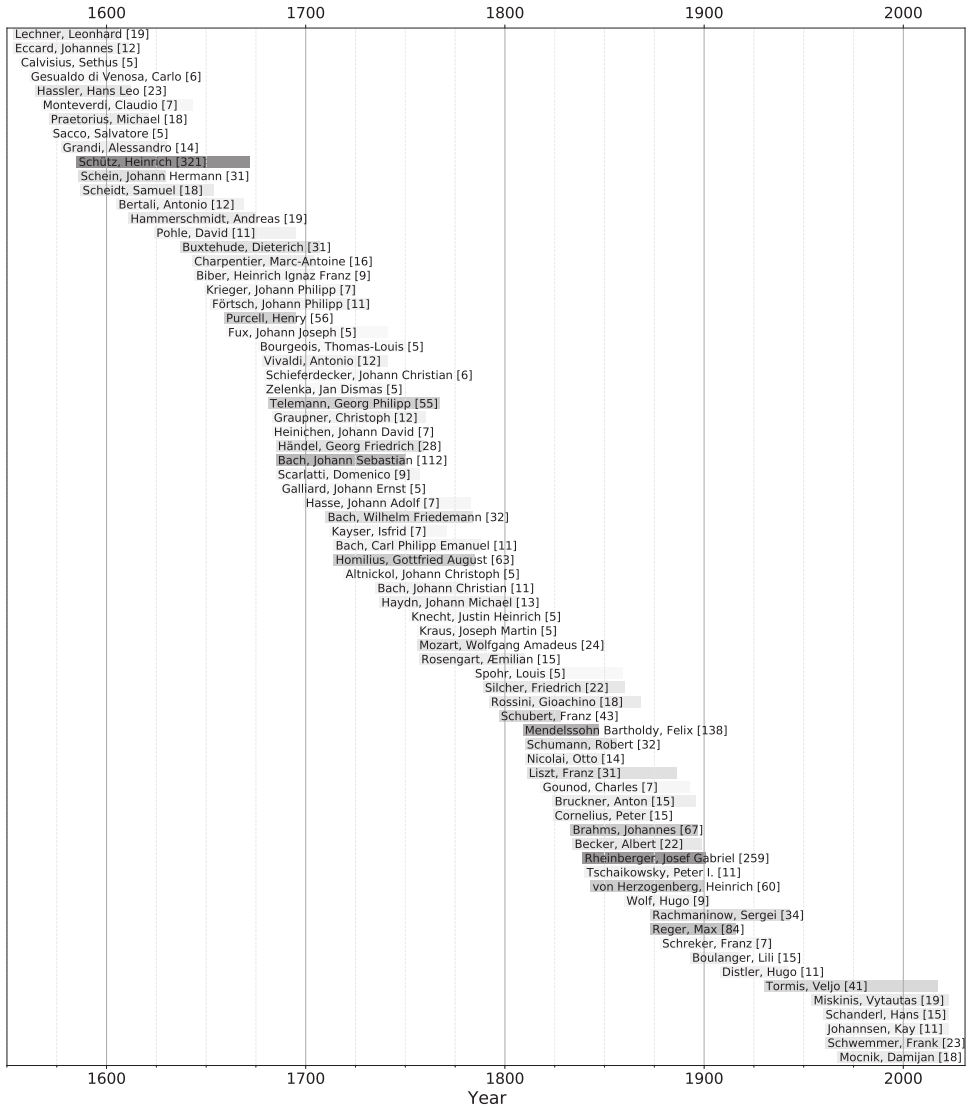

Fig. 3 Historical view of CAC considering all composers with at least five works. The number of works by each composer is indicated in square brackets and encoded by the darkness of the bars.

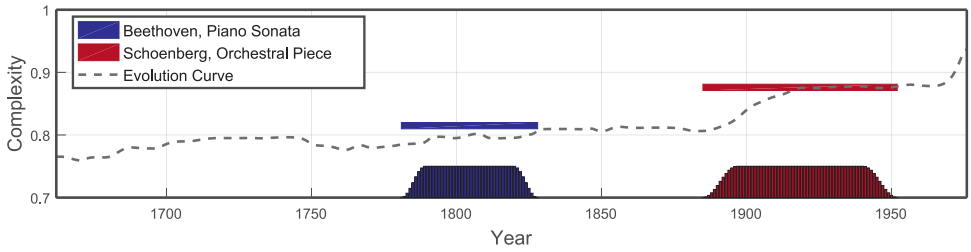

Fig. 4 Approximating evolution of tonal complexity based on composer dates, schematic example.

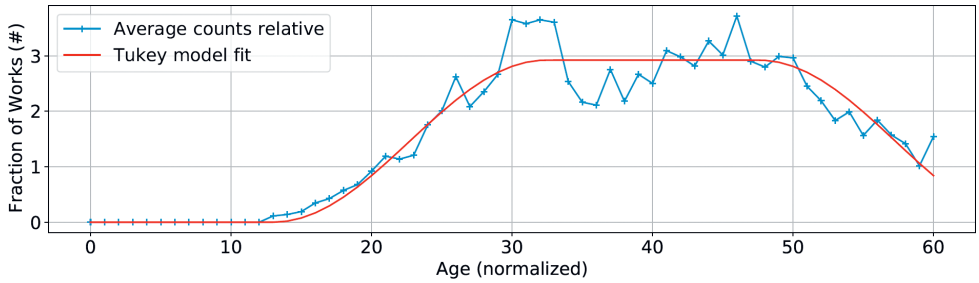

Fig. 5 Curve fitting procedure to determine the optimal window parameters.

puted a roughly flat distribution with smooth edges, using a so-called *Tukey window* (shown in Fig. 4). In Weiß et al. (2019), the parameters (start age and Tukey parameter α) were heuristically chosen. Since CAC contains such annotations for roughly half of the works (cf. Tab. 2), we can validate the approximation strategy and search for optimal values of the parameters. We do this in a stepwise fashion, by first determining the optimal start age, obtaining a value of 13. Then, we determine the optimal Tukey parameter to $\alpha = 0.72$. For details on the Tukey window and the fitting strategy, we refer to Weiß & Müller (2023).

The resulting curve is shown in Fig. 5 for an example hypothetic composer having died at the age of 60. With these optimized window parameters, we now validate the approximation strategy for the work count curve. To this end, we first compute the reference curve using the work dates for 1151 works that have these annotations. We post-process the curve with an average filter of length 15 years (red curve in Fig. 6). We then compare this reference curve with our approximation curve based on composer dates and our optimized Tukey window (blue curve in Fig. 6). Overall, the approximation seems to be suitable. Only in some periods (e.g., around 1680), the approximation curve is ahead, for others (e.g., at 1770), it lags behind the reference curve. We conclude that the approximation based on Tukey windows is a suitable strategy to compensate for missing work date annotations.

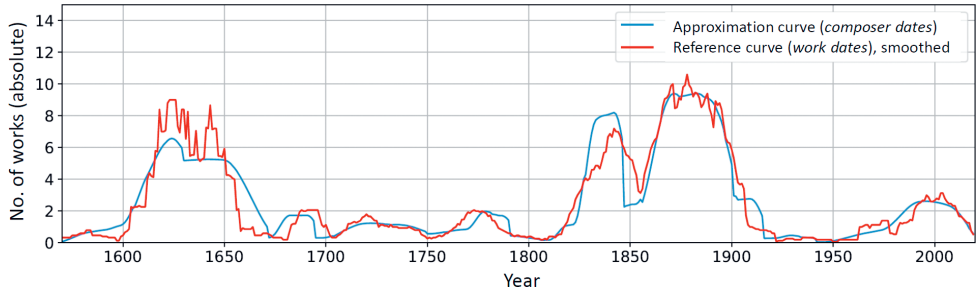

Fig. 6 Work count curves based on composer dates (approximation curve, blue) and based on work dates (reference curve, red), respectively.

3.3 Studying the Tonal Evolution of Sacred Music

With the presented strategy, we now investigate the tonal evolution of choral music in the CAC. To this end, we consider a computational approach for measuring tonal complexity from audio recordings (Weiß et al. 2019). Musical complexity is a highly relevant notion for analysis, which comprises several aspects such as acoustic, timbral, or rhythmic properties (Streich 2007). In our previous work (Weiß & Müller 2014), we introduced tonal complexity measures that locally describe distributions of energy across the twelve chromatic pitch classes used in the Western tonal system. Here, we considered the variety of pitch classes 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 Fig. 7). To this end, we again rely on a chroma representation of the audio recording. For computing the complexity values, we map each chromagram frame (chroma vector) $c \in \mathbb{R}^{12}$ onto the circle of fifth. To this end, we first re-order the chroma values according to perfect fifth intervals. Based on the reordered vector, we compute circular statistics using the resultant vector. Then, our complexity measure $\Gamma(c) \in [0, 1]$ relates to the inverse length of this resultant vector. This measure describes the spread of the pitch classes around the circle of fifths, thus considering also the tonal relationship of active pitch classes. Fig. 7 illustrates the definition of the complexity measure and the resultant vector (in red) showing examples for three input chroma vectors. For a sparse vector (left), the complexity is minimal. For a flat vector (middle), we obtain maximal complexity. Other chroma vectors yield intermediate complexity values. We note that there are different strategies of aggregation to track-wise values. A local measure is obtained by calculating the complexity of each chromagram frame and then averaging over these values. A global chroma statistics can be computed by averaging chromagram frames first and then calculating a single complexity value for each track (movement). Aggregation to works is then done by averaging over the complexity values for all movements. In Fig. 6, we have studied the total number of works in CAC over the course of the years (work count curves)

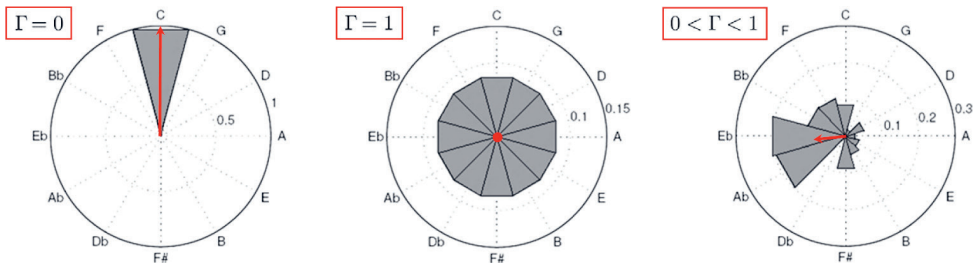

Fig. 7 Complexity measure Γ based on the circle of fifths. Values for a sparse chroma vector (left), a flat chroma vector (middle), and a more realistic chroma vector (right) are shown. The red arrows denote the resultant vectors.

using the work dates or our approximation strategy based on composer dates. We now apply these strategies to our measurements of tonal complexity. 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 of the respective work for weighting. We sum up all weighted windows and divide by the respective work count curve for normalization. We obtain a so-called *evolution* curve (EC) that indicates the average complexity of the works 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 composers' life dates (for the approximation curve).

We now apply this mixed strategy for investigating the evolution of the tonal complexity in CAC. Fig. 8 shows the resulting ECs, one based on the local and one on the global complexity. Looking at the global EC (black), we observe an increase in complexity over the course of the 17th and 18th century. Interestingly, we do not observe a drop around 1750, in contrast to (Weiß et al. 2019) where the demand for more "simplicity" after the Baroque era was clearly visible. On the other hand, the increase during the 19th century observed in Weiß et al. (2019) is not visible for CAC. Even more remarkable, 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 A. Schoenberg or I. 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.

We finally show how the corpus analysis can be used for hypothesis-driven research, investigating the hypothesis that instrumental music is more complex than vocal music. We might 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 usually claim that compositional "revolutions" were often happening in compact instrumental genres such as the string quartet. To test our hypothesis, we use the instrumentation annotations and

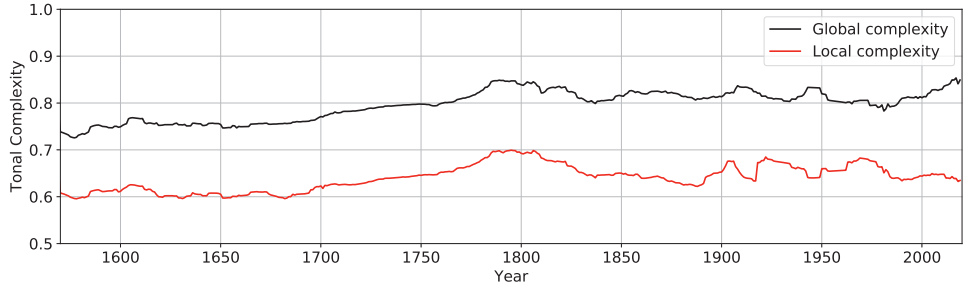

Fig. 8 Comparing ECs for global and local complexity.

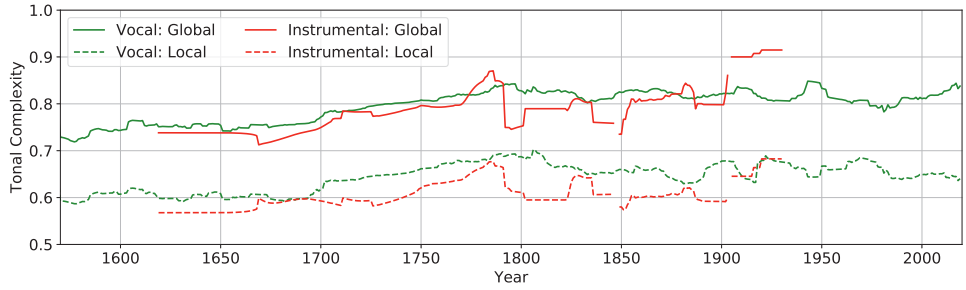

Fig. 9 Comparing ECs for global and local complexity separated into vocal and instrumental music.

compute a vocal as well as an instrumental evolution curve (Fig. 9). As a downside of CAC, we find an unbalanced situation (compare Tab. 2), resulting in a small number of works available for the instrumental EC. Nevertheless, we observe a clear tendency that contradicts our hypothesis: Vocal music seems to be more complex than instrumental music for most time periods. In particular, the offset is large for the local complexity (dashed lines). This may point to an interesting observation, but could also have a technical reason. Our chromagrams are based on a signal processing approach, which maps frequency components extracted from audio recordings to the twelve pitch classes. When dealing with recorded vocal music, this process often leads to substantial artifacts since pitch stability is much lower than for instruments and effects such as vibrato or portamento substantially blur the chromagrams. This may lead to quasi-chromatic artifacts that push the complexity measurements even locally. To overcome such issues, more recent chroma extraction strategies based on deep neural networks are very promising since they have shown to be successful for deriving tonal information from vocal recordings by reducing typical artifacts (Weiß & Peeters 2021).

4. Conclusions

These insights of corpus-based strategies demonstrate the high potential of audio recordings for research in computational theology. Summarizing this chapter, we emphasize the challenges that come along with the analysis of raw audio. Exact recognition of symbols (transcription) is hard and often infeasible. Nevertheless, there are computational approaches for deriving soft, probabilistic ("mid-level") descriptions that strongly correlate with human understanding and intuition. In two case studies, we showed how such approaches may be employed to investigate large-scale musical works (as the *Johannespassion*) or whole corpora spanning several centuries (as for CAC). With the rapid development of processing techniques based on deep neural networks, a considerable improvement of such strategies can be expected in the near future. Nevertheless, in order to obtain reliable insights into theological and humanities question, an interdisciplinary dialogue between experts in both fields (theology and computer science) will remain essential.

Acknowledgments

To a large part, this chapter relies on previous work together with Meinard Müller and colleagues at the AudioLabs Erlangen. C.W. thanks the Carus-Verlag Stuttgart (Johannes Graulich and Ester Petri) for enabling the study of the Carus audio corpus.

References

Bäckström, T., Räsänen, O., Zewoudie, A., Zarazaga, P.P., Koivusalo, L., Das, S., Mellado, E.G., Mansali, M.B., Ramos, D., Kadiri, S., & Alku, P. (2022) Introduction to Speech Processing. 2. ed. [Computer Software]. *Zenodo.* DOI: https://doi.org/10.5281/zenodo.6821775 (Accessed: 21 June 2024).

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K., & Frommolt, K. (2010). Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring, *Pattern Recognition Letters,* 31, 1524–1534. DOI: https://doi.org/10.1016/j.patrec.2009.09.014 (Accessed: 21 June 2024).

Bellmann, H.G. (2012). *Categorization of Tonal Music Style. A Quantitative Investigation* [Diss.]. London: LAMBERT Academic Publishing.

Bello, J.P., Silva, C.T., Nov, O., DuBois, R.L., Arora, A., Salamon, J., Mydlarz, Ch., & Doraiswamy, H. (2019). SONYC. A system for monitoring, analyzing, and mitigating urban noise pollution, *Communications of the ACM,* 62, 68–77. DOI: https://doi.org/10.1145/3224204 (Accessed: 21 June 2024).

Benetos, E., Dixon, S., Duan, Z., & Ewert, S. (2019). Automatic music transcription. An overview, *IEEE Signal Processing Magazine,* 36, 20–30. DOI: https://doi.org/10.1109/MSP.2018.2869928 (Accessed: 21 June 2024).

Calvo-Zaragoza, J., Hajič Jr., J., & Pacha, A. (2020). Understanding optical music recognition, *ACM Computing Surveys,* 53(4), 1–35. DOI: https://doi.org/10.1145/3397499 (Accessed: 21 June 2024).

Mauch, M., MacCallum, R.M., Levy, M., & Leroi, A.M. (2015). The evolution of popular music. USA 1960–2010, *Royal Society Open Science,* 2, 1–10. DOI: https://doi.org/10.1098/rsos.150081 (Accessed: 21 June 2024).

Müller, M. (2021). *Fundamentals of Music Processing. Using Python and Jupyter Notebooks.* 2. ed. Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-69808-9 (Accessed: 21 June 2024).

Nakamura, E., & Kaneko, K. (2019). Statistical evolutionary laws in music styles, *Scientific Reports,* 9, no pag. DOI: https://doi.org/10.1038/s41598-019-52380-6 (Accessed: 21 June 2024).

Scherbaum, F., Müller, M., & Rosenzweig, S. (2017). Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966. In *Proceedings of the 7th International Workshop on Folk Music Analysis* (FMA) (pp. 29–36). Málaga.

Smaragdis, P. (2004). Non-negative matrix factor deconvolution. Extraction of multiple sound sources from monophonic inputs. In C.G. Puntonet & A. Prieto (Eds.), *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation ICA* (pp. 494–499). Berlin/Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-540-30110-3_63 (Accessed: 21 June 2024).

Schneider, S., Baevski, A., Collobert, R., & Auli, M. (2019). wav2vec. Unsupervised pre-training for speech recognition. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)* (pp. 3465–3469). Graz. DOI: https://doi.org/10.21437/Interspeech.2019-1873 (Accessed: 21 June 2024).

Streich, S. (2007) *Music Complexity. A Multi-Faceted Description of Audio Content* [Diss.]. Barcelona: University Pompeu Fabra.

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

Weiß, Ch., Balke, S., Abeßer, J., & Müller, M. (2018). Computational corpus analysis. A case study on jazz solos. In *Proceedings of the 19th International Society for Music Information Retrieval Conference* (ISMIR) (pp. 416–423). Paris. DOI: https://doi.org/10.5281/zenodo.1492439 (Accessed: 21 June 2024).

Weiß, Ch., Mauch, M., Dixon, S., & Müller, M. (2019). Investigating style evolution of Western classical music. A computational approach, *Musicae Scientiae,* 23, 486–507. DOI: https://doi.org/10.1177/1029864918757595 (Accessed: 21 June 2024).

Weiß, Ch., & Müller, M. (2014). Quantifying and visualizing tonal complexity. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM 14)* (pp. 184–187). Berlin. URL: https://www.audiolabs-erlangen.de/content/05_fau/

professor/00_mueller/03_ publications/2014_WeissMueller_TonalComplexity_CIM.pdf (Accessed: 21 June 2024).

Eid. (2021). Computergestützte Visualisierung von Tonalitätsverläufen in Musikaufnahmen. Möglichkeiten für die Korpusanalyse. In S. Klauk (Ed.), *Instrumentalmusik neben Haydn und Mozart. Analyse, Aufführungspraxis und Edition* (pp. 107–130). Würzburg: Königshausen & Neumann [= *Saarbrücker Studien zur Musikwissenschaft,* 20].

Eid. (2023). Studying tonal evolution of Western choral music. A corpus-based strategy. In A. Šela, F. Jannidis & I. Romanowska (Eds.), *Proceedings of the Computational Humanities Research Conference* (pp. 687–702). Paris. URL: https://ceur-ws.org/Vol-3558/paper7862.pdf (Accessed: 21 June 2024).

Weiß, Ch., & Peeters, G. (2021). Training deep pitch-class representations with a multi-label CTC loss. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)* (pp. 754–761). Online: Zenodo. DOI: https://doi.org/10.5281/zenodo.5624358 (Accessed: 21 June 2024).

White, Ch. W. (2013). *Some Statistical Properties of Tonality, 1650–1900* [Diss.]. New Haven, Connecticut: Yale University.

Figure Credits

Fig. 1–9 were generated by the author. Fig. 1 f. were published here for the first time. The others have already been published in previous publications as follows:

Fig. 3: Weiß & Müller 2023, 691.

Fig. 4: Weiß et al. 2019, 8.

Fig. 5: Weiß & Müller 2023, 694.

Fig. 6: Weiß & Müller 2023, 695.

Fig. 7: Weiß et al. 2018, 417.

Fig. 8: Weiß & Müller 2023, 699.

Fig. 9: Weiß & Müller 2023, 699.